

Aus dem Arbeitsbereich Biomedizinische Technik und  
Krankenhaustechnik der Medizinischen Hochschule Hannover

# Analyse von Servicedaten medizin-technischer Geräte in Krankenhäusern

Dissertation  
zur Erlangung des Doktorgrades der Humanbiologie  
der Medizinischen Hochschule Hannover

vorgelegt von  
Dipl. Phys. Thorsten Förstemann  
aus Großburgwedel

Hannover 2004

Angenommen vom Senat der Medizinischen Hochschule Hannover am 15.12.2004  
Gedruckt mit Genehmigung der Medizinischen Hochschule Hannover

Präsident: Prof. Dr. Dieter Bitter-Suermann  
Betreuer: Prof. Dr. Ing. Christoph Hartung  
Referent: Prof. Dr. Berthold Schneider  
Koreferent: Prof. Dr. Siegfried Piepenbrock  
Koreferent: Prof. Dr. J.-Matthias von der Schulenburg

Tag der mündlichen Prüfung 24.02.2005

Promotionsausschussmitglieder:  
Prof.'in Dr. Renate Wrbitzky  
Prof. Dr. Ing. Christoph Hartung  
Prof. Dr. Berthold Schneider  
Prof. Dr. Siegfried Piepenbrock

## Danksagung

Diese Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter im Arbeitsbereich für Biomedizinische Technik und Krankenhaustechnik der Medizinischen Hochschule Hannover. Die Durchführung und Niederschrift dieser Arbeit wäre nicht möglich gewesen ohne die Unterstützung und Förderung zahlreicher Personen und Institutionen.

Mein besonderer Dank gilt dem Leiter des Arbeitsbereiches für Biomedizinische Technik und Krankenhaus Technik, Herrn Prof. Dr.-Ing. Christoph Hartung, für die hervorragende fachliche Betreuung dieser Arbeit sowie für zahlreiche Anregungen und Diskussionen.

Herrn Dipl.-Ing. Rainer Jabs, Leiter des Bereiches Technik / Gebäude der Medizinischen Hochschule Hannover, danke ich für seine vielfältige Förderung. Aus der Medizintechnik der Medizinischen Hochschule möchte ich insbesondere Herrn Rainer Gau für die gute und lehrreiche Zusammenarbeit danken.

Auch möchte ich allen Mitarbeitern der Biomedizinischen Technik für die freundliche Arbeitsatmosphäre Danken, an die ich mich immer gerne zurückerinnern werde.

Nicht zuletzt gilt mein Dank natürlich auch allen Geschäftsführungen und Leitungen der Medizintechnik der hier dargestellten Krankenhäuser für die Bereitstellung der Daten, die die Grundlage meiner Arbeit bilden. Ohne ihre Unterstützung wäre diese Arbeit nicht möglich gewesen.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Gegenwärtige Situation . . . . .	1
1.1.1	Gründe für den Einsatz von Datenbanken . . . . .	1
1.1.2	Gründe gegen den Einsatz von Datenbanken . . . . .	1
1.2	Positionierung . . . . .	2
1.2.1	Ziel dieser Arbeit . . . . .	2
1.2.2	Randbedingungen . . . . .	2
1.3	Aufgabenbeschreibung . . . . .	3
<b>2</b>	<b>Methoden</b>	<b>4</b>
2.1	Datenbanktechnik . . . . .	4
2.1.1	Das Datenbanksystem Microsoft Access . . . . .	4
2.1.2	Begriffsdefinitionen . . . . .	4
2.1.3	Codepages . . . . .	5
2.2	Methoden des Data Mining . . . . .	6
2.2.1	Aufgabenstellung und Projektplanung . . . . .	7
2.2.2	Auswahl der relevanten Datenquellen . . . . .	7
2.2.3	Datenaufbereitung . . . . .	8
2.2.4	Auswahl der Data Mining Methoden . . . . .	9
2.2.5	Anwendung der Data Mining Methoden . . . . .	9
2.2.6	Evaluation, Interpretation und Auswertung der Ergebnisse . . . . .	9
2.3	Statistische Methoden . . . . .	10
2.3.1	Grundlagen . . . . .	10
2.3.2	Deskriptive Statistik . . . . .	13
2.3.3	Prüfende Statistik . . . . .	14
2.3.4	Assoziationsanalyse . . . . .	18
2.3.5	Markowsche Ketten . . . . .	20
2.3.6	Quantitative Linguistik . . . . .	20
2.4	Grundlagen der Instandhaltung . . . . .	21
2.4.1	Begriffs- und Aufgabendefinition . . . . .	21
2.4.2	Kenngrößen der Instandhaltung . . . . .	22
2.4.3	Statistik der Instandhaltung . . . . .	23
2.5	Visuelle Darstellung von Daten . . . . .	27
2.5.1	Visualisierung hierarchischer Daten . . . . .	27
2.5.2	Visualisierung von Kreuztabellen . . . . .	30
2.5.3	Visualisierung von Ungleichverteilungen . . . . .	32

<b>3</b>	<b>Erhebung und Vorbereitung der Daten</b>	<b>34</b>
3.1	Datenerhebung . . . . .	34
3.1.1	Das Anschreiben . . . . .	34
3.1.2	Auswahl der Krankenhäuser . . . . .	34
3.1.3	Eingesandte Antworten . . . . .	34
3.1.4	Eingesandte Daten . . . . .	35
3.2	Datentransfer . . . . .	35
3.2.1	Transfer der Originaldaten ins Textformat . . . . .	35
3.2.2	Codepage-Säuberungen, Teil I . . . . .	36
3.3	Struktur- oder Syntaxanalyse der Daten . . . . .	36
3.3.1	Vorbereitungen . . . . .	37
3.3.2	Analyse und Interpretation . . . . .	38
3.3.3	Manuelle Auswahl und Gruppierung der Spalten . . . . .	41
3.4	Erstellung der Gesamtdatenbank . . . . .	41
3.4.1	Import von Relationen . . . . .	43
3.4.2	Codepage-Säuberungen, Teil II . . . . .	43
3.5	Normierung der Daten . . . . .	45
3.5.1	Normierung metrischer Merkmale . . . . .	46
3.5.2	Normierung kategorialer Merkmale . . . . .	47
3.5.3	Sonderfall: Normierung der Gerätebezeichnung . . . . .	50
3.5.4	Integration der normierten Daten . . . . .	53
3.6	Skalentransformationen . . . . .	54
3.6.1	Absenken des Skalenniveaus . . . . .	54
3.6.2	Anheben des Skalenniveaus . . . . .	54
3.7	Datenanreicherung . . . . .	55
3.7.1	Deutsches Krankenhaus Adressbuch . . . . .	55
3.7.2	EMTEC-Gerätecatalog . . . . .	55
3.7.3	Sonstige Datenquellen . . . . .	55
<b>4</b>	<b>Analyse der Daten</b>	<b>56</b>
4.1	Vorbemerkungen . . . . .	56
4.1.1	Auswahl der Merkmale . . . . .	56
4.1.2	Unvollständige Datensätze . . . . .	56
4.1.3	Einfache und robuste Kennzahlen . . . . .	58
4.1.4	Ordnung der Analysen . . . . .	59
4.1.5	Verwendete Diagrammartentypen . . . . .	60
4.2	Untersuchung der Krankenhäuser . . . . .	61
4.3	Untersuchung der Gerätedaten . . . . .	63
4.3.1	Verteilung der Häufigkeiten . . . . .	63
4.3.2	Verteilung der Investitionen . . . . .	66
4.3.3	Verteilung der Kaufpreise . . . . .	74
4.3.4	Verteilung der Erfassungszeitpunkte . . . . .	79
4.4	Untersuchung der Leistungsdaten . . . . .	80
4.4.1	Verteilung der Häufigkeiten . . . . .	80
4.4.2	Verteilung der Ausgaben . . . . .	85
4.4.3	Verteilung der Leistungskosten . . . . .	91
4.4.4	Verteilung der Durchführungszeitpunkte . . . . .	97

4.5	Analyse von Ungleichverteilungen . . . . .	111
4.5.1	ABC-Analysen . . . . .	111
4.5.2	Lorenz-Kurven . . . . .	113
4.6	Assoziations-Analysen . . . . .	116
4.6.1	Analyse von Koinzidenzen . . . . .	116
4.6.2	Analyse von Sequenzen . . . . .	117
4.7	Analysen der quantitativen Linguistik . . . . .	118
4.7.1	Zipf's Gesetz . . . . .	119
4.7.2	Wortentropie . . . . .	119
<b>5</b>	<b>Schlussteil</b>	<b>127</b>
5.1	Leiterkenntnisse und Verbesserungsvorschläge . . . . .	127
5.1.1	Methodenteil . . . . .	127
5.1.2	Vorbereitungsteil . . . . .	131
5.1.3	Analyseteil . . . . .	135
5.1.4	Unterstützung bei der Gerätebewirtschaftung der Medizinischen Hochschule Hannover . . . . .	143
5.2	Zusammenfassung . . . . .	143
5.3	Schlusswort . . . . .	144
5.4	Ausblick . . . . .	145
<b>A</b>	<b>Das Anschreiben im Wortlaut</b>	<b>147</b>
<b>B</b>	<b>Vollständige Regelsätze für Normierungen</b>	<b>148</b>
B.1	Merkmal Arbeitsbeschreibung . . . . .	148
B.2	Merkmal Abteilung . . . . .	150
<b>C</b>	<b>Gerätelisten der Häufigkeitsanalysen</b>	<b>151</b>
<b>D</b>	<b>Hauptanteile je MT-Untergerätegruppen für verschiedene Merkmale</b>	<b>154</b>
<b>E</b>	<b>Bemerkungen zu Regression der Kaufpreise</b>	<b>158</b>
E.1	Johnson Transformationen . . . . .	158
E.2	Problematische Kaufpreisverteilungen . . . . .	159
<b>F</b>	<b>Maßnahmen in Abhängigkeit der Betriebszeit je Geräteart</b>	<b>161</b>
<b>G</b>	<b>Vollständige Gerätelisten der ABC-Analyse</b>	<b>166</b>

# Tabellenverzeichnis

2.1	Codepage ISO 8859-1 / Latin-1 . . . . .	6
3.1	Zellinhalte nach Häufigkeit sortiert . . . . .	37
3.2	Zellinhalte alphabetisch sortiert . . . . .	38
3.3	Spalteneigenschaften und Histogramm der Zellinhalte . . . . .	39
3.4	Spalteneigenschaften bestimmter Merkmale . . . . .	42
3.5	Zuordnung der Merkmale zu den Spalten . . . . .	43
3.6	Merkmale der Entität „Gerät“ mit Anteil leerer Zellen . . . . .	44
3.7	Merkmale der Entität „Maßnahme“ mit Anteil leerer Zellen . . . . .	45
3.8	Teilzeichenketten der Arbeitsbeschreibung nach Häufigkeit sortiert . . . . .	49
3.9	Teilzeichenketten der Arbeitsbeschreibung nach Häufigkeitsdifferenz sortiert . . . . .	49
3.10	Vorgeschlagene EMTEC-Gerätebezeichnungen zu einer originalen Gerätebezeichnung . . . . .	53
4.1	Untersuchte Merkmale . . . . .	57
4.2	Absolute Gerätezahlen von MT-Geräten . . . . .	63
4.3	Absolute Maßnahmenzahlen an MT-Geräten . . . . .	80
5.1	Jahresinstandhaltungs- und Eigenservicequote je MT-Geräte- und MT-Untergeräteart, Teil I . . . . .	140
5.2	Jahresinstandhaltungs- und Eigenservicequote je MT-Geräte- und MT-Untergeräteart, Teil II . . . . .	141
C.1	Gerätezahlen und Kaufpreise, Teil I . . . . .	151
C.2	Gerätezahlen und Kaufpreise, Teil II . . . . .	152
C.3	Gerätezahlen und Kaufpreise, Teil III . . . . .	153
D.1	Hauptanteile der Gerätezahlen je MT-Untergeräteart . . . . .	154
D.2	Hauptanteile der Kaufpreissummen je MT-Untergeräteart . . . . .	155
D.3	Hauptanteile der Maßnahmenzahlen je MT-Untergeräteart . . . . .	156
D.4	Hauptanteile der Kostensummen je MT-Untergeräteart . . . . .	157
G.1	Gerätelisten der ABC-Analyse, Teil I . . . . .	166
G.2	Gerätelisten der ABC-Analyse, Teil II . . . . .	167
G.3	Gerätelisten der ABC-Analyse, Teil III . . . . .	168

# Abbildungsverzeichnis

2.1	Verteilungsfunktion und Wahrscheinlichkeitsdichte . . . . .	11
2.2	Eigenschaften der log-Normalverteilung . . . . .	15
2.3	Veranschaulichung der Schnittmenge . . . . .	19
2.4	Verlauf des Nutzungsvorrates . . . . .	22
2.5	Bestimmung der Ausfallrate . . . . .	24
2.6	Eigenschaften der Weibull-Verteilung . . . . .	26
2.7	Squarified treemap Algorithmus . . . . .	28
2.8	Benennung der Höhendifferenzen in Balkendiagrammen . . . . .	31
2.9	Vergleich von ungruppierten, gruppierten und sortierten Balkendiagrammen . . . . .	31
2.10	Beispielhafte Lorenzkurven . . . . .	33
3.1	Punktendiagramm: Anteile leerer und einmaliger Zellen je Spalte . . . . .	42
3.2	Histogramm: Häufigkeiten von Teilzeichenketten . . . . .	50
4.1	Punktendiagramm: Gerätezahl über Bettenzahl je Krankenhausart . . . . .	62
4.2	Punktendiagramm: Gerätezahl pro Bett über Bettenzahl je Krankenhausart . . . . .	62
4.3	Tortendiagramm: relative Gerätezahl je Hauptgeräteart und Geräteart . . . . .	64
4.4	Balkendiagramm: relative Gerätezahl je MT-Geräteart und Krankenhaus . . . . .	66
4.5	Treemap: relative Gerätezahl je MT-Geräteart und -Untergeräteart für Universitätskliniken . . . . .	67
4.6	Treemap: relative Gerätezahl je MT-Geräteart und -Untergeräteart für nicht- Universitätskliniken . . . . .	68
4.7	Treemap: relative Gerätezahl je MT-Geräteart und Hersteller . . . . .	69
4.8	Balkendiagramm: relative Kaufpreissummen je MT-Geräteart und Krankenhaus . . . . .	71
4.9	Treemap: relative Kaufpreissummen je MT-Geräteart und -Untergeräteart für Universitätskliniken . . . . .	72
4.10	Treemap: relative Kaufpreissummen je MT-Geräteart und -Untergeräteart für nicht Universitätskliniken . . . . .	73
4.11	Lorenzkurve: Kaufpreise je Krankenhausart . . . . .	75
4.12	Punktendiagramm: kumulierte relative Kaufpreise über Kaufpreis je Krankenhausart . . . . .	75
4.13	Punktendiagramm: Beispielhafte Regression der Kaufpreise . . . . .	76
4.14	Punktendiagramm: Regressionsparameter der log-Normalverteilung der Kaufpreise je Krankenhausart . . . . .	77
4.15	Punktendiagramm: Regressionsparameter der log-Normalverteilung der Kaufpreise je MT-Geräteart . . . . .	77

4.16	Punktendiagramm: Regressionsparameter der log-Normalverteilung der Kaufpreise je MT-Untergeräteart . . . . .	78
4.17	Punktendiagramm: kumulative normierte Inbetriebnahmen je Krankenhaus	79
4.18	Balkendiagramm: relative Maßnahmenzahlen je MT-Gerätearten und Krankenhaus . . . . .	81
4.19	Treemap: relative Maßnahmenzahlen je MT-Geräteart und -Untergeräteart für Universitätskliniken . . . . .	82
4.20	Treemap: relative Maßnahmenzahlen je MT-Geräteart und -Untergeräteart für nicht Universitätskliniken . . . . .	83
4.21	Balkendiagramm: relative Maßnahmenzahlen je MT-Geräteart und Arbeitsbeschreibung . . . . .	84
4.22	Balkendiagramm: relative Maßnahmenzahlen je Leistungsart und MT-Geräteart . . . . .	86
4.23	Balkendiagramm: relative Maßnahmenzahlen je Leistungsart und Krankenhausart . . . . .	86
4.24	Balkendiagramm: relative Kostensummen je MT-Geräteart und Krankenhaus . . . . .	87
4.25	Treemap: relativen Kostensummen je MT-Geräteart und -Untergeräteart für Universitätskliniken . . . . .	88
4.26	Treemap: relativen Kostensummen je MT-Geräteart und -Untergeräteart für nicht Universitätskliniken . . . . .	89
4.27	Balkendiagramm: relative Kostensummen je MT-Geräteart und Arbeitsbeschreibung . . . . .	90
4.28	Balkendiagramm: relative Kostensummen je MT-Geräteart und Leistungsart	91
4.29	Balkendiagramm: relative Kostensummen je Krankenhausart und Leistungsart . . . . .	92
4.30	Lorenz-Diagramm: Kosten je Krankenhausart . . . . .	92
4.31	Punktendiagramm: kumulierte relative Kosten über Kosten je Krankenhausart	93
4.32	Beispielhafte Regression der Kosten . . . . .	94
4.33	Punktendiagramm: Regressionsparameter der log-Normalverteilung der Kosten je Krankenhausart . . . . .	95
4.34	Punktendiagramm: Regressionsparameter der log-Normalverteilung der Kosten je Arbeitsbeschreibung und Leistungsart . . . . .	96
4.35	Punktendiagramm: Regressionsparameter der log-Normalverteilung der Kosten je MT-Geräteart . . . . .	97
4.36	Punktendiagramm: Regressionsparameter der log-Normalverteilung der Kosten je MT-Untergeräteart . . . . .	98
4.37	Punktendiagramm: Regressionsparameter der log-Normalverteilung der Kosten je Kaufpreis . . . . .	99
4.38	Histogramm und Balkendiagramm: Histogramm der Kosten und relative Maßnahmenzahl je Kosten und Leistungsart . . . . .	99
4.39	Punktendiagramm: kumulierte relative Maßnahmenzahlen über Durchführungszeitpunkt . . . . .	100
4.40	Histogramm: Maßnahmenzahlen über der Betriebszeit . . . . .	100
4.41	Balkendiagramm: relative Maßnahmenzahlen je MT-Geräteart über der Betriebszeit . . . . .	101

4.42	Balkendiagramm: relative Maßnahmenzahlen je Arbeitsbeschreibung über der Betriebszeit . . . . .	102
4.43	Balkendiagramm: relative Maßnahmenzahlen je Leistungsart über der Betriebszeit . . . . .	103
4.44	Histogramm: Kostensummen über der Betriebszeit . . . . .	104
4.45	Balkendiagramm: relative Kostensummen je MT-Geräteart über der Betriebszeit . . . . .	104
4.46	Balkendiagramm: relative Kostensummen je Arbeitsbeschreibung über der Betriebszeit . . . . .	105
4.47	Balkendiagramm: relative Kostensummen je Leistungsart über der Betriebszeit . . . . .	106
4.48	Punktendiagramm: Regressionsparameter der Weibull-Verteilungen für Durchführungszeitpunkte je MT-Geräteart . . . . .	107
4.49	Punktendiagramm: Regressionsparameter der Weibull-Verteilungen für Durchführungszeitpunkte je MT-Untergeräteart . . . . .	108
4.50	Punktendiagramm: Regressionsparameter der Weibull-Verteilungen für Durchführungszeitpunkte je Kaufpreis . . . . .	109
4.51	Punktendiagramm: Regressionsparameter der Weibull-Verteilungen für Durchführungszeitpunkte je Krankenhausart . . . . .	109
4.52	Balkendiagramm: Aktivgerätezahl über der Betriebszeit . . . . .	110
4.53	Balkendiagramm: Maßnahmenzahl pro Gerät und Jahr über der Betriebszeit	110
4.54	Punktendiagramm: ABC-Analyse je MT-Geräteart . . . . .	112
4.55	Punktendiagramm: ABC-Analyse je MT-Untergeräteart . . . . .	112
4.56	Punktendiagramm: ABC-Analyse je Krankenhausart . . . . .	114
4.57	Punktendiagramm: ABC-Analyse je Kaufpreis . . . . .	114
4.58	Lorenzdiagramme für verschiedene Merkmale . . . . .	115
4.59	Assoziations-Diagramm: Arbeitsbeschreibung = Akku . . . . .	117
4.60	Assoziations-Diagramm: Arbeitsbeschreibung = bedienungsbedingt . . . . .	118
4.61	Assoziations-Diagramm: Leistungsart = Eigenleistung . . . . .	119
4.62	Assoziations-Diagramm: Leistungsart = Fremdleistung . . . . .	120
4.63	Assoziations-Diagramm: Arbeitsbeschreibung = Elektrik . . . . .	121
4.64	Assoziations-Diagramm: Arbeitsbeschreibung = Mechanik . . . . .	121
4.65	Assoziations-Diagramm: Arbeitsbeschreibung = Prüfung . . . . .	122
4.66	Assoziations-Diagramm: Arbeitsbeschreibungen als Markoff-Ketten . . . . .	122
4.67	Punktendiagramm: Zipf-Verteilung der originalen Gerätebezeichnungen . . . . .	123
4.68	Punktendiagramm: Zipf-Verteilung der originalen Arbeitsbeschreibungen . . . . .	123
4.69	Wortentropie der Gerätebezeichnung über Inbetriebnahme . . . . .	124
4.70	Wortentropie der Arbeitsbeschreibung über Durchführungszeitpunkt . . . . .	126
4.71	Wortentropie der Arbeitsbeschreibung über Betriebszeit . . . . .	126
5.1	Punktendiagramm: Wartungs- / Instandhaltungs- und Eigenservicequote über der Betriebszeit . . . . .	139
5.2	Punktendiagramm: Jahresinstandhaltungs- und Eigenservicequote je Krankenhaus . . . . .	139
A.1	Anschreiben im Wortlaut . . . . .	147

E.1	Beispielhafte Kaufpreis-Verteilungen im $(\beta_1, \beta_2)$ -Raum . . . . .	159
E.2	Beispielhafte Kaufpreis-Verteilungen für Blutdruckmessgeräte . . . . .	160
F.1	Balkendiagramm: Maßnahmenzahl pro Gerät und Jahr und Aktivgerätenzahl über Betriebszeit für Bildgebende Systeme . . . . .	161
F.2	Balkendiagramm: Maßnahmenzahl pro Gerät und Jahr und Aktivgerätenzahl über Betriebszeit für Chirurgie-/Endoskopiegeräte . . . . .	161
F.3	Balkendiagramm: Maßnahmenzahl pro Gerät und Jahr und Aktivgerätenzahl über Betriebszeit für Desinfektion- / Sterilisationsgeräte . . . . .	162
F.4	Balkendiagramm: Maßnahmenzahl pro Gerät und Jahr und Aktivgerätenzahl über Betriebszeit für Fachärztliches Gerät . . . . .	162
F.5	Balkendiagramm: Maßnahmenzahl pro Gerät und Jahr und Aktivgerätenzahl über Betriebszeit für Funktionsdiagnostik . . . . .	163
F.6	Balkendiagramm: Maßnahmenzahl pro Gerät und Jahr und Aktivgerätenzahl über Betriebszeit für Laborgeräte . . . . .	163
F.7	Balkendiagramm: Maßnahmenzahl pro Gerät und Jahr und Aktivgerätenzahl über Betriebszeit für Patientenüberwachungsgeräte . . . . .	164
F.8	Balkendiagramm: Maßnahmenzahl pro Gerät und Jahr und Aktivgerätenzahl über Betriebszeit für Stoffaustauschgeräte . . . . .	164
F.9	Balkendiagramm: Maßnahmenzahl pro Gerät und Jahr und Aktivgerätenzahl über Betriebszeit für Strahlentherapiegeräte . . . . .	165
F.10	Balkendiagramm: Maßnahmenzahl pro Gerät und Jahr und Aktivgerätenzahl über Betriebszeit für Vitalfunktion- / Intensivmedizinische Geräte . . . . .	165

# Symbole

Symbol	Beschreibung
$A, B, D$	Ereignismengen, Ereignisse
$\#(A)$	Anzahl der Elemente (Elementarereignisse) der Menge (des Ereignisses) $A$
$\alpha, \beta$	Parameter der Weibull-Verteilung
$\alpha_L, \beta_L$	Parameter der log-Normalverteilung
$\beta_1$	Schiefe (skewness)
$\beta_2$	Wölbung (kurtosis)
$C$	Kontingenzkoeffizient von Pawlik
$C_{max}$	maximaler Kontingenzkoeffizient von Pawlik
$C_{korr}$	korrigierter Kontingenzkoeffizient von Pawlik
$E(X)$	Erwartungswert der Zufallsvariablen $X$
$f$	Anzahl der Freiheitsgrade
$f(x)$	Wahrscheinlichkeitsdichte
$f_X$	Wahrscheinlichkeitsdichte der Zufallsvariablen $X$
$F_X$	Verteilungsfunktion der Zufallsvariablen $X$
$f_i, f(t)$	Anzahl ausgefallener Geräte pro Zeiteinheit
$F_i, F(t)$	kumulierter Anteil ausgefallener Geräte
$h$	Häufungspunkt, Entropie
$h_R$	relative Entropie
$H_0$	Nullhypothese
$H_A$	Alternativhypothese
$k_{n;\alpha}$	kritischer Wert für $n$ Freiheitsgrade und Irrtumswahrscheinlichkeit (Signifikanzniveau) $\alpha$
$m$	Median
$\mu$	Mittelwert
$p(A)$	Wahrscheinlichkeit für das Eintreten des Ereignisses $A$
$P_i, p_i$	absolute und relative Worthäufigkeit
$R_i, R(t)$	Anteil der noch nicht ausgefallen Geräte
$r_S$	Spearman'scher Rang-Korrelationskoeffizient
$T$	Testgröße
$Var(X)$	Varianz der Zufallsvariablen $X$
$X, Y$	Zufallsvariable
$x_i, y_i$	Stichprobenwerte der Zufallsvariablen $X$ und $Y$
$z_i, z(t)$	Ausfallrate



# 1 Einleitung

## 1.1 Gegenwärtige Situation

### 1.1.1 Gründe für den Einsatz von Datenbanken

Trotz des Mehraufwandes setzt sich eine EDV der Geräte-Stamm- und -Bewegungsdaten im Krankenhaus immer mehr durch. Zum einen wecken kostspielige Geräte und Vorgänge, die zahlreich in diesem Bereich auftreten, Hoffnung auf umfangreiche Einsparmöglichkeiten bei effektiver, EDV gestützter Gerätebewirtschaftung, und zum anderen ist der Betreiber auch durch Auflagen gezwungen, in einem Maße die Bewirtschaftungsvorgänge zu dokumentieren, das nur noch unter Zuhilfenahme von EDV zu bewältigen ist. Zu nennen sind hier das Medizin Produkte Gesetz (MPG) und die Medizinprodukte Betreiber Verordnung (MPBetrVO). Weitere Informationen finden sich in der DIN 31051 und der Richtlinie VDI / VDE 2426. Für diese Aufgabe stehen bereits seit längerer Zeit spezielle Softwarelösungen verschiedener Hersteller zur Auswahl (z.B. MT-DATA). Grundsätzlich sind diese Softwarelösungen auf modernen Standard-PCs einsetzbar. Die Hard- und Software- Voraussetzungen zum Einsatz von EDV-gestützter Gerätebewirtschaftung stellen also heutzutage keine Hürde mehr da. Darüber hinaus ist die nötige Software mittlerweile an vielfältige Spezialfälle anpassbar. Nicht nur die Eingabe von Daten ist intuitiv möglich, auch intuitiv zu bedienende Analyse-Werkzeuge sind mittlerweile verfügbar.

### 1.1.2 Gründe gegen den Einsatz von Datenbanken

Die eigentliche Durchführung der Gerätebewirtschaftung hat vor der EDV Erfassung natürlich Priorität, und diese Erfassung wiederum vor der Auswertung der erfassten Daten. Da die Ressourcen knapp sind, wird die Analyse durch eine lückenhafte Erfassung erschwert oder unterbleibt sogar ganz.

Zwar sind die Hard- und Software-Voraussetzungen zum Einsatz von EDV-gestützter Gerätebewirtschaftung heute keine Herausforderung mehr, jedoch erfordert die Bedienung und Pflege der Erfassungssysteme teilweise spezielle Kenntnisse der Nutzer. Spätestens bei einer Störung des Datenbank-Systems oder eines nicht vorgesehenen Anwendungsfalls ist der Bediener ohne spezielle Kenntnisse überfordert. Oft zeigt sich, je einfacher ein System im Regelfall zu bedienen ist, desto schwieriger ist eine Problemlösung im Ausnahmefall. Das im Problemfall nötige Wissen erfordert entweder eine Schulung eines oder mehrerer Mitarbeiter – was zu einer Abhängigkeit vom genutzten System führt, da das Wissen der Mitarbeiter nur mit hohem Aufwand an ein neues System angepasst werden kann. Die andere Möglichkeit des Einkaufs des nötigen Wissens durch externe Fachkräfte führt zwar zu einer geringeren Abhängigkeit von einem Datenbank-System, jedoch tritt gegenüber den mit der Gerätebewirtschaftung

befassten Technikern neben den medizinischen Personal als Nutzer, der kaufmännischen Verwaltung des Krankenhauses und dem Hersteller der Geräte noch eine weitere Gruppe, mit der das Vorgehen abgestimmt werden muss: die Datenbankfachleute. Beide Alternativen sind der Akzeptanz von EDV gestützter Gerätebewirtschaftung nicht förderlich.

## 1.2 Positionierung

### 1.2.1 Ziel dieser Arbeit

Die EDV-technische Erfassung und Analyse von Gerätedaten im Krankenhaus erfolgt zur Zeit sicher nicht optimal. Dies ist jedoch weniger auf das Fehlen effektiver Konzepte zurückzuführen als auf das Fehlen der nötigen Ressourcen, diese zu realisieren. Ziel dieser Arbeit ist nicht, neue Konzepte zur Gerätebewirtschaftung mit EDV im Krankenhaus zu entwickeln, sondern die Gerätedaten, die bis jetzt unter meist nicht optimalen Bedingungen digital erfasst wurden, aus verschiedenen Krankenhäusern zu sammeln, aufzuarbeiten, zu analysieren und in interpretierbarer Form zusammenzufassen. Um Empfehlungen zur Gerätebewirtschaftung oder deren EDV Erfassung aussprechen zu können, bedarf es oft zusätzlicher Informationen, die nicht in den Datenbanken vorliegen und auch nicht im Rahmen dieser Arbeit beschafft werden können. Trotzdem soll am Ende der Arbeit versucht werden, aus den gewonnen Erkenntnissen Empfehlungen abzuleiten, obwohl schwerpunktmäßig das Ziel dieser Arbeit darin besteht, die Erkenntnisse so darzustellen, dass der Leser die Ableitung von Empfehlungen selber vornehmen kann und dabei eigene Informationen mit integrieren kann.

### 1.2.2 Randbedingungen

An die Daten, die zu dieser Untersuchung herangezogen werden können, ist nur eine Anforderung zu stellen: sie müssen in elektronischer Form vorliegen, denn sonst ist der Aufwand der Analyse im Rahmen dieser Untersuchung nicht zu bewältigen. Darüber hinaus werden an die Daten jedoch keine weiteren Anforderungen gestellt, und die damit gegebenenfalls notwendigen, vorbereitenden Maßnahmen sind explizit Teil dieser Untersuchung.

Auch können im Rahmen dieser Arbeit aus den Daten heraus keine Nachforschungen zurück zu den abgebildeten Vorgängen unternommen werden, d.h. es können keine Vergleiche zwischen abgebildeter Wirklichkeit und vorliegenden Datensätzen vorgenommen werden, da auch dies einen nicht zu bewältigenden Aufwand bedeutet. Durch dieses Vorgehen soll die sich ergebende Unsicherheit bezüglich der Vertrauenswürdigkeit der vorliegenden Daten durch erhebliche Vergrößerung der betrachteten Datenmenge kompensiert werden. Darüber hinaus macht sich diese Untersuchung auch unabhängig von nicht oder nur schlecht dokumentierten Zusatzinformationen zu den vorliegenden Datenbanken, wie sie sich zum Beispiel aus Gesprächen mit den Bedienern ergeben können.

## 1.3 Aufgabenbeschreibung

Elektronisch gespeicherte Datenbestände über insbesondere medizin-technische und sonstige Geräte in verschiedenen Krankenhäusern sollen ausgewertet werden. Die Datenbestände umfassen im Idealfall alle Stamm-, und Bewegungsdaten der Geräte. Die Auswertungen haben das Ziel, die in den Datenbeständen enthaltenen Informationen zu sammeln, zusammenzufassen und in interpretierbare Form zu bringen. Mögliche Konsequenzen dieser Erkenntnisse sollen kurz umrissen werden.

## 2 Methoden

Da die Methoden, die in dieser Arbeit zur Anwendung kommen, hauptsächlich nach ihrem Nutzen für die zu bearbeitende Fragestellung ausgewählt werden, kommen Methoden verschiedener Fachbereiche zum Einsatz. Das für das Verständnis der angewandten Methoden notwendige Grundwissen soll in diesem Abschnitt kurz vorgestellt werden.

### 2.1 Datenbanktechnik

Datenbanken sind das Hilfsmittel um Daten, wie sie z.B. bei der Gerätebewirtschaftung anfallen, effektiv zu erfassen und zu verwalten. Aus diesem Grunde werden in diesem Abschnitt kurz die Grundlagen der in dieser Arbeit verwendeten Datenbanktechnik dargestellt. Für umfassende Darstellungen sei auf z.B. [1] verwiesen.

#### 2.1.1 Das Datenbanksystem Microsoft Access

Als Datenbanksystem kam in dieser Arbeit Microsoft Access 2000 zum Einsatz. Vorteile von Microsoft Access sind die intuitive Bedienungsoberfläche und die Integration in die anderen Microsoft Office Komponenten. Als Nachteil stellten sich die 2GB-Grenze von Access heraus sowie die teilweise geringe Verarbeitungsgeschwindigkeit bei Großen Datenbeständen. Auch das Eigenleben der Importassistenten von Access machte sich negativ bemerkbar. Da Access in der Lage ist, SQL-Befehle (Structured Query Language) auszuführen, ist es mächtig genug, um die in dieser Arbeit notwendigen Datenbankabfragen durchzuführen.

Access ist ein relationales Datenbanksystem, d.h. die Daten werden in Tabellen abgelegt, die miteinander verknüpft werden können. Durch die Möglichkeit Tabellen zu verknüpfen, können Darstellungsformen (sog. „Normalformen“) für die in der Datenbank abzulegenden Daten gefunden werden, die weitgehend redundanzfrei sind. Dies erleichtert die Datenbestände konsistent zu halten und verkleinert den benötigten Speicherplatz (siehe z.B. [1] Seite 216ff).

Tabellen können als ASCII-Text-Dateien (American Standard Code for Information Interchange) von Microsoft Access ex- und importiert werden. Da mit diesem Dateiformat viele Programme umgehen können, eignet es sich hervorragend zum Datenaustausch zwischen verschiedenen Anwendungen.

#### 2.1.2 Begriffsdefinitionen

Da im weiteren viel mit in Datenbanken und Tabellen strukturierten Daten gearbeitet wird, sollen hier wichtige Begriffe definiert werden:

**Datenbank** Eine Datenbank stellt eine Sammlung von Tabellen dar. Sie ist die oberste Strukturebene der eingesandten Daten. In einer Datenbank können alle eingesandten Informationen abgelegt werden.

**Tabelle** Eine Tabelle stellt eine in Zeilen und Spalten strukturierte Sammlung von Zeilen dar. In einer Tabelle können alle Informationen zu einer Gruppe von Betrachtungseinheiten (z.B. Geräte oder Maßnahmen) abgelegt werden, die den gleichen Satz an Merkmalen (z.B. Gewicht oder Durchführungszeitpunkt) aufweisen.

**Zelle** Die Zelle ist die kleinste Struktureinheit der eingesandten Daten. Sie kann auch als Feld bezeichnet werden. In einer Zelle kann die Ausprägung (z.B. rot) eines Merkmals (z.B. Farbe) einer Betrachtungseinheit (z.B. Gerät) abgelegt werden.

**Zeile** Eine Zeile ist eine Sammlung von Zellen. Sie ist Teil einer Tabelle. In einer Zeile können Informationen zu allen Merkmalen einer Betrachtungseinheit abgelegt werden.

**Spalte** Eine Spalte ist eine Sammlung von Zellen. Sie ist Teil einer Tabelle. In einer Spalte können Informationen zu einem Merkmal aller Betrachtungseinheiten einer Tabelle abgelegt werden.

**Betrachtungseinheit** Alle in den Datenbanken abgelegten Informationen sind Betrachtungseinheiten (z.B. Gerät, Maßnahme) zugeordnet. Eine Betrachtungseinheit ist mit verschiedenen Merkmalen (z.B. Farbe, Gewicht) in der Datenbank repräsentiert. Jede Betrachtungseinheit kann einer Zeile zugeordnet werden.

**Merkmal** Merkmale (z.B. Farbe, Gewicht) sind Eigenschaften von Betrachtungseinheiten (z.B. Gerät, Maßnahme). Ein Merkmal einer Betrachtungseinheit kann in einer bestimmten Ausprägung (z.B. rot) realisiert sein. Jedem Merkmal kann man eine Spalte zuordnen. Merkmale können auch als Attribute bezeichnet werden.

**Ausprägung** Ausprägungen (z.B. rot, grün) sind individuelle Realisierung eines Merkmals (z.B. Farbe) einer bestimmten Betrachtungseinheit (z.B. Gerät). Jede Ausprägung kann man einer Zelle zuordnen. Die Menge der verschiedenen Ausprägungen eines Merkmal soll als Ausprägungssatz bezeichnet werden. Ausprägungssätze können durchaus unendliche Mengen sein (z.B. zum Merkmal Gewicht in Kilogramm).

**Index** Zur eindeutigen Identifikation von Zeilen enthalten Tabellen oft eine Spalte, die eine eindeutige Kennung enthält. Eine solche Spalte wird als Index bezeichnet.

**Relation** Zwei Indizes aus verschiedenen Tabellen können miteinander verknüpft werden. Dazu wird eine neue zweispaltige Tabelle angelegt, die die verbunden Indizes enthält. Diese Tabelle nennt man Relation. Auf diese Weise können z.B. Geräte mit Maßnahmen verbunden werden.

### 2.1.3 Codepages

Da die Ursprungsdaten dieser Arbeit hauptsächlich im Textformat vorliegen, ist es wichtig zu verstehen, wie Text im Computer verarbeitet wird.

Tabelle 2.1: Codepage ISO 8859-1 oder Latin 1. In der ersten Spalte und Zeile sind jeweils die erste und zweite Ziffer in hexadezimaler Darstellung des jeweiligen Zeichens angegeben. Die Steuerzeichen sind nicht druckbar.

	-0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-A	-B	-C	-D	-E	-F
0-	Steuerzeichen															
1-	Steuerzeichen															
2-		!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
3-	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4-	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5-	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6-	'	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7-	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
8-	Steuerzeichen															
9-	Steuerzeichen															
A-		ı	¢	£	¤	¥	¦	§	¨	©	ª	«	¬	–	®	¯
B-	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
C-	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D-	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	/O	Û	Ü	Û	Ü	Ý	Þ	ß
E-	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F-	ð	ñ	ò	ó	ô	õ	ö	÷	/o	ù	ú	û	ü	ý	þ	ÿ

Texte werden in Textdateien zeichenweise als Bytes gespeichert. Ein Byte besteht aus 8 Bit oder Ja- / Nein-Informationen. Ein Byte kann also  $2^8 = 256$  verschiedene Zustände annehmen. Die Zuordnung dieser Byte-Zustände zu den dargestellten Buchstaben nennt man Codepage.

Es gibt verschiedene genormte Codepages und damit das Problem der Vertauschung (Codepage-Konflikt) – gerade bei heterogenen Datenquellen. Einerseits kann man das Problem dadurch lösen, dass man darauf achtet, mit welcher Codepage die Textdateien erstellt wurden. Andererseits ist es auch möglich das Problem dadurch zu beseitigen, dass man sich auf die wichtigsten Zeichen beschränkt, die von fast allen Codepages gleich kodiert werden.

In der Tabelle 2.1 ist die in dieser Arbeit verwendete Codepage (ISO 8859-1 oder Latin 1) dargestellt. Diese ist eine Erweiterung des ASCII-Standards. Die dargestellten 256 Zeichen sind hexadezimal numeriert. Auf diese Numerierung wird im Laufe dieser Arbeit wieder zurückgegriffen. Die Zeichen 00-1F und 7F-9F sind nicht druckbare Steuerzeichen.

## 2.2 Methoden des Data Mining

Data Mining (DM) ist ein Sammelbegriff für unterschiedliche Methoden verschiedener Fachbereiche. Aus den vielen verschiedenen Definitionen, was Data Mining ist, hat sich bis jetzt noch keine durchgesetzt. Aus [2] (Seite 95) wird hier beispielhaft eine

Definition von Hagedorn et al. (1997, Seite 601) zitiert, der den Begriff Data Mining mit Datenmustererkennung übersetzt und anhand zweier Eigenschaften charakterisiert:

1. Gegenstand der Datenmustererkennung sind große, strukturierte Bestände numerischer, ordinal oder nominal skaliertes Daten, in denen interessante aber schwer aufzuspürende Zusammenhänge vermutet werden.
2. Das Forschungsziel der Datenmustererkennung sind allgemein verwendbare, effiziente Methoden, die autonom aus großen Datenmengen die bedeutsamsten und aussagekräftigsten Muster identifizieren und sie dem Anwender als interessantes Wissen präsentieren.

Beim Data Mining ist für eine Analyse der Daten fast immer eine Vorbereitung nötig. Die Erfahrung zeigt, dass nur 10% der Zeit in die Analyse jedoch 90% in die Datenaufbereitung einfließen [2] (Seite 21).

In [2] (Seiten 21ff und 94ff) ist das Vorgehen beim Data Mining strukturiert dargestellt. Im folgenden werden dieser Arbeitsplan dargestellt und die in dieser Arbeit angefallenen Arbeitspakete zu den entsprechenden Positionen zugeordnet. Dies ermöglicht einen guten Überblick über das konzeptionelle Vorgehen dieser Arbeit.

### 2.2.1 Aufgabenstellung und Projektplanung

- Beschreibung der Aufgabenstellung (siehe 1.3)
- Bei der Projektplanung sind insbesondere zu berücksichtigen
  - Ressourcen : Personal, Qualifikationen, Hardware, Software
  - Ergebnisanforderungen (siehe 5.1.1)
  - Beschränkungen
    - \* Datenvolumen (siehe 3.1.4)
    - \* Datenqualität (allgemeine Betrachtungen: 4.1.2, sachliche Datenfehler: 3.3.2 und 3.3.3, zeitliche Datenfehler: 4.3.4 und 4.4.4)
    - \* Rechtsvorschriften der Datennutzung (siehe 3.1.1)
  - Projektrisiken

### 2.2.2 Auswahl der relevanten Datenquellen

- Formale Eigenschaften der Daten sind zu berücksichtigen, wie
  - Datenquelle, Datengenerierungsprozess, Datenstandort und Datenzugriffsmethoden (siehe 3.1 und 3.2)
  - Dateiformat (siehe 3.1.4), Datenstruktur (siehe 3.3), Datenvolumen: Anzahl der Datensätze und Merkmale (siehe 3.3.2)
  - Merkmalsdefinitionen, Kodierung der Merkmale (siehe 3.3.3)
- Inhaltliche Eigenschaften der Daten sind zu berücksichtigen, wie
  - Vollständigkeit der relevanten Merkmale (siehe 3.3.3 und 4.1.1)

- Art und Häufigkeit von Datenfehlern und fehlenden Daten (allgemeine Betrachtungen: 4.1.2, sachliche Datenfehler: 3.3.2 und 3.3.3, zeitliche Datenfehler: 4.3.4 und 4.4.4)
- Verfügbarer Zeithorizont der Daten (siehe 4.3.4 und 4.4.4)
- Sachliche (siehe 5.1.2) und zeitliche (siehe 4.3.4 und 4.4.4) Granularität der Daten
- Heterogenität der Datenquelle (syntaktisch siehe 3.3, semantisch siehe 5.1.2)
- unformatierte Daten (siehe 1.2.2)
- Auf Datenbestände, die gerne übergangen werden, ist zu achten, wie
  - Daten in Papierform (siehe 1.2.2 und 3.1.1)
  - Ausgelagerte Daten bei externen Dienstleistern (siehe 3.7.3)
  - Daten ohne direkten Bezug
  - unformatierte Daten
  - implizites Wissen

### 2.2.3 Datenaufbereitung

- Transformation der Daten in das Standarddatenformat des Data Mining
  - Querschnittanalyse: Tabelle enthalten Angaben zu Merkmalen (Spalten) verschiedener Betrachtungseinheiten (Zeilen) ohne Informationen des zeitlichen Verlaufes. Zeitliche Verläufe können jedoch implizit dargestellt werden, indem man z.B. als Betrachtungseinheit „Maßnahme“ wählt und als Merkmale „betroffenes Gerät“ und „Durchführungszeitpunkt“ (siehe 3.3.3 und 3.4).
  - Längsschnittanalyse: Tabelle enthalten Angaben zu Merkmalen bestimmter Betrachtungseinheiten (Spalten) zu verschiedenen Zeitpunkten (Zeilen). Diese Darstellung wird nicht eingesetzt, da die Erfassung der Merkmale nicht zu festen Zeiten erfolgte.
  - Kombinierte Analyse: wird auch aus oben genanntem Grund nicht eingesetzt.
- Explorative Datenanalyse: Hiermit ist eine sondierende Untersuchung der vorliegenden Daten gemeint. Diese soll die Informationen über die Qualität und Quantität der Daten liefern, die für die Entscheidungen über das weitere Vorgehen bei der Datenaufbereitung nötig sind (siehe 3.3 und 5.1.2).
- Datenanreicherung: Damit ist ein Rückgriff auf andere Datenbestände gemeint, die in die zu untersuchenden Daten integriert werden (siehe 3.7).
- Datenreduktion
  - Stichproben
  - Aggregation: z.B. Summe, Mittelwert (siehe 4.1.3)
  - Reduktion der Dimensionalität (siehe 3.3.3 und 4.1.1)

- Behandlung fehlender Merkmalswerte (siehe 4.1.2)
- Behandlung fehlerhafter Merkmalswerte (siehe 5.1.2, insbesondere 3.5.1)
- Kodierung der Merkmale
  - Skalentransformation (siehe 2.3.1 und 3.6)
  - Transformation und Normierung (siehe 5.1.2)
  - Abgeleitete Merkmale (siehe 4.3.4: Betriebszeit und Aktivgerätezahl; ansatzweise 5.1.3)

## 2.2.4 Auswahl der Data Mining Methoden

- Für Beschreibungsprobleme stehen zur Verfügung
  - Deskription (siehe 4.2, 4.3, 4.4 und 4.5.1)
  - Assoziation (siehe 4.6.1)
  - Gruppenbildung (siehe ansatzweise 5.1.3)
- Für Prognoseprobleme stehen zur Verfügung
  - Klassifikation
  - Wirkungsprognosen
- Auswahlkriterien für Data Mining Methoden
  - Approximations- und Prognosegenauigkeit (siehe E.1)
  - Generalisierungsfähigkeit (siehe 4.1.3, 4.3.3 und 4.4.3)
  - Interpretierbarkeit, Robustheit (siehe 4.1.3)

## 2.2.5 Anwendung der Data Mining Methoden

- Vorbereitende Methoden (siehe Abschnitt 3)
- Analysierende Methoden (siehe Abschnitt 4)

## 2.2.6 Evaluation, Interpretation und Auswertung der Ergebnisse

- Filtern handlungsrelevanter Data Mining Ergebnisse nach (siehe 4.1.4 und 5.1.3)
  - Validität
  - Neuheitsgrad
  - Nützlichkeit
  - Kompaktheit / Verständlichkeit
- Betriebswirtschaftliche Bewertung der Data Mining Ergebnisse (siehe 5.1.3)
- Anwendung der Data Mining Ergebnisse

## 2.3 Statistische Methoden

In diesem Abschnitt werden kurz die Methoden der Statistik vorgestellt, die in dieser Arbeit verwendet werden. Spezielle Anwendungen der Statistik im Bereich der Instandhaltung von Geräten werden im Abschnitt 2.4.3 vorgestellt.

### 2.3.1 Grundlagen

#### Zufällige Ereignisse

Es gibt Ereignisse, die nicht vorhersagbar und daher zufällig sind: z.B. das Erhalten einer ungeraden Augenzahl nach einem Würfelwurf oder der Ausfall mindestens einer Glühlampe in einer leuchtenden Lichterkette in einer bestimmten Zeitspanne. Beiden Beispielen ist gemeinsam, dass man

1. ein Experiment durchführt (Werfen eines Würfels, Brennenlassen einer Lichterkette eine bestimmte Zeit lang),
2. am Ende genau ein Ergebnis aus einer Anzahl verschiedener möglicher Ergebnisse erhält (Augenzahl ist gerade oder ungerade, mindestens eine Lampe brennt durch oder alle bleiben heile) und
3. und alle möglichen Ergebnisse bekannt sind.

Kann man jetzt noch das Zufallsexperiment beliebig oft durchführen, so kann man nun zu jedem Ergebnis, das ab jetzt als Ereignis  $A$  bezeichnet wird, die Wahrscheinlichkeit  $p(A)$  einführen, mit der das Ereignis  $A$  eintritt. Wir schreiben:

$$p(A) = \lim_{N \rightarrow \infty} \frac{\#(A)}{N} \quad (2.1)$$

Dabei bedeutet  $\#(A)$  die Anzahl der Experimente mit Ergebnis  $A$  und  $N$  die Gesamtzahl der durchgeführten Experimente.

In der Praxis muss man sich oft mit Näherungen bzw. Schätzwerten der Wahrscheinlichkeit  $p(A)$  begnügen, da man Experimente nicht beliebig häufig wiederholen kann. Nach dem schwachen Gesetz der großen Zahl (z.B. [3] Seite 129) wird  $p(A)$  aber mit steigender Zahl an Experimenten beliebig genau approximiert. Aus der obigen Definition folgt, dass  $0 \leq p(A) \leq 1$ . Außerdem gilt für sich ausschließende Ereignisse  $A$  und  $B$ :

$$p(A \cup B) = p(A) + p(B) \quad (2.2)$$

Im folgenden wird auch noch der Begriff der „Bedingten Wahrscheinlichkeit“ benötigt:

$$p(A | B) = \frac{p(A \cap B)}{p(B)} \quad (2.3)$$

Dabei gibt dieser Wert die Wahrscheinlichkeit an, dass das Ereignis  $A$  eintritt unter der Voraussetzung, dass das Ereignis  $B$  eintritt.

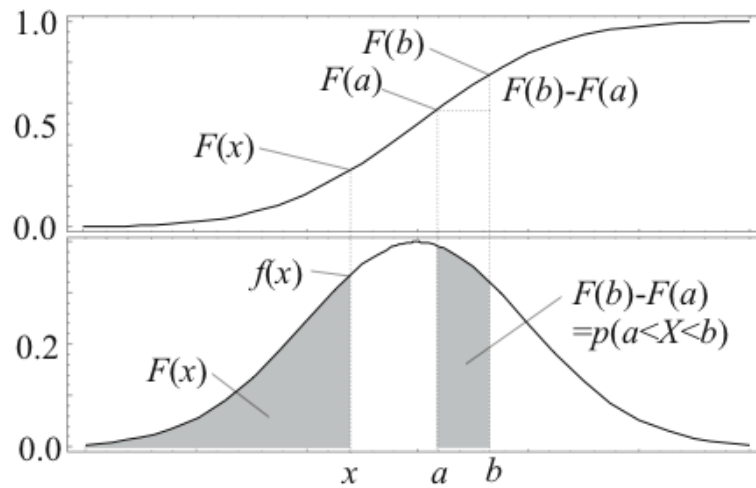


Abbildung 2.1: Oben ist die Verteilungsfunktion  $F(x)$  und unten die Wahrscheinlichkeitsdichte  $f(x)$  angegeben. Die Darstellung soll den Zusammenhang dieser beiden Funktionen verdeutlichen: Flächen unter der Wahrscheinlichkeitsdichte entsprechen Funktionswerten der Verteilungsfunktion. Mit Veränderungen entnommen aus [3] Seite 92.

### Zufällige Variable

Neben zufälligen Ereignissen werden in dieser Arbeit auch zufällige Variable untersucht. Zum Beispiel ist die Zeit zwischen zwei Ausfällen eines Gerätes eine so genannte Zufallsvariable. Die Wahrscheinlichkeit, bei einer Bestimmung der Zufallsvariablen  $X$  einen Wert  $x$  zu erhalten, für den  $a < x \leq b$  gilt, wird als  $p(a < X \leq b)$  bezeichnet.

Als reellwertig wird eine Zufallsvariable bezeichnet, wenn ihre Wertemenge reell ist. Als diskret / stetig wird sie bezeichnet, wenn ihr Definitionsbereich diskret / stetig ist. Alle im folgenden betrachteten Zufallsvariablen sind stetig. Die Zusammenhänge lassen sich aber leicht auf diskrete Zufallsvariablen übertragen.

Zu einer stetigen Zufallsvariablen  $X$  lässt sich eine Verteilungsfunktion  $F_X(a)$  angeben:

$$F_X(a) = p(X \leq a) = p(-\infty \leq X \leq a) \quad (2.4)$$

Diese Verteilungsfunktion ist monoton steigend und hat einen Wertebereich von 0 bis 1.

Wenn eine Verteilungsfunktion existiert, so kann man auch eine Wahrscheinlichkeitsdichte  $f_X(a)$  angeben. Für stetige Zufallsvariablen gilt:

$$F_X(a) = \int_{-\infty}^a f_X(t) dt \quad \text{wobei} \quad \int_{-\infty}^{\infty} f_X(t) dt = 1 \quad (2.5)$$

Mit der Wahrscheinlichkeitsdichte können wir nun schreiben:

$$p(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(t) dt \quad (2.6)$$

Die Zusammenhänge für stetige Verteilungsfunktionen und Wahrscheinlichkeitsdichten sind in Abbildung 2.1 angegeben.

## Empirische Verteilungsfunktion

Seien  $x_i$  mit  $i \in \{1 \dots n\}$  eine Gruppe bzw. Stichprobe von Werten einer diskreten Zufallsvariablen  $X$ . Die empirische Verteilungsfunktion  $\hat{F}_n(X)$  lautet dann (siehe [3] Seite 94)

$$\hat{F}_n(x) = \frac{\#(X \leq x)}{n} \quad (2.7)$$

mit der Anzahl  $\#(X \leq x)$  der Stichprobenwerte  $x_i \leq x$  sind. Von dieser Funktion wird im Folgenden noch häufig Gebrauch gemacht.

Analog zum Gesetz der großen Zahl bei der Wahrscheinlichkeit gilt hier der Satz von Glivenko und Cantelli: Die Differenz  $|\hat{F}_n(x) - F(x)|$  strebt für große  $n$  mit der Wahrscheinlichkeit 1 gegen Null (siehe [3] Seite 94), wenn  $F(x)$  die Verteilung der Zufallsvariablen  $X$  ist.

## Skalenniveau

Zufallsvariablen und deren Verteilungsfunktionen kann man sowohl stetig als auch diskret formulieren, um den mit ihnen verbunden Formalismus auf möglichst viele Anwendungsfälle übertragen zu können. Zufallsvariablen können in drei (ein stetiges und zwei diskrete) verschiedene Gruppen – so genannte Skalenniveaus – eingeteilt werden:

**Metrisch skalierte Zufallsvariablen** besitzen als Wertebereich eine Teilmenge der reellen Zahlen. Zwischen je zwei Elementen des Wertebereiches kann ein Abstand einfach als Differenz der beiden Elemente definiert werden. Beispiele sind Längen, Gewichte oder Zeiten.

**Ordinal skalierte Zufallsvariablen** besitzen einen Wertebereich, der eindeutig auf eine Teilmenge der natürlichen Zahlen abgebildet werden kann. Durch die Eindeutigkeit der Abbildung wird die Reihenfolge der natürlichen Zahlen auf den Wertebereich der Zufallsvariablen übertragen. Es gibt also eine natürliche Reihenfolge der möglichen Werte der Zufallsvariablen, ohne dass man einen Abstand bestimmen kann. Beispiele sind subjektive Klassifizierungen wie laut / mittel / leise.

**Kategorial skalierte Zufallsvariablen** besitzen weder einen Abstand noch eine natürliche Ordnung. Eigennamen sind hier ein allgegenwärtiges Beispiel.

Diese Reihenfolge der Skalenniveaus ist nicht zufällig gewählt. Durch Fortlassen der Eigenschaft einen Abstand bestimmen zu können (Metrik), erhält man im Normalfall aus einer metrischen eine ordinal skalierte Zufallsvariablen. Durch Fortlassen der Ordnung einer ordinalen Zufallsvariablen erhält man eine kategorial skalierte Zufallsvariable.

Man kann die Skalenniveaus also anhand ihrer Eigenschaften anordnen und von höheren und niedrigeren Skalenniveaus sprechen. Die Transformation von höheren zu niedrigeren Skalenniveaus und umgekehrt wird in Abschnitt 3.6 kurz vorgestellt.

## 2.3.2 Deskriptive Statistik

### Mittelwert und Standardabweichung

Zur groben Charakterisierung von Zufallsvariablen sind Mittel- bzw. Erwartungswert und Varianz bzw. Standardabweichung geeignet. Der Erwartungswert  $E(X)$  einer diskreten Verteilungsfunktion  $X$  lautet (siehe [3] Seite 95)

$$E(X) = \sum_i x_i p(X = x_i) = \sum_i x_i \frac{\#(X = x_i)}{n} \quad (2.8)$$

Die  $x_i$  bezeichnen hier die verschiedenen Werte, die von der diskreten Zufallsvariablen angenommen werden und die  $\#(X = x_i)$  die zugehörige Anzahl. Für den Erwartungswert gelten umfangreiche Rechenregeln (z.B. [3] Seiten 95 und 96).

Für die Varianz  $Var(X)$  und die Standardabweichung  $\sigma$  gilt in diesem Fall (siehe [3] Seite 95):

$$Var(X) = \sigma^2 = E(X - E(X))^2 = \sum_i [x_i - E(X)]^2 p(X = x_i) \quad (2.9)$$

### Lineare und nichtlineare Regression

Hat man die Messwerte  $x_1, x_2, \dots, x_n$  der Zufallsvariablen  $X$  und die Messwerte  $y_1, y_2, \dots, y_n$  der Zufallsvariablen  $Y$  gepaart als Messpunkte  $(x_i, y_i)$  zur Verfügung, so besteht oft der Wunsch bei geeigneter Verteilung der Messpunkte eine Gerade  $y = bx + a$  durch diese Messpunkte zu legen, die die Darstellung des Zusammenhangs möglichst gut an die Messwerte angepasst und in einfacher Form darstellt. Die einfache Darstellung ist durch die unkomplizierte Geradengleichung gegeben. Für die optimale Anpassung dieser Geraden an die Messwerte fordert man folgende Minimalbedingung (siehe [5] Seiten 1028 und 1029):

$$E(Y - (bX + a))^2 = \min! \quad (2.10)$$

Damit ergibt sich folgende Regressionsgerade als Lösung:

$$y = E(Y) + r \frac{\Delta X}{\Delta Y} (x - E(X)) \quad \text{mit} \quad (2.11)$$

$$\Delta X = \frac{1}{n} \sum_{i=1}^n x_i \quad , \quad \Delta Y = \frac{1}{n} \sum_{i=1}^n Y_i \quad \text{und} \quad (2.12)$$

$$r = \frac{1}{(n-1)\Delta X \Delta Y} \sum_{i=1}^n (x_i - E(X))(y_i - E(Y)) \quad (2.13)$$

Oft ist es nicht sinnvoll, den Zusammenhang zweier Zufallsvariablen  $X$  und  $Y$  durch eine Gerade zu beschreiben, da sich dann die oben geforderte Minimalbedingung nicht gut befriedigen lässt. In einem solchen Fall kann es hilfreich sein, eine Regression der Abbildungen  $f_1(X)$  und  $f_2(Y)$  vorzunehmen, wobei diese Funktionen durch Ausprobieren oder externe Informationen ergeben.

Führt auch die Transformation der Messpunkte zu keinem ausreichend linearem Verlauf, oder ist eine Transformation aus anderen Gründen nicht erwünscht, so kann man auf die nichtlineare Regression zurückgreifen. Die in dieser Arbeit vorgenommen nicht-linearen Regressionen wurden mit der Software Mathematica 5.0 vorgenommen.

### Beispiel: log-Normalverteilung

Im Laufe der Arbeit werden einige Regressionen mit der log-Normalverteilung durchgeführt. Um die dabei gewonnenen Ergebnisse besser interpretieren zu können, sei hier die log-Normalverteilung etwas genauer untersucht.

Für Wahrscheinlichkeitsdichte  $f(x)$  der log-Normalverteilung gilt

$$f(x) = \frac{1}{\sqrt{2\pi} \beta_L} \frac{1}{x} e^{-\frac{(\log(x) - \alpha_L)^2}{2\beta_L^2}} \quad (2.14)$$

Dabei wird der Parameter  $\alpha_L$  Häufungspunkt und den Parameter  $\beta_L$  Streuung der log-Normalverteilung genannt. Die einzigen Unterschiede der log-Normalverteilung zur Normalverteilung sind der veränderte Normierungsfaktor und die Transformation der Argumente durch die log-Funktion.

In Abbildung 2.2 sind die Verteilungsfunktion  $F$  und die Wahrscheinlichkeitsdichte  $f$  der log-Normalverteilung dargestellt. Zusätzlich sind auch Häufungspunkt  $h$ , Median  $m$  und Mittelwert  $\mu$  eingezeichnet. Es gelten

$$h = \alpha_L \quad , \quad m = e^{\alpha_L} \quad \text{und} \quad \mu = e^{\alpha_L + \frac{\beta_L^2}{2}} \quad (2.15)$$

Durch die Transformation der Argumente fallen Häufungspunkt, Median und Mittelwert der log-Normalverteilung nicht wie bei der Normalverteilung zusammen. In diesem Fall spricht man auch von einer Schiefe (skewness) der Verteilung. Der Mittelwert der log-Normalverteilung hängt also im Gegensatz zur Normalverteilung nicht nur vom Häufungspunkt sondern auch von der Streuung ab.

Durch die Definition der Schiefe  $\beta_1$  im Anhang E.1 kann man diese quantifizieren. Für die log-Normalverteilung ergibt sich damit

$$\beta_1 = \sqrt{-1 + e^{\beta_L^2}} \left( 2 + e^{\beta_L^2} \right) \quad (2.16)$$

Da die Schiefe  $\beta_1$  monoton vom Parameter  $\beta_L$  abhängt, kann man aus Gründen der Einfachheit zum Vergleich der Schiefen zweier log-Normalverteilungen den Parameter  $\beta_L$  heranziehen.

### 2.3.3 Prüfende Statistik

Im Normalfall sollte eine statistisch testende Analyse folgendermaßen ablaufen:

1. Die Fragestellung wird in eine prüfbare Hypothese umformuliert.
2. Man überlegt sich, mit welchen Verfahren die Hypothese geprüft wird.
3. Die Daten werden nach den Erfordernissen des ausgewählten Verfahrens erhoben.
4. Das Testverfahren wird durchgeführt und das Ergebnis liegt vor.

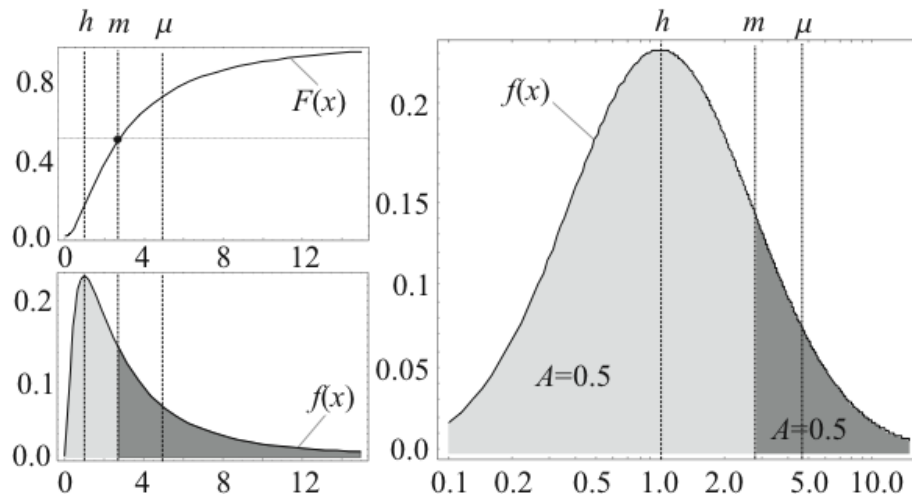


Abbildung 2.2: Verteilungsfunktion  $F(x)$  und Wahrscheinlichkeitsdichte  $f(x)$  der log-Normalverteilung. Zusätzlich sind Häufungspunkt  $h$ , Median  $m$  und Mittelwert  $\mu$  gekennzeichnet. Die Fläche unter der Wahrscheinlichkeitsdichte ist 1. Sie wird durch den Median in 2 gleiche Flächen geteilt. Die Verteilungsfunktion besitzt an der Stelle des Medians den Wert 0.5. Durch die logarithmische Skalierung der rechten Abbildung ergibt sich ein symmetrischer Verlauf wie bei der Normalverteilung. Die Parameter der log-Normalverteilung lauten  $\alpha_L = \beta_L = 1$ .

Das Vorgehen bei dieser Untersuchung unterscheidet sich zwangsläufig stark von dem oben genannten Vorgehen:

- Die Daten liegen bereits vor und sind auf unbekannte Weise und mit unbekannter Qualität erhoben worden.
- Sie sind nicht in der Absicht erhoben worden, mit einem bestimmten Verfahren ausgewertet zu werden.
- Das Formulieren der Hypothesen vor der Auswahl des Testverfahrens ist ineffektiv, da die meisten Hypothesen in Gegenwart des vorliegenden Datenmaterials nicht überprüfbar sind.

Es können also verschiedene Verfahren nur auf die schon erhobenen Daten angewendet und dann entschieden werden, ob die gewonnenen Erkenntnisse von Interesse sind – also eine Umkehrung des normalen Vorgehens. Aufgrund dieser Randbedingungen ist es nicht unkritisch, testende Statistik zu betreiben – man muss also hauptsächlich mit deskriptiver Statistik auskommen.

In einigen Fällen wird jedoch auch in dieser Arbeit testenden Statistik verwendet. Zur Beurteilung der Regressionen an eine willkürlich ausgewählte log-Normalverteilung wird der Kolmogoroff-Smirnow-Test für die Güte von Anpassungen verwendet. Für die Untersuchung von Mehrfeldtafeln bzw. Kreuztabellen wird der  $\chi^2$ -Test und der Kontingenzkoeffizient von Pawlik verwendet. Korrelationen werden mit dem Spearmanschen Rang-Korrelationskoeffizienten untersucht.

## Grundlagen

Aus den vorliegenden Daten  $X$ , die als Vektor oder auch als Matrix vorliegen können, wird eine Testgröße  $T(X)$  berechnet. Anhand dieser Testgröße kann man entscheiden, ob die Nullhypothese  $H_0$  angenommen oder verworfen wird und damit die Alternativhypothese  $H_A$  angenommen wird. Testgröße, Nullhypothese und Alternativhypothese werden durch den gewählten Test festgelegt.

Bei der Entscheidung zwischen Null- und Alternativhypothese anhand der ermittelten Testgröße kann der Test zwei prinzipielle Fehler machen. Die Nullhypothese kann abgelehnt werden, obwohl sie richtig ist – diesen Fehler nennt man Fehler 1. Art oder  $\alpha$ -Fehler. Wird die Nullhypothese jedoch nicht abgelehnt, obwohl die Alternativhypothese vorliegt, so spricht man vom Fehler 2. Art bzw.  $\beta$ -Fehler.

Es hat sich eingebürgert, Entscheidungen zwischen Null- und Alternativhypothese als signifikant zu bezeichnen, wenn der  $\alpha$ -Fehler kleiner als 5% ist. Der  $\alpha$ -Fehler kann durch Wahl des kritischen Wertes festgelegt werden. Der kritische Wert ist dabei die Grenze zwischen Werten der Testgröße, die die Nullhypothese bestätigen bzw. widerlegen. Eine zu starke Verkleinerung des  $\alpha$ -Fehlers führt im Allgemeinen zu einer unerwünschten Erhöhung des  $\beta$ -Fehlers.

## Anpassungstest von Kolmogoroff-Smirnow

Bei diesem Test soll untersucht werden, ob eine empirische Verteilungsfunktion  $F_E$  (siehe Abschnitt 2.3.1) durch eine Verteilungsfunktion  $F_0$  dargestellt werden kann.

Die Nullhypothese  $H_0$  lautet  $F_E = F_0$ , die Alternativhypothese  $H_A$  lautet damit  $F_E \neq F_0$ . Die Testgröße  $T$  lautet

$$T = \sqrt{n} \max(|F_0 - F_E|) \quad (2.17)$$

wobei  $n$  die Anzahl der Werte der empirischen Verteilungsfunktion  $F_E$  ist.

Die Nullhypothese  $H_0$  wird mit dem  $\alpha$ -Fehler abgelehnt, wenn  $T \geq k_{n;\alpha}$  gilt. Die kritischen Werte  $k_{n;\alpha}$  können aus Tabellen entnommen werden (z.B. [4] Seite 184).

Anschaulich bedeutet dies, dass die empirische Verteilungsfunktion  $F_E$  dann durch eine andere Verteilungsfunktion  $F_0$  angenähert werden kann, wenn  $F_E$  sich innerhalb eines Gebietes der Breite  $k_{n;\alpha}/\sqrt{n}$  um  $F_0$  befindet. In den Abbildungen 4.13 und 4.32 ist dieses Gebiet zur Beurteilung der Regression eingezeichnet.

## $\chi^2$ -Test für Mehrfeldtafeln

Mehrfeldtafeln sind matrizenförmig angeordnete Häufigkeiten. Solche Tabellen werden auch als Kreuz- oder Kontingenztabellen bezeichnet. Die einzelnen Werte der  $r \times c$ -Matrix werden mit  $n_{ij}$  bezeichnet.

Dieser Test untersucht, ob die  $n_{ij}$  gleich verteilt, d.h. die  $n_{ij}$  weder von  $i$  noch von  $j$  abhängen. Dazu wird folgende Testgröße berechnet:

$$T = -n + n \sum_{i=1}^r \sum_{j=1}^c \frac{n_{ij}^2}{n_{i.} \cdot n_{.j}} \quad (2.18)$$

Dabei ist

$$n_{i.} = \sum_{j=1}^r n_{ij} \quad , \quad n_{.j} = \sum_{i=1}^c n_{ij} \quad \text{sowie} \quad n = \sum_{i=1}^r \sum_{j=1}^c n_{ij} \quad (2.19)$$

Die Nullhypothese  $H_0$  besagt, dass die  $n_{ij}$  stochastisch unabhängig sind. Die Nullhypothese wird mit dem  $\alpha$ -Fehler abgelehnt, wenn  $T \geq k_{f;\alpha}$ .  $k_{f;\alpha}$  ist dabei der kritische Wert, der von  $\alpha$ -Fehler und Anzahl der Freiheitsgrade der Matrix abhängt. Die Anzahl der Freiheitsgrade lautet  $f = (r - 1)(c - 1)$ , wenn die Randsummen  $n_{i.}$  und  $n_{.j}$  vorgegeben sind.

Die kritischen Werte  $k_{f;\alpha}$  sind hier genau die  $\alpha$ -Quantile der  $\chi^2$ -Verteilung für  $f$  Freiheitsgrade. Diese Werte sind leicht verfügbar (z.B. tabelliert in [3] oder Berechnung mit Mathematica).

### Kontingenzkoeffizient von Pawlik

Mit Hilfe des Kontingenzkoeffizienten von Pawlik soll die Stärke einer Korrelation innerhalb einer Kreuztabelle untersucht werden. Zum Vergleich verschiedener Kreuztabellen ist die im vorherigen Abschnitt vorgestellte Testgröße  $T$  nicht geeignet, da sie proportional zu  $n$  ist. Diese Abhängigkeit weist der folgende Kontingenzkoeffizient  $C$  von Pawlik nicht auf (siehe [3] Seite 601ff):

$$C = \sqrt{\frac{T}{n + T}} \quad (2.20)$$

Bei völliger Unabhängigkeit ist  $C = 0$ . Für den maximalen Kontingenzkoeffizienten gilt

$$C_{max} = \sqrt{\frac{r}{n + r}} \quad \text{wobei bitte} \quad r > c \quad (2.21)$$

Um die Kontingenzkoeffizienten von den Kreuztabellen unabhängig und damit vergleichbar zu machen, wird der korrigierte Kontingenzkoeffizient  $C_{korr}$  eingeführt. Es gilt

$$C_{korr} = \frac{C}{C_{max}} \quad (2.22)$$

### Der Spearmansche Rang-Korrelationskoeffizient

Soll eine Stichprobe gepaarter Werte  $(x_i, y_i)$  zweier Zufallsvariablen  $X$  und  $Y$  auf einen Zusammenhang von  $X$  und  $Y$  hin untersucht werden, so kann man auf Korrelationsmaße zurückgreifen. Der normale Korrelationskoeffizient kann jedoch nur lineare Zusammenhänge nachweisen und ist auf normalverteilte Zufallsvariablen  $X$  und  $Y$  angewiesen (siehe [3] Seite 495).

Der Spearmansche Rang-Korrelationskoeffizient kann dagegen bei Zufallsvariablen mit unbekannter Verteilung auch nichtlineare Zusammenhänge nachweisen. Er ist damit wesentlich vielseitiger einsetzbar (siehe [3] Seite 511f).

Bei Rang-Korrelationskoeffizienten werden nicht die  $x_i$  und  $y_i$  direkt untersucht sondern nur ihre Ränge. Der Rang  $d_{x_i}$  eines Wertes  $x_i$  ist einfach die Position des Wertes in der sortierten Liste aller  $x_1, x_2, \dots, x_{n-1}, x_n$ . Kommen einige  $x_i$  mehrfach vor, so spricht man von Bindungen. Der Rang von gebundenen  $x_i$  ist der Mittelwert ihrer Positionen.

Die Bestimmung der Rangzahlen für eine metrische Zufallsvariable kommt einer Skalentransformation einer metrischen Zufallsvariablen in eine ordinal skalierte Zufallsvariable gleich (siehe Abschnitt 2.3.1). Damit ist auch klar, warum mit diesem Korrelationskoeffizienten auch nichtlineare Zusammenhänge nachgewiesen werden können.

Der Spearmansche Rang-Korrelationskoeffizient berechnet sich für  $n \geq 6$  als

$$r_S = 1 - \frac{6 \sum_{i=1}^n (d_{x_i} - d_{y_i})^2}{n(n^2 - 1)}$$

Für Zufallsvariablen mit vielen Bindungen (mehr als 20 % der Beobachtungen) steht noch ein korrigierter Koeffizient zur Verfügung (siehe [3] Seite 513), der hier aber nicht benötigt wird.

Um nun zwischen der Nullhypothese  $H_0$  – zwischen  $X$  und  $Y$  besteht kein Zusammenhang – oder der Alternativhypothese  $H_A$  – es besteht ein Zusammenhang – zu entscheiden, wird die Testgröße  $T = r_S$  herangezogen. Die Nullhypothese wird auf dem Signifikanzniveau  $\alpha$  abgelehnt, wenn  $T > k_{\alpha;n}$  gilt. Dabei können die kritischen Werte  $k_{\alpha;n}$  aus [3] (Seite 511) entnommen werden.

### 2.3.4 Assoziationsanalyse

Ein Ereignis  $A$  und ein Ereignis  $B$  können gemeinsam auftreten. Tun sie dies, so spricht man von einer Koinzidenz. Ist diese Koinzidenz nicht zufällig, so spricht man von einer Korrelation. Eine Korrelation kann aus verschiedenen Gründen vorliegen: zum einen kann das Ereignis  $A$  das Ereignis  $B$  hervorrufen oder umgekehrt. Es besteht auch die Möglichkeit, dass  $A$  und  $B$  eine gemeinsame Ursache haben.

Die Assoziationsanalyse versucht nun die Korrelation von Koinzidenzen zu bestimmen, d.h. sie versucht zu bestimmen, ob zwei Ereignisse zufällig gemeinsam aufgetreten sind oder ob das eine Ereignis das andere hervorgerufen oder beide Ereignisse eine gemeinsame Ursache haben. Beispielsweise könnten Ereignis  $A$  „das Gerät gehört zur Gerätegruppe der Infusionspumpen“ und Ereignis  $B$  „das Gerät ist günstiger als 5000 €“ lauten. Dann könnte man jeweils die Anzahl der Geräte, für die Ereignisse  $A$ ,  $B$  und beide Ereignisse  $A$  und  $B$  eintreten sowie die Anzahl aller Geräte bestimmen. Aus diesen Anzahlen lassen sich zur Beurteilung der Korrelation der Ereignisse  $A$  und  $B$  verschiedene Assoziationsmaße berechnen, die die Assoziationsanalyse bereit stellt. Im Folgenden sind einige Assoziationsmaße aufgeführt, die in dieser Arbeit verwendet wurden. Eine ausführliche Darstellung findet sich in [2] ab Seite 427.

#### support

Er gibt an, wie häufig ein Ereignis im Verhältnis zu allen Ereignissen auftritt. Insbesondere ist der support des Ereignissen  $A$  und  $B$  treten gemeinsam auf für die Analysen wichtig. Wir schreiben

$$\text{support}(A \cap B) = p(A \cap B) = \frac{\#(A \cap B)}{\#(D)} \quad (2.23)$$

Dabei sind mit  $p(A \cap B)$  die Wahrscheinlichkeit des gleichzeitigen Eintretens von  $A$  und  $B$  bzw. die Anzahl der Koinzidenzen von  $A$  und  $B$  gemeint und mit  $\#(D)$  die Anzahl

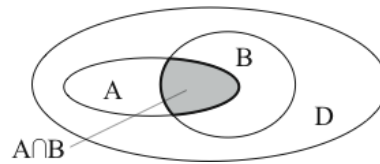


Abbildung 2.3: Zur Veranschaulichung sind beispielhaft die Mengen  $A$ ,  $B$ ,  $A \cap B$  und  $D$  dargestellt.

aller Ereignisse.

Dieses Maß ist zwar noch kein Assoziationsmaß, aber es gibt an, wie viele Ereignisse von einer noch eventuell zu entdeckenden Assoziation überhaupt betroffen sind, und ist damit für die Analysen nicht uninteressant. Außerdem wird dieses Maß als Baustein für die folgenden Maße wieder verwendet. Die Ereignismengen sind zur Veranschaulichung in Abbildung 2.3 dargestellt.

### confidence

Sie ist nichts anderes als die bedingte Wahrscheinlichkeit des Auftretens des Ereignisses  $B$  unter der Voraussetzung des Eintretens des Ereignisses  $A$ . Wir schreiben

$$\text{confidence}(A, B) = \frac{\text{support}(A \cap B)}{\text{support}(A)} = \frac{\#(A \cap B)}{\#(A)} = p(B | A) \quad (2.24)$$

In ungünstigen Fällen, kann die  $\text{confidence}(A, B)$  hohe Werte annehmen, obwohl in der Ereignismenge kein Zusammenhang zwischen den Ereignissen  $A$  und  $B$  besteht. Dies ist dann der Fall, wenn  $\text{support}(B)$  groß ist - dann ist auch  $\text{confidence}(A, B)$  bei nicht vorhandener Korrelation der Ereignisse  $A$  und  $B$  groß. Diesen Nachteil versucht der lift zu umgehen.

### lift

Dieser ergibt sich als Quotient von  $\text{confidence}(A, B)$  und  $\text{support}(B)$ . Wir schreiben

$$\text{lift}(A, B) = \frac{\text{confidence}(A, B)}{\text{support}(B)} = \frac{\text{support}(A \cap B)}{\text{support}(A) \text{support}(B)} \quad (2.25)$$

Der lift gibt damit das Verhältnis von beobachteten Koinzidenzen  $\text{support}(A \cap B)$  und von den bei Gleichverteilung zu erwartenden Koinzidenzen als Produkt von  $\text{support}(A)$  und  $\text{support}(B)$ . Also ist der lift ein Maß dafür wie viel eine beobachtete Koinzidenz häufiger bzw. seltener ist, als sie bei Gleichverteilung zu erwarten wäre. Er ist damit weniger ein direktes Maß für die Abhängigkeit zwischen  $A$  und  $B$ , als vielmehr ein Maß für die Abweichung der Unabhängigkeit.

Leider lassen sich für den lift auch Fälle konstruieren, an denen er als Assoziationsmaß versagt: wenn  $\text{confidence}(A, B)$  und  $\text{support}(B)$  etwa gleich groß sind, dann ergibt sich auch bei hoher confidence ein lift von etwa eins – was trotz der hohen confidence hier auf einen geringen Zusammenhang hindeutet.

Man könnte noch ein weiteres Assoziationsmaß einführen – die conviction –, das diesen Nachteil nicht besitzt. Jedoch besitzt diese Maß eine nicht mehr so anschauliche Interpretation wie die des liftes und wird deshalb in dieser Arbeit nicht verwendet.

### 2.3.5 Markowsche Ketten

Betrachten wir ein System, das sich zu diskreten Zeitpunkten  $t_n$  beliebig in einem der Zustände  $Z_i$  befinden kann. Die Zustände, in denen sich das System zum Zeitpunkt  $t_n$  befindet, werden mit  $Z_n$  bezeichnet.

Nun kann man eine Übergangswahrscheinlichkeit  $p_{ij}$  als bedingte Wahrscheinlichkeit einführen, dass auf dem Zustand  $Z_i$  der Zustand  $Z_j$  folgt. Diese Wahrscheinlichkeiten lassen sich übersichtlich in einer Übergangsmatrix  $p_{ij}$  darstellen.

Hängen die  $p_{ij}$  nicht von der Zeit ab (Homogenität der Zeit), so kann man die  $Z_n$  als markowsche Kette auffassen (siehe [5] Seite 1084).

### 2.3.6 Quantitative Linguistik

#### Grundlagen

Erst seit gut 50 Jahren wird Sprache nicht nur qualitativ sondern auch mit quantitativen Methoden untersucht – in der Linguistik kann nun also wie bei anderen Wissenschaften auch „gemessen“ werden. Gemessen werden dabei hauptsächlich Häufigkeiten von sprachlichen Objekten wie Wörtern und Buchstaben. Aus diesen Daten wird dann mit Hilfe statistischer Methoden versucht, neue Erkenntnisse zu gewinnen oder bekannte Vermutungen zu begründen. Als Geburtsstunde dieser Entwicklung wird das Zipf'sche Gesetz angesehen.

#### Zipf's Gesetz

Das Zipf'sche Gesetz beschreibt in sehr einfacher Weise die Häufigkeitsverteilung von Worten in Texten. Dazu formuliert das Gesetz einen Zusammenhang zwischen der Häufigkeit  $P_i$  eines Wortes in einem (ausreichend langen) Text und seines nach Häufigkeit sortierten Ranges  $i$  als (siehe [6])

$$P_i = \frac{c}{i^a} \quad (2.26)$$

Dabei passt der Parameter  $c$  die Verteilung an verschiedene Textumfänge an und der Formparameter  $a$  bestimmt die Form der Verteilung. Für gewöhnliche Texte liegt  $a$  meist etwas über 1 und kann näherungsweise ganz entfallen.

Durch Logarithmieren erhält man mit  $\log(P_i) = \log(c) - a \log(i)$  eine einfach zu untersuchende Geradengleichung.

Trotz seiner einfachen Form ist das Gesetz erstaunlich gut in der Lage, Häufigkeitsverteilungen verschiedenster Objekte zu beschreiben. Der Grund dafür ist bis heute nicht verstanden. Da es sich beim Zipf'schen Gesetz um ein empirisches Gesetz handelt, gilt es nicht exakt. Oft liefert aber auch die Abweichung einer Verteilung vom Zipf'schen Gesetz wertvolle Informationen.

Die dem Zipf'schen Gesetz zugrunde liegende Verteilungsfunktion wird als Zeta-Verteilung bezeichnet.

#### Wortentropie

Als mächtiges Werkzeug bei der quantitativen Analyse von Texten erweist sich die Entropie. Die Entropie kann als Maß der Unordnung einer Verteilung angesehen wer-

den. Für die oben genannten Worthäufigkeiten  $P_i$  ergibt sich die Entropie  $h$  mit  $p_i = P_i / \sum_i P_i$  zu

$$h = - \sum_i p_i \log(p_i) \quad (2.27)$$

Bemerkenswert ist, dass die Entropie die Unordnung als Informationsgehalt pro betrachtetem Element angibt – also hier pro Wort. Benutzt man zur Berechnung der Entropie den Logarithmus zur Basis 2, so gibt die Entropie den Informationsgehalt pro Wort in Bit an. Je größer also die Unordnung in einem Text ist, desto mehr Information wird pro Wort übertragen.

Ist der absolute Wert der Entropie nicht von Interesse, so kann man eine relative Entropie  $h_R$  einführen. Es gilt

$$h_R = \frac{h}{h_{max}} \quad (2.28)$$

Dabei ist  $h_{max}$  die maximale Entropie, die die betrachtete Wortgruppe annehmen kann. Man kann sich leicht überlegen, dass die Entropie bei Gleichverteilung maximal wird, d.h. alle  $p_i$  sind gleich groß, und es gilt  $p_i = p = 1/N$ , wenn  $N$  die Anzahl der verschiedenen Wörter ist. Damit wird

$$h_{max} = - \sum_i^N \frac{1}{N} \log\left(\frac{1}{N}\right) = \log(N) \quad (2.29)$$

## 2.4 Grundlagen der Instandhaltung

### 2.4.1 Begriffs- und Aufgabendefinition

Die DIN 31051 unterteilt die Aufgaben der Instandhaltung in drei Bereiche [7] (Seite 22ff):

1. Die Vorsorge dafür, dass der Abbau des Nutzungsvorrates während der nutzbaren Lebenszeit durch geeignete Maßnahmen so gering wie möglich gehalten wird, d.h. Wartung, Maßnahmen zur Bewahrung des Sollzustandes.
2. Erkennen, wie und warum der Abbau fortschreitet, d.h. Inspektion, Maßnahmen zur Feststellung und Beurteilung des Istzustandes.
3. Eingetretenen Abbau wieder ausgleichen, den Nutzungsvorrat wieder auffüllen, d.h. Instandsetzung, Maßnahme zur Wiederherstellung des Sollzustandes.

In der Abbildung 2.4 sind die oben genannten Begriffe an einem Ausschnitt eines Gerätelebenslaufes dargestellt. Bei der Inbetriebnahme eines neuen Gerätes beträgt der Nutzungsvorrat 100% – das Gerät befindet sich im Sollzustand. Während der Nutzung des Gerätes sinkt der Nutzungsvorrat. Dieser kann durch Inspektionen ermittelt werden. Durch Wartungen kann die Abnahme des Nutzungsvorrates vermindert werden. Ein idealisierter Verlauf des Nutzungsvorrates ohne Wartung ist durch die dünne Linie angedeutet. Unterschreitet der Nutzungsvorrat die Schadensgrenze, so ist eine volle

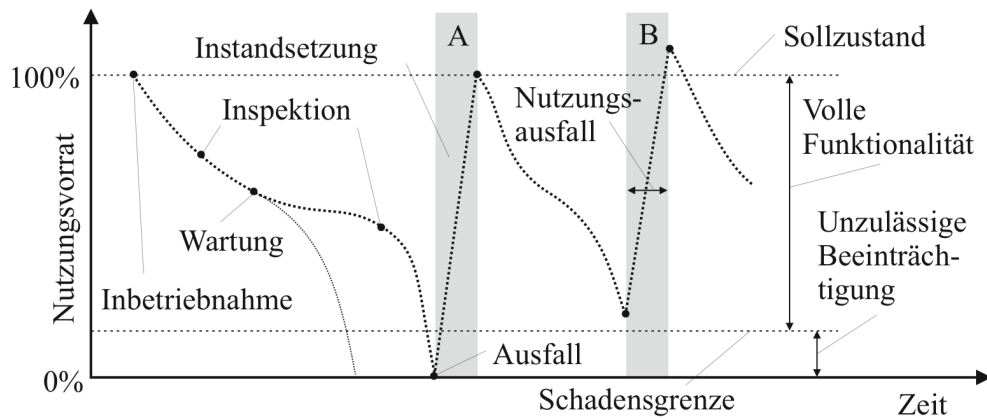


Abbildung 2.4: Beispielhafter Verlauf des Nutzungsvorrates über der Zeit. Mit Änderungen entnommen aus [7] (Seite 23)

Funktionalität des Gerätes nicht mehr gewährleistet – ab jetzt muss mit unzulässigen Beeinträchtigungen gerechnet werden, bis schließlich der Ausfall des Gerätes eintritt. Durch eine Instandsetzung kann ein ausgefallenes Gerät wieder in den Sollzustand versetzt werden (A). Durch geeignete Maßnahmen kann der Nutzungsvorrat auch auf über 100% erhöht werden (B). Instandsetzungen können selbstverständlich auch vor Ausfall eines Gerätes erfolgen. Der zeitliche Verlauf des Nutzungsvorrates muß nicht, wie in der Darstellung angegeben, stetig verlaufen.

## 2.4.2 Kenngrößen der Instandhaltung

Bei der EDV gestützten Instandhaltung fallen viele Informationen an. Um diese Informationen für nötige Entscheidungen in Hinblick auf eine wirtschaftliche Bewirtschaftung nutzbar zu machen, müssen die oft verstreuten Informationen in brauchbarer Form zusammengefasst werden. Dies sollen Kenngrößen leisten. Die wichtigsten, etablierten Kenngrößen lauten [8]:

**Investitionskosten / Reinvestitionskosten** §2 Punkt 2 Krankenhausfinanzierungsgesetz definiert:

- Investitionskosten: Kosten der Errichtung (Neubau, Umbau, Erweiterungsbau) von Krankenhäusern und der Anschaffung der zum Krankenhaus gehörenden Wirtschaftsgüter, ausgenommen der zum Verbrauch bestimmten Güter (Verbrauchsgüter)
- Reinvestitionskosten: Kosten der Wiederbeschaffung der Güter des zum Krankenhaus gehörenden Anlagevermögens (Anlagegüter)

### Instandhaltungskosten

- Lohnkosten (intern)
- Sachkosten (Ersatzteile)
- Externe Kosten (Kundendienst)

## Instandhaltungsquoten

- Jahresinstandhaltungsquote: Definiert als Verhältnis von Instandhaltungskosten pro Jahr und Anschaffungspreis ([9] Seite 297).
- Gesamteinstandhaltungsquote: Definiert als Verhältnis von Summe der Instandhaltungskosten zu Summe der Reinvestitionskosten. Diese Größe liegt normalerweise zwischen 4% und 7% [8].

**Eigenservicequote:** Definiert als Verhältnis der von eigenen Personal aufgewendeten Arbeitszeit zu der insgesamt aufgewendeten Arbeitszeit für die Instandhaltung (sowohl intern als auch extern). In [8] wird eine optimale Instandhaltungsquote von 65% angegeben.

## Zeitliche Kriterien

- Abschreibungszeitpunkt: Dieser spielt jedoch keine große Rolle, da medizintechnische Geräte nach Ablauf der Abschreibungszeit oft weiter genutzt werden.
- Wartungs- / Instandhaltungsquote: Definiert als Anteil der Wartungskosten an den Gesamtkosten (also Wartungs- und Instandsetzungskosten). In [10] wird als optimales Verhältnis 50% angegeben.

### 2.4.3 Statistik der Instandhaltung

#### Ausfallrate

Zum Zeitpunkt  $t_1$  mögen sich  $N$  Geräte in Betrieb befinden. Zu diskreten Zeitpunkten  $i\Delta t$  wird festgestellt, wie viele Geräte seit der letzten Untersuchung zum Zeitpunkt  $(i-1)\Delta t$  ausgefallen sind. Die Zahl nennen wir  $f_i$ . Dann können wir den Anteil  $F_i$  der bis jetzt ausgefallenen und den Anteil der noch laufenden Geräte  $R_i$  an allen  $N$  Geräten bestimmen. Dabei gilt:

$$F_i = \frac{1}{N} \sum_{j=1}^i f_j \quad \text{und} \quad R_i = 1 - F_i \quad (2.30)$$

nun kann die Ausfallrate  $z_i$  bestimmt werden:

$$z_i = - \frac{(R_i - R_{i-1})/\Delta t}{R_i} \quad (2.31)$$

Für den Divisor  $R_i$  wäre  $\frac{R_i + R_{i-1}}{2}$  zwar korrekter, da auch die numerische Ableitung in der Mitte des Zeitintervalls  $[(i-1)\Delta t, i\Delta t]$  bestimmt wird, aber für kleine  $\Delta t$  ist der dadurch zu erwartende Fehler vertretbar. Die Ausfallrate gibt die Wahrscheinlichkeit an, mit der ein Gerät bis zur nächsten Untersuchung ausgefallen ist. Dies ist eine wichtige Information für die Gerätebewirtschaftung. Durch die Funktion ([7] Seite 129)

$$F(t) = p(T \leq t) \quad (2.32)$$

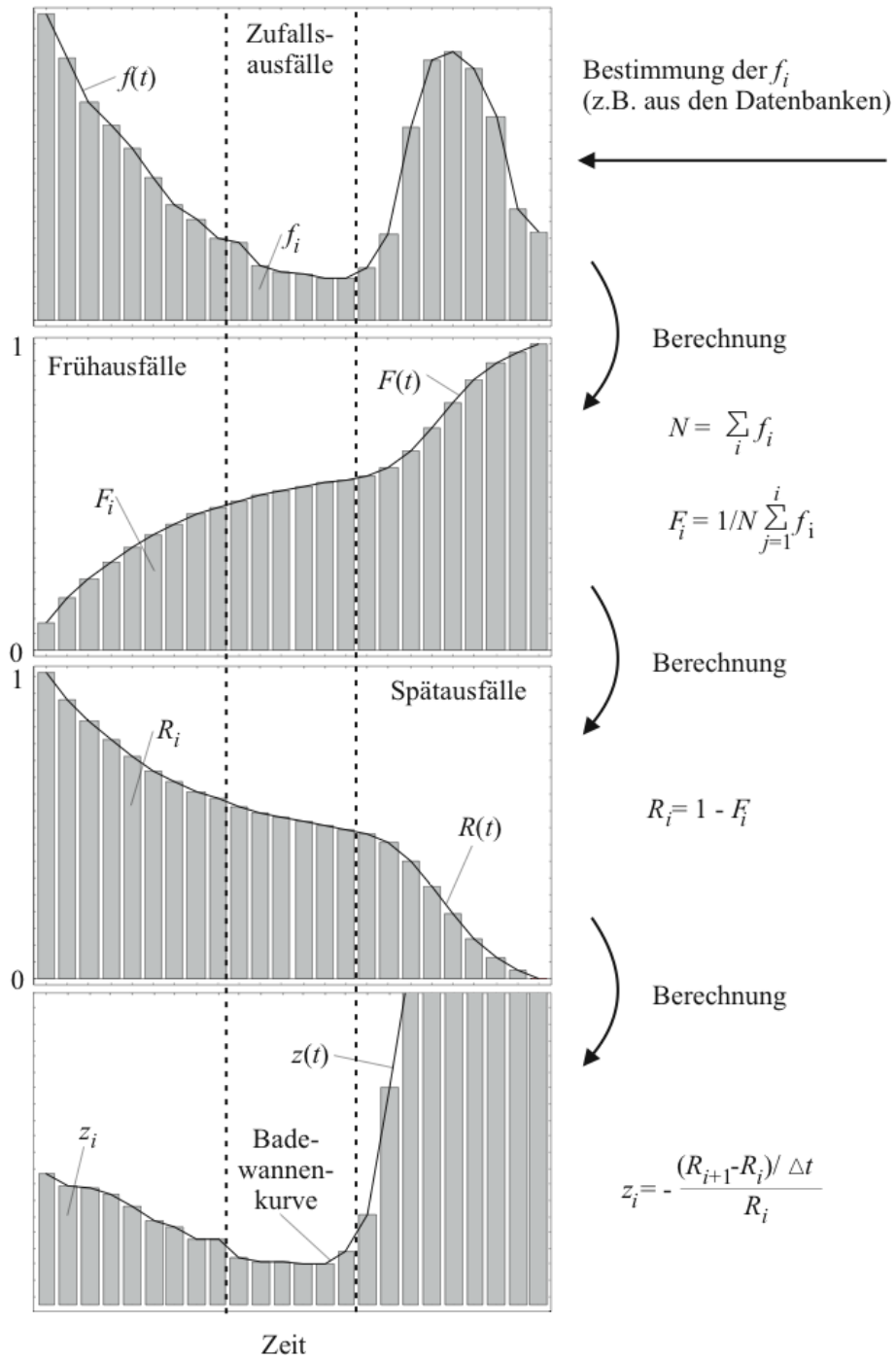


Abbildung 2.5: Graphische Veranschaulichung der Bestimmung der Ausfallrate  $z_i$  bzw.  $z(t)$ . Erläuterungen siehe Text.

erhalten wir eine stetige Verteilungsfunktion der reellwertigen Zufallsvariablen  $F$ . Hierbei ist  $p(T \leq t)$  die Wahrscheinlichkeit, dass der Ausfallzeitpunkt  $T$  vor dem Betrachtungszeitpunkt  $t$  liegt. Damit ergibt sich

$$f(t) = \frac{dF(t)}{dt} \quad (2.33)$$

und analog zu den oben eingeführten diskreten Größen kann man schreiben:

$$R(t) = 1 - F(t) \quad \text{und} \quad (2.34)$$

$$z(t) = -\frac{dR(t)/dt}{R(t)} \quad (2.35)$$

Die vorgestellten diskreten und stetigen Zusammenhänge sind zur Veranschaulichung in Abbildung 2.5 dargestellt. In der obersten Grafik ist die diskrete Anzahl  $f_i$  bzw. kontinuierliche Anzahl  $f(t)$  der ausgefallenen Geräte pro Zeit aufgetragen. In der zweiten Grafik von oben ist die kumulierte Anteil  $F_i$  bzw.  $F(t)$  der ausgefallenen Geräte angegeben. Die oberste Verlauf ergibt sich somit als Ableitung des zweiten Verlaufes von oben. Die dritte Grafik gibt die Anzahl  $R_i$  bzw.  $R(t)$  der Geräte ohne Ausfall an. Aus diesem Zusammenhang lässt sich leicht die in der unteren Grafik angegebene Ausfallrate  $z_i$  bzw.  $z(t)$  ermitteln. Die zahlen in diesem Beispiel sind so gewählt, dass sich in der untersten Grafik die sogenannte Badewannenkurve ergibt. Nach einer „Einbrennphase“ fällt in der „Nutzungsphase“ die Ausfallrate zunächst ab, um dann in der „Verschleissphase“ wieder anzusteigen.

In [7] (Seite 130) wird der Verlauf der Ausfallrate  $z(t)$  in drei Bereiche eingeteilt:

**Sinkende Ausfallraten** kennzeichnen Frühausfälle („Kinderkrankheiten“); es wird vielfach auch von der „Einbrennphase“ gesprochen. Diese Ausfälle können häufig auf Konstruktions-, Werkstoff- oder Montagefehler o.ä. zurückgeführt werden.

**Konstante Ausfallraten** charakterisieren Zufallsausfälle, wobei die Zufälligkeit darin zu sehen ist, dass Ausfälle durch das Zusammenwirken vieler statistisch voneinander unabhängiger Faktoren zustande kommen (typische Beispiele: Einfahren des Werkzeuges in die Spannvorrichtung, Ausfall elektronischer Bauelemente); die Phase einer konstanten Ausfallrate wird vielfach als Nutzungsphase bezeichnet.

**Steigende Ausfallraten** beschreiben im wesentlichen Verschleiß- und Alterungsausfälle (Beispiele: Materialermüdung, Korrosion, Strukturänderungen des Werkstoffes etc.); diese Phase wird als Verschleissphase bezeichnet.

Dabei ist zu beachten, dass die Zeit  $t$  bei den bisherigen Betrachtungen immer seit dem letzten Ausfall gemessen wurde. Nach einem Ausfall und darauf folgender Instandsetzung wird die Zeit wieder von Null an gemessen. Es wird so getan, als ob das Gerät nach einer Instandsetzung wieder neu ( $t = 0$ ) ist – also kein Gedächtnis besitzt. Diese Annahme muss durchaus kritisch beurteilt werden.

Möchte man die Ausfallrate nicht in Abhängigkeit der Zeit seit dem letzten Ausfall untersuchen sondern seit Inbetriebnahme des betroffenen Gerätes, so kann man eine Ersatzausfallrate

$$z'_i = \frac{f'_i / \Delta t}{R'_i} \quad (2.36)$$

bestimmen, bei der die  $f'_i$  die Anzahl der Ausfälle im Zeitraum  $[(i-1)\Delta t, i\Delta t]$  nach Inbetriebnahme des betroffenen Gerätes und  $R'_i$  die Anzahl der noch laufenden Geräte zum Zeitpunkt  $i\Delta t$  nach ihrer Inbetriebnahme ist.  $z'_i$  gibt dann die Anzahl der Ausfälle pro Gerät und Zeitraum in Abhängigkeit der Betriebszeit an – und nicht in Abhängigkeit der Zeit seit der letzten Maßnahme.

### Mean Time Between Failures (MTBF)

Die Mean Time Between Failures (MTBF) bezeichnet den Mittelwert der Lebensdauer. Für stetige Zufallsvariablen kann man die MTBF folgendermaßen ermitteln

$$MTBF = \int_0^{\infty} t f(t) dt = \int_0^{\infty} R(t) dt \quad (2.37)$$

Für diskrete Zufallsvariablen ergibt sich

$$MTBF = N \sum_{i=1}^n i R_i \Delta t \quad (2.38)$$

Der dabei begangene Fehler durch Lage der Betrachtungszeitpunkte der  $f_i$  und  $F_i$  am Anfang der Zeitintervalle  $[i\Delta t, (i+1)\Delta t]$  statt in der Mitte ist für kleine  $\Delta t$  vernachlässigbar.

### Exponential- und Weibull-Verteilung

Betrachten wir den einfachen Fall konstanter Ausfallquote  $z(t) = \lambda$ . Dann ergibt sich aus Formel 2.35

$$\int_0^{R(t)} \frac{dR(t)}{R(t)} = - \int_0^t \lambda dt \quad (2.39)$$

und durch Integration

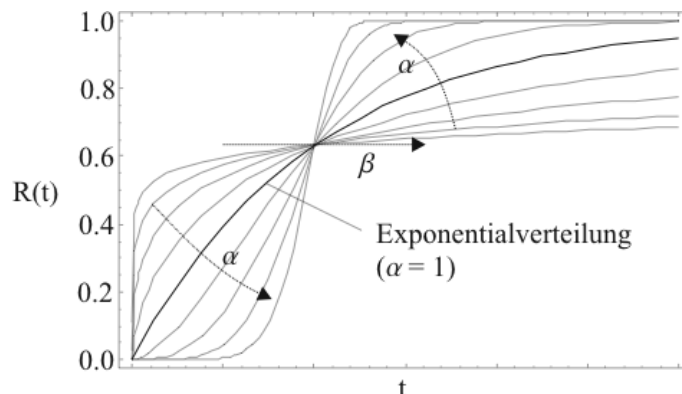


Abbildung 2.6: Verteilungsfunktion der Weibull-Verteilung für verschiedene Parameterwerte  $\alpha$  und  $\beta$ . Erläuterungen siehe Text.

$$R(t) = e^{-\lambda t} \quad (2.40)$$

Für Konstante Ausfallraten  $z(t) = \lambda$  ergibt sich also die Überlebenswahrscheinlichkeit  $R(t)$  als Exponentialverteilung. Für die MTBF ergibt sich

$$MTBF = \int_0^{\infty} R(t) dt = \frac{1}{\lambda} \quad (2.41)$$

Neben der Exponentialverteilung gibt es noch eine Reihe anderer Verteilungen. Einen bewährten Kompromiss zwischen einfacher Darstellung und Anpassungsfähigkeit stellt die Weibull-Verteilung dar:

$$R(t) = e^{-(t/\alpha)^\beta} \quad (2.42)$$

Die Wirkung der beiden Parameter ist in Abbildung 2.6 dargestellt. Durch  $\alpha$  kann man festlegen ob die Ausfallquote mit der Zeit ansteigt (überwiegend Verschleissausfälle,  $\alpha > 1$ ), konstant bleibt (überwiegend Zufallsausfälle,  $\alpha = 1$ ) oder abfällt (überwiegend Frühausfälle,  $\alpha < 1$ ). Gesondert eingezeichnet ist der Spezialfall konstanter Ausfallrate, der sich für  $\alpha = 1$  ergibt. Mit  $\beta$  kann man die Lage des Häufungspunktes festlegen.

## 2.5 Visuelle Darstellung von Daten

### 2.5.1 Visualisierung hierarchischer Daten

#### Problemstellung

Hierarchische Strukturen werden oft in Form von Baumgraphen (treemaps) dargestellt. Ein typisches Beispiel für eine hierarchische Struktur ist das Dateisystem einer Festplatte mit Hauptverzeichnis, Unterverzeichnissen und Dateien. Oft wird diese Struktur in Datei-Browsern als Baum dargestellt, wie z.B. auch im Explorer von Microsoft, der der Standard-Datei-Browser des Betriebssystems Windows ist.

Um sich einen Überblick über eine gesamte Festplatte zu verschaffen – zum Beispiel um die Frage zu beantworten, warum sie schon wieder so voll ist – bedarf es schon in normalen Fällen vieler Manöver, um alle Unterverzeichnisse auf ihren Inhalt hin durchzusehen. Ein Anzeigen aller Dateien absteigend der Größe nach sortiert muss nicht zum Ziel führen, wenn der überwiegende Teil des Festplattenplatzes durch kleine Dateien, die aber sehr zahlreich sind und sich ggf. nur in wenigen Unterordner befinden, repräsentiert wird. Die Darstellung einer hierarchischen Struktur durch gewöhnliche Baumdiagramme verschafft einen also nur in glücklichen Fällen einen schnellen Überblick. Die Zahl der Fälle, in denen ein schneller Überblick gelingt, kann durch Verwendung spezieller treemaps, sogenannter squarified treemaps stark vergrößert werden [11].

#### Algorithmus mit Beispiel

Treemaps setzen die hierarchische Struktur statt in eine Baumstruktur in eine Karte um. Dabei werden die hierarchischen Ebenen durch Rechtecke repräsentiert, die durch ihre Schachtelung die hierarchische Struktur widerspiegeln. Die Fläche der Rechtecke kann dabei, um beim Beispiel des Dateisystems zu bleiben, die Größe oder Anzahl der

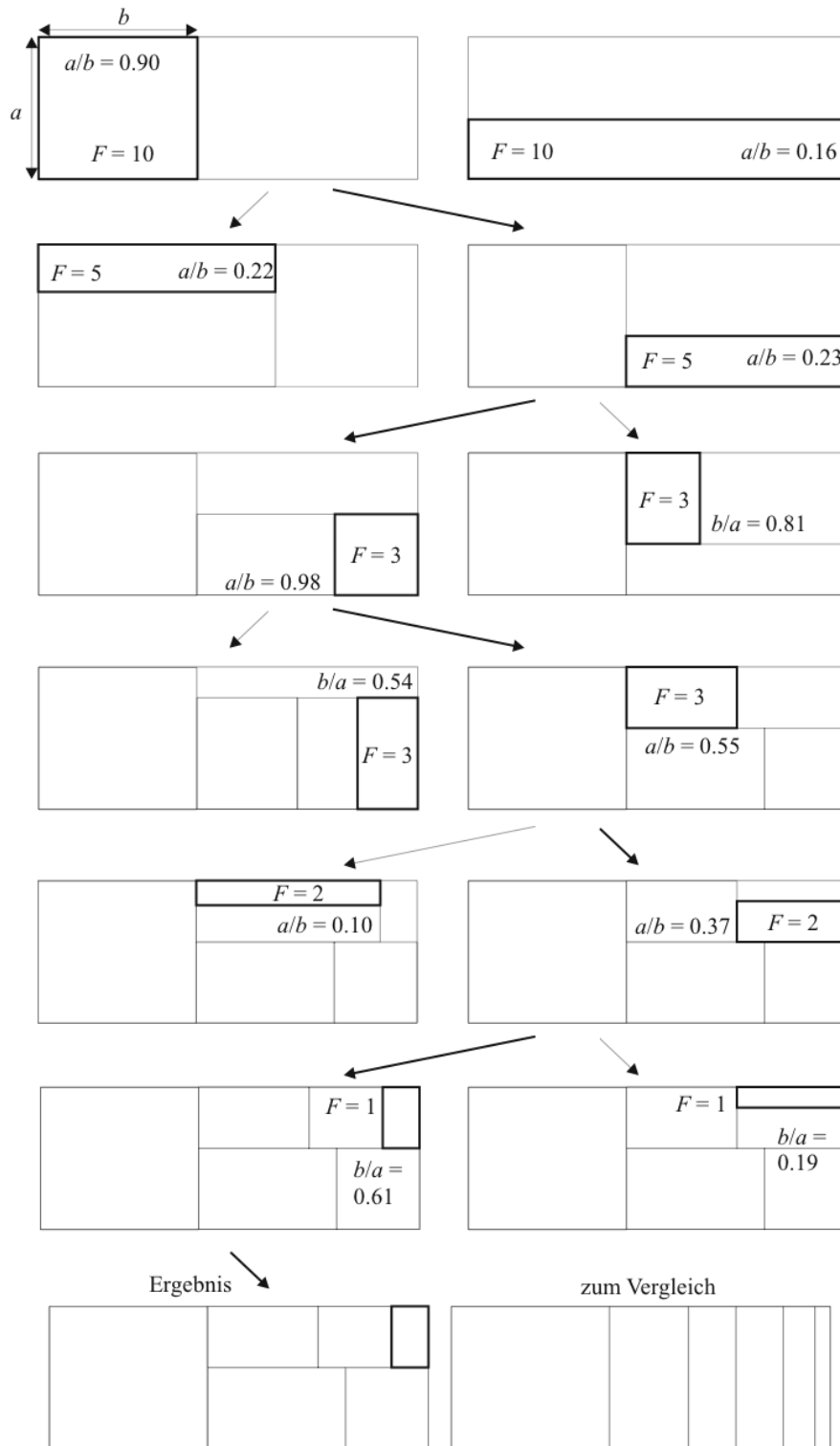


Abbildung 2.7: Ablauf des Algorithmus zur Erstellung einer squarified treemap. Die Flächen der 6 einzupassenden Rechtecke lauten 10, 5, 3, 3, 2 und 1. Das Seitenverhältnis des Umfassenden Rechtecks beträgt  $8/3$ . Der Algorithmus wird im Text beschrieben.

Dateien darstellen. Wichtig ist, dass diese Größe über die Hierarchieebenen additiv ist, d.h. diese Größe muss für ein Element in einer bestimmten Hierarchieebene die Summe aller Elemente aus der nächsten tiefer liegenden Hierarchieebene sein. Damit ist z.B. die Darstellung der durchschnittlichen Dateigröße durch treemaps nicht darstellbar.

Bei squarified treemaps wird versucht die Form der Rechtecke möglichst wenig von Quadraten abweichen zu lassen, da dies sich günstig auf die visuelle Erscheinung der treemap auswirkt (siehe Abbildung 2.7 unten). Diese Optimierung wird durch einen Algorithmus durchgeführt, der zwar keine optimale Lösung garantiert, aber im Normalfall unter vertretbarem Aufwand brauchbare Ergebnisse liefert ([11] ab Seite 33).

Abbildung 2.7 zeigt an einem Beispiel, wie der Algorithmus arbeitet. In diesem Beispiel werden nur die Rechtecke einer Hierarchieebene in ein großes Rechteck eingepasst. Das Beispiel lässt sich aber leicht auf eine hierarchische Struktur mit mehreren Ebenen erweitern, indem man das Vorgehen des Beispiels iterativ auf die erzeugten Teilrechtecke anwendet und so eine Darstellung einer hierarchischen Struktur erzeugt.

In diesem Beispiel soll eine Hierarchieebene mit 6 Elementen dargestellt werden, denen die Größen 10, 5, 3, 3, 2 und 1 zugeordnet sind. Die Rechtecke sollen in ein Rechteck mit Seitenverhältnis  $8/3$  eingepasst werden. Die Fläche aller Teilrechtecke ergibt die Fläche des Gesamtrechtecks. Beispielhaft sind an diesem Rechteck auch noch die Seitenlängen  $a$  und  $b$  eingezeichnet, aus denen das später benötigte Seitenverhältnis bestimmt werden kann. Dazu werden die Elemente der Größe nach absteigend in das Rechteck eingepasst. In der ersten Zeile der Abbildung 2.7 wird das größte Teilrechteck mit der Fläche 10 entsprechend seiner Größe in das Gesamtrechteck eingezeichnet. Es gibt zwei Möglichkeiten, dies zu tun. Hier wird die Möglichkeit mit dem ausgewogensten Verhältnis von  $a$  und  $b$  gewählt.

Nun wird das nächste kleinere Rechteck eingepasst. In Zeile zwei der Abbildung 2.7 sind dafür zwei Möglichkeiten angegeben: einmal kann das nächste Rechteck über dem Vorherigen eingepasst werden – beide Rechtecke haben dann die gleiche horizontale Kantenlänge, oder es kann daneben eingepasst werden. Für beide Alternativen ergeben sich unterschiedliche Seitenverhältnisse  $a/b$ . Für den Fall, dass  $a/b > 1$  ist wird im folgenden einfach der Kehrwert  $b/a$  betrachtet. Die Seitenverhältnisse sind also immer kleiner oder gleich eins. Um zu entscheiden, welche Alternative gewählt wird, um das neue Rechteck anzufügen, wird das Seitenverhältnis herangezogen, da es ermöglicht optisch leichter zu erfassende Rechtecke mit ausgeglichenem Seitenverhältnis nahe bei 1 gegenüber länglicheren zu bevorzugen. In diesem Auswahlschritt findet die „squarification“ statt – die Rechtecke sollen so eingepasst werden, dass sie möglichst eine quadratische Form erhalten.

Dieser Schritt wird für die folgenden und kleiner werdenden Rechtecke wiederholt, wie in den folgenden Zeilen der Abbildung 2.7 dargestellt. Rechtecke können dabei sowohl horizontal als auch vertikal angefügt werden.

In der letzten Zeile ist das Endergebnis dargestellt und zum Vergleich das Ergebnis, das man erhält, wenn man die Flächenaufteilung wie in einem gewöhnlichen Balkendiagramm vornimmt. Insbesondere bei den sehr schmalen Teilrechtecken ist das Einpassen einer weiteren tieferen Hierarchieebene in optische ansprechender Form kaum noch möglich.

## 2.5.2 Visualisierung von Kreuztabellen

### Problemstellung

Soll die Anzahl der Geräte nach kategorialen Merkmalen wie z.B. „Krankenhaus“ und „Gerätebezeichnung“ gruppiert angegeben werden, so eignen sich zur Darstellung dieser Zahlen Balkendiagramme. Ein Balken repräsentiert dann z.B. ein Krankenhaus, und die Rechtecke, aus denen ein Balken zusammengesetzt ist, repräsentieren die Anzahl der Geräte in den betrachteten Gerätegruppen des Krankenhauses. Möchte man nicht die absoluten sondern die relativen Geräteanzahlen vergleichen, so normiert man die Gesamtzahl der Geräte in einem Krankenhaus auf 100%. Die tabellarische Anordnung dieser Zahlen nennt man Kreuz- oder Pivot-Tabelle.

Ein Problem bei der Darstellung einer Kreuztabelle durch ein Balkendiagramm ist, dass es sehr viele verschiedene Möglichkeiten gibt die Kreuztabelle darzustellen. Da die Kategorien keine natürliche Ordnung besitzen, können sie beliebig angeordnet werden. Für eine Kreuztabelle mit dem Merkmal A ( $n_A$  verschiedene Kategorien) und Merkmal B ( $n_B$  verschiedene Kategorien) ergibt sich eine Anzahl von  $n_A!n_B!$  verschiedenen Anordnungen der Kategorien mit jeweils verschiedenen Balkendiagrammen. Welches Diagramm ist für eine übersichtliche Darstellung der Zahlen zu wählen?

### Algorithmus

Günstig auf die visuelle Wahrnehmung wirkt es sich aus, wenn das Balkendiagramm einen möglichst geordneten Eindruck macht – insbesondere lassen sich so auch leicht Gruppen von Kategorien erkennen. Der Eindruck eines geordneten Balkendiagramms lässt sich zumindest näherungsweise dadurch erreichen, dass man die Sortierung der Kategorien so wählt, dass die Rechteckkanten zu gleichen Kategorien benachbarter Spalten möglichst dicht beieinander liegen (siehe Abbildung 2.8). Diese Minimalbedingung lässt sich mit den Bezeichnungen aus Abbildung 2.8 schreiben als

$$Min! = \sum_{i=1}^{n_A-1} \sum_{j=1}^{n_B} \Delta_{ij} \quad (2.43)$$

Die Spalten sind dabei von  $i = 1$  bis  $i = n_A$  und die Rechtecke der Spalten mit  $j = 1$  bis  $j = n_B$  numeriert. Für auf 100% normierte Balken sind die  $\Delta_{in_B} = 0$ , da die Spalten alle gleich hoch sind.

Diese Summe ist nun für alle Permutationen der Kategorien der Merkmale A und B zu bestimmen. Für  $n_A = n_B = 6$  ergeben sich ca. 0.5 Millionen Möglichkeiten. Für  $n_A = n_B = 7$  sind es schon ca. 25 Millionen und für  $n_A = n_B = 8$  sind es gar 16 000 Millionen. Will man mittels „brute-force“ das Minimum ermitteln, so ist das nur noch bei sieben Kategorien pro Merkmal mit zeitlich vertretbarem Aufwand möglich. Sollen Merkmale untersucht werden, die über mehr als sieben Kategorien verfügen, so können bis auf die sechs häufigsten Kategorien alle restlichen in der siebten Kategorie „sonstige“ zusammengefasst werden (siehe Abbildung 2.9).

Einen Algorithmus, der auch für größere Anzahlen von Kategorien eine wenn auch nicht optimale aber dennoch brauchbare Lösung bei vertretbarer Laufzeit ermittelt, konnte in dieser Arbeit nicht gefunden werden. Das Problem hat Ähnlichkeit zum

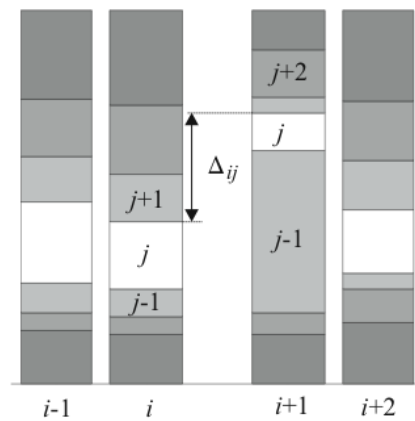


Abbildung 2.8: Beispielhaft ist eine Differenz  $\Delta_{ij}$  zwischen zwei Rechteckkanten gepaarter Rechtecke zweier Spalten eingezeichnet. Die Spalten werden mit  $j$  die Rechtecke mit  $i$  durchnummeriert. Die Spaltenlängen sind auf die gleiche Länge normiert.

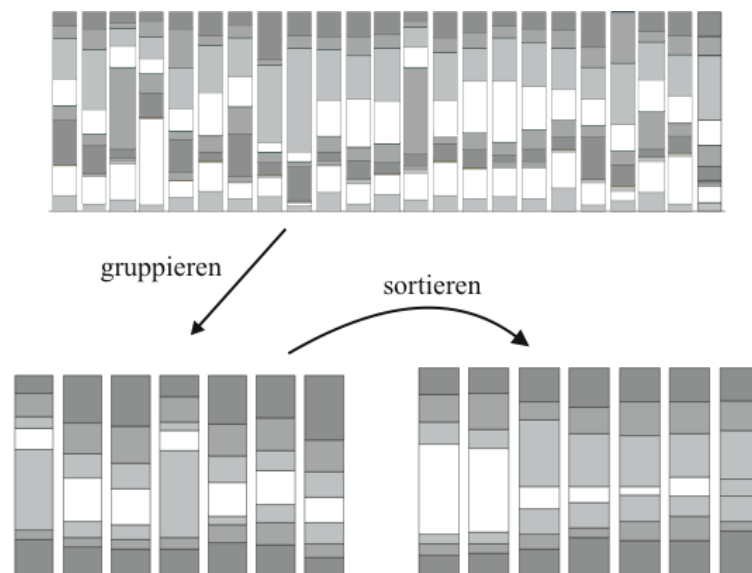


Abbildung 2.9: In der oberen Abbildung sind beispielhaft die Daten einer Kreuztabelle als normiertes Balkendiagramm dargestellt. Die Spalten- und Zeilen der Kreuztabelle sind zufällig (zum Beispiel alphabetisch) sortiert. In der linken unteren Abbildung wurde eine Gruppierung dergestalt vorgenommen, dass nur die 6 häufigsten Zeilen- und Spalten-Kategorien übernommen wurden und die restlichen Kategorien jeweils für Spalten und Zeilen zusammengefasst wurden. In der rechten unteren Abbildung wurde nun eine Sortierung der Kategorien vorgenommen, wie sie im Text beschrieben ist. Es ist deutlich zu erkennen, dass durch den ruhigeren Bildaufbau gegenüber der unteren linken Abbildung Zusammenhänge sichtbar werden, die vorher so nicht erkennbar waren. Insbesondere sei hier hervorgehoben, dass beide Abbildungen die gleichen Daten darstellen.

Traveling-Salesman-Problem, für das eine Reihe solcher Lösungen existiert. Da bei diesem Problem jedoch nicht nur ein Merkmal sondern zwei zu sortieren sind, können die Algorithmen zur Lösung von Traveling-Salesman-Problemen nicht verwendet werden. Das entwickeln eines neuen Algorithmus ist im Rahmen dieser Arbeit nicht gelungen. Es zeigt sich aber auch häufig, dass ab der sechst-häufigsten Kategorie die Häufigkeiten schon so stark abgefallen sind, dass das Diagramm davon kaum noch beeinflusst wird.

### 2.5.3 Visualisierung von Ungleichverteilungen

#### Problemstellung

Hat man Entitäten oder Gruppen von Entitäten  $E_i$ , denen zwei metrische Merkmale  $A_i$  und  $B_i$  zugeordnet werden können, so ist zunächst die Frage von Interesse, ob zwischen diesen Merkmalen ein Zusammenhang (Korrelation) besteht. In diesen Fällen hilft die Korrelations- oder Regressionsanalyse weiter. Zum Beispiel können die  $E_i$  Gerätegruppen sein, Merkmal  $A_i$  kann die Anzahl der Geräte und Merkmal  $B_i$  kann die Summe der Kaufpreise der Geräte sein jeweils für die Gruppe  $E_i$ . Findet sich nun ein linearer Zusammenhang zwischen  $A_i$  und  $B_i$ , so kann man die Steigung dieses linearen Zusammenhangs als über die Gerätegruppen hinweg existenten mittleren Kaufpreises pro Gerät interpretieren, den man durch eine Regressionsanalyse bestimmen kann.

Lässt sich jedoch keine Korrelation zwischen den Merkmalen finden, so kann man dennoch ggf. wertvolle Informationen gewinnen: Es liegt eine ausreichend deutliche Ungleichverteilung der Merkmale  $A_i$  und  $B_i$  in den Gruppen  $E_i$  vor, die eine Einteilung in verschiedene neue Gruppen erlaubt. Es können im obigen Beispiel sich für die verschiedenen Gerätegruppen deutlich unterschiedliche mittlere Kaufpreise ergeben, die Anzahl der Geräte also unabhängig von der Summe der Kaufpreise sein. Gegebenenfalls sind die Unterschiede der mittleren Kaufpreise sogar so groß, dass sie eine neue Einteilung der Gerätegruppen rechtfertigen.

Zur Beurteilung der Stärke der Ungleichverteilung sollen die im folgenden vorgestellten Lorenzkurve und der Gini-Koeffizient verwendet werden (z.B. [12] Seite 52).

#### Lorenzkurve und Gini-Koeffizient

Zum Erstellen einer Lorenzkurve sortiert man die Gruppen  $E_i$  mit  $i \in 1 \dots n$  aufsteigend oder absteigend nach dem Quotienten  $A_i/B_i$ . Dann ordnet man den sortierten  $E_i$  jeweils die kumulierten

$$A_i^K = \sum_{j=1}^i A_j \quad \text{und} \quad B_i^K = \sum_{j=1}^i B_j \quad (2.44)$$

zu. Dann kann man die kumulierten  $A_i^K$  und  $B_i^K$  noch als

$$A_i^N = \frac{A_i^K}{\sum_{j=1}^n A_j} \quad \text{und} \quad B_i^N = \frac{B_i^K}{\sum_{j=1}^n B_j} \quad (2.45)$$

auf eins normieren und als sog. „Lorenz- oder Konzentrationskurve“ mit Definitions- und Wertebereich  $[0, 1]$  auftragen. Als Beispiel sind in Abbildung 2.10 zwei Lorenzkurven zu den Werten  $((5,45), (7,91), (3,20), (6,62), (5,57))$  und zu den Werten  $((10,21), (1,71),$

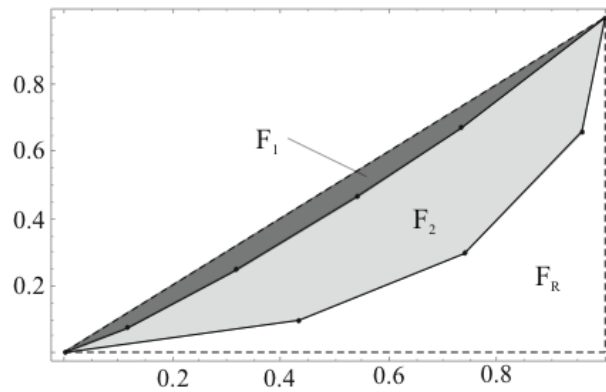


Abbildung 2.10: Beispielhafte Darstellung zweier Lorenzkurven. Es sind die Gini-Koeffizienten  $F_1$  und  $F_2$  eingezeichnet. Es gilt  $F_R + F_2 = 0.5$  und nicht  $F_R + F_2 + F_1 = 0.5$ . Details siehe Text.

(5,77), (7,12)) angegeben. Die größere Ungleichverteilung in der zweiten Wertreihe lässt sich anhand der stärker gebogenen Lorenzkurve deutlich erkennen. Zur objektiven Beurteilung der Wölbung kann man die in Abbildung 2.10 eingezeichneten Flächen heranziehen: je größer die Ungleichverteilung desto größer die Wölbung der Lorenzkurve desto größer die Fläche zwischen Lorenzkurve und der gedachten Linie durch die Punkte (0,0) und (1,1). Diese Fläche wird als „lorenzisches Konzentrationsmaß“ oder „Gini-Koeffizient“ bezeichnet.

### Andere Konzentrationsmaße

Neben dem Gini-Koeffizienten gibt es noch eine Reihe anderer Ungleichverteilungskoeffizienten. Ein Vorteil des Gini-Koeffizienten – seine anschauliche Interpretation als Fläche der Lorenzkurve kann auch als Nachteil gedeutet werden: der Zahlenwert hat keine praktische Bedeutung und besitzt eine komplizierte Berechnungsvorschrift. Diese Nachteile besitzt der Hoover-Ungleichverteilungskoeffizient  $h$  nicht [13]. Er berechnet sich zu

$$h = \sum_{i=1}^n |A_i^N - B_i^N| \quad (2.46)$$

Die Berechnung des Koeffizienten ist einfach und damit transparenter als beim Gini-Koeffizienten. Außerdem kann man den Zahlenwert des Hoover-Koeffizienten als den Anteil interpretieren, der umverteilt werden muss, um eine Gleichverteilung zu erhalten.

Andere Koeffizienten quantifizieren die Ungleichverteilung anhand von Entropiebetrachtungen der zu untersuchenden Verteilung. Diese Koeffizienten lassen sich zwar schlüssig aus der Theorie ableiten, setzen aber auf den wenig anschaulichen Begriff der Entropie auf und werden deswegen in dieser Arbeit nicht zur Untersuchung herangezogen.

# 3 Erhebung und Vorbereitung der Daten

## 3.1 Datenerhebung

In diesem Kapitel werden das Anschreiben, die erhaltenen Antworten und insbesondere die eingesandten Daten vorgestellt. Im Anschluss daran erfolgt eine erste sondierende Analyse und schließlich eine Normierung der Daten.

### 3.1.1 Das Anschreiben

Der zu erwartende Umfang der Daten zur Gerätebewirtschaftung eines Krankenhauses macht eine Berücksichtigung von nicht elektronisch erfassten Daten im Rahmen dieser Untersuchung unmöglich. Eine Erhebung z.B. durch Fragebogen ist damit nicht möglich. Die einzige Möglichkeit, Daten mit vertretbarem personellem Aufwand zu erfassen, ist die ausschließliche Berücksichtigung von Daten in elektronischer Form.

Sollte ein Krankenhaus nicht bereit sein, alle erfassten Daten zur Verfügung zu stellen, ist eine Prioritätenliste der gewünschten Informationen angegeben. Selbstverständlich erfolgt ein Hinweis, dass die Daten nur anonymisiert veröffentlicht werden. Das Anschreiben ist im Anhang [A](#) dargestellt.

### 3.1.2 Auswahl der Krankenhäuser

Angeschrieben wurden alle 306 Krankenhäuser aus der Referenzliste der Firma Loy & Hutz. Ein viertel Jahr später wurde noch eine zufällige Auswahl von 163 Krankenhäusern aus dem Deutschen Krankenhaus-Adressbuch (DKA 2000) angeschrieben. Insgesamt wurden also 469 Krankenhäuser angeschrieben.

### 3.1.3 Eingesandte Antworten

Von den 469 angeschriebenen Krankenhäusern (100 %) haben 68 geantwortet (14.5 %); 23 haben zusätzlich Daten eingeschickt (5.00 %). Die Begründungen für das Nichteinsenden von Daten sind mit Häufigkeit im folgenden angegeben. Die restlichen Absagen erfolgten ohne Begründung.

- Daten nicht elektronisch erfasst: 11
- Daten werden nicht herausgegeben: 4
- keine Zeit: 7
- Großschaden: 3

### 3.1.4 Eingesandte Daten

Die eingeschickten 23 Datenbanken haben einen Umfang von etwa 1.3 GB. Als Datenformate wurden Text-, Access- sowie proprietäre Formate verwendet. Bei letzteren bot die mitgelieferte Software den Export der Daten in übliche Dateiformate. Der Datentransfer erfolgte über Disketten, CD's, und Downloads per Email-Anhang oder per FTP.

Die Daten sind ausschließlich in Tabellen strukturiert, die teilweise wiederum zu Datenbanken zusammengefasst sind. Die Tabellen bestehen aus Zellen, die zu Spalten und Zeilen zusammengefasst werden können. Die Zellen können Zahlen oder Text beinhalten. Die Zellen einer Spalte enthalten ähnliche Informationen (z.B. Firmennamen oder Zahlen mit zwei Nachkommastellen). Die Zellen einer Zeile hingegen enthalten sehr unterschiedliche Informationen. Man kann leicht erkennen, dass jede Zeile einer Betrachtungseinheit (z.B. Gerät, Person, Maßnahme) und jede Spalte einem Merkmal (z.B. Hersteller, Vorname, Durchführungszeitpunkt) zugeordnet werden kann. In Tabellen werden Betrachtungseinheiten zusammengefasst, die den gleichen Satz an Merkmalen besitzen.

Alle Tabellen haben ausnahmslos diese Struktureigenschaften. Die Tabellen und Spalten haben nur zum Teil aussagekräftige Bezeichnungen, aus denen hervorgeht, welche Daten in ihnen abgelegt wurden.

## 3.2 Datentransfer

### 3.2.1 Transfer der Originaldaten ins Textformat

Jede Tabelle der Originaldaten wird in einer Textdatei exportiert. Zeilen werden durch das Zeilenende-Zeichen oder Return und Zellen durch Tabulatoren getrennt. Returns und Tabulatoren werden in Zellen durch Leerzeichen ersetzt. Dies kommt zwar relativ selten vor, macht dann aber oft Probleme. Falls das Ersetzen vor dem Export nicht möglich sein sollte, muss dies nach dem Export erfolgen. Dabei helfen folgende Regeln:

- Zwischen den meisten Returns kommt eine bestimmte Anzahl von Tabulatoren – dies ist die reguläre Breite oder Spaltenzahl der Tabelle.
- Folgt ein Return schon nach einer kleineren Anzahl von Tabulatoren, so wird dieses Return durch ein Leerzeichen ersetzt und anschließend geprüft, ob sich nun zwischen den Returns die passende Anzahl an Tabulatoren befinden. Ist das nicht der Fall, wird die Zeile markiert oder gelöscht.
- Folgt ein Return erst nach einer größeren Anzahl von Tabulatoren, so wird die Zeile markiert oder gelöscht.
- Die markierten Zellen werden manuell bearbeitet. Dies sind im allgemeinen nur wenige Zeilen.

Durch diese Vorarbeiten haben die Daten nun folgende Struktur: Jede Tabelle ist für sich in einer Textdatei gespeichert. Diese Dateien sind eindeutig nach Krankenhaus bzw.

Datenbank und Tabellenname benannt („Krankenhaus-Tabellenname.txt“). Jede Zeile einer Tabelle hat nun die gleiche Anzahl von Zellen. Das Datenvolumen von insgesamt 310 MB an Textdateien verteilt sich auf:

- 23 Datenbanken
- 470 Tabellen (ca. 20 Tabellen pro Datenbank)
- 11.284 Spalten (ca. 24 Spalten pro Tabelle)
- 52 Mio. Zellen (ca. 4300 Zellen pro Spalte)

#### 3.2.2 Codepage-Säuberungen, Teil I

Die für die folgenden Analysen interessanten Informationen lassen sich mit weniger Textzeichen darstellen, als in den Textdateien vorhanden sind. Ein Beschränken des Zeichensatzes (Codepage, siehe auch 2.1.3) hat die Vorteile, dass es weniger unterschiedliche Schreibweisen gibt – was die Analyse vereinfacht – und dass es auch keine Sonderzeichen mehr gibt – was den Ex- und Import vereinfacht.

In diesem ersten Teil der Codepage-Säuberung geht es hauptsächlich darum, bei der Verarbeitung störende (Sonder-)Zeichen zu ersetzen. Nur bei offensichtlich für den Informationstransport überflüssigen Variationen (Groß- / Kleinschreibung und Umlaute / Umlaute ausschreiben) werden Zeichen ersetzt. Im später folgenden zweiten Teil werden zusätzlich auch andere Zeichen entfernt, die kaum zum Informationstransport beitragen, auch wenn sie die Datenverarbeitung nicht stören. Dies erfolgt aber erst, wenn genauere Informationen über die zu untersuchenden Daten vorliegen.

Folgende Ersetzungsregeln werden für die Zellinhalte angewendet:

- Großbuchstaben zu Kleinbuchstaben
- Umlaute und ß ausschreiben
- Sonderzeichen 00-1F und 7F-9F (siehe Tabelle 2.1) löschen

In den folgenden Tabellen sind ausschnittsweise Zellinhalte dargestellt, jeweils nach Häufigkeit (Tabelle 3.1) und Alphabet (Tabelle 3.2) sortiert. Man kann schon an diesen Ausschnitten erkennen, dass die Informationen nicht durchgehend systematisch abgelegt wurden, was bei einer Analyse zu berücksichtigen ist. Zum Beispiel enthält eine Zelle gemischte (nicht atomare) Informationen: Gerätebezeichnung und Identifikationsnummer – dies erschwert eine Verarbeitung der Daten, ohne dass diese Art der Speicherung einen anderen Vorteil bietet. Auch kann man ein Beispiel für unterschiedliche Schreibweisen einer Bezeichnung finden – auch das ist für Analysen problematisch.

### 3.3 Struktur- oder Syntaxanalyse der Daten

Nachdem die Daten vorbereitet sind, können sie auf ihre Struktur untersucht werden. Dabei geht es jetzt noch nicht um Bedeutungen (Semantik), sondern nur um die Darstellung (Syntax). Das Ergebnis dieser Analysen würde sich nicht ändern, wenn man

Tabelle 3.1: In der folgenden Tabelle sind alle Zellinhalte aller Datenbanken absteigend nach Häufigkeit sortiert. Leere Zellen sind mit Abstand am häufigsten. Die hohe Anzahl von Nullen in den Daten ist darauf zurückzuführen, dass in vielen Spalten leere Zellen verboten sind und durch Nullen repräsentiert werden. Zellen mit lesbaren Textinhalten folgen erst ab Platz 30. Am Fuß der Tabelle ist erkennbar, dass es 454 992 verschiedene Zellinhalte in allen Datenbanken gibt.

Nr	Zellinhalt	Anzahl
1		21637422
2	0	2746884
3	x	1626973
4	0000	885950
5	j	420169
6	f	334295
7	0,00 DM	333439
8	16	282803
9	0,00	275740
10	r	268341
		254955
		252076
28	n	252076
29	0407020	70007
30	en	68734
31	elektriker-niederspannung	68629
32	gb9905	68491
33	2	68464
34	10	65281
35	reparatur	60031
36	1997	58164
37	15	57614
38	gb9910	56978
39	0409022	56076

jeden Buchstaben (jedoch bitte eindeutig) durch einen anderen ersetzen würde: die semantischen Informationen würden dabei verloren gehen, der Syntax der Daten bleibt jedoch unberührt.

Der Hauptvorteil dieser Analysen ist gleichzeitig auch der Hauptnachteil: dadurch, dass man die Bedeutungen der Daten ausblendet, ignoriert man viele Informationen, begeht aber garantiert keinen Interpretationsfehler bei der Zuordnung von Bedeutungen zu den Daten. Die Analyse ist also ungenau, aber robust.

### 3.3.1 Vorbereitungen

Zu jeder Datenbank, zu jeder Tabelle und zu jeder Spalte aus den Originaldaten werden Struktur-Informationen ermittelt und in einer Tabelle jeweils in einer Zeile abgelegt. Ein Ausschnitt dieser Daten ist in Tabelle 3.3 angegeben. Die Struktur-Informationen sind dabei im einzelnen:

**Spalte „DB-Nr“:** Datenbanknummer

**Spalte „Tab-Name“:** Tabellen-Name (nur bei Tabellen und Spalten)

Tabelle 3.2: Beispielhafter Ausschnitt der alphabetisch sortierten Zellinhalte. Man erkennt, gleiche Dinge werden unterschiedlich bezeichnet (z.B. Gerätebezeichnung „ueberwachungszentrale“). Zellen enthalten nicht atomare, d.h. gemischte Informationen (z.B. Gerätebezeichnung „ueberwachungsmonitor“ und vermutliche Seriennummer „t44bc96“). Der Name wurde nachträglich unkenntlich gemacht.

Nr	Zellinhalt	Anzahl
443841	ueberwachungsmonitor t44bc96	2
443842	ueberwachungsschrank	4
443843	ueberwachungsturm	5
443844	ueberwachungszentrale	7
443845	ueberwachungs-zentrale	5
443846	ueberwuerfe einbauen	2
443847	ueberwuerfe erneuern (umkleideschr.)	3
443848	ueberwurfmutter wandlerkombination nicht festgezogen	2
443849	ueckt	3
443850	uefen	2
443851	ueft	12
443852	ueft.	3
443853	uehl*****, frau	26
443854	uehl*****, frau, frau	25
443855	u-einheit def.	4
443856	ue-kamera def.	3
443857	uellt.	3
443858	uelt	2
443859	uemruestung auf alle sprizentypen stk	9

**Spalte „Sp-Name“:** Spaltenbezeichnung aus den Originaldaten (nur bei Spalten)

**Spalte „Zellen“:** Anzahl der enthaltenen Zellen

**Spalte „diff“:** Anzahl der enthaltenen, unterschiedlichen Zellen

**Spalte „einmal“:** Anzahl der enthaltenen, einmaligen Zellen

**Spalte „leer“:** Anzahl der leeren Zellen

Zusätzlich werden in einer weiteren Tabelle zu jeder internen Referenznummer aus der vorherigen Tabelle alle unterschiedlichen Zellen jeweils mit Häufigkeit abgelegt. Beispielhaft sind dazu die unterschiedlichen Zellinhalte der Spalte „art“ aus der Tabelle „DURCH120“ aus der Datenbank „22“ in der Untertabelle von Tabelle 3.3 absteigend nach Häufigkeit sortiert dargestellt. Diese beiden Tabellen werden als leicht zugängliches Zwischenergebnis in eine Access-Datenbank importiert und verknüpft und dienen damit als Arbeitsgrundlage für das weitere Vorgehen.

### 3.3.2 Analyse und Interpretation

Aus den oben erstellten Tabellen lässt sich leicht ermitteln, dass 50 % der Zellen, 42 % aller Spalten und 43 % aller Tabellen leer sind. Die 22 Mio. nicht leeren Zellen enthalten

Tabelle 3.3: Beispielhafter Ausschnitt aus der Tabelle, die Strukturinformationen der Datenbanken, Tabellen und Spalten enthält. Die Tabelle wurde nach Datenbanknummer (Spalte „DB-Nr“) und Tabellennamen (Spalte „Tab-Name“) sortiert. Zum Beurteilen von Inhalten der Datenbank, Tabelle oder Spalte sind Informationen über Anzahl der enthaltenen Zellen (Spalte „Zellen“), Anzahl der unterschiedlichen Zellen (Spalte „diff“), Anzahl der einmaligen Zellen (Spalte „einmal“) und Anzahl der leeren Zellen (Spalte „leer“) hilfreich. Zusätzlich enthalten ist ein beispielhafter Ausschnitt aus der Tabelle, die die unterschiedlichen Zellinhalte mit Häufigkeiten enthält. Der Ausschnitt enthält nur Zellinhalte der Spalte „art“ aus der Tabelle „DURCH120“ aus der Datenbank „22“. Die Tabelle ist absteigend nach der Häufigkeit der Zellinhalte sortiert. Man kann erkennen wie sich die Zellinhalte auf die unterschiedlichen Zellinhalte verteilen.

	DB-Nr	Tab-Name	Sp-Name	Zellen	diff	leer	einmal	REF
+	22	DURCH120	zeit1	20353	8900	573	6813	11014
+	22	DURCH120	von	20353	2031	10513	1131	11015
+	22	DURCH120	vbemerk	20353	7834	5615	6647	11016
+	22	DURCH120	aufnr	20353	13612	1	13277	11017
+	22	DURCH120	faufnr	20353	51	20277	44	11018
+	22	DURCH120	art	20353	9	117	5	11019
				<b>Zellinhalt</b>		<b>Anzahl</b>		
▶	fremdleistung					12528		
	eigenleistung					7696		
						117		
	reparatur					7		
	ueberpruefung					1		
	stk 11 (26 wochen)					1		
	betriebswartung					1		
	art					1		
	0					1		
	*							
+	22	DURCH120	fristbis	20353	3	20351	2	11020
+	22	DURCH120	fehlertext	20353	2892	12184	2618	11021
+	22	DURCH120	persschad	20353	4	1	2	11022
+	22	DURCH120	meldepfl	20353	4	1	2	11023
+	22	DURCH120	m_zubeh	20353	316	19170	216	11024
+	22	DURCH120	standort	20353	230	299	44	11025

450 000 verschiedene Zellinhalte. Man sieht, in den Daten ist eine Menge „Luft“. Um Spalten zu finden, die interessante Informationen bieten, kann man zunächst Spalten auswählen, die mehr als 10 Zellen enthalten. Dies sind 5137 Spalten (ca. 51 % aller Spalten). Davon enthalten

- 2971 Spalten mehr als 10 % ausgefüllte Zellen (ca. 26 % aller Spalten)
- 2436 Spalten mehr als 50 % ausgefüllte Zellen (ca. 21 % aller Spalten)
- 2059 Spalten mehr als 90 % ausgefüllte Zellen (ca. 18 % aller Spalten)

Von diesen 2059 Spalten, die mehr als 10 Zellen enthalten und wiederum zu weniger als 10 % leer sind, enthalten

- 391 Spalten mehr als 95 % verschiedene Zelleninhalte (ca. 3.5 % aller Spalten). Dies sind meist datenbankinterne Spalten oder Indizes,
- 854 Spalten weniger als 5 % verschiedene Zelleninhalte (ca. 7.5 % aller Spalten). Dies sind meist Spalten, die kaum genutzt werden, aber automatisch eine Eingabe erzeugen.

Also bleiben  $2059 - 391 - 854 = 814$  Spalten (ca.  $18\% - 3.5\% - 7.5\% = 7\%$  aller Spalten, also ca. 35 Spalten pro Datenbank), die

- mehr als 10 Zellen enthalten,
- zu mehr als 90 % ausgefüllt sind,
- weniger als 95 % verschiedene Zellen enthalten und
- mehr als 5 % verschiedene Zellen enthalten.

Die Datenbanken enthalten also viele ungenutzte Spalten. Außerdem beträgt der Anteil der leeren Zellen bei obiger Auswahl ca. 23%. D.h. der Anteil leerer Zellen ist in dieser Auswahl halb so groß wie bei allen Zellen.

Durch diese Spalten-Anforderungen schrumpft die Anzahl der interessanten Spalten auf ein Maß zusammen, das mit vertretbarem Aufwand manuell weiter untersuchbar ist. So können – wie im Abschnitt 3.3.3 dargestellt – Spalten zu Gruppen zusammengefasst werden, wobei jede Gruppe nur aus Spalten besteht, die Informationen zu einem bestimmten Merkmal enthalten (wie z.B. Gerätebezeichnung).

Unterstellt sei, dass diese Informationen jetzt schon vorliegen: In der Tabelle 3.4 sind Informationen einiger ausgewählter kategorialer Merkmale von Geräten und Tätigkeiten aufgeführt. Jeweils angegeben sind die Mittelwerte mit Standardabweichungen aller Spalten, die Informationen zu einem Merkmal enthalten. Die Anzahl der „unterschiedlichen Zellen“ pro Spalte und „einmaligen Zellen“ sind als prozentuale Anteile aller Zellen einer Spalte dargestellt.

Aus dieser Tabelle ist ableitbar, dass jede 5. (20.53 %) Eintragung im Feld „Anlagenbezeichnung“ vorher noch nicht verwendet wurde. Jede 2. (10.77 % / 20.53 %) neu eingeführte Bezeichnung wird danach nicht mehr verwendet.

Für die anderen Merkmale gelten die selben Überlegungen. Daraus folgt: es werden bedenklich viele unterschiedliche Bezeichnungen verwendet, insbesondere bei den Merkmalen „Anlagenbezeichnung“ und „Tätigkeitsart“. Bei dieser großen Zahl von verschiedenen Bezeichnungen ist zu vermuten, dass auch als gleich anzusehende Dinge unterschiedlich bezeichnet werden. Dies erschwert eine Analyse der Daten erheblich.

Die hohen Standardabweichungen des Mittelwertes bei den Merkmalen „Tätigkeitsart“ und „Arbeitsbeschreibung“ deuten auf eine Unabhängigkeit der Anzahl der unterschiedlichen und einmaligen Zellen aus der Anzahl aller Zellen hin. Bei der hohen Anzahl von Tätigkeiten werden bei Neueintragungen kaum noch neue Begriffe eingeführt (Sättigung).

Bei dem Merkmal „E/F-Leistung“ ist jeder 1000. Eintrag ein neuer Eintrag. In den Tabellen überwiegen 3 verschiedene Eintragungen: jeweils für „Eigenleistung“, „Fremdleistung“ und „leer“. Daraus folgt: Bei einfachen Eingabealternativen wird vermutlich durch Eingabemasken ein hohes „Begriffs-Recycling“ erreicht.

Trägt man für jede der über 10 000 Spalten jeweils den Anteil der leeren Felder und der einmaligen Felder an allen Feldern einer Spalte auf, so erhält man die Abbildung 3.1. Es ist zu erkennen, dass sich viele Spalten in den Randbereichen befinden. Durch Ellipsen sind die Bereiche angedeutet, in denen sich gehäuft Spalten bestimmter Merkmale befinden.

### 3.3.3 Manuelle Auswahl und Gruppierung der Spalten

Mit Hilfe der im vorherigen Abschnitt gewonnenen Informationen (welche Spalten sind interessant, was sind die häufigsten Zellenhalte) können nun die Spalten der einzelnen Tabellen zu Gruppen zusammengefasst werden. Eine Gruppe enthält dann alle Spalten eines Merkmales einer Entität. Der Auswahlprozess findet manuell statt und hat ca. 30 Stunden gedauert.

Ob zu einem Merkmal eine Gruppe gebildet wird, die alle Spalten mit dem Merkmal enthält, hängt davon ab, ob das betreffende Merkmal für die weiteren Analysen potentiell interessant ist und ob überhaupt ausreichend viele Spalten in den Daten vorhanden sind, die dann auch noch ausreichend mit sinnvollen Informationen erfüllt sind.

Am Ende des Auswahlprozesses ist also zu jeder Spalte bekannt, zu welchem Merkmal welcher Entität sie gehört. Dies wird in der Tabelle 3.5 unter der Spalten „Entität“ und „Merkmal“ vermerkt. Als Entitäten haben sich Geräte und Maßnahmen heraus kristallisiert. Andere Entitäten, wie z.B. Personen, werden nicht weiter untersucht, da sie für die in dieser Arbeit untersuchten Fragestellungen nicht interessant sind.

Die für die Entität Gerät (z.B. Infusionspumpe, Fahrstuhl etc) gruppierten Merkmale sind in Tabelle 3.6 mit Anzahl der Spalten und Anteil der leeren Zellen angegeben. Für die Betrachtungseinheit Maßnahme (z.B. Störungsmeldung, Maßnahme etc) sind die entsprechenden Größen in Tabelle 3.7 angegeben.

## 3.4 Erstellung der Gesamtdatenbank

Hier wird wieder auf die Ursprungsdaten zurückgegriffen. Tabellen, die Spalten enthalten, die weiter verwendet werden sollen, werden in Excel importiert, da mit Excel ein intuitives Verschieben von Zellen in Tabellen möglich ist. Leider kann Excel nur Tabellen mit maximal ca. 65 000 Zeilen verarbeiten, was dazu führt, dass man große Tabellen in kleinere aufteilen muss. Dieser Nachteil wird aber durch die leichte Bedienung mehr als aufgewogen.

Für jedes Krankenhaus bzw. für jede eingesandte Datenbank wird eine Geräte- und eine Maßnahmen-Tabelle erstellt, die alle eben ausgewählten Merkmale als Spalten enthält. In die Spalten dieser Tabelle werden dann die zugeordneten Spalten der Originaldaten kopiert. Werden mehrere Spalten in eine Spalte kopiert, so werden die ursprünglichen Zellenhalte durch Semikolons getrennt. Da diese Arbeiten manuell durchgeführt werden, sind hier etwa 5 Stunden pro Datenbank an Zeitaufwand nötig.

Um es deutlich zu sagen: wenn hier von manuellem Vorgehen gesprochen wird, so ist nicht gemeint, dass die Zellen einzeln per Hand kopiert werden. Das Kopieren geschieht mit Hilfe von vordefinierten Makros oder Funktionen – anders wäre das Kopieren der vielen 100 000 Zellen nicht zu bewältigen. Manuell bedeutet hier nur, dass das Auslösen eines Makros oder das Anwenden einer Funktion vom Benutzer entschieden wird

Tabelle 3.4: In dieser Tabelle sind Informationen aller Spalten, die Informationen zum gleichen Merkmal (z.B. „Gerätebezeichnung“) in einer Zeile zusammengefasst. Beschreibung siehe Text.

Merkmal (Geräte)	unterschiedliche Zellen		einmalige Zellen	
	Mittelwert	StdAbw	Mittelwert	StdAbw
Gerätebezeichnung	20,53%	6,91%	10,77%	4,94%
Typ/Modell	42,30%	15,40%	31,26%	16,41%
Hersteller	14,21%	5,32%	6,71%	2,55v
Standort, Abteilung	4,59%	3,14%	1,27%	1,69%
Merkmal (Maßnahmen)				
Tätigkeitsart	8,37%	20,73%	5,26%	15,96%
Arbeitsbeschreibung	25,68%	21,04%	23,38%	19,96%
E/F-Leistung	0,10%	0,07%	0,03%	0,02%
Firma	2,68%	1,56%	0,97%	0,66%

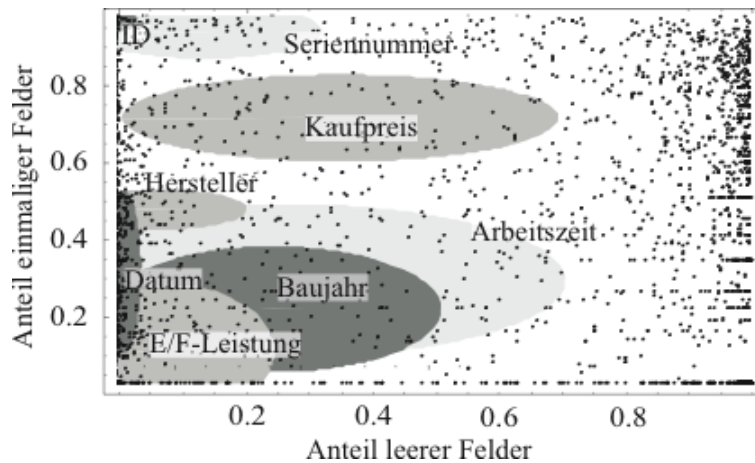


Abbildung 3.1: Jeder Punkt der oberen Abbildung entspricht einer Spalte. Die Abbildung enthält alle 11.284 Spalten der eingesandten Daten. Die x-Koordinate eines Punktes stellt den Anteil der leeren Felder der zugehörigen Spalte dar, die Y-Koordinate den Anteil der einmaligen Felder je Spalte. Durch Ellipsen sind Bereiche gekennzeichnet in denen gehäuft Spalten bestimmter Merkmale auftreten. Am deutlichsten konzentriert sind die Spalten mit dem Merkmal „ID“, d.h. der Spalte, die die eindeutige Kennung jedes Datensatzes enthält, da diese Spalte kaum leere und fast nur einmalige Zelleninhalte enthält. Der Anteil der einmaligen Zelleninhalte beim Merkmal „Kaufpreis“ ist höher als beim Merkmal „Baujahr“ – auch dies ist klar: es gibt deutlich mehr Kaufpreise als Baujahre. Die große Streuung beim Anteil der leeren Felder deutet darauf hin, dass in den verschiedenen Datenbanken dieses Merkmal unterschiedlich konsequent erfasst wurde. Die durchgängig gute Erfassung des Merkmales „Datum“ ist ggf. darauf zurückzuführen, dass der Inhalt dieser Zellen automatisch beim Anlegen eines Datensatzes ausgefüllt wird.

Tabelle 3.5: Beispielhafter Ausschnitt der Zuordnungen der Spalten zu den Entitäten und deren Merkmalen.

DB-Nr	Tab-Name	Sp-Name	Entität	Merkmal
22	DURCH120	sernum	Maßnahme	
22	DURCH120	garantie	Maßnahme	
22	DURCH120	abtlg	Maßnahme	
22	DURCH120	aufanfang	Maßnahme	Datum
22	DURCH120	zeit1	Maßnahme	
22	DURCH120	von	Maßnahme	
22	DURCH120	vbemerk	Maßnahme	Beschreibung
22	DURCH120	aufnr	Maßnahme	
22	DURCH120	faufnr	Maßnahme	
22	DURCH120	art	Maßnahme	E/F-Leistung
22	DURCH120	fristbis	Maßnahme	
22	DURCH120	fehlertext	Maßnahme	Beschreibung
22	DURCH120	persschad	Maßnahme	
22	DURCH120	meldepfl	Maßnahme	
22	DURCH120	m_zubeh	Maßnahme	
22	DURCH120	standort	Maßnahme	
22	DURCH120	wdurch	Maßnahme	
22	DURCH120	wdate	Maßnahme	
22	DURCH120	beri_gedr	Maßnahme	
22	DURCH120	an	Maßnahme	Firma
22	DURCH120	anint	Maßnahme	
22	DURCH120	kostvor	Maßnahme	

Datensatz: 10525 von 11284

und nicht von einem Algorithmus. Diese Vorbereitung ist nicht in wenigen Tagen zu bewältigen!

Alle Geräte- und Maßnahmen-Tabellen haben nun die gleiche Form und können leicht in eine Gesamtdatenbank kopiert werden.

### 3.4.1 Import von Relationen

Neben dem Kopieren der Spalten ist es an dieser Stelle auch wichtig, eine ggf. vorhandene Beziehung von Maßnahmen bzw. Ereignissen und Geräten mit in die neuen Tabellen zu übernehmen. Dazu bekommt jedes Gerät und jede Maßnahme bzw. jedes Ereignis eine die Gesamtdatenbank übergreifende, eindeutige Kennung. Die Geräte-kennungen aus den Originaldaten werden durch diese neuen Kennungen ersetzt. Damit ist sichergestellt, dass alle Kennungen das gleiche Format haben und in der gesamten Datenbank einmalig sind.

### 3.4.2 Codepage-Säuberungen, Teil II

Da, wie oben geschrieben, auf die Originaldaten zurückgegriffen wird, liegen wieder alle Zellinhalte als Text vor, der Sonderzeichen enthält. Da diese Sonderzeichen kaum Informationen transportieren aber den In- und Export erschweren, werden sie gelöscht oder ersetzt. Im einzelnen werden folgende Regeln der Reihe nach angewendet:

1. Umlaute und „ß“ ausschreiben

Tabelle 3.6: Merkmale der Entität „Gerät“ mit Anzahl der Zuordnungen und Anteil der leeren Zellen. Problematisch sind sowohl Merkmale, die zwar häufig Spalten zugeordnet werden konnten, aber viele leere Zellen enthalten, als auch Merkmale, die nur wenigen Spalten zugeordnet werden konnten, obwohl sie ggf. wenige leere Zellen enthalten.

Entität	Merkmal	Anzahl	leer
Gerät		625	63,30%
Gerät	Anlagenbezeichnung	26	0,04%
Gerät	Modell	25	18,42%
Gerät	Hersteller	24	26,52%
Gerät	Kaufdatum	23	6,18%
Gerät	Standort, Abteilung	22	31,69%
Gerät	Kaufpreis	21	17,67%
Gerät	ID	18	0,00%
Gerät	Seriennummer	16	22,01%
Gerät	Bemerkung	15	86,52%
Gerät	Baujahr	15	25,46%
Gerät	Service	15	56,61%
Gerät	Lieferant	14	41,31%
Gerät	MedGV	13	9,02%
Gerät	Standort, Gebäude	11	11,50%
Gerät	AfA, Stanz	11	45,51%
Gerät	Garantie, von	11	78,40%
Gerät	Standort, Raum	10	12,50%
Gerät	Spannung	10	81,65%
Gerät	Besitzstatus	9	19,93%
Gerät	Hauptgerät	9	95,42%
Gerät	Strom	9	76,03%
Gerät	Schutzart	9	97,93%
Gerät	VDE-Zeichen	9	84,06%
Gerät	Garantie, bis	9	84,65%
Gerät	Ortsfest	9	84,01%
Gerät	Leistung	9	83,74%
Gerät	Schutzklasse	8	97,18%
Gerät	Anlagenstatus	8	3,13%
Gerät	FTZ	8	97,72%
Gerät	AFA, Nutzungszeit	9	15,95%
Gerät	Neupreis	7	95,05%
Gerät	MPG	3	83,02%
Gerät	Garantie, Dauer	3	0,01%
Gerät	Betriebsstunden	2	0,07%
Gerät	Zählerstand	2	0,07%

2. Sonderzeichen und Trennzeichen ersetzen (angegeben sind die Zeichennummern aus Tabelle 2.1)

- 00-1F (Sonderzeichen) → 20(Leerzeichen)
- 20 (Leerzeichen) belassen
- 21-2F (Trennzeichen) → 20 (Leerzeichen)
- 30-39 (Ziffern) belassen

Tabelle 3.7: Merkmale der Entität „Maßnahme“ mit Anzahl der Zuordnungen und Anteil der leeren Zellen. Problematisch sind sowohl Merkmale, die zwar häufig Spalten zugeordnet werden konnten aber viele leere Zellen enthalten, als auch Merkmale, die nur wenigen Spalten zugeordnet werden konnten, obwohl sie ggf. wenige leere Zellen enthalten.

Entität	Merkmal	Anzahl	leer
Maßnahme		1304	47,77%
Maßnahme	Beschreibung	49	82,30%
Maßnahme	Tätigkeitsart	20	30,87%
Maßnahme	Arbeitszeit	20	4,45%
Maßnahme	Datum	19	0,12%
Maßnahme	Firma	18	10,15%
Maßnahme	Fehler	18	66,53%
Maßnahme	Kosten	16	6,69%
Maßnahme	E/F-Leistung	15	4,09%
Maßnahme	REF	13	0,03%
Maßnahme	ID	10	0,00%
Maßnahme	Auszeit	9	73,51%
Maßnahme	Betriebsstunden	7	47,63%
Maßnahme	Zählerstand	7	47,55%

- 3A-40 (Trennzeichen) → 20 (Leerzeichen)
- 41-5A (Großbuchstaben) → 61-7A (Kleinbuchstaben)
- 5B-60 (Trennzeichen) → 20 (Leerzeichen)
- 61-7A (Kleinbuchstaben) belassen
- 7B-7F (Trennzeichen) → 20 (Leerzeichen)
- 80-FF (Sonderzeichen) → 20 (Leerzeichen)

3. Schließlich werden mehrfache Leerzeichen durch einfache Leerzeichen ersetzt.

## 3.5 Normierung der Daten

Bei normierten Daten werden gleiche oder genauer als gleich angesehene Dinge auch gleich bezeichnet. Zum Beispiel schreibt der eine bei nicht normierten Daten „Mausefalle“ und der andere „Falle, Klapp- für Nagetier, grau“.

Hierbei sind zwei Entscheidungen zu treffen: 1. welche Begriffe sollen zusammengefasst werden und 2. mit welchem Begriff soll diese Zusammenfassung benannt werden. Man kann die eine Frage nicht beantworten, ohne über die Antwort der anderen Frage zu verfügen. Die Fragen lassen sich also nicht nacheinander beantworten – vielmehr müssen die Antworten in einem iterativen Prozess gefunden werden. Mehr jedoch dazu später.

Wünschenswert ist natürlich auch, dass nicht nur gleiche Dinge „gleich“ bezeichnet werden, sondern auch Unterschiedliche „unterschiedlich“. Handelt es sich zum Beispiel bei einem Monitor um einen Patienten- oder um einen Computer-Monitor? Obwohl

grundsätzlich gilt, dass dieses Fehlen an Detailinformation in den Daten nicht im Rahmen dieser Untersuchung behoben werden kann, so kann man es jedoch manchmal dadurch etwas abfangen, dass man Begriffe verwendet, die über eine hierarchische Struktur verfügen – für die es also Ober- und Unterbegriffe gibt. Dann kann bei ungenauer Beschreibung einfach ein ungenauere Oberbegriff verwendet werden, und diese vorhandene Teilinformation kann dann genutzt und muss nicht verworfen werden.

Da es durchaus vorkommen kann, dass zu einem Zelleninhalt der Originaldaten mehrere verschiedene normierte Begriffe zugeordnet werden, werden sämtliche Zuordnungen im folgenden Format gespeichert: Zu jedem gefundenen normierten Begriff werden in einer neuen Tabelle das zugehörige Merkmal (bzw. die zugehörige Spalte), die zugehörige Zeile und die zugehörige Tabelle gespeichert. Dies hat auch den Vorteil, dass nicht zu jeder Zelle der Originaldaten ein Datensatz angelegt werden muss, wenn kein normierter Begriff zugeordnet werden kann – dies spart Speicherplatz. Der Nachteil, dass Datenbankabfragen der neuen Tabellen länger brauchen, fällt bei der Erstellung der Tabellen noch nicht ins Gewicht.

Das eben Gesagte gilt nur für kategoriale bzw. nominal skalierte Merkmale. Wenn im folgenden von der Normierung metrischer Größen gesprochen wird, so sind eher ein Import in ein einheitliches Datenformat und eine gleichzeitige Validitätsprüfung der Daten gemeint.

#### 3.5.1 Normierung metrischer Merkmale

Metrische Merkmale lassen sich leichter analysieren als nominal skalierte Merkmale, da für diese arithmetische Operationen zur Verfügung stehen. Diesen Vorteil kann man aber nur dann nutzen, wenn die Daten in einem Variablenformat gespeichert sind, das Rechenoperationen zulässt. Deshalb ist es nötig, die im Textformat vorliegenden metrischen Größen in ein neues Variablenformat zu transformieren.

##### Preisangaben

Kaufpreise oder Durchführungskosten von Maßnahmen, die in den Originaldaten noch in Deutsche Mark angegeben sind, werden ins Integerformat oder Ganzzahlformat gewandelt und in Euro umgerechnet. Die Ungenauigkeit durch das nötige Runden sind vertretbar. Beim Umwandeln ist auf wechselnde Dezimaltrennungszeichen (Punkt oder Komma) zu achten. Kaufpreise unter 100 € oder über 10 Mio. € werden ignoriert. Bei Kosten von Maßnahmen werden Beträge von 10 € bis 1 Mio. € akzeptiert.

##### Datumsangaben

Datumsangaben wie das Kaufdatum oder der Durchführungszeitpunkt können durch einen Trick verlustfrei in das Integerformat gewandelt werden: statt der unterschiedlichen Schreibweisen „28-06-1993“ oder „19. Juni. 98“ etc werden einfach die vergangenen Tage nach dem 01.01.1900 ganzzahlig abgespeichert. Dies ist für die direkte Lesbarkeit der Daten zwar nicht förderlich, vereinfacht die Analyse jedoch erheblich. Es werden nur Datumsangaben zwischen dem 01.01.1900 und dem 01.01.2002 akzeptiert. Bei dieser Umwandlung ist viel Handarbeit nötig.

### 3.5.2 Normierung kategorialer Merkmale

Die hier zu normierenden kategorialen Merkmale liegen in Freitext-Form vor, d.h. oft sind es Teilzeichenketten oder Stichwörter, die über die Einordnung entscheiden. Zunächst geht es darum, diese Stichwörter zu finden, um sie dann in Gruppen einzuteilen. Auf diese Weise löst sich auch das Problem der unterschiedlichen Bezeichnung bei Verwendung von Synonymen.

Zunächst werden die Spalten „Abteilung“, „Hersteller“ und „Arbeitsbeschreibung“ in getrennte Textdateien exportiert. Die Spalte „Gerätebezeichnung“ wird später gesondert behandelt. Dies verkleinert die Größe der zu verarbeitenden Dateien. Die Zellinhalte werden jeweils als einzelne Zeile gespeichert.

In diesen Dateien werden nun alle Leerzeichen durch Tiefstriche („\_“) ersetzt und zusätzlich am Ende und am Anfang jeder Zeile ein Tiefstrich eingefügt. Das Einfügen am Ende und am Anfang einer Zeile führt dazu, dass ein Wortende oder -anfang am Anfang bzw. am Ende einer Zeile im folgenden nicht anders behandelt wird als ein Wortanfang oder -ende mitten in einer Zeile.

Im nächsten Schritt wird nun jede Zeile in überlappende Teilzeichenketten der Längen 4 bis 25 Zeichen zerlegt, die zeilenweise in eine neue Datei geschrieben werden. Aus der Zeile `_pumpe_` z.B. werden so die folgende Zeilen (hier durch Komma getrennt) generiert:

```
_pumpe_, _pumpe, pumpe_, _pump, pumpe, umpe_, _pum, pump, umpe, mpe_
```

Teilzeichenketten, die länger als 25 oder kürzer als 4 Zeichen sind, werden nicht generiert, weil die langen Zeichenketten zu selten auftreten und die kurzen kaum Information transportieren. Außerdem würden die kurzen Teilzeichenketten wegen ihrer großen Anzahl die Verarbeitung erheblich verlangsamen.

Nun wird zu jeder eben erstellten Teilzeichenkettenliste eine Liste aller vorkommenden Teilzeichenketten jeweils mit Häufigkeit erstellt. Aus technischen Gründen (Access kann keine Datenbanken größer als 2 GB verarbeiten) werden Teilzeichenketten, die seltener als 10 mal auftauchen, nicht weiter berücksichtigt. Die weiteren Analysen werden dadurch kaum beeinflusst.

Mit Hilfe dieser Häufigkeitstabellen kann nun die Häufigkeit der häufigsten Zeichenkette ermittelt werden, die eine bestimmte Zeichenkette enthält. Hierzu ein Beispiel: die Zeichenkette „umpe\_“ tritt 11 284 mal auf. Die häufigste Zeichenkette, die die Zeichenkette „umpe\_“ enthält ist „pumpe“ mit Häufigkeit 11 230. Diese häufigste Zeichenkette ist immer um ein Zeichen länger als die ursprüngliche Zeichenkette. Diese Zahlen werden nun für jede Teilzeichenkette bestimmt.

Bildet man nun die Differenz zwischen der ermittelten Häufigkeit einer Teilzeichenkette und der Häufigkeit der häufigsten umfassenden Teilzeichenkette, so zeigt sich, dass Teilzeichenketten, die eine große Differenz aufweisen, für die weiteren Analysen von Interesse sind. Woran liegt das?

In der Abbildung 3.2 sind ausgehend von der Teilzeichenkette „umpe“ jeweils die häufigsten umfassenden Teilzeichenketten mit Häufigkeit angegeben. Man sieht, dass die Häufigkeiten von Teilzeichenketten, die interessante Teilzeichenketten („pumpe“, „spritzenpumpe“) umfassen, stärker abfallen als andere.

Im Deutschen werden Begriffe aus verschiedenen Worten, Präpositionen oder Wortteilen zusammengesetzt. Wird ein zusammengesetzter Begriff nun Buchstabe für Buch-

stabe aufgebaut, so gibt es beim Anfügen von Buchstaben innerhalb der Wortteile kaum Variationsmöglichkeiten. Wenn jedoch ein Wortteil fertig aufgebaut ist, dann stehen beim nächsten Buchstaben viele Möglichkeiten zur Verfügung, weil der Wortteil mit vielen anderen Wortteilen kombiniert werden kann. Dadurch, dass viele Möglichkeiten bestehen, ist die Anzahl, mit der die einzelnen Möglichkeiten eintreten, kleiner und damit die Differenz zwischen den Anzahlen höher.

#### **Merkmal „Arbeitsbeschreibung“**

Für das Merkmal „Arbeitsbeschreibung“ ergibt sich damit Tabelle 3.8. Es sind die Teilzeichenketten nach Häufigkeit sortiert. Wie man sieht, ist die Häufigkeit kein gutes Kriterium für interessante Teilzeichenketten. In der nächsten Tabelle 3.9 sind die Teilzeichenketten nach Differenz sortiert. Schon an dieser kleinen Auswahl ist deutlich zu sehen, dass der Anteil interessanter Teilzeichenketten größer ist.

Die Einteilung der gefundenen Begriffe in die Gruppen Objekt, Verb und Attribut erleichtert die Gruppierung von Synonymen und die anschließende Vergabe eines gemeinsamen, normierten Begriffes. Insgesamt wurde ein Zeitaufwand von ca. 100 Tagen in diese Arbeiten investiert. 60% der Zeit flossen in das Finden eines oben beschriebenen Algorithmus, der die Datenmengen überhaupt verarbeitbar macht; 30% der Zeit wurde für das Finden von Oberbegriffen benötigt, die die vorhandenen Informationen knapp aber ausreichend detailliert zusammenfassen; die restlichen 10% der Zeit wurden für das Zuordnen der normierten Begriffe zu den vorhandenen Teilzeichenketten benötigt.

Die Trennung zwischen dem Finden normierter Begriffe und dem Zuordnen dieser ist nicht so scharf möglich, wie die obigen Zeitangaben vielleicht vermuten lassen. Wie schon im vorherigen Abschnitt angedeutet wurde, ist das Finden und Zuordnen von Begriffen ein iterativer Prozess. Ohne Begriffe kann man nichts zuordnen – jedoch kann man ohne Zuordnungen auch nicht entscheiden, welche Begriffe die vorliegenden Daten gut zusammenfassen.

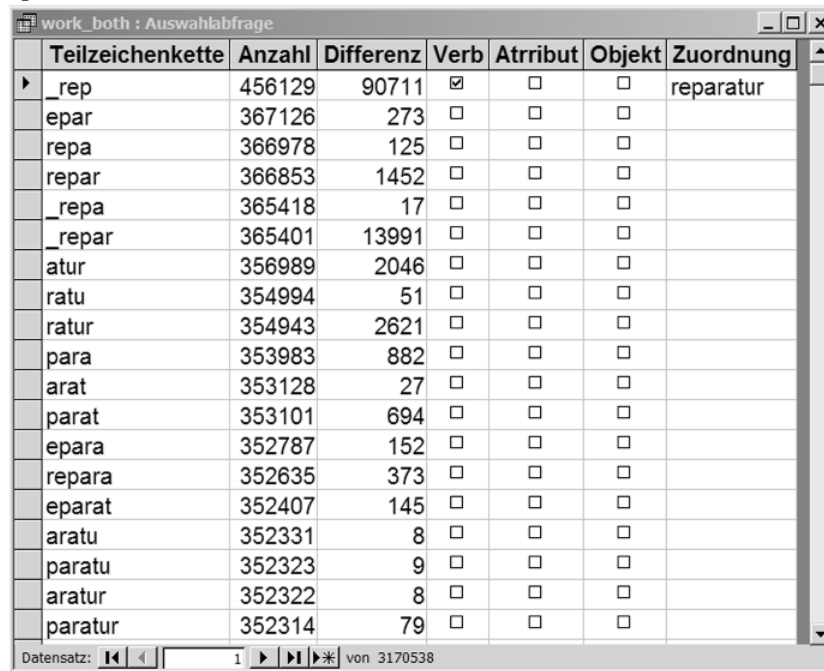
Versucht man, die normierten Begriffe zu systematisieren, so stellt man bei der Zuordnung fest, dass man sehr unterschiedlich und schwach besetzte Gruppen erhält. Orientiert man sich an den Häufigkeiten der Nennungen in den Originaldaten, so erhält man sehr unsystematische normierte Begriffe. Eine Normierung kann immer nur ein Kompromiss zwischen diesen Extremen sein.

Die in dieser Arbeit verwendeten normierten Arbeitsbeschreibungen sind im Anhang B.1 mit den ihnen zugeordneten Teilzeichenketten angegeben. Eine Maßnahme kann selbstverständlich mehreren normierten Arbeitsbeschreibungen zugeordnet werden.

#### **Merkmal „Hersteller“**

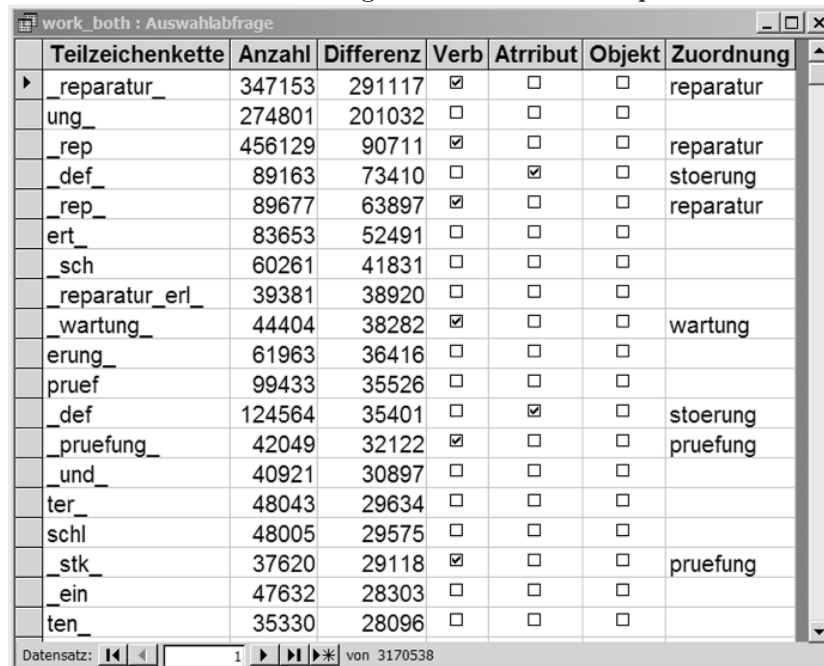
Beim Merkmal „Hersteller“ kann genau wie beim Merkmal „Arbeitsbeschreibung“ vorgegangen werden. Der einzige vereinfachende Unterschied ist, dass interessante Teilzeichenketten praktisch immer begrenzt durch „\_“ vorliegen, was die Anzahl der durchzusehenden Teilzeichenketten erheblich verringert. Insgesamt wurden etwa 800 unterschiedliche Firmen zugeordnet.

Tabelle 3.8: In dieser Tabelle wurden die Teilzeichenketten nach Häufigkeit sortiert. Es ist deutlich zu erkennen, dass die Häufigkeit einer Teilzeichenkette nicht geeignet ist, interessante Teilzeichenketten heraus zu filtern. Erläuterungen zu den weiteren Spalten siehe Text.



Teilzeichenkette	Anzahl	Differenz	Verb	Attribut	Objekt	Zuordnung
_rep	456129	90711	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	reparatur
epar	367126	273	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
repa	366978	125	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
repar	366853	1452	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
_repa	365418	17	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
_repar	365401	13991	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
atur	356989	2046	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
ratu	354994	51	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
ratur	354943	2621	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
para	353983	882	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
arat	353128	27	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
parat	353101	694	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
epara	352787	152	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
repara	352635	373	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
eparat	352407	145	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
aratu	352331	8	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
paratu	352323	9	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
aratur	352322	8	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
paratur	352314	79	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Tabelle 3.9: In dieser Tabelle wurden die Teilzeichenketten nach Differenz zwischen Häufigkeit der Teilzeichenkette und Häufigkeit der häufigsten umfassenden Teilzeichenkette sortiert. Es ist deutlich zu erkennen, dass diese Differenz besser geeignet ist, interessante Teilzeichenketten herauszufiltern, als die bloße Häufigkeit einer Teilzeichenkette. Erläuterungen zu den weiteren Spalten siehe Text.



Teilzeichenkette	Anzahl	Differenz	Verb	Attribut	Objekt	Zuordnung
_reparatur_	347153	291117	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	reparatur
ung_	274801	201032	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
_rep	456129	90711	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	reparatur
_def_	89163	73410	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	stoerung
_rep_	89677	63897	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	reparatur
ert_	83653	52491	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
_sch	60261	41831	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
_reparatur_erl_	39381	38920	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
_wartung_	44404	38282	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	wartung
erung_	61963	36416	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
pruef	99433	35526	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
_def	124564	35401	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	stoerung
_pruefung_	42049	32122	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	pruefung
_und_	40921	30897	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
ter_	48043	29634	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
schl	48005	29575	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
_stk_	37620	29118	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	pruefung
_ein	47632	28303	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
ten_	35330	28096	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

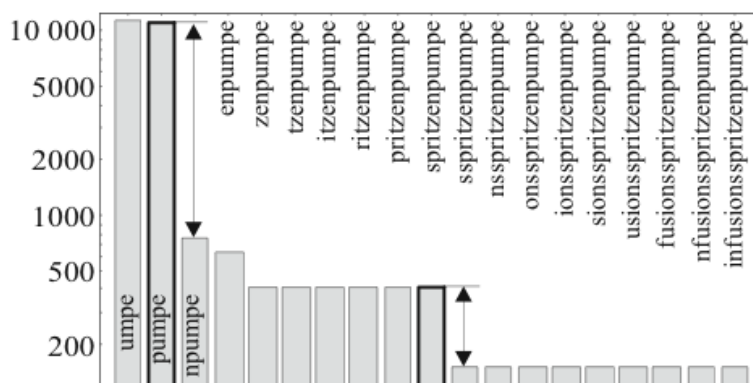


Abbildung 3.2: Häufigkeiten von Teilzeichenketten als Balkendiagramm dargestellt. Die häufigste umfassende Teilzeichenkette steht rechts neben der enthaltenen Teilzeichenkette. Große Sprünge zwischen den Anzahlen deuten auf interessante Teilzeichenketten hin (Erläuterung im Text). Diese sind markiert.

### Merkmal „Abteilung“

Auch beim Merkmal „Abteilung“ kann genau wie beim Merkmal „Arbeitsbeschreibung“ vorgegangen werden. Auch hier sind viele wichtige Stichworte abgekürzt oder liegen in zusammengesetzter Form („kinderchirurgie“ statt „paediatric“ und „chirurgie“) vor. Damit ist klar, dass eine originale Abteilungsbezeichnung mehreren normierten Abteilungen zugeordnet werden kann. Insgesamt wurden 76 unterschiedliche Abteilungen zugeordnet.

### Nachbearbeitung der Zuordnungen

Die bis jetzt erstellten Zuordnungen von Teilzeichenketten zu den einheitlichen Begriffen müssen nun noch auf die Daten angewendet werden. Dazu werden die Zuordnungen in Textdateien exportiert, die die Teilzeichenkette und die Zuordnung als Spalte enthalten.

Diese Dateien werden auf überflüssige oder widersprüchliche Zuordnungen durchsucht. Eine überflüssige Zuordnung liegt vor, wenn es eine identische Zuordnung mit kürzerer Teilzeichenkette gibt (z.B. ist die Zuordnung „\_repara“ → „Reparatur“ überflüssig, wenn es eine Zuordnung „\_rep“ → „Reparatur“ gibt).

Eine widersprüchliche Zuordnung liegt vor, wenn es zwei unterschiedliche Zuordnungen mit gleicher Teilzeichenkette gibt.

### 3.5.3 Sonderfall: Normierung der Gerätebezeichnung

Im Rahmen dieser Arbeit wurde erst versucht, mit dem eben geschilderten Vorgehen, einen normierten Gerätekatalog aufzustellen. Diese Versuche mussten jedoch aufgegeben werden, da ein solcher Katalog zu viele Begriffe enthält und darüber hinaus auch noch eine Gruppierung in Gerätearten zulassen sollte – dies war vom Arbeitsaufwand her nicht im Rahmen dieser Arbeit zu bewältigen.

Aus diesem Grunde wurde versucht, mit bereits fertigen Gerätekatalogen zu arbeiten. Aus den getesteten Katalogen konnte besonders der EMTEC-Katalog überzeugen, da

er schon von Haus aus Synonyme mitbringt und eine sehr übersichtliche Struktur in vier Hierarchieebenen besitzt.

Zwar macht das Vorhandensein eines Kataloges die mühselige, kaum zu systematisierende Tätigkeit beim Erstellen eines neuen Kataloges überflüssig, aber dafür wird das Assoziieren der Gerätebezeichnungen mit den vorhandenen Daten schwieriger, weil der normierte Begriffskatalog nicht mehr aus dem Datenbestand gewachsen ist, auf den er angewendet werden soll. Aus diesem Grunde unterscheidet sich das Vorgehen beim Normieren des Merkmales „Gerätebezeichnung“ von den anderen kategorialen Merkmalen.

### Vorbereitung

Zunächst muss der EMTEC-Katalog für die weitere Verarbeitung importiert werden. Da er auch als Textdatei vorliegt, ist der Import problemlos möglich. Als zweckmäßiges Format, das alle für die folgenden Analysen notwendigen Informationen enthält, hat sich eine Textdatei herausgestellt, die die Spalten

- EMTEC-Katalog-ID (enthält auch die Hierarchieebene)
- Bezeichnung
- Synonym (Ja / Nein)

enthält. Alle weiteren Informationen des Kataloges werden im folgenden nicht benötigt.

Eine erste Analyse dieser Textdatei zeigt, dass die vier Hierarchieebenen des EMTEC-Katalogs folgende Anzahlen an Gerätearten besitzt (die Anzahl der Gerätearten aus der Medizintechnik (MT) sind in Klammern angegeben):

1. **Ebene:** 13 Gerätearten (MT: 1)
2. **Ebene:** 101 Gerätearten (MT: 13)
3. **Ebene:** 553 Gerätearten (MT: 155)
4. **Ebene:** 1760 Gerätearten (MT: 924)

Diese Anzahl an Gerätearten steht einer Anzahl von über 150 000 Geräte-Entitäten aus der Datenbank gegenüber. Manuell sind Zuordnungen dieser Größenordnung im Rahmen dieser Arbeit nicht mehr möglich. Aus diesem Grunde wird auch hier das Verfahren weitgehend automatisiert. Dies führt auch zu einer weniger subjektiven Zuordnung.

Zunächst werden gleiche Gerätebezeichnungen in den Originaldaten zusammengefasst und mit Häufigkeit versehen in einer neuen Textdatei gespeichert. Die Anzahl der verschiedenen Gerätebezeichnungen beträgt immerhin nur noch 11 105.

Wie beim Normieren der anderen kategorialen Merkmale werden die Gerätebezeichnungen des EMTEC-Kataloges und der Originaldaten nach Säubern und Begrenzen mit Tiefstrichen („\_“, siehe Normierung der Gerätebezeichnung) jeweils in überlappende Dreier-Teilzeichenketten (Trigramme) zerlegt. Dann werden die übereinstimmenden Dreier-Teilzeichenketten für jede Kombination von Originalgerätebezeichnung und EMTEC-Gerätebezeichnung bestimmt ([14] Seite 386).

Mit Hilfe dieser Informationen sollen die besten Übereinstimmungen von EMTEC-Bezeichnungen und originalen Gerätebezeichnungen gefunden werden. Als Gütekriterium die Anzahl der übereinstimmenden Dreier-Teilzeichenketten zu verwenden, erweist sich als ungünstig, da längere Begriffe dabei bevorzugt werden – was nicht immer erwünscht ist. Auch das Heranziehen der Anzahl der untereinander unterschiedlichen, aber in beiden Begriffen übereinstimmenden Teilzeichenketten zeigt nicht so gute Ergebnisse wie folgendes Vorgehen: Es wird die Anzahl der Buchstaben der EMTEC-Bezeichnung ermittelt. Dann wird ermittelt, welche Buchstaben in identischen Dreier-Teilzeichenketten aus der Original-Bezeichnung enthalten sind. Ist ein Buchstabe in mehreren Teilzeichenketten enthalten, so wird er nur einmal gezählt. Das Verhältnis von betroffenen zu allen Buchstaben der EMTEC-Bezeichnung ist ein brauchbares Kriterium zur Beurteilung der Ähnlichkeit.

Es werden mit dieser Methode auch die EMTEC-Bezeichnungen durchsucht, die als Synonym gekennzeichnet sind. Die Synonyme werden jedoch nicht selber als normierte Begriffe verwendet, sondern nur den Synonymen zugeordnete EMTEC-Hauptbegriffe.

Die jeweils 20 besten Treffer aus dem EMTEC-Katalog werden je einer Gerätebezeichnung aus den Originaldaten zugeordnet und mit Quotient in einer Textdatei gespeichert. Diese Textdatei kann man in Access importieren und unter Verwendung von Unterdatenblättern übersichtlich darstellen (siehe Tabelle 3.10).

#### **Manuelle Zuordnung**

Durch die in der Access-Datenbank abgelegten Informationen lassen sich Gerätebezeichnungen effektiv normalisieren. Zuerst werden per Datenbank-Abfrage Gerätebezeichnungen mit einer bestimmten Zeichenfolgenkombination (wie z.B. „defi“ und „moni“ in Tabelle 3.10) per Auswahlabfrage aus allen Gerätebezeichnungen selektiert, um dann gemeinsam über eine der gefundenen EMTEC-Bezeichnungen zugeordnet zu werden – man muss also nicht für jede Gerätebezeichnung nach der besten Zuordnung einzeln suchen. Die Zuordnung zu den EMTEC-Bezeichnungen kann Dank der im Unterdatenblatt angegebenen 20 besten Treffer schnell und effizient erfolgen.

Insgesamt waren hierfür ca. 60 Tage Arbeit nötig. Etwa 70% der Zeit wurde für das Finden eines geeigneten Verfahrens benötigt, das diese Datenmengen verarbeitbar macht. In der restlichen Zeit wurde dann die eigentliche Normierung durchgeführt.

#### **Nachbereitung der Zuordnungen**

Wie bei den anderen kategorialen Merkmalen müssen die erstellten Zuordnungen von Originalbegriffen zu den EMTEC-Begriffen nun noch auf die Daten angewendet werden. Dazu werden auch hier die Zuordnungen in Textdateien exportiert, die die Originalbezeichnung und die Zuordnung als Spalte enthalten.

Eine Überprüfung dieser Dateien auf überflüssige oder widersprüchliche Zuordnungen ist hier nicht nötig. Jeder Zelle der Originaldaten wird nach passenden Zuordnungen durchsucht. Wird eine passende gefunden, so wird die Fundstelle wieder durch Angabe der eindeutigen Kennung (Zeile) und Merkmal (Spalte) gekennzeichnet und in einer neuen Textdatei abgelegt.

Tabelle 3.10: Darstellung einer Auswahlabfrage der originalen Gerätebezeichnungen mit den Teilzeichenketten „defi“ und „monitor“. Beispielhaft sind die Zuordnungen der EMTEC-Gerätebezeichnungen in einem Unterdatenblatt dargestellt. Das Unterdatenblatt ist absteigend nach der Übereinstimmung von EMTEC- und Originalbezeichnung (Spalte „fit“) sortiert. In der Spalte „EMTEC-Gruppe“ werden die zugeordneten Normierungen abgelegt. Die Spalte „Häufigkeit“ enthält die Anzahl der Vorkommen der originalen Gerätebezeichnung in den Daten.

Zellinhalt					Häufigkeit	EMTEC_Gruppe																																																																																				
*_monitor_defi_monitor_defi_					1	WEAC																																																																																				
*_ekg_monitor_defibrillator_mit_ekg_monitor_inv_nr_088411140_					1	WEAC																																																																																				
*_defibrillator_monitor_					1	WEAC																																																																																				
*_defibrillator_defibrillator_m_monitor_lp5_					1	WEAC																																																																																				
*_defibrillator_defibrillator_lp6s_facht_st_na_ekg_monitor_0018399_					1	WEAC																																																																																				
*_defibrillator_defibrillator_lp6s_ekg_monitor_0032814_					1	WEAC																																																																																				
<table border="1"> <thead> <tr> <th>EMTEC_Name</th> <th>Synonym</th> <th>EMTEC_Gruppe</th> <th>fit</th> </tr> </thead> <tbody> <tr><td>Fibrillator</td><td>J</td><td>WEAH</td><td>0,82</td></tr> <tr><td>Monitor</td><td>J</td><td>WCQH</td><td>0,71</td></tr> <tr><td>Monitor</td><td>J</td><td>PDMB</td><td>0,71</td></tr> <tr><td>Monitor</td><td>J</td><td>PDHD</td><td>0,71</td></tr> <tr><td>Monitor</td><td>J</td><td>P</td><td>0,71</td></tr> <tr><td>Monitor</td><td>J</td><td>GAJB</td><td>0,71</td></tr> <tr><td>Monitor</td><td>J</td><td>GADB</td><td>0,71</td></tr> <tr><td>Defibrillator, mit Monitor</td><td>N</td><td>WEAC</td><td>0,62</td></tr> <tr><td>Defibrillator, ohne Monitor</td><td>N</td><td>WEAB</td><td>0,59</td></tr> <tr><td>Notfalldefibrillator</td><td>J</td><td>WEAC</td><td>0,55</td></tr> <tr><td>Notfalldefibrillator</td><td>J</td><td>WEAB</td><td>0,55</td></tr> <tr><td>EKG-Monitor</td><td>J</td><td>WDBN</td><td>0,55</td></tr> <tr><td>Handmonitor</td><td>J</td><td>WHAE</td><td>0,45</td></tr> <tr><td>EMG-Monitor</td><td>N</td><td>WDED</td><td>0,45</td></tr> <tr><td>EEG-Monitor</td><td>N</td><td>WDEB</td><td>0,45</td></tr> <tr><td>N<sup>2</sup>O-Monitor</td><td>J</td><td>WDDE</td><td>0,45</td></tr> <tr><td>HMV-Monitor</td><td>J</td><td>WDBO</td><td>0,45</td></tr> <tr><td>Farbmonitor</td><td>J</td><td>PDHD</td><td>0,45</td></tr> <tr><td>S/W-Monitor</td><td>J</td><td>PDHB</td><td>0,45</td></tr> <tr><td>Defibrillator, implantierbar</td><td>N</td><td>WEAD</td><td>0,43</td></tr> </tbody> </table>					EMTEC_Name	Synonym	EMTEC_Gruppe	fit	Fibrillator	J	WEAH	0,82	Monitor	J	WCQH	0,71	Monitor	J	PDMB	0,71	Monitor	J	PDHD	0,71	Monitor	J	P	0,71	Monitor	J	GAJB	0,71	Monitor	J	GADB	0,71	Defibrillator, mit Monitor	N	WEAC	0,62	Defibrillator, ohne Monitor	N	WEAB	0,59	Notfalldefibrillator	J	WEAC	0,55	Notfalldefibrillator	J	WEAB	0,55	EKG-Monitor	J	WDBN	0,55	Handmonitor	J	WHAE	0,45	EMG-Monitor	N	WDED	0,45	EEG-Monitor	N	WDEB	0,45	N <sup>2</sup> O-Monitor	J	WDDE	0,45	HMV-Monitor	J	WDBO	0,45	Farbmonitor	J	PDHD	0,45	S/W-Monitor	J	PDHB	0,45	Defibrillator, implantierbar	N	WEAD	0,43		
EMTEC_Name	Synonym	EMTEC_Gruppe	fit																																																																																							
Fibrillator	J	WEAH	0,82																																																																																							
Monitor	J	WCQH	0,71																																																																																							
Monitor	J	PDMB	0,71																																																																																							
Monitor	J	PDHD	0,71																																																																																							
Monitor	J	P	0,71																																																																																							
Monitor	J	GAJB	0,71																																																																																							
Monitor	J	GADB	0,71																																																																																							
Defibrillator, mit Monitor	N	WEAC	0,62																																																																																							
Defibrillator, ohne Monitor	N	WEAB	0,59																																																																																							
Notfalldefibrillator	J	WEAC	0,55																																																																																							
Notfalldefibrillator	J	WEAB	0,55																																																																																							
EKG-Monitor	J	WDBN	0,55																																																																																							
Handmonitor	J	WHAE	0,45																																																																																							
EMG-Monitor	N	WDED	0,45																																																																																							
EEG-Monitor	N	WDEB	0,45																																																																																							
N <sup>2</sup> O-Monitor	J	WDDE	0,45																																																																																							
HMV-Monitor	J	WDBO	0,45																																																																																							
Farbmonitor	J	PDHD	0,45																																																																																							
S/W-Monitor	J	PDHB	0,45																																																																																							
Defibrillator, implantierbar	N	WEAD	0,43																																																																																							
*_defibrillator_defibrillator_lp6s_ekg_monitor_0019966_					1	WEAC																																																																																				
*_defibrillator_defibrillator_lp6s_ekg_monitor_0018419_					1	WEAC																																																																																				
*_defibrillator_defibrillator_lp6s_ausgemustert_ekg_monitor_0018394_					1	WEAC																																																																																				
*																																																																																										

Datensatz: 1 von 11

### 3.5.4 Integration der normierten Daten

Alle Zuordnungen von Originaldaten und normierten Begriffen liegen nun in einem einheitlichen Format vor: immer, wenn ein Originalbegriff durch einen normierten Begriff ersetzt werden kann, so sind in einer Tabelle die Zeile, Spalte und der normierte Begriff angegeben. Aus diesen Informationen lassen sich Tabellen konstruieren, die genauso aufgebaut sind wie die Originaldaten, jedoch die normierten Begriffe enthalten.

Diese Tabellen werden als Grundlage für die weiteren Analysen in einer Access-Datenbank abgelegt.

## 3.6 Skalentransformationen

Wie in Abschnitt 2.3.1 dargestellt, können Zufallsvariable metrische, ordinale und nominale Größen sein. Bei den bisherigen Normierungen wurden zwar der Wertebereich der Zufallsvariablen in einen anderen abgebildet, jedoch blieb stets das Skalenniveau erhalten. Hier sollen Abbildungen des Wertebereiches einer Zufallsvariablen vorgestellt werden, die das Skalenniveau verändern.

In Abschnitt 2.3.1 wurde auch dargestellt, dass die Skalenniveaus eine natürliche Reihenfolge besitzen und man deswegen von höheren und niedrigeren Niveaus sprechen kann.

### 3.6.1 Absenken des Skalenniveaus

Von Absenken des Skalenniveaus spricht man meistens, wenn metrisch skalierte in kategorial skalierte Zufallsvariable umgewandelt werden sollen. Diesen Verlust von Informationen geht man im allgemeinen nur ein, weil die Daten aus anderen Gründen mit einem Verfahren untersucht werden sollen, das metrische Größen nicht verarbeiten kann.

Zum Beispiel wurden in dieser Arbeit metrische Merkmale wie der Kaufpreis kategorial umskaliert, um diese Merkmale in Kreuztabellen darstellen oder mit Hilfe der Assoziationsanalyse untersuchen zu können.

Das eigentliche Umskalieren läuft dabei als Einteilung der metrischen Größen in unterschiedlich benannte Intervalle ab. Als günstig erweist es sich, die Intervallgrenzen so zu wählen, dass die Besetzungszahlen der einzelnen Intervalle nicht zu unterschiedlich sind. Die Bezeichnung des Intervalls kann man so wählen, dass die Grenzen daraus hervorgehen.

### 3.6.2 Anheben des Skalenniveaus

Das Anheben des Skalenniveaus gestaltet sich sehr viel schwieriger, da hierbei nicht Informationen der beteiligten Zufallsvariablen verworfen werden können sondern neu aus anderen Quellen hinzugewonnen werden müssen. Eine solche Quelle ist häufig die Forderung, dass die umzuskalierenden Zufallsvariablen nach einer bestimmten Funktion verteilt sind.

Zum Beispiel werden im Rahmen der in dieser Arbeit erprobten Lancaster-Skalierung gleichzeitig zwei kategorial skalierte Merkmale auf metrisches Skalenniveau angehoben. Die dazu zusätzlich nötigen Informationen ergeben sich aus der willkürlichen Forderung, dass die kategorialen Daten – die in Form einer Kontingenztabelle vorliegen – möglichst gut an eine Binormalverteilung angepasst werden (siehe [4] ab Seite 282).

Zwar stehen durch die Skalentransformation der Daten nun eine Vielzahl von neuen Analysen zur Verfügung, jedoch zeigte sich im Verlauf dieser Arbeit, dass nicht das Fehlen geeigneter Analysen sondern die Qualität der Daten mit sehr wahrscheinlich systematischen Fehlern das Hauptproblem ausmachen. Hinzu kommt noch, dass die gewonnen Ergebnisse durch die aufwendige Berechnung nicht mehr ohne Weiteres nachvollziehbar und plausibel sind. Da den genannten Nachteilen kaum Vorteile gegenüberstehen, wurde im Rahmen dieser Arbeit ein Anheben des Skalenniveaus der zu untersuchenden Zufallsvariablen nicht weiter verfolgt.

## 3.7 Datenanreicherung

Unter Datenanreicherung versteht man das Aufspüren und Integrieren externer Datenquellen in die vorhandenen Daten. In dieser Arbeit wurden zwei weitere Datenquellen in die vorliegenden Daten integriert.

### 3.7.1 Deutsches Krankenhaus Adressbuch

Für diese Arbeit interessante Informationen kommen nicht ausschließlich aus den eingesandten Daten. Die eingesandten Daten enthalten meist keine Informationen über das Krankenhaus selber, wie zum Beispiel die Krankenhausart oder die Bettenzahl. Diese Informationen können aber leicht aus dem Deutschen Krankenhaus Adressbuch (DKA) entnommen werden. Daten für diese Arbeit wurden aus dem Jahrgang 2000 entnommen. Im Einzelnen wurden die Krankenhausart und die Bettenzahl für jedes Krankenhaus ermittelt und den erhobenen Daten zugeordnet.

### 3.7.2 EMTEC-Gerätecatalog

Wie schon im Abschnitt 3.5.3 erwähnt, wird zur Normierung der Gerätebezeichnungen der EMTEC-Gerätecatalog verwendet. Durch die Normierung werden die EMTEC-Daten zwar nicht auf einen gewöhnlichen Import in die erhobenen Daten integriert, aber dennoch werden die vorhandenen Daten durch Zugriff auf externe Daten angereichert – darum auch hier dessen Erwähnung.

### 3.7.3 Sonstige Datenquellen

Andere Quellen interessanter Informationen könnten Datenbanken der Gerätehersteller oder externer Servicedienstleister sein. Oft machen die Datensätze den Eindruck, als ob in den einzelnen Krankenhäusern Daten über Geräte in mühevoller Einzelarbeit erfasst werden, die beim Hersteller schon vorliegen.

Auch ist im Allgemeinen die Detaillierung der Abbildung von Fremdleistungen im Gegensatz zu Eigenleistungen in den Daten nicht so hoch – viele interessante Informationen verschwinden in den Datenbanken der Servicedienstleister und sind dem Betreiber nicht zugänglich.

Da neben dem Aufspüren und Integrieren weiterer externer Datenquellen auch noch rechtliche Aspekte zu beachten sind, wurde im Rahmen dieser Arbeit auf die Integration weiterer Datenquellen verzichtet. Das bedeutet aber nicht, dass hier noch große Informationspotentiale zu erwarten sind.

# 4 Analyse der Daten

## 4.1 Vorbemerkungen

In diesem Abschnitt werden Erkenntnisse aus den folgenden Analysen vorgezogen, die die Analysen selbst beeinflussen. Oft stellte sich erst beim analysieren der Daten heraus, dass bestimmte Methoden aufgrund der vorliegenden Datenqualität nicht oder nur mit Vorsicht anwendbar sind. Die Analyse ist also ein iterativer Prozess. Dieser Teil ist notwendig, da die Analysen nicht in chronologischer sondern inhaltlicher Ordnung dargestellt werden sollen.

### 4.1.1 Auswahl der Merkmale

Ob ein Merkmal interessant ist, stellt sich ggf. erst nach einer Analyse heraus. Da nicht alle eingeschickten Daten eingehend untersucht werden konnten bzw. an allen Daten die für die Analysen notwendigen Vorbereitungen vorgenommen werden konnten, muss eine Vorauswahl erfolgen. Diese erfolgte in Abschnitt 3.3.3. Die in Tabellen 3.6 und 3.7 angegebenen Merkmale wurden in der Hoffnung aufgenommen, einen Kompromiss zwischen noch zu bewältigendem Datenvolumen und vertretbar geringer Menge verworfener Informationen zu finden.

Leider stellte sich bei einigen ursprünglich aufgenommenen Merkmalen heraus, dass sie nur in wenigen Datenbanken nennenswert erfasst wurden. Da diese Untersuchung einen Schwerpunkt in Krankenhaus übergreifenden Untersuchungen setzt, werden diese Merkmale nicht weiter untersucht. Merkmale, die nach diesem zweiten Auswahlprozess übriggeblieben, sind in Tabelle 4.1 angegeben.

Zur deutlichen Kennzeichnung und Abgrenzung von in dieser Arbeit verwendeten Merkmalen und kategorialen Ausprägungen werden beide in dieser **Schriftart** gesetzt. Im Gegensatz zu Laborgerät im Allgemeinen ist hier also z.B. **Laborgerät** eine Ausprägung des Gerätemerkmales **Geräteart**. Man darf annehmen, dass sich hinter der **Geräteart Laborgerät** meistens Geräte der Art Laborgerät verbergen. Jedoch soll hier nicht diese Trennung verwischt werden.

### 4.1.2 Unvollständige Datensätze

Die Daten sind in zweifacher Hinsicht unvollständig erfasst. Zum Teil fehlen Daten komplett oder sind unstetig erfasst, wie man an den Abbildungen 4.17 und 4.39 sehen kann. Daraus ergibt sich, dass absolute Zahlen problematisch sind – zum Beispiel verlieren Quartalsvergleiche in Gegenwart von Erfassungsschwankungen ihre Aussagekraft.

Unter der Voraussetzung, dass Auslassungen in Hinblick auf die zu untersuchenden Daten zufällig erfolgt sind, sind Verhältniszahlen gegenüber Auslassungen unempfindlich. Sie können also mit einem gewissen Vorbehalt auch bei unvollständigen Daten verwendet werden.

Tabelle 4.1: In dieser Tabelle sind alle Krankenhaus-, Geräte- und Maßnahmenmerkmale mit Beschreibung aufgelistet, die in den Analysen untersucht werden.

Merkmal	Beschreibung
Krankenhausmerkmale	
Krankenhausart	Art des Krankenhauses. Ausprägungen: <b>privat</b> , <b>frei</b> / <b>gemeinnützig</b> , <b>öffentlich</b> und <b>öffentlich</b> / <b>universitär</b> . Ermittelt aus dem DKA 2000.
Bettenzahl	Anzahl der Betten im Krankenhaus. Ermittelt aus DKA 2000.
Gerätemerkmale	
Hauptgeräteart	Bezeichnung des Gerätes nach EMTEC-Katalog Ebene 1. Ausprägungen siehe Abbildung 4.3 oben.
MT-Geräteart	Bezeichnung medizin-technischer Geräte nach EMTEC-Katalog Ebene 2. Ausprägungen siehe Tabellen C.1, C.2 und C.3.
MT-Untergeräteart	Bezeichnung medizin-technischer Geräte nach EMTEC-Katalog Ebene 3. Ausprägungen siehe Tabellen C.1, C.2 und C.3.
Hersteller	Normierter Herstellername
Kaufpreis	Ganzzahliger Gerätekaufpreis in €
Abteilung	Standort des Gerätes. Ausprägungen siehe Anhang B.2
Inbetriebnahme	Inbetriebnahme- oder Kaufdatum des Gerätes.
Maßnahmenmerkmale	
Arbeitsbeschreibung	Normierte Beschreibung der durchgeführten Tätigkeiten innerhalb einer Maßnahme. Ausprägungen siehe Anhang B.1
Durchführungszeitpunkt	Datum der Maßnahme
Betriebszeit	Zeit zwischen Durchführungszeitpunkt und Inbetriebnahme des betroffenen Gerätes.
Leistungsart	Eigen- oder Fremdleistung
Kosten	Kosten der Maßnahme in €

Auslassungen liegen aber auch in der Art vor, dass Ereignisse zwar stetig und ausnahmslos dokumentiert seien mögen, jedoch wurden nicht immer alle wichtigen Eigenschaften der Ereignisse erfasst. Zum Beispiel können alle Geräte in einer Datenbank verzeichnet sein, aber die Angabe der Kaufpreise ist nur bei einem Teil vorhanden (siehe Tabelle 3.6).

Man könnte diese Unvollständigkeit als sachliche Unvollständigkeit bezeichnen – im Gegensatz zur oben vorgestellten, die dann analog als zeitliche Unvollständigkeit zu bezeichnen wäre.

Auch sachliche Unvollständigkeit verfälscht absolute Zahlen. Wenn man wie bei zeitlicher Unvollständigkeit relative Größen verwendet, verwirft man gegebenenfalls wertvolle Informationen. Man kann zusätzlich die vorhandenen Angaben auswerten, um diese dann auf alle vorhandenen Datensätze hochzurechnen. Bei dieser Hochrechnung

fließen etwas mehr Daten in die Rechnung ein, als bei der reinen Berechnung von relativen Größen.

Am Beispiel von Kaufpreisen sei dies erläutert. Sollen die Summen der Kaufpreise von Gerätegruppen bestimmt und verglichen werden, so werden die einfach gebildeten Summen aus Gründen von zeitlicher (Geräte fehlen) und sachlicher (Kaufpreise fehlen) Unvollständigkeit fehlerbehaftet sein. Zur Verringerung des sachlichen Fehlers wird von den angegebenen Kaufpreisen je Gerätegruppe ein Mittelwert bestimmt, und dieser dann mit der Gerätezahl der Gruppe multipliziert. Unter der nicht unkritischen Annahme, dass die Kaufpreise der Geräte, bei denen der Kaufpreis nicht angegeben wurde, genauso verteilt sind wie bei Geräten mit angegebenem Kaufpreis, erhält man so eine korrigierte Kaufpreissumme. In Abschnitt 4.3.2 erfolgt noch einmal eine Erläuterung des Vorgehens im konkreten Fall. Wie oben beschrieben, kann man nun noch, um zeitliche Fehler zu verkleinern, relative Kaufpreissummen verwenden.

Werden im weiteren Verlauf dieser Arbeit Summen betrachtet, so werden diese immer mit eben beschriebenen Verfahren bestimmt.

### 4.1.3 Einfache und robuste Kennzahlen

Um die in den Datenbanken vorhandene Fülle an Informationen übersichtlich darzustellen, ist oft erforderlich Informationen zusammenzufassen. Diese zusammengefaßten (aggregierten) Informationen nennen wir Kennzahlen.

Zur deutlichen Kennzeichnung werden diese wie Merkmale und kategoriale Ausprägungen auch in dieser **Schriftart** gesetzt.

Eine Kennzahl ist einfach, wenn der Ursprung der zugrunde liegenden Informationen und die Berechnung der Kennzahl nachvollziehbar ist. Eine Kennzahl ist robust, wenn trotz Datenfehler und -lücken der zugrunde liegenden Informationen die Berechnung der Kennzahl weitgehend numerisch stabil ist. Einfache und robuste Kennzahlen, die nützliche Informationen liefern und fast überall einsetzbar sind, sind

**Gerätezahl:** Anzahl der Geräte

**Kaufpreissumme:** Summe der Kaufpreise der Geräte

**Maßnahmenzahl:** Anzahl der durchgeführten Maßnahmen

**Kostensumme:** Summe der Kosten der durchgeführten Maßnahmen

Diese Kennzahlen können für beliebige Gruppen von Geräten oder Maßnahmen bestimmt werden. Nicht so einfache und robuste Kennzahlen, die auch nützliche Informationen liefern, sind

**Kaufpreiserwartungswert:** Erwartungswert bzw. Mittelwert des Kaufpreises pro Gerät

**Kostenerwartungswert:** Erwartungswert der Kosten je Maßnahme

**MTBF:** Mittlere Zeit zwischen Durchführungszeitpunkten aufeinanderfolgender Maßnahmen einem Gerät (Mean Time Between Failures)

**Aktivgerätezahl:** Anzahl der Geräte zu einem bestimmten Zeitpunkt, vor und nach diesem Zeitpunkt mindestens einmal in der Datenbank in Erscheinung treten – also Geräte, von denen angenommen werden kann, dass sie von der Datenbank erfasst werden.

Auch diese Kennzahlen können für beliebige Gruppen von Geräten oder Maßnahmen bestimmt werden.

Die geringere Robustheit dieser Kennzahlen ist zumindest bei den oberen drei auf die zusätzlich nötige Division zurückzuführen. Außerdem können diese Kennzahlen nicht so gut als relative Größen interpretiert werden.

## Beziehungen zu bekannten Kennzahlen

In Abschnitt 2.4.2 wurden einige Instandhaltungskennzahlen aufgelistet. Nur zum Teil lassen sich aus den hier aufgestellten Kennzahlen und Diagrammen die in Abschnitt 2.4.2 vorgestellten Kennzahlen bestimmen.

Eigenservicequoten können nur indirekt bestimmt werden, da Arbeitszeiten nur in Ausnahmefällen erfasst sind. Hier werden stellvertretend **Maßnahmenzahlen** oder **Kostensummen** verwendet.

Die Bestimmung der Wartungs- / Instandhaltungsquoten geschieht durch Auswertung der **Arbeitsbeschreibung**. Neben möglichen Mehrfachzuordnungen von Maßnahmen sind diese Quoten mit Unsicherheiten behaftet.

Die Jahresinstandhaltungsquote kann durch Auswerten der Anzahl der Maßnahmen pro Jahr multipliziert mit den **Kosten** je Maßnahme und geteilt durch den **Kaufpreis** ermittelt werden für verschiedene Gerätegruppen ermittelt werden. Da hier schon die eben genannten Größen nur mit Unsicherheiten ermittelt werden können ist also auch diese Werte mit Fehlern behaftet.

Einige Kennzahlen lassen sich nicht aus den vorliegenden Daten gewinnen. Hauptsächlich liegt das an der lückenhaften Erfassung oder dem vollständigen Fehlen der nötigen Informationen. Keine Aussagen lassen sich zu Investitions- / Reinvestitionskosten und damit zu den Gesamtinvestitionskosten machen, da zu dem überwiegenden Teil der Kosten in den Daten nicht angegeben ist, ob sie investiv sind. Im Nachhinein ist diese Ermittlung im Rahmen dieser Arbeit nicht mehr durchführbar.

### 4.1.4 Ordnung der Analysen

Aus den vorangegangenen Abschnitten folgt, dass gegenüber gleichverteilten, statistisch unabhängigen Datenfehlern Verhältnisse von Kennzahlen robust sind und dass wir 4 robuste Kennzahlen gem. Abschnitt 4.1.3 zur Verfügung haben.

Nach Tabelle 4.1 gibt es 2 Krankenhausmerkmale, 7 Gerätemerkmale und 5 Maßnahmenmerkmale. Bedenkt man, dass Krankenhausmerkmale auch Gerätemerkmale sind, da die Geräte eindeutig einem Krankenhaus zugeordnet werden können, und Gerätemerkmale wiederum Tätigkeitsmerkmale sind, da genauso Tätigkeiten eindeutig Geräten zugeordnet werden können, ergeben sich also insgesamt 2 Krankenhausmerkmale, 9 Gerätemerkmale und 14 Tätigkeitsmerkmale. Mit den 4 Kennzahlen ergeben sich schon  $4(2 + 9 + 14) = 100$  einfache Verteilungen.

Möchte man die Abhängigkeiten der Verhältnisse der Kennzahlen eines Merkmales von einem anderen Merkmal der gleichen Entität untersuchen, so ergeben sich schon  $4(\sum_i^{2-1} i + \sum_i^{9-1} i + \sum_i^{14-1} i) = 512$  Darstellungen. Dazu stelle man sich die Möglichkeiten zwei Merkmale gegeneinander aufzutragen in einer quadratischen Matrix vor. Die Hälfte der Kombinationen bzw. der Matrixelemente ist uninteressant, da nur die Merkmale vertauscht sind. Auch die Diagonale ist uninteressant, da dann ein Merkmal gegen sich selbst aufgetragen wird. Damit ergeben sich für eine  $n \times n$ -Matrix  $\sum_i^{n-1} i = \frac{n-1}{2}((n-1) + 1)$  Darstellungen.

Aus dieser großen Menge an Darstellungen werden im folgenden eine Auswahl getroffen. Folgende Auswahlkriterien sollen erfüllt sein:

- einfach, d.h. nachvollziehbar in Bezug auf Erstellung und Datenquelle
- weitgehend robust in Bezug auf Datenfehler und Numerik und
- entnommen aus einem in Bezug auf Variationsmöglichkeiten vollständigen Satz von Darstellungen

Diese Auswahl zu treffen, ist nicht immer leicht. Stellt man nur die aussagekräftigen Zusammenhänge dar, so könnte der Eindruck aufkommen, man habe bei den ausgelassenen Darstellungen gegebenenfalls interessante Zusammenhänge unbeabsichtigt übergangen. Stellt man, auch z.B. aus Gründen der Vollständigkeit, uninteressante Daten dar, dann verschwendet man Platz und Aufmerksamkeit. Es muss gewissermaßen Kompaktheit der Darstellung gegen ihre Objektivität abgewogen werden.

Als Beispiel für diese Problematik kann Abschnitt 4.4.2 dienen, der keine besonderen Ergebnisse hervorgebracht hat, jedoch dessen Fehlen als Unvollständigkeit der Analysen erschienen wäre.

#### 4.1.5 Verwendete Diagrammart

Zur Visualisierung der in dieser Arbeit untersuchten Daten werden verschiedene Diagrammart verwendet. Einige sind Standarddarstellungen, andere wurden speziell an die Anforderungen dieser Arbeit angepasst. Im Folgenden werden die Diagrammart kurz vorgestellt.

**Punktogramme** werden verwendet, wenn zwei gepaarte metrische Größen in Abhängigkeit einer kategorialen oder ordinalen Größe untersucht werden sollen. Gepaarte Größen können auch in Form von freien und abhängigen Variablen einer diskreten Funktion vorliegen. Damit werden auch empirische Verteilungsfunktionen hier durch Punktogramme dargestellt.

**Balkendiagramme** werden verwendet, wenn eine metrische Größe in Abhängigkeit von zwei gepaarten kategorialen oder ordinalen Merkmalen untersucht werden soll. Bei kategorialen Merkmalen werden die Kategorien nach dem in Abschnitt 2.5.2 Algorithmus sortiert.

**Histogramme** sind spezielle Balkendiagramme. Sie geben eine metrische Größe (die Häufigkeit) in Abhängigkeit eines kategorialen oder ordinalen Merkmales an.

**Treemaps** sind Diagramme die metrische Größen in Abhängigkeit von hierarchischen kategorialen Merkmalen untersuchen. Die metrischen Größen müssen dabei für die Hierarchieebenen additiv sein (siehe Abschnitt 2.5.1).

**Assoziationsdiagramme** sind Grafen von Ausprägungen kategorialer Merkmale als Knoten und nach Stärke der Assoziation der beteiligten Merkmale ausgewählter Kanten. Die Stärke der Assoziation wird hier durch den lift bestimmt (siehe Abschnitt 2.3.4).

**Lorenzdiagramme** (auch „Paretodiagramme“ genannt) sind spezielle Punktdiagramme. Dargestellt wird dabei ein diskreter funktionaler Zusammenhang, wobei die Abhängige Variable den Anteil an einer Wertemenge angibt, der die Werte enthält, die kleiner als die freie Variable sind (siehe Abschnitt 2.5.3).

## 4.2 Untersuchung der Krankenhäuser

Untersuchte Krankenhausmerkmale sind Bettenzahl und Krankenhausart. In den folgenden zwei Abbildungen 4.1 und 4.2 ist noch zusätzlich die Gerätezahl dargestellt.

Es ist keine deutliche Abhängigkeit der Gerätezahl von der Bettenzahl zu erkennen – auch innerhalb einer Krankenhausart nicht. Ausnahme bilden hier die beiden untersuchten Universitätskliniken, die durch hohe Bettenzahlen und hohe Gerätezahlen pro Bett gemeinsam auffallen.

Um den Zusammenhang statistisch zu untersuchen, wird hier der Spearmansche Rang-Korrelationskoeffizient nach Abschnitt 2.3.3 verwendet. Für Abbildung 4.1 ergibt sich  $r_S = 0.794$ . Der kritische Wert zum Signifikanzniveau  $\alpha = 0.05$  lautet  $k_{0.05;23} = 0.415$ . Die Nullhypothese wird also mit Irrtumswahrscheinlichkeit  $\alpha = 0.05$  abgelehnt – Bettenzahl und Gerätezahl hängen also signifikant zusammen.

Für Abbildung 4.2 ergibt sich  $r_S = 0.377$ . Der kritische Wert zum Signifikanzniveau  $\alpha = 0.05$  lautet wieder  $k_{0.05;23} = 0.415$ . Die Nullhypothese kann also mit Irrtumswahrscheinlichkeit  $\alpha = 0.05$  nicht abgelehnt werden – Bettenzahl und Gerätezahl pro Bett hängen also nicht signifikant zusammen.

An diesen Darstellungen wird ein Nachteil absoluter Zahlen (hier der Gerätezahl je Krankenhaus) deutlich: unterschiedliche Erfassungen der Krankenhäuser schlagen voll auf die Zahlen durch. Dabei muss ein Krankenhaus nicht nur einen Teil seiner Geräte erfasst haben, auch kann in einem Krankenhaus die Erfassung von Untergeräten detaillierter erfolgt sein als in einem anderen. Bei Abbildungen, die relative Zahlen, wie z.B. den Anteil der Infusionspumpen an allen Geräten je Krankenhaus darstellen, ist diese Fehlerquelle abgeschwächt.

Diese Darstellung mischen also Informationen über die Situation der Krankenhäuser mit Informationen über die Art der Erfassung. Welcher dieser beiden Einflüsse die Zahlen mehr beeinflusst, lässt sich an diesen Darstellungen nicht ablesen.

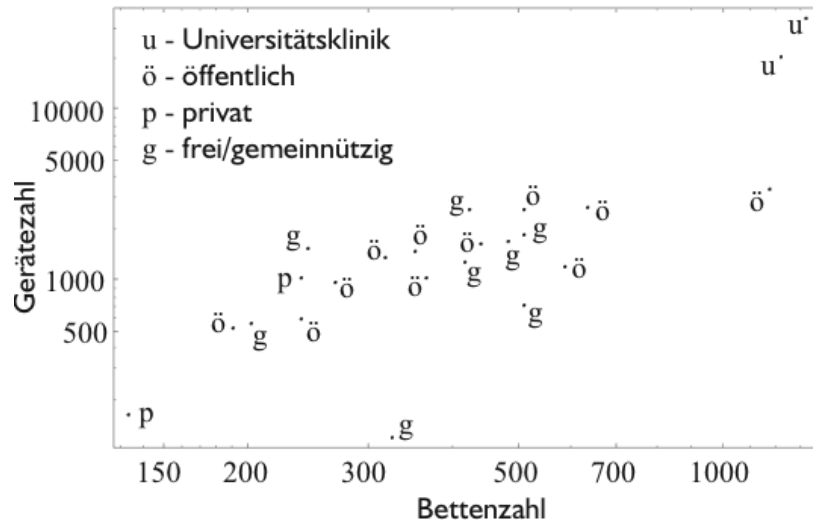


Abbildung 4.1: Jeder Punkt entspricht einem Krankenhaus. Zu jedem Punkt ist die Krankenhausart angegeben. Auf der x-Achse ist die Bettanzahl und auf der y-Achse die Gerätezahl des jeweiligen Krankenhauses logarithmisch aufgetragen. Zwischen Bettanzahl und Gerätezahl besteht nur ein schwacher Zusammenhang. Universitätskliniken fallen bei hoher Bettanzahl durch stark erhöhte Gerätezahl auf. Bei den sonstigen Krankenhausarten fallen keine Gruppenbildungen auf.

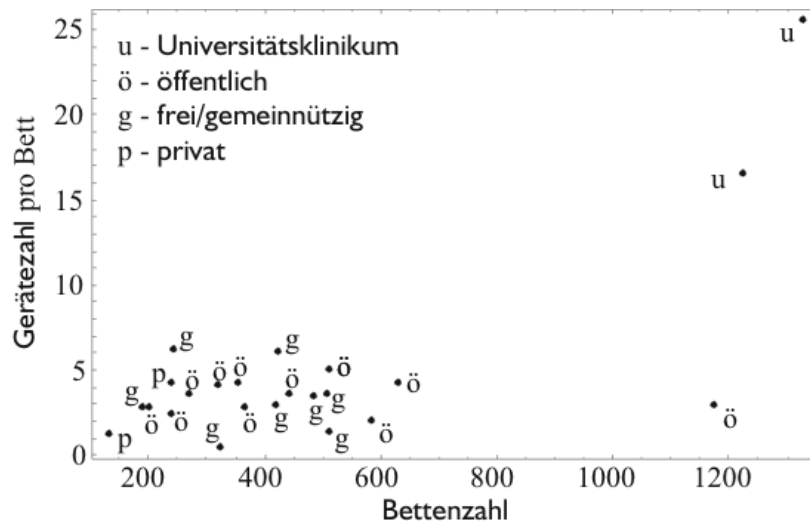


Abbildung 4.2: Jeder Punkt entspricht einem Krankenhaus. Zu jedem Punkt ist die Krankenhausart angegeben. Auf der x-Achse ist die Bettanzahl und auf der y-Achse die Gerätezahl pro Bett des jeweiligen Krankenhauses aufgetragen. Zwischen Bettanzahl und Gerätezahl pro Bett besteht kein Zusammenhang. Universitätskliniken fallen bei hoher Bettanzahl durch erhöhte Gerätezahl pro Bett auf. Bei den sonstigen Krankenhausarten fallen keine Gruppenbildungen auf. Für nicht-Universitätskliniken liegt die Gerätezahl pro Bett zwischen 1 und 7, für Universitätskliniken zwischen 15 und 25.

Tabelle 4.2: In dieser Tabelle sind absolute **Gerätezahlen** medizinisch-technischer Geräte angegeben. Unterschieden werden Geräte mit erfasster **Inbetriebnahme** oder **Kaufpreis**.

Gerätezahl	alle Geräte	Geräte mit Kaufpreis
alle Geräte	82 205 (100 %)	32 513 (40.6 %)
Geräte mit Inbetriebnahme	66 641 (81.1 %)	30 973 (37.7 %)

## 4.3 Untersuchung der Gerätedaten

### 4.3.1 Verteilung der Häufigkeiten

In diesem Abschnitt werden Anteile von **Gerätearten** am gesamten Gerätebestand untersucht. Die Anteile werden dabei in Abhängigkeit von der **Krankenhausart** und **Bettenzahl** sowie des **Herstellers** ermittelt. Im wesentlichen wird hier die Frage beantwortet: Wo sind die meisten Geräte?

In Tabelle 4.2 sind die absoluten **Gerätezahlen** medizinisch-technischer Geräte angegeben, die in dieser Arbeit zur Verfügung stehen. Sobald man jedoch fordert, dass zu den Geräten auch z.B. die **Inbetriebnahme** oder der **Kaufpreis** angegeben sind, nehmen die **Gerätezahlen** deutlich ab.

Da in den folgenden Abbildungen meist relative **Gerätezahlen** verwendet werden, kann man anhand der Tabelle 4.2 sehen, auf wie viele Geräte sich diese absolut beziehen.

#### Tortendiagramm der relativen Gerätezahlen

Durch die Normierung der Gerätebezeichnungen nach dem EMTEC-Katalog besitzt dieses kategoriale Merkmal nun eine hierarchische Struktur nach den Gerätearten. Ein Histogramm dieser Hierarchie mit gewöhnlichen Balken- oder Tortendiagrammen übersichtlich darzustellen ist gerade bei zahlenmäßig inhomogen besetzten Kategorien kaum möglich. Die Grenzen eines Tortendiagramms zeigt die folgende Abbildung 4.3. Ein direkter Größenvergleich der Tortenstücke z.B. für die **Untergeräteart Personenrufanlage** und **Patientenüberwachung** macht keinen Sinn, da sie sich auf unterschiedlich besetzte **Gerätearten** beziehen. Diesen Nachteil weisen die im Folgenden verwendeten treemaps nicht auf, die in Abschnitt 2.5.1 vorgestellt werden.

14 % der Geräte konnten keiner EMTEC-Geräteart zugeordnet werden. Dies müssen nicht immer Bezeichnungen sein, bei denen eine Zuordnung im Einzelfall nicht möglich ist. Viele Geräte wurden nicht zugeordnet, da sie bei keiner Auswahlabfrage (siehe Abschnitt 3.5.3) aufgrund ihrer außergewöhnlichen oder sehr seltenen Bezeichnung dabei waren. Jede Bezeichnung konnte nicht mit vertretbarem Aufwand einzeln zugeordnet werden. Das bedeutet aber auch, dass durch mehr Aufwand sich dieser nicht zugeordnete Anteil noch verkleinern ließe.

Unter Berücksichtigung der nicht zugeordneten Geräte kann man in etwa zusammenfassen, dass zwei Drittel der Geräte medizin-technische Geräte sind und dass sich das restliche Drittel je zur Hälfte auf EDV- und sonstige Geräte aufteilt. Ein Drittel der EDV-Geräte sind Kommunikationsgeräte, ein anderes Drittel entfällt auf Personal Computer.

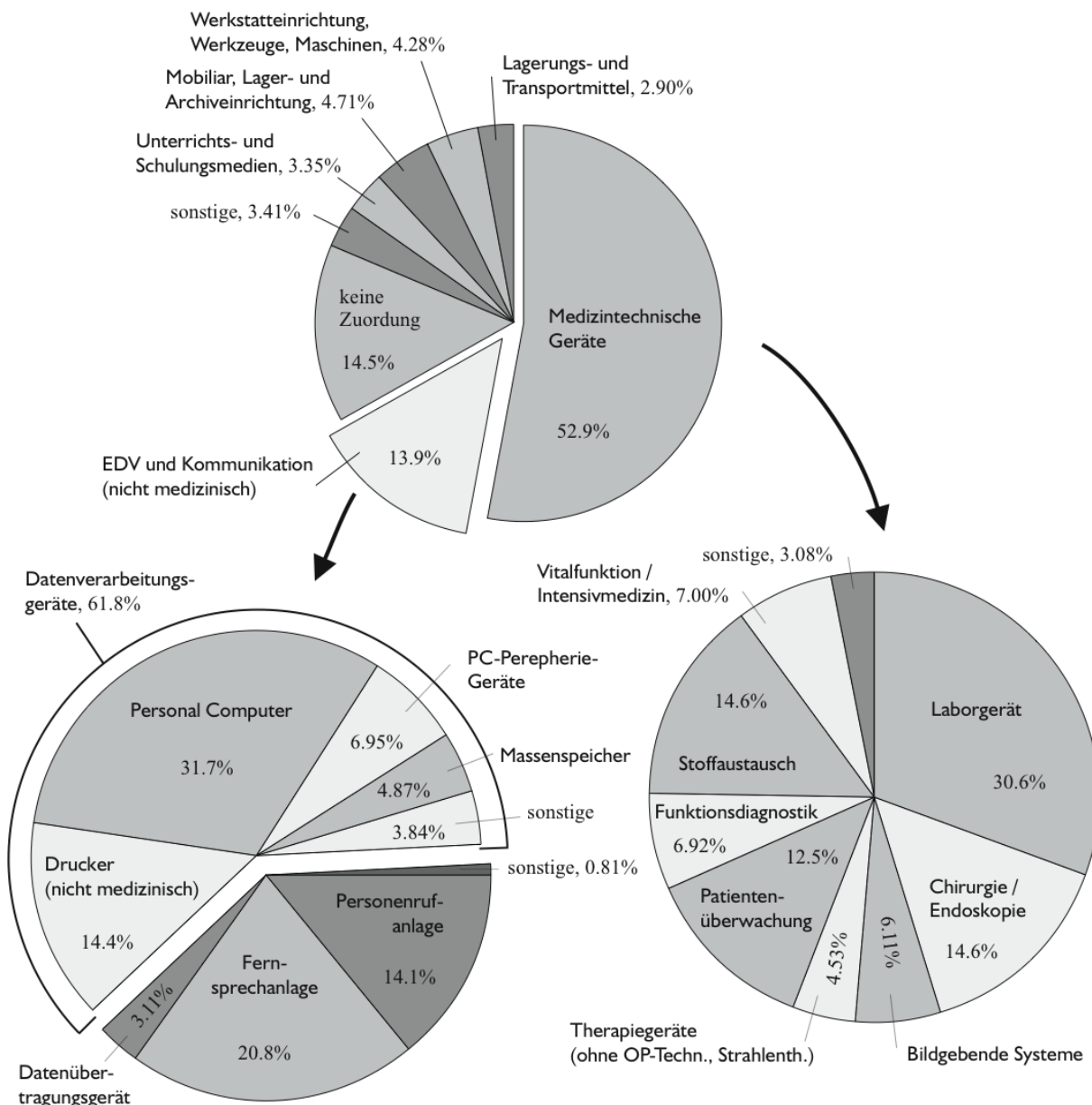


Abbildung 4.3: Tortendiagramm der Gerätezahl je Hauptgeräteart und Geräteart. In diesem Diagramm sind alle in den Daten gefundenen Geräte verzeichnet. Die Gruppe keine Zuordnung besteht aus Geräten, die keiner Bezeichnung aus dem EMTEC-Katalog zugeordnet werden konnte. Unter der Bezeichnung „sonstige“ sind für die Darstellung zu kleine Gruppen zusammengefasst. Dabei kann auch die Gruppe „keine Zuordnung“ in der Gruppe „sonstige“ subsumiert werden, wenn sie klein genug ist. Die obere Torte stellt die Ausprägungen des Merkmals Hauptgeräteart dar. Die rechte untere Torte stellt die Ausprägungen für das Merkmal MT-Geräteart dar. Die linke untere Torte stellt für die Geräteart EDV und Kommunikation (nicht medizinisch) und die Untergeräteart Datenverarbeitungsgeräte dar.

## Abhängigkeit der relativen Gerätezahlen der MT-Gerätearten vom Krankenhaus

Es zeigt sich, dass die relative Gerätezahl der MT-Gerätearten von der Krankenhausart abhängt (siehe Abbildung 4.4). Dabei ist bei Universitätskliniken der Anteil der Laborgeräte deutlich gegenüber anderer Krankenhausarten erhöht. Diese Erhöhung geht hauptsächlich auf Kosten von Anteilen der MT-Gerätearten Stoffaustausch und Vitalfunktion / Intensivmedizin.

Um diesen Sachverhalt statistisch zu prüfen, kann man die Kreuztabelle, die der Abbildung 4.4 zugrunde liegt, mit Hilfe des  $\chi^2$ -Tests auf Unabhängigkeit prüfen. Das Testverfahren wird in Abschnitt 2.3.3 vorgestellt. Für die dargestellten Daten ergibt sich eine Testgröße von  $T = 12\,757$ . Der kritische Wert für  $\alpha = 0.05$  und 36 Freiheitsgrade lautet  $k_{36;0.05} = 51.00$ . Damit wird die Nullhypothese mit überwältigender Deutlichkeit abgelehnt – die Gerätezahlen der MT-Gerätearten hängen also hoch signifikant von den Krankenhäusern ab.

Betrachtet man nur die beiden Universitätskliniken, so ergibt sich eine Testgröße von  $T = 278.4$ . Der kritische Wert lautet für  $\alpha = 0.05$  und 6 Freiheitsgrade  $t_{6;0.05} = 12.59$ . Für nicht-Universitätskliniken ergibt sich  $T = 688$  und  $t_{24;0.05} = 36.42$ . Damit sind auch die Verteilungen innerhalb der Universitätskliniken und nicht-Universitätskliniken vom Krankenhaus abhängig.

Zur Beurteilung der Stärke der Abhängigkeit sind nach Abschnitt 2.3.3 die Testgrößen nicht geeignet. Aus diesem Grunde wird hier der korrigierte Kontingenzkoeffizient  $C_{korr}$  von Pawlik verwendet (siehe Abschnitt 2.3.3). Für alle Daten der Abbildung 4.4 ergibt sich  $C_{korr} = 39.59\%$ , für Universitätskliniken  $C_{korr} = 10.17\%$  und für nicht-Universitätskliniken  $C_{korr} = 17.10\%$ . Hier bestätigt sich der optische Eindruck, dass die Schwankungen innerhalb der Universitätskliniken und nicht-Universitätskliniken geringer sind als innerhalb aller Krankenhäuser. Aufgrund dieses deutlichen Unterschieds werden im folgenden Abbildungen 4.5 und 4.6 Universitätskliniken und nicht-Universitätskliniken getrennt betrachtet.

Prüft man die Abhängigkeit vom Krankenhaus des Anteils der Laborgeräte gegen den der restlichen Geräte, so ergibt sich eine Testgröße von  $T = 10\,128$ . Der kritische Wert lautet  $k_{6;0.05} = 12.59$ , und es ergibt sich  $C_{korr} = 46.9\%$ .

Prüft man die Abhängigkeit des Summenanteils der Labor-, Stoffaustausch- und Vitalfunktions- / Intensivmedizingeräte gegen den der restlichen Geräte, so ergibt sich eine Testgröße von  $T = 1872$ . Der kritische Wert lautet wieder  $k_{6;0.05} = 12.59$ , und es ergibt sich  $C_{korr} = 14.9\%$ .

Weiterhin ist eine leichte Zunahme des Anteils von Stoffaustauschgeräten bei gleichzeitiger Abnahme von Patientenüberwachungsgeräten mit steigender Bettenzahl von nicht-Universitätskliniken zu erkennen. Eine statistische Prüfung dieser Aussage wird hier nicht vorgenommen.

## Treemaps relativer Gerätezahlen für MT-Gerätearten und -Untergerätearten

Eine alternative zu den Torten-Diagrammen stellen squarified treemaps dar, wie sie in Abschnitt 2.5.1 vorgestellt wurden. Anhand dieser Darstellung können Hierarchien inhomogen besetzter Kategorien – wie sie mit den MT-Geräte- und -Untergerätearten vorliegen – übersichtlich dargestellt werden.

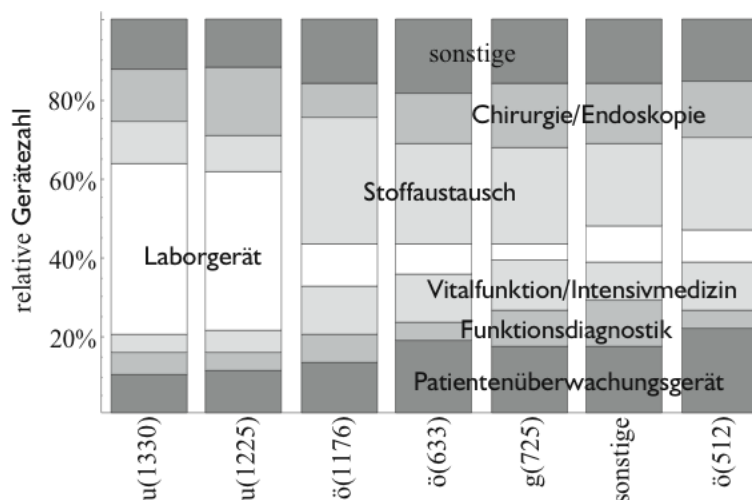


Abbildung 4.4: Balkendiagramm der relativen Gerätezahl je MT-Geräteart und Krankenhaus. Bei den Krankenhäusern sind Krankenhausart als Buchstabe und die Bettenzahl dahinter in Klammern angegeben. Die Kodierung der Krankenhausarten durch Buchstaben ist in Abbildung 4.1 angegeben. Das Diagramm wurde durch den Algorithmus aus Abschnitt 2.5.2 sortiert – damit sind Kategorien ab der sechst-häufigsten Kategorie zusammengefasst. Deutung siehe Text.

In den folgenden Abbildungen 4.5 und 4.6 sind die relativen Gerätezahlen der einzelnen MT-Gerätearten und MT-Untergerätearten jeweils für Universitätskliniken und nicht-Universitätskliniken angegeben. Zusammen mit den ähnlichen Abbildungen 4.9 und 4.10 über die relativen Kaufpreissummen, Abbildungen 4.19 und 4.20 über die relativen Maßnahmenzahlen und Abbildungen 4.9 und 4.10 über die relativen Kostensummen ergibt sich ohne komplizierte Berechnungen ein guter Überblick über die „dicken Brocken“ in der Gerätebewirtschaftung. Tabellen hierzu finden sich im Anhang D insbesondere in Tabelle D.1.

### Homogenität der Hersteller innerhalb der MT-Gerätearten

Abbildung 4.7 zeigt die Segmentierung der MT-Gerätearten nach verschiedenen Herstellern. Bei Stoffaustauschgeräten dominieren zwei große Hersteller. Bei Patientenüberwachungsgeräten ist der Markt nach Herstellern stärker aufgesplittet. Vitalfunktion / Intensivmedizin-Geräte werden durch einen Hersteller dominiert.

Aus standhalterischer Sicht mag es sinnvoll sein, einen Gerätebestand mit möglichst wenigen Herstellern bzw. möglichst viele Geräte eines Herstellers zu bewirtschaften. Aus kaufmännischer Sicht kann eine solche Monokultur jedoch wegen der Abhängigkeit Nachteile mit sich bringen.

### 4.3.2 Verteilung der Investitionen

In diesem Abschnitt werden die Anteile der Kaufpreissummen – also die Anteile der Investitionen – für verschiedene MT-Gerätearten und -Untergerätearten untersucht. Die Darstellungen sind analog zu denen aus Abschnitt 4.3.1, nur dass die Gerätezahl durch die Kaufpreissummen ersetzt ist.

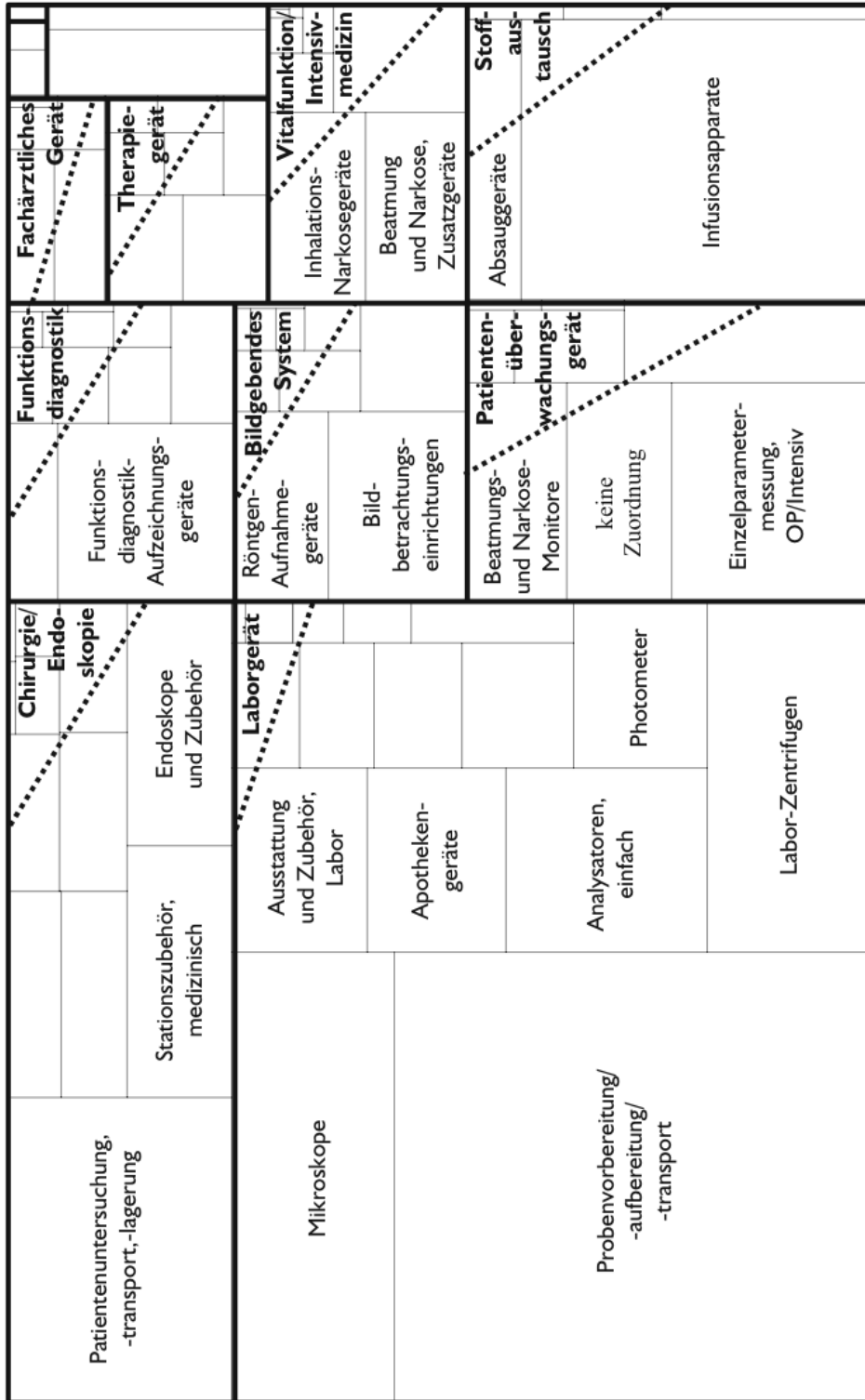


Abbildung 4.5: Nach Algorithmus aus Abschnitt 2.5 erstellte squarified treemap. Die Flächen entsprechen der Gerätezahl der jeweiligen MT-Gerätearten und -Untergearten. Es sind die Anteile für Geräte aus Universitätskliniken angegeben. Deutung siehe Text.





Die Untersuchung der Verteilung der Kaufpreise innerhalb bestimmter MT-Gerätearten und -Untergerätearten wird im Abschnitt 4.3.3 untersucht. Hier werden nur Summen betrachtet, nicht die Summanden.

### Abhängigkeit der relativen Kaufpreissummen je MT-Geräteart vom Krankenhaus

In Abbildung 4.8 sind die Anteile der Investitionen je MT-Geräteart für verschiedene Krankenhäuser aufgetragen. Im Vergleich zu Abbildung 4.4 sind die Unterschiede insbesondere bei Laborgeräten zwischen Universitätskliniken und nicht-Universitätskliniken nicht so deutlich. Dies ist auch darauf zurückzuführen, dass ein öffentliches Krankenhaus hohe Kaufpreissummen bei Laborgeräten aufweist.

Auch fällt die allgemein höhere Variation der Anteile auf. Dies ist wohl auf die Variation der Kaufpreise der Geräte zurückzuführen, die noch zur Variation der Gerätezahlen aus Abbildung 4.4 hinzutritt. In Abbildung 4.4 hatten alle Geräte praktisch den gleichen Kaufpreis 1.

Um die Abhängigkeit der relative Kaufpreissummen zu prüfen, kann man die Kreuztabelle, die der Abbildung 4.8 zugrunde liegt, mit Hilfe des  $\chi^2$ -Tests auf Unabhängigkeit prüfen. Das Testverfahren wird in Abschnitt 2.3.3 vorgestellt. Für die dargestellten Daten ergibt sich eine Testgröße von  $T = 7.6010^7$ . Der kritische Wert für  $\alpha = 0.05$  und 36 Freiheitsgrade lautet  $k_{36;0.05} = 51.00$ . Damit wird die Nullhypothese mit überwältigender Deutlichkeit abgelehnt – die Kaufpreissummen der MT-Gerätearten hängen also hoch signifikant von den Krankenhäusern ab.

In Anbetracht dieses Ergebnisses erscheint eine Untersuchung, zusammengefasst nach Universitätskliniken und nicht-Universitätskliniken wie in Abschnitt 4.3.1, nicht sinnvoll. Trotzdem sollen zur besseren Vergleichbarkeit auch hier die Kaufpreissummen für Universitätskliniken und nicht-Universitätskliniken getrennt dargestellt werden.

Wie in Abschnitt 4.3.1 wird zur Beurteilung der Stärke der Abhängigkeit der korrigierte Kontingenzkoeffizient  $C_{korr}$  von Pawlik verwendet (siehe Abschnitt 2.3.3). Für die Daten der Abbildung 4.8 ergibt sich  $C_{korr} = 33.25 \%$ .

### Fehlerbetrachtungen

Schon die Berechnung der Kaufpreissummen einer Gerätegruppe kommt nicht ohne Berücksichtigung der Datenqualität aus. Nach Tabelle 3.6 sind zwar 21 Spalten, die Kaufpreise enthalten und Geräten zugeordnet werden konnten, vorhanden – fast jede der 22 eingesandten Datenbanken enthält also Informationen über den Kaufpreis. Jedoch sind diese Spalten nur zu 17.67 % ausgefüllt. Der Anteil gültiger Kaufpreise (siehe Abschnitt 3.5.1) ist damit noch geringer. Hier werden jedoch fast ausschließlich medizintechnische Geräte untersucht. Bei diesen beträgt der Anteil der Geräte mit gültigem Kaufpreis immerhin noch 40 % (siehe Tabelle 4.2).

Addiert man nun nur die gültigen Kaufpreise einer Gerätegruppe, so hängt die Summe zwar wie erwünscht von der Höhe und Anzahl der Summanden – also der Kaufpreise – ab, jedoch wird diese Summe auch maßgeblich vom Anteil der gültigen Kaufpreise beeinflusst. Dies ist nicht erwünscht. Zwar ist die Darstellung der Investitionsanteile (statt der absoluten Kaufpreissummen) gegenüber einer Gleichverteilung der ungültigen Kaufpreise über die betrachteten Gerätegruppen und Kaufpreise unempfindlich (ver-

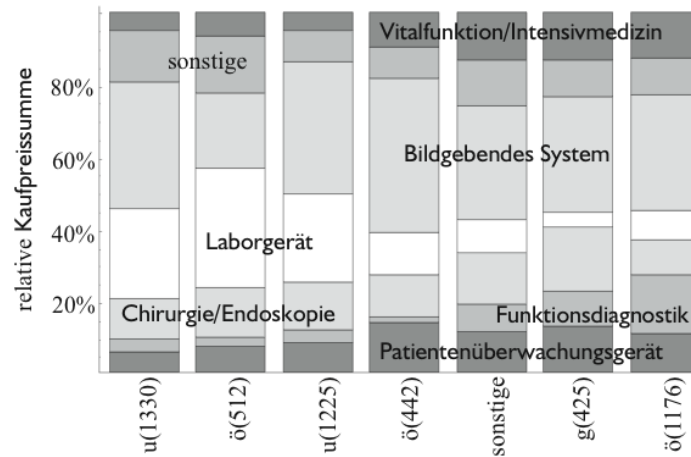


Abbildung 4.8: Balkendiagramm der relativen Kaufpreissummen verschiedener MT-Gerätearten und Krankenhäuser. Das Diagramm wurde durch den Algorithmus aus Abschnitt 2.5.2 sortiert – damit sind auch ab der sechst-häufigsten Kategorie Krankenhäuser und MT-Gerätearten zusammengefasst. Deutung siehe Text.

gleiche Abschnitt 4.1.2), jedoch kann im allgemeinen Fall von dieser Gleichverteilung nicht ausgegangen werden.

Diesem Problem kann teilweise vorgebeugt werden, in dem der mittlere Kaufpreis der Gerätegruppe aus den gültigen Kaufpreisen berechnet wird und dann mit der Anzahl der in der Gerätegruppe vorhandenen Geräten multipliziert wird. Immerhin kann man sich so gegenüber einer Ungleichverteilung der gültigen Kaufpreise über die Gerätegruppen absichern. Sind die ungültigen Kaufpreise jedoch auch über die Kaufpreise ungleich verteilt, so muss, da die Verteilung der Kaufpreise in den Gerätegruppen nicht als gleich angesehen werden kann, mit einem Fehler gerechnet werden.

Es zeigt sich immer wieder in dieser Arbeit: bei Untersuchung der Daten beobachtet man nie ausschließlich die ihnen zugrunde liegenden Größen sondern auch immer die Abbildung dieser Größen in die Daten, die z.B. von der Qualität der Erfassung abhängt. Nimmt die Qualität der Erfassung ab, so wird der Einfluss der Datenabbildung gegenüber dem Einfluss der zugrunde liegenden Größen immer größer – man beobachtet die Erfassung, nicht die erfassten Größen selber.

### Treemaps relativer Kaufpreissummen für MT-Gerätearten und -Untergerätearten

Gegenüber den Darstellungen 4.5 und 4.6 aus Abschnitt 4.3.1 sind in den Abbildungen 4.9 und 4.10 Geräte, die keiner Geräteart zugeordnet werden konnten, auffälliger vertreten. Dies ist auf eine nicht optimale Normierung zurückzuführen. Vor Beginn der Normierung konnten die Daten nur rudimentär untersucht werden, und es wurde davon ausgegangen, dass die Gerätekaufpreise nicht so ungleich verteilt sind. Diese Annahme ermöglichte eine Optimierung der Normierungen auf die Anzahl der Geräte, d.h. häufige Geräte wurden vorrangig normiert. Da nicht gleichzeitig teure Geräte vorrangig normiert wurden, es aber seltene und gleichzeitig teurere Geräte gibt, ist die Wahrscheinlichkeit erhöht, dass nicht normierte Gerätegruppen bei der Betrachtung

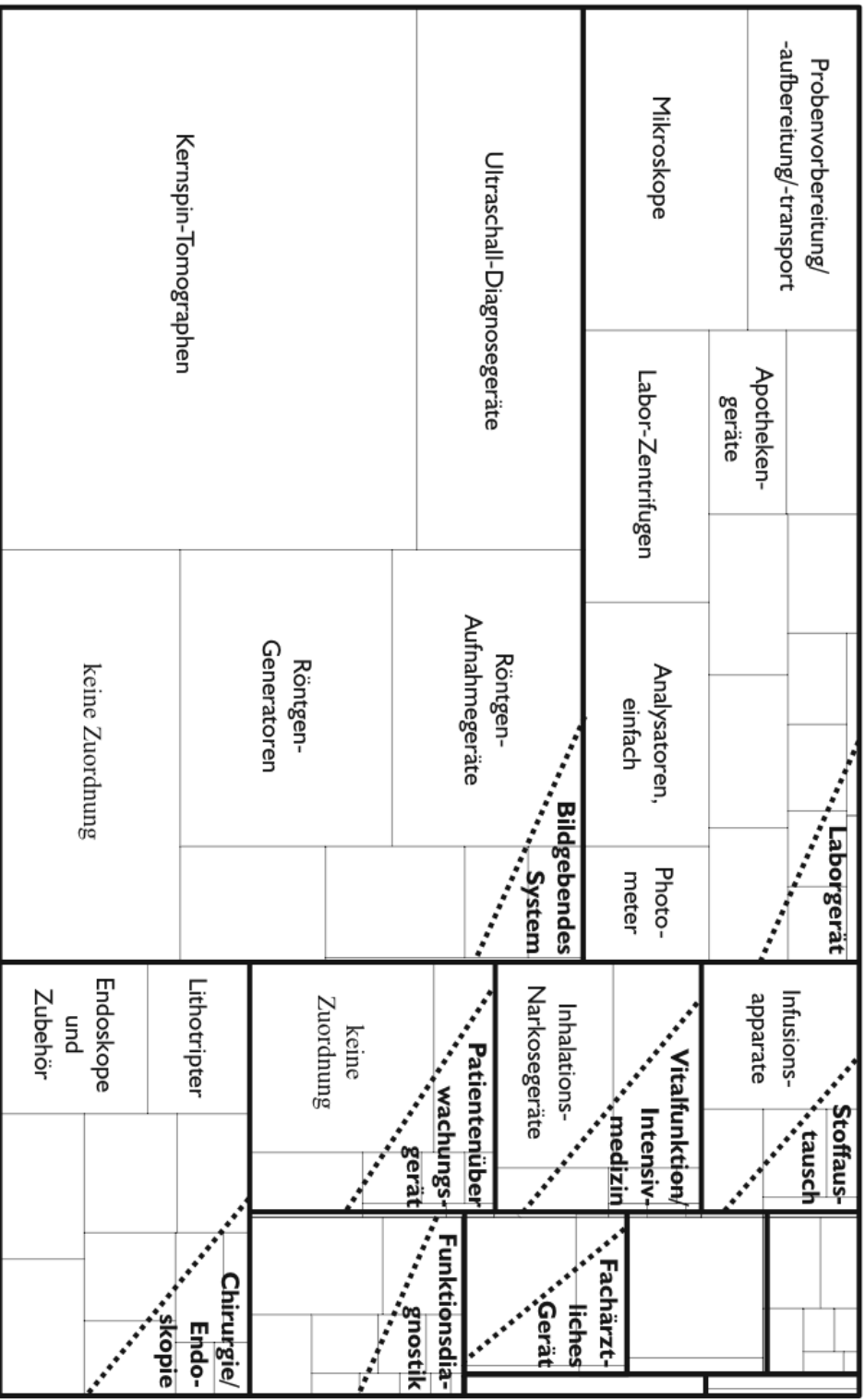


Abbildung 4.9: Nach Algorithmus aus Abschnitt 2.5 erstellte squarified treemap. Die Flächen entsprechen den Kaufpreissummen der jeweiligen MT-Geräteart und -Untergeräteart. Es sind die Anteile für Geräte aus Universitätskliniken angegeben. Deutung siehe Text.

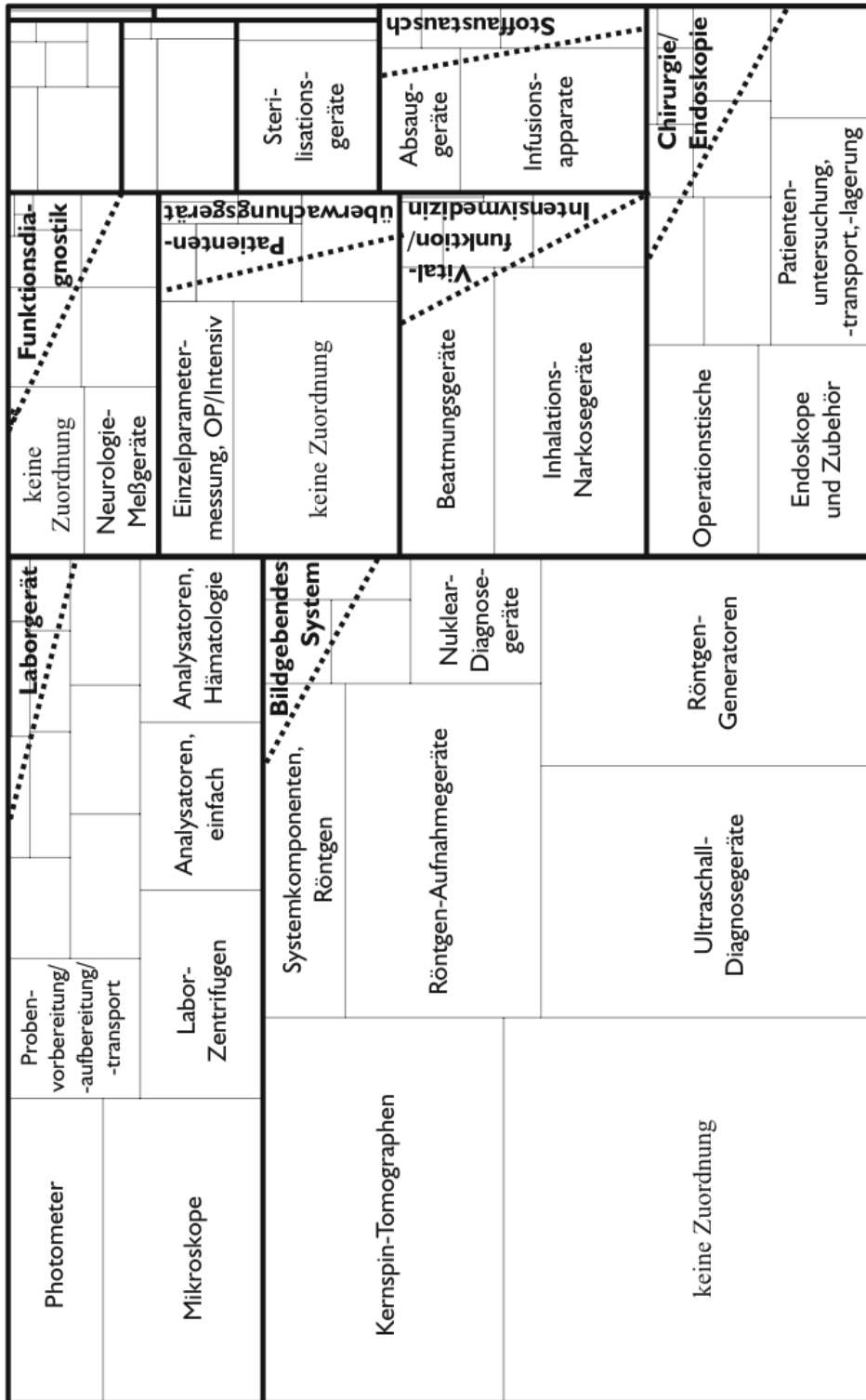


Abbildung 4.10: Nach Algorithmus aus Abschnitt 2.5 erstellte squarified treemap. Die Flächen entsprechen den Kaufpreissummen der jeweiligen MT-Geräteart und -Untergeräteart. Es sind die Anteile für Geräte aus nicht-Universitätskliniken angegeben. Deutung siehe Text.

der Kaufpreissummen auf die vorderen Plätze verschoben werden.

Bei allen Krankenhäusern ist eine deutliche Dominanz der Bildgebenden Systeme zu verzeichnen (siehe Abbildungen 4.8, 4.9 und 4.10), wobei bei Universitätskliniken diese Dominanz hauptsächlich durch Kernspintomographen hervorgerufen wird. Tabellen hierzu finden sich im Anhang D insbesondere in Tabelle D.2.

Aus Gerätezahl und Kaufpreissumme können Informationen über die Preise pro Gerät gewonnen werden. Diese werden in Abschnitt 4.3.3 vorgestellt.

### 4.3.3 Verteilung der Kaufpreise

Wurden im Abschnitt 4.3.2 die Kaufpreissummen analysiert, so wird jetzt die Verteilung der einzelnen Kaufpreise einer Gerätegruppe analysiert.

#### Darstellung der Verteilungsfunktion

Es zeigt sich, dass die Kaufpreise ungleich verteilt sind: es gibt viele günstige und wenige sehr teure Geräte. Diese Ungleichverteilung ist bei privaten Kliniken am geringsten und bei frei / gemeinnützigen Kliniken am höchsten, wie Abbildung 4.11 zeigt. Allgemein kann man von einer 80/20-Regel sprechen: 20 % der teuersten Geräte machen 80 % der Investitionen aus, und 80 % der günstigsten Geräte machen 20 % der Investitionen aus.

Aus Abbildung 4.12 lässt sich ablesen, dass private Kliniken viele Geräte im Bereich zwischen 1000 € und 3000 € besitzen, dass Universitätskliniken und öffentliche Kliniken – die einen sehr ähnlichen Verlauf zeigen – relativ viele Geräte im Bereich von einer Million Euro besitzen. Freie / gemeinnützige Krankenhäuser liegen dazwischen.

#### Regression der Kaufpreise

Auch, wenn nach Abbildung 4.13 ( $N=498$  sind nicht unterschiedliche Geräte sondern unterschiedliche Kaufpreise) die beispielhaft dargestellte Regression von Kaufpreisen das Konfidenzintervall von  $\alpha=0.5$  knapp verfehlt, so sollen im folgenden doch die Verteilung von Kaufpreisen durch log-Normalverteilungen angenähert werden. Gründe hierfür sind:

- Der Fehler der durch eine unpassende Approximation zu erwarten ist, ist im Vergleich mit anderen zu erwartenden Fehlern vertretbar. Dazu zählen fehlende Kaufpreisangaben, wahrscheinlich fehlerhafte Kaufpreisangaben und fragliche Handhabung der Kaufpreise bei Untergeräten.
- Die log-Normalverteilung ist eine einfache zu handhabende Verteilung, die mit 2 leicht zu interpretierenden Parametern auskommt. Unter gewissem Vorbehalt darf diese Auswahl auch für minimal willkürlich gehalten werden, da der offensichtlich einseitig begrenzte Wertebereich der Kaufpreise durch Transformation mit der log-Funktion in einen unbegrenzten Wertebereich verwandelt wird, um dann die naheliegendste Verteilung – die Normalverteilung – anzuwenden.
- Im Johnson-System liegen die Verteilungen zwar eher im Bereich der begrenzten Verteilungen als im Bereich der log-Normalverteilungen (Grenzbereich zu den unbegrenzten Verteilungen) (siehe Anhang E.1). Der Gewinn durch eine genauere

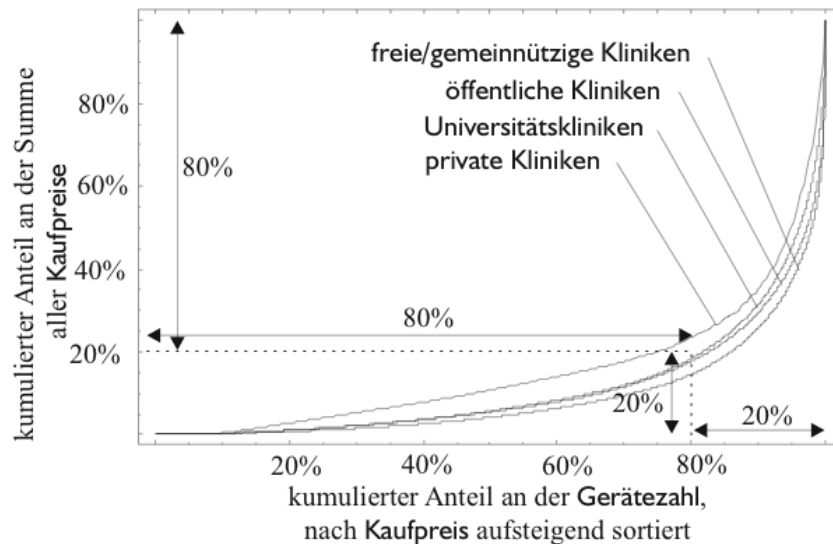


Abbildung 4.11: In dieser Abbildung ist der kumulierte Anteil der Kaufpreise über der kumulierten und relativen Gerätezahl aufgetragen. Die Geräte sind dabei aufsteigend nach Kaufpreis sortiert. Eine solche Kurve nennt man auch Lorenzkurve. Die Fläche unter der gedachten  $(0,0)$ - $(1,1)$ -Linie ist ein Maß für die Ungleichverteilung in diesem Fall der Kaufpreise. Man kann ablesen, dass 20 % der teuersten Geräte 80 % der Investitionen ausmachen und dass 80 % der günstigsten Geräte 20 % der Investitionen ausmachen.

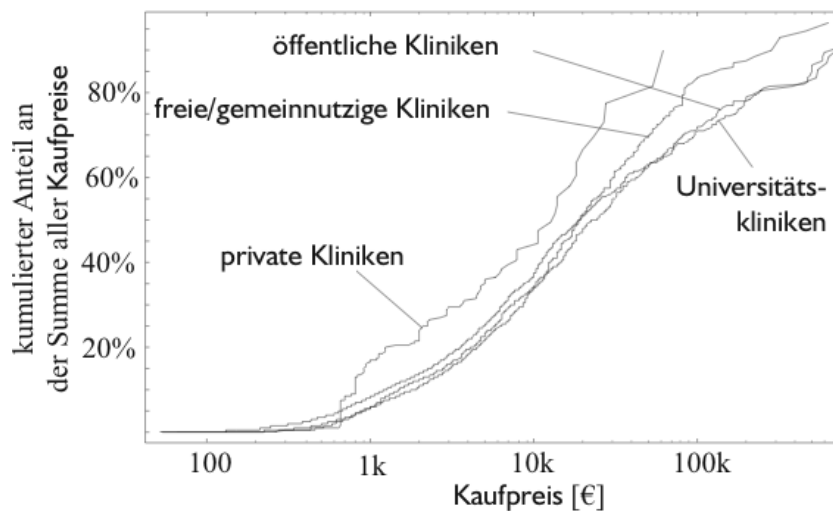


Abbildung 4.12: In dieser Abbildung ist der kumulierte Anteil der Kaufpreise statt über der Gerätezahl (vgl. Abbildung 4.11) über dem Kaufpreis logarithmisch aufgetragen. Wieder sind die Kaufpreise aufsteigend sortiert. Man kann erkennen, in welchen Kaufpreisregionen unterschiedliche Anteile der Investitionen der einzelnen Krankenhäuser vorliegen. Deutung siehe Text.

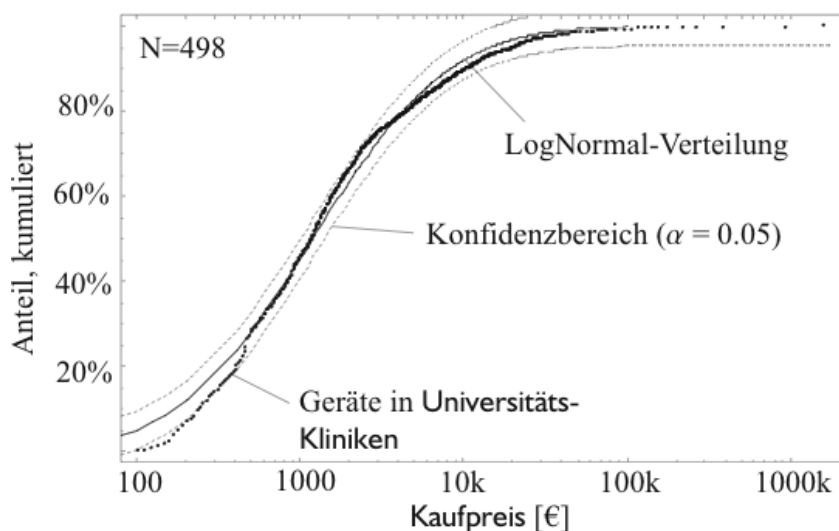


Abbildung 4.13: Beispielhafte Regression einer log-Normalverteilung an die empirische Verteilung der Kaufpreise aus Universitätskliniken. Eingezeichnet ist auch das Konfidenzintervall für  $\alpha=0.05$  nach Kolmogoroff-Smirnow (siehe Abschnitt 2.3.3). Zur Diskussion siehe Text.

Approximation der vorliegenden Daten steht insbesondere vor dem Hintergrund der schlechten Qualität der Daten in keinem Verhältnis zum Aufwand.

### Darstellung der ermittelten Regressionsparameter

In den folgenden Abbildungen sind die ermittelten Parameter für die Verteilungen der Kaufpreise für verschiedene Gerätegruppen angegeben. Die Grafiken geben nicht nur den Erwartungswert des Kaufpreises an sondern auch sein Zustandekommen. Daher ist der Erwartungswert im Vergleich zu anderen Gerätegruppen eher durch den Häufungspunkt der Kaufpreise oder die Schiefe der Verteilung beeinflusst (siehe auch Abschnitt 2.3.2)?

In Abbildung 4.14 ist zu erkennen, dass obwohl der Häufungspunkt der Kaufpreise von Geräten aus privaten Krankenhäusern höher liegt als bei Universitätskliniken, der Erwartungswert jedoch niedriger liegt. Dies ist auf die stärkere Schiefe der Verteilung bei Universitätskliniken zurückzuführen – d.h. einige teure Geräte ziehen den Erwartungswert nach oben. Die Kaufpreise sind im Vergleich zu privaten Kliniken weniger gleich verteilt (vergleiche dazu auch Abbildung 4.11). Frei / gemeinnützige Kliniken und öffentliche Kliniken haben ähnliche Häufungspunkte und Streuungen, was zu ähnlichen Erwartungswerten der Kaufpreise führt.

Da die vier wesentlich verschiedenen Krankenhausarten untersucht werden, ist eine Untersuchung der Abhängigkeit der ermittelten log-Normalverteilungsparameter  $\alpha_L$  und  $\beta_L$  für die niedrige Zahl verschiedener Krankenhausarten kaum sinnvoll. Insbesondere ist der in Abschnitt 2.3.3 vorgestellte Spearmansche Rang-Korrelationskoeffizient erst ab einer Anzahl von 6 Wertepaaren einsetzbar. Aus diesen Gründen ist hier eine Untersuchung der ermittelten Parameterwerte  $\alpha_L$  und  $\beta_L$  auf eine gegenseitige Abhängigkeit nicht möglich.

Alle MT-Gerätearten in Abbildung 4.15 befinden sich in einem Rechteck, dass durch die MT-Gerätearten Stoffaustausch (niedriger Häufungspunkt, niedrige Streuung), La-

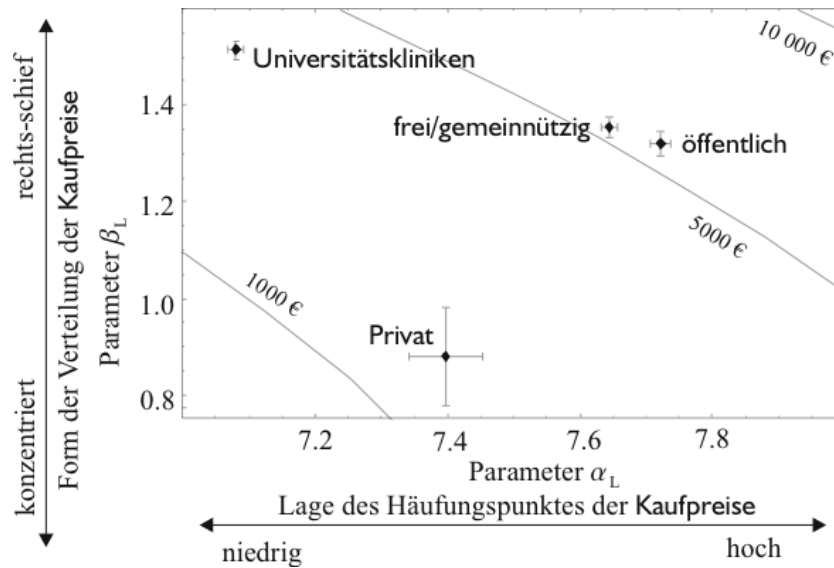


Abbildung 4.14: Aufgetragen sind hier die beiden Parameter der log-Normalverteilung für die Verteilungen der Kaufpreise aus verschiedenen Krankenhausarten. Parameter  $\alpha$  gibt die Lage des Häufungspunktes an, Parameter  $\beta$  die Streuung der logarithmierten Kaufpreise und damit die Schiefe der Verteilung. Durch gestrichelte Linien sind beispielhaft Erwartungswerte einiger Verteilungen angegeben. Deutung siehe Text.

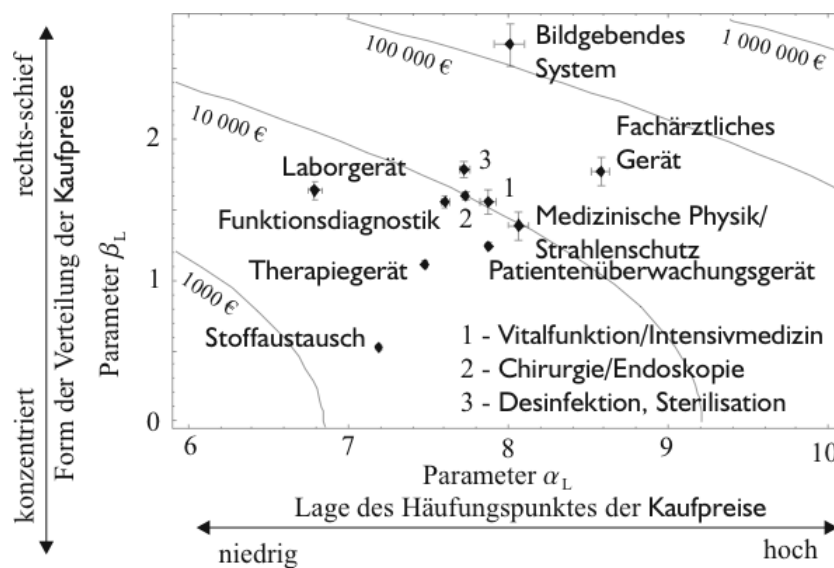


Abbildung 4.15: Aufgetragen sind hier die beiden Parameter der log-Normalverteilung für die Verteilungen der Kaufpreise für verschiedene MT-Gerätearten. Parameter  $\alpha$  gibt die Lage des Häufungspunktes an, Parameter  $\beta$  die Streuung der logarithmierten Kaufpreise und damit die Schiefe der Verteilung. Durch gestrichelte Linien sind beispielhaft Erwartungswerte einiger Verteilungen angegeben. Deutung siehe Text.

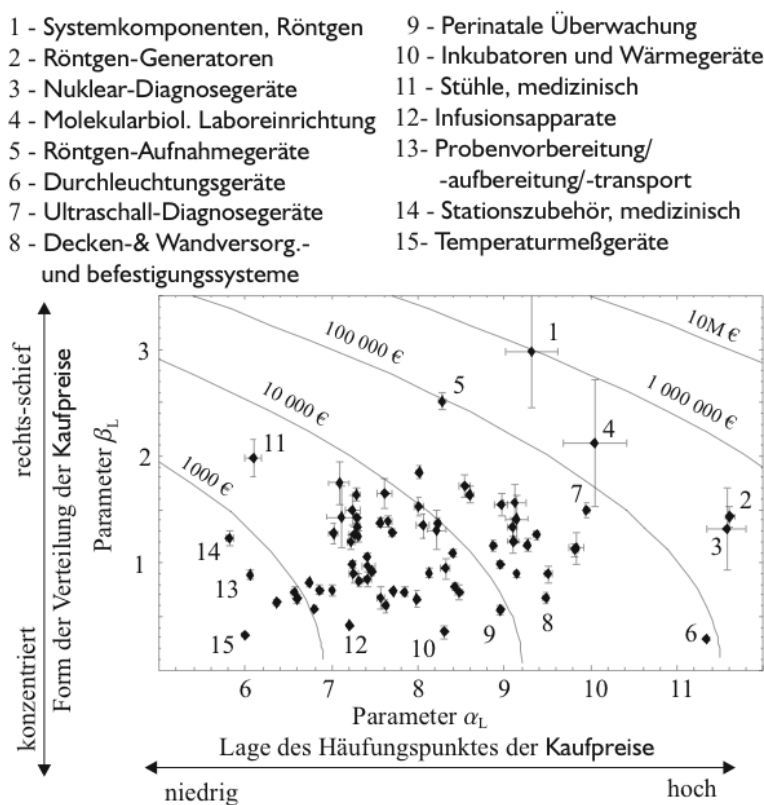


Abbildung 4.16: Aufgetragen sind hier die beiden Parameter der log-Normalverteilung für die Verteilungen der Kaufpreise für verschiedenen MT-Untergerätearten. Parameter  $\alpha$  gibt die Lage des Häufungspunktes an, Parameter  $\beta$  die Streuung der logarithmierten Kaufpreise und damit die Schiefe der Verteilung. Durch gestrichelte Linien sind beispielhaft Erwartungswerte einiger Verteilungen angegeben. Deutung siehe Text.

borgerät (niedriger Häufungspunkt, hohe Streuung), Bildgebende Systeme (hoher Häufungspunkt, hohe Streuung) und Fachärztliches Gerät (hoher Häufungspunkt, niedrige Streuung) markiert wird. Die Drehung des Rechtecks ist auf die leichte Korrelation von Streuung und Lage des Häufungspunktes zurückzuführen.

Zur Untersuchung einer Abhängigkeit der ermittelten Parameter  $\alpha_L$  und  $\beta_L$  wird nach Abschnitt 2.3.3 der Spearmansche Rang-Korrelationskoeffizient  $r_S$  ermittelt. Es ergibt sich  $r_S = 0.436$ . Der kritische Wert lautet  $k_{11;0.05} = 0.618$ . Die Nullhypothese kann also nicht auf dem Signifikanzniveau  $\alpha = 0.05$  abgelehnt werden – d.h. es besteht kein signifikanter Zusammenhang zwischen den ermittelten Regressionsparameter. Dies kann man als Hinweis darauf werten, dass die verwendete log-Normalverteilung nicht zu viele Parameter benötigt, um die Daten zu beschreiben.

In Abbildung 4.16 sind im Gegensatz zu Abbildung 4.15 nicht die MT-Gerätearten sondern die MT-Untergerätearten angegeben. Für den Spearmansche Rang-Korrelationskoeffizient ergibt sich hier  $r_S = 0.162$ . Der kritische Wert lautet  $k_{73;0.05} = 0.231$ . Die Nullhypothese kann also nicht auf dem Signifikanzniveau  $\alpha = 0.05$  abgelehnt werden – d.h. es besteht kein signifikanter Zusammenhang zwischen den ermittelten Regressionsparameter.

### 4.3.4 Verteilung der Erfassungszeitpunkte

In Abbildung 4.17 sind die kumulierten und normierten Inbetriebnahmen je Krankenhaus aufgetragen. Man kann erkennen, dass die Geräteerfassung über der Zeit für die verschiedenen Datenbanken sehr unterschiedlich verläuft. Vier Hauptformen des Erfassungsverlaufes mit wahrscheinlicher Ursache sind jedoch auch in Mischformen erkennbar:

- Knick: schlagartige Einführung der Erfassung
- Bogen: schleichende Einführung der Erfassung
- „Zick-Zack“-Verlauf: unstetige / nicht zeitnahe Erfassung
- linearer Verlauf: stetige / zeitnahe Erfassung

Dabei sind die Anteile unter den Krankenhäusern jeweils für schlagartige / schleichende Einführung und „Zick-Zack“ / linearer Verlauf etwa 50 % / 50 %.

Die Datenbanken unterscheiden sich jedoch nicht nur in Art und Weise der Erfassung (stetig / unstetig) und Einführung der Erfassung (schlagartig / schleichend), sondern auch durch das Datum der Einführung.

Vor dem Hintergrund dieser Zeitabhängigkeit der Erfassung (zeitliche Datenfehler, siehe Abschnitt 4.1.2) macht die Untersuchung von absoluten Zahlen in Abhängigkeit von der Zeit keinen Sinn, bzw. muss sehr kritisch beurteilt werden. Dienen z.B. absolute Zahlen in Quartalsvergleichen nun dazu, Veränderungen im Gerätebestand zu ermitteln oder nur dazu Veränderungen der Erfassung zu dokumentieren?

Wie schon in Abschnitt 4.1.2 dargestellt können durch Verwendung von relativen Größen die Fehler verkleinert werden, wenn vorausgesetzt wird, dass die Lücken in der Erfassung nicht von den zu untersuchenden Größen abhängen.

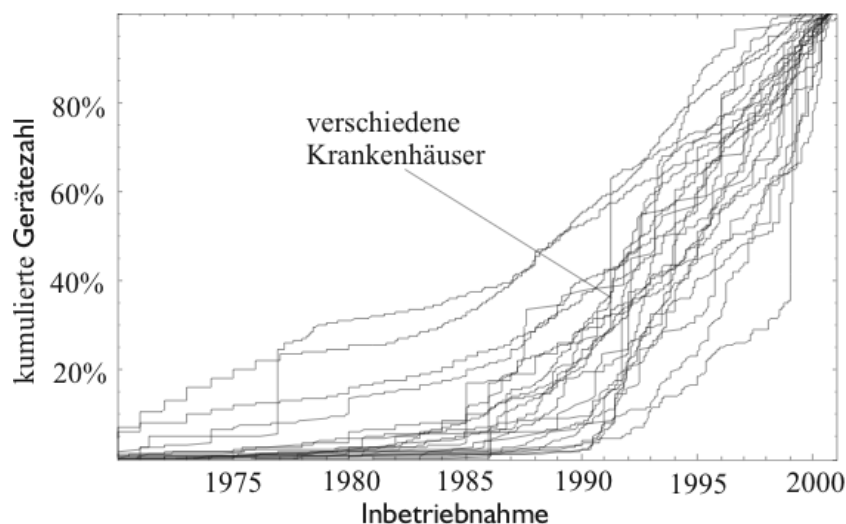


Abbildung 4.17: Kumulative Darstellung der Inbetriebnahmen für verschiedene Krankenhäuser. Details siehe Text.

Tabelle 4.3: In dieser Tabelle sind absolute Maßnahmenzahlen an medizin-technischen Geräten angegeben, von allen Maßnahmen, Maßnahmen mit erfasstem Durchführungszeitpunkt, Kosten oder Inbetriebnahme des betroffenen Gerätes.

Maßnahmenzahl	alle Maßnahmen	Maßnahmen mit Kosten
alle Maßnahmen	219 029 (100 %)	88 681 (40.5 %)
	betroffenes Gerät mit Inbetriebnahme 184 899 (84.4 %)	betroffenes Gerät mit Inbetriebnahme 73 269 (33.5 %)
Maßnahmen mit Durchführungszeit- punkt	215 967 (98.6%)	86 251 (39.4%)
	betroffenes Gerät mit Inbetriebnahme 182 234 (83.2 %)	betroffenes Gerät mit Inbetriebnahme 71 061 (32.4 %)

## 4.4 Untersuchung der Leistungsdaten

### 4.4.1 Verteilung der Häufigkeiten

In diesen Abschnitt werden die relativen Häufigkeiten von Maßnahmen untersucht. Zeitliche Informationen werden dabei ausgeblendet. Diese werden in Abschnitt 4.4.4 eingehend untersucht. Die Darstellungen sind bis auf einen Unterschied zu denen aus Abschnitt 4.3 identisch: sie zeigen statt der Gerätezahl die Maßnahmenzahl.

In Tabelle 4.3 sind die absoluten Maßnahmenzahlen an medizin-technischen Geräten angegeben, die in dieser Arbeit zur Verfügung stehen. Sobald man jedoch fordert, dass zu den Maßnahmen auch z.B. die Kosten oder zum betroffenen Gerät die Inbetriebnahme anzugeben sind, nehmen die Maßnahmenzahlen deutlich ab. Da in den folgenden Abbildungen meist relative Maßnahmenzahlen verwendet werden, kann man anhand der Tabelle 4.3 sehen, auf wie viele Geräte sich diese absolut beziehen.

### Abhängigkeit der relativen Maßnahmenzahl vom Krankenhaus

Da in Abbildung 4.4 festgestellt wurde, dass die Anteile der Gerätezahl für verschiedene MT-Gerätearten von der Krankenhausart abhängt, wird hier mit Abbildung 4.18 geprüft, ob diese Unterschiede auch bezüglich der Maßnahmenzahl vorliegen. Man erkennt zwar, dass in Universitätskliniken die Anteile der Maßnahmenzahlen an Laborgeräten – wie zu erwarten – gegenüber anderen Krankenhausarten leicht angehoben sind. Jedoch tritt diese Abweichung nicht so stark in Erscheinung wie in Abbildung 4.4. Insgesamt ist keine deutliche Tendenz aus Abbildung 4.18 ablesbar.

Statistisch kann man die Kreuztabelle, die der Abbildung 4.4 zugrunde liegt mit Hilfe des  $\chi^2$ -Tests auf Unabhängigkeit prüfen. Das Testverfahren wird in Abschnitt 2.3.3 vorgestellt. Für die dargestellten Daten ergibt sich eine Testgröße von  $T = 16\,504$ . Der kritische Wert für  $\alpha = 0.05$  und 36 Freiheitsgrade lautet  $k_{36;0.05} = 51.00$ . Damit wird die Nullhypothese mit überwältigender Deutlichkeit abgelehnt – die relativen Maßnahmenzahlen der MT-Gerätearten hängen also hoch signifikant von den Krankenhäusern ab, obwohl keine deutlichen Tendenzen ablesbar sind.

Betrachtet man nur die beiden Universitätskliniken, so ergibt sich eine Testgröße von

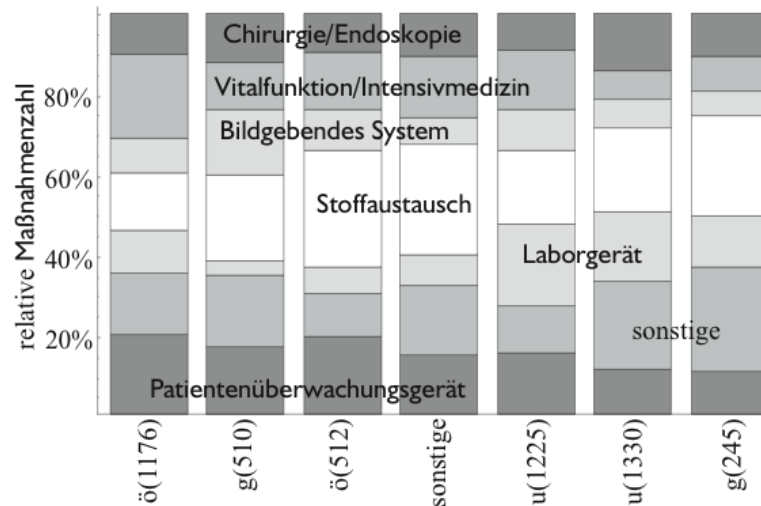


Abbildung 4.18: Balkendiagramm der relativen relative Maßnahmehzahlen je MT-Gerätearten für verschiedene Krankenhäuser. Das Diagramm wurde durch den Algorithmus aus Abschnitt 2.5.2 sortiert – damit sind auch ab der siebthäufigsten Kategorie Krankenhäuser und Gerätebezeichnungen zusammengefasst. Deutung siehe Text.

$T = 6069$ . Der kritische Wert lautet für  $\alpha = 0.05$  und 6 Freiheitsgrade  $t_{6;0.05} = 12.59$ . Für nicht-Universitätskliniken ergibt sich  $T = 4881$  und  $t_{24;0.05} = 36.42$ . Damit sind auch die Verteilungen innerhalb der Universitätskliniken und nicht-Universitätskliniken vom Krankenhaus abhängig.

Zur Beurteilung der Stärke der Abhängigkeit sind nach Abschnitt 2.3.3 die Testgrößen nicht geeignet. Aus diesem Grunde wird hier der korrigierte Kontingenzkoeffizient  $C_{korr}$  von Pawlik verwendet (siehe Abschnitt 2.3.3). Für die alle Daten der Abbildung 4.4 ergibt sich  $C_{korr} = 28.56\%$ , für Universitätskliniken  $C_{korr} = 20.06\%$  und für nicht-Universitätskliniken  $C_{korr} = 24.70\%$ . Alleine aus diesen Daten wäre eine Gruppierung dieser Daten wie in Abschnitt 4.3.1 nicht zu rechtfertigen. Dennoch sollen im Folgenden Universitätskliniken und nicht-Universitätskliniken aus Gründen der Vergleichbarkeit der Abbildungen getrennt betrachtet werden.

### Treemaps relativer Maßnahmehzahlen für MT-Geräteart und -Untergeräteart

Aus Gründen der Vergleichbarkeit zu Abbildungen 4.5 und 4.6 aus Abschnitt 4.3.1 sind in den Abbildungen 4.19 und 4.20 auch hier die Anteile der Maßnahmehzahlen gruppiert nach MT-Geräteart und -Untergeräteart jeweils für Universitätskliniken und nicht-Universitätskliniken dargestellt. Wie man schon an Abbildung 4.18 erkennen kann, fallen die Unterschiede zwischen den Krankenhausarten bezüglich der Maßnahmehzahlen nicht so deutlich aus wie bezüglich der Gerätezahlen. Tabellen hierzu finden sich im Anhang D insbesondere in Tabelle D.3.

### Relative Maßnahmehzahlen je Arbeitsbeschreibung und MT-Geräteart

In Abbildung 4.21 ist erkennbar, dass die relativen Maßnahmehzahlen mit der Arbeitsbeschreibung Reparatur und Prüfung über verschiedene MT-Gerätearten konstant bleiben

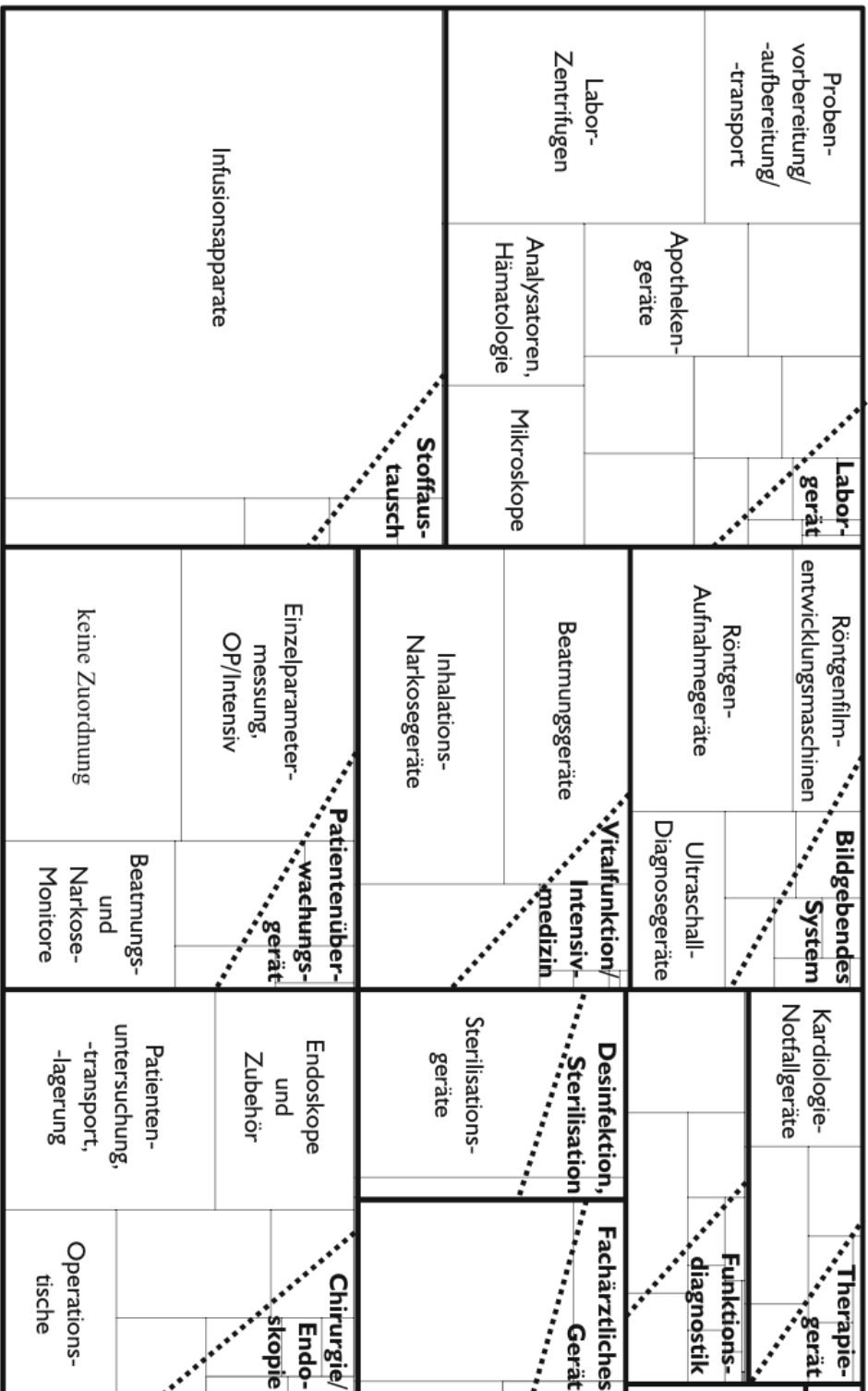


Abbildung 4.19: Squarified Treemap nach Algorithmus aus Abschnitt 2.5 Die Flächen entsprechen der relativen Maßnahmenzahlen je MT-Geräteart und -Untergeräteart. Es sind die Anteile für Geräte aus Universitätskliniken angegeben. Deutung siehe Text.

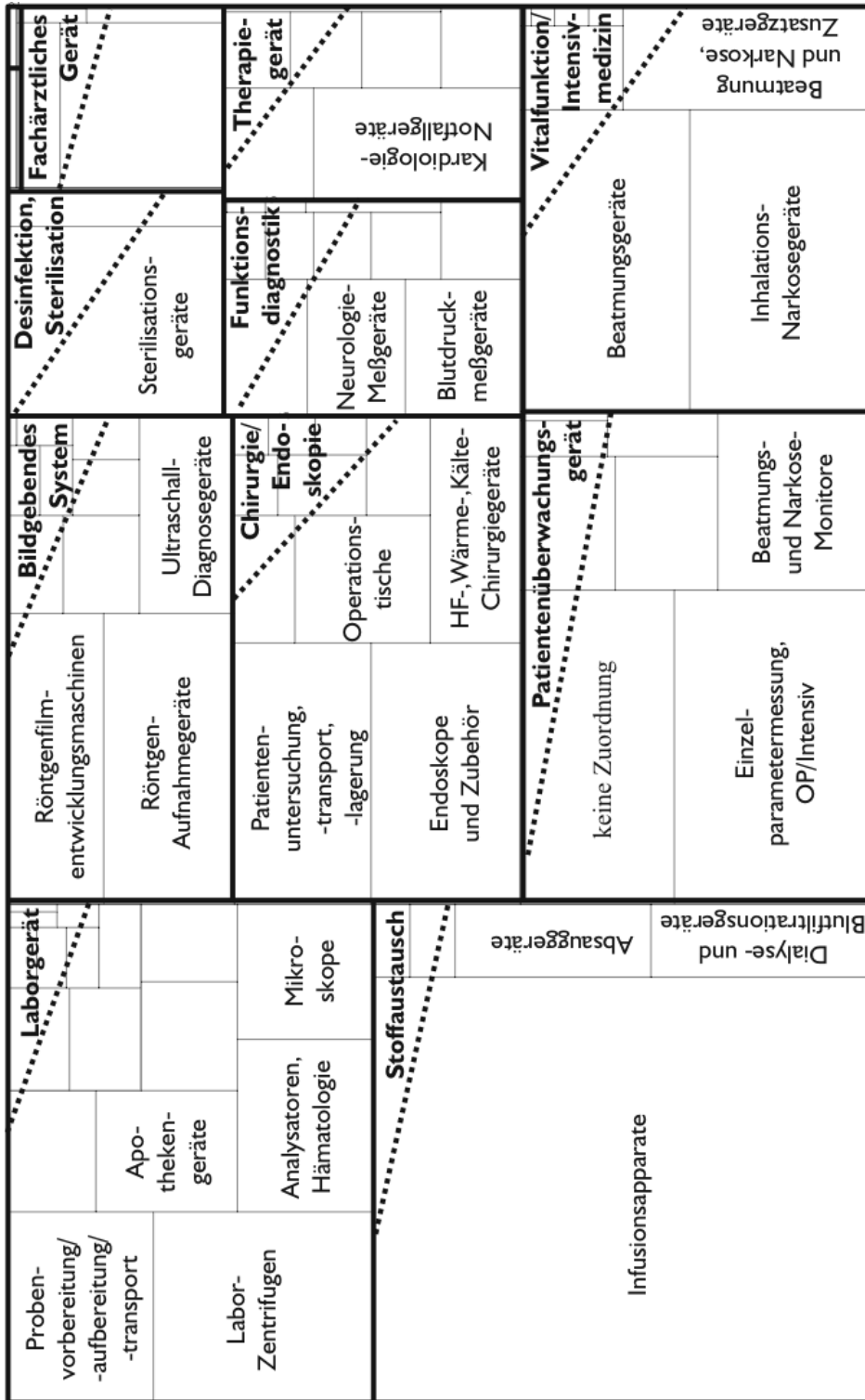


Abbildung 4:20: Squarified Treemap nach Algorithmus aus Abschnitt 2.5 Die Flächen entsprechen der relativen Maßnahmenzahlen je MT-Geräteart und -Untergeräteart. Es sind die Anteile für Geräte aus nicht- Universitätskliniken angegeben. Deutung siehe Text.

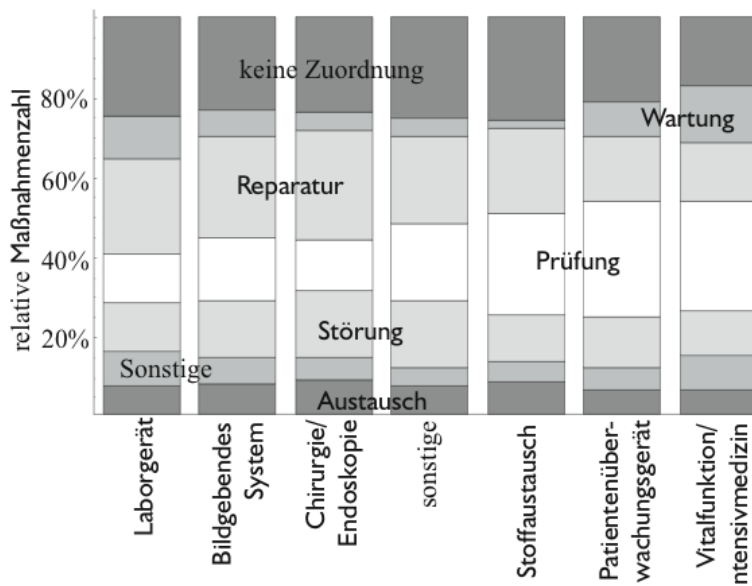


Abbildung 4.21: Balkendiagramm der relativen Maßnahmehzahlen für verschiedene MT-Gerätearten und Arbeitsbeschreibungen. Das Diagramm wurde durch den Algorithmus aus Abschnitt 2.5.2 sortiert – damit sind auch ab der siebthäufigsten Kategorie Arbeitsbeschreibungen und MT-Gerätearten zusammengefasst. Deutung siehe Text.

– die Summanden ändern sich jedoch. Relativ viele Reparaturen und wenige Prüfungen liegen bei MT-Gerätearten Laborgerät, Bildgebendes Systemen und Chirurgiegerät vor. Relativ viele Prüfungen und wenige Reparaturen liegen bei Stoffaustausch-, Patientenüberwachungsgeräten und Vitalfunktion / Intensivmedizin vor. Außerdem fällt auf, dass die sonstigen MT-Gerätearten in der Mitte liegen.

In Hinblick auf die in Abschnitt 2.4.2 vorgestellten Kennzahlen liefert diese Abbildung am ehesten Informationen zur Wartungs- / Instandhaltungsquote. Auch wenn hier nur die Anzahl der Maßnahmen und nicht die mit ihnen verbundenen Kosten betrachtet werden. Auch ist hier eine Einteilung der Maßnahmen anhand der Arbeitsbeschreibungen nicht immer eindeutig in Wartungs- oder Instandsetzungsmaßnahmen möglich. Der Anteil der Wartungsmaßnahmen schwankt für verschiedene MT-Gerätearten zwischen 5 und 10 %.

Für die dargestellten Daten ergibt sich eine Testgröße von  $T = 24\,557$ . Der kritische Wert für  $\alpha = 0.05$  und 36 Freiheitsgrade lautet  $k_{36;0.05} = 51.00$ . Damit wird die Nullhypothese mit überwältigender Deutlichkeit abgelehnt – die relativen Maßnahmehzahlen der MT-Gerätearten hängen also hoch signifikant von den Arbeitsbeschreibungen ab. Es ergibt sich ein  $C_{korr} = 26.21\%$ .

Untersucht man nun die Maßnahmehzahlen von Reparatur und Prüfung gegen die restlichen Arbeitsbeschreibungen in Abhängigkeit der MT-Gerätearten, so ergibt sich eine Testgröße von  $T = 4036$ . Der kritische Wert für  $\alpha = 0.05$  und 6 Freiheitsgrade lautet  $k_{6;0.05} = 12.59$ . Es ergibt sich ein  $C_{korr} = 14.27\%$ . Obwohl die Abhängigkeit nach dem  $\chi^2$ -Test wieder hoch signifikant ist, fällt der kleine Kontingenzkoeffizient auf – was den optischen Eindruck der Konstanz dieser Anteilssumme bestätigt.

Untersucht man nun die Maßnahmehzahlen von Reparatur gegen Prüfung in Abhän-

gigkeit der MT-Gerätearten, so ergibt sich eine Testgröße von  $T = 6347$ . Der kritische Wert für  $\alpha = 0.05$  und 6 Freiheitsgrade lautet  $k_{6,0.05} = 12.59$ . Es ergibt sich ein  $C_{korr} = 45.48\%$ . Anhand des Kontingenzkoeffizienten zeigt sich also – in Übereinstimmung mit dem optischen Eindruck – dass die Maßnahmenzahlen der Reparaturen und die der Prüfungen deutlich mehr von der MT-Geräteart abhängen als die Maßnahmenzahl der Reparaturen und der Prüfungen.

### Relative Maßnahmenzahlen je Leistungsart und MT-Geräteart

Untersucht man die Abhängigkeit der relativen Maßnahmenzahlen der Leistungsarten von der MT-Geräteart, so fällt beim Betrachten von Abbildung 4.22 zunächst der große Anteil von nicht zugeordneten Maßnahmen auf. Unter der nicht unkritischen aber minimal willkürlichen Annahme, dass die nicht zugeordneten Maßnahmen zu gleichen Teilen aus Eigen- und Fremdleistungen zusammensetzen, geben die gestrichelten Linien an, wie sich die Anteile auf die beiden Leistungsarten verteilen. Die geringen Anteile der Maßnahmenzahlen von Eigenleistungen der Laborgeräten, Bildgebenden Systemen und Geräten der Vitalfunktion / Intensivmedizin müssen insbesondere vor dem Hintergrund des hohen Anteils nicht zugeordneter Maßnahmen als unsicher angesehen werden. Aus diesem Grund entfällt auch eine statistische Analyse.

Für die in Abschnitt 2.4.2 vorgestellten Kennzahlen liefert diese Abbildung am ehesten Informationen zur Eigenservicequote. Auch wenn hier nur die Anzahl der Maßnahmen und nicht die mit ihnen verbundenen Kosten betrachtet werden. Auch ist hier eine Einteilung der Maßnahmen anhand der Leistungsarten nur sehr lückenhaft vorhanden. Dennoch bietet die Abbildung einen Anhaltspunkte für den Wert der Eigenservicequote. Der Anteil der Eigenleistungen beträgt für verschiedene MT-Gerätearten ca. 30 % bis 70 % und hängt deutlich von der MT-Geräteart ab.

### Relative Maßnahmenzahlen je Leistungsart und Krankenhausart

Auch bei Betrachtung von Abbildung 4.23 fällt zunächst der große Anteil von nicht zugeordneten Maßnahmen auf. Die Schwankung der Anteile muss insbesondere vor dem Hintergrund des hohen Anteils nicht zugeordneter Maßnahmen als vernachlässigbar angesehen werden. Eine statistische Analyse entfällt.

Auch diese Abbildung liefert am ehesten Informationen zur Eigenservicequote aus Abschnitt 2.4.2. Die Eigenservicequote (nach Maßnahmenzahl und nicht nach Kostensumme) beträgt für verschiedene Krankenhausarten ca. 50 %.

## 4.4.2 Verteilung der Ausgaben

In diesen Abschnitt werden Maßnahmenzahlen statt Kostensummen untersucht. Auch hier werden, wie in Abschnitt 4.4.1, zeitliche Informationen ausgeblendet. Diese werden in Abschnitt 4.4.4 eingehend untersucht. Die Verteilung der einzelnen Kosten – also nicht der Summe wie hier – erfolgt in Abschnitt 4.4.3.

### Relative Kostensummen je MT-Geräteart und Krankenhaus

Obwohl in Abbildung 4.24 von Krankenhaus zu Krankenhaus deutliche Unterschiede in den Anteilen der Kostensummen je MT-Geräteart zu erkennen sind, lässt sich keine ein-

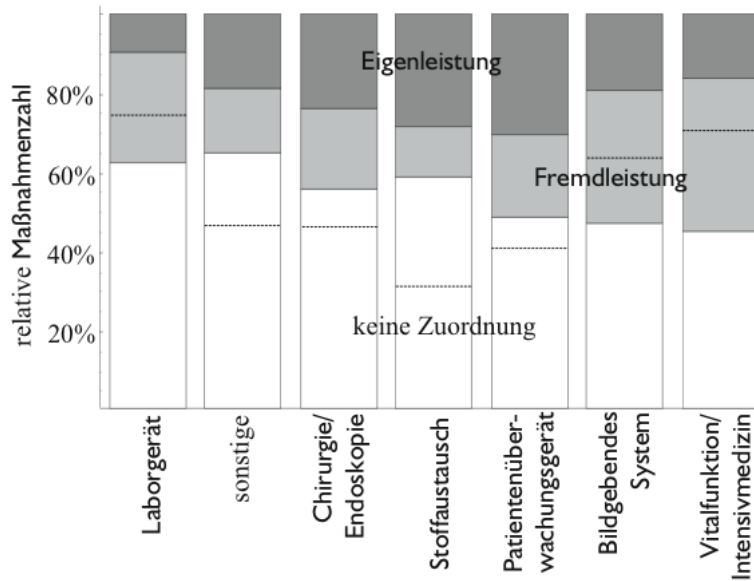


Abbildung 4.22: Balkendiagramm der relative Maßnahmenzahlen für verschiedene Leistungsarten und MT-Gerätearten. Das Diagramm wurde durch den Algorithmus aus Abschnitt 2.5.2 sortiert – damit sind auch ab der siebt-häufigsten Kategorie MT-Gerätearten zusammengefasst. Die gestrichelten Linien deuten die Anteile an, die sich bei Auslassen der nicht zugeordneten Maßnahmen ergeben. Deutung siehe Text.

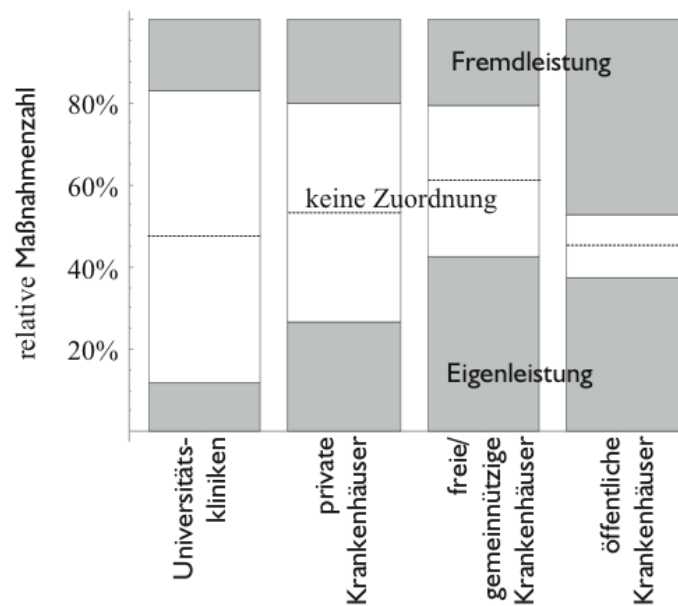


Abbildung 4.23: Balkendiagramm der relativen Maßnahmenzahlen für verschiedene Leistungsarten und Krankenhausarten. Das Diagramm wurde durch den Algorithmus aus Abschnitt 2.5.2 sortiert. Die gestrichelten Linien deuten die Anteile an, die sich bei Auslassen der nicht zugeordneten Maßnahmen ergeben. Deutung siehe Text.

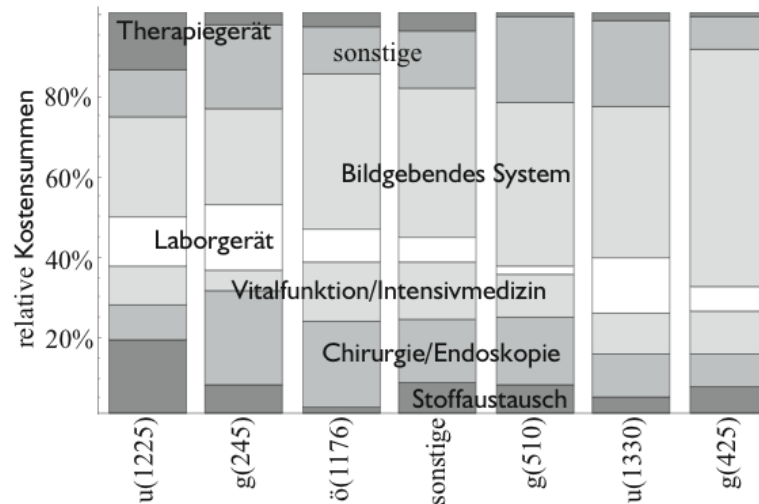


Abbildung 4.24: Balkendiagramm der Summenanteile der angefallenen Kosten für Maßnahmen an verschiedenen Gerätegruppen für verschiedene Krankenhäuser. Das Diagramm wurde durch den Algorithmus aus Abschnitt 2.5.2 sortiert – damit sind auch ab der siebt-häufigsten Kategorie Krankenhäuser und Gerätebezeichnungen zusammengefasst. Deutung siehe Text.

deutige Tendenz feststellen. Bemerkenswert ist jedoch die Zunahme dieser Variabilität ausgehend von der Gerätezahl in Abbildung 4.4 über die Maßnahmenzahl in Abbildung 4.18 bis jetzt zur Kostensumme der Maßnahmen.

Diese Zunahme an Variabilität ist darauf zurückzuführen, dass bei Vorliegen der Verhältnisse in Abbildung 4.4 nur die Gerätezahlen schwanken, in Abbildung 4.18 Gerätezahlen und die Maßnahmenzahlen und schließlich in Abbildung 4.24 Gerätezahlen, die Maßnahmenzahlen und deren Kosten.

Für die Abbildung 4.24 ergibt sich eine Testgröße von  $T = 32.28 \cdot 10^6$ . Der kritische Wert für  $\alpha = 0.05$  und 36 Freiheitsgrade lautet  $k_{36;0.05} = 51.00$ . Damit wird die Nullhypothese mit überwältigender Deutlichkeit abgelehnt – die relativen Kostensummen der MT-Gerätearten hängen also hoch signifikant von den Krankenhäusern ab. Es ergibt sich ein  $C_{korr} = 37.57 \%$ .

Für Universitätskliniken ergeben sich  $T = 11.39 \cdot 10^6$ ,  $k_{6;0.05} = 12.59$  und  $C_{korr} = 35.15 \%$ . Für nicht-Universitätskliniken erhält man  $T = 6.45 \cdot 10^6$ ,  $k_{24;0.05} = 36.42$  und  $C_{korr} = 35.15 \%$ .

### Treemaps relativer Kostensummen je MT-Geräteart und -Untergeräteart

Aus Gründen der Vergleichbarkeit mit den Abschnitten 4.3.1, 4.3.2 und 4.4.1 sind in den Abbildungen 4.25 und 4.26 die relativen Kostensummen jeweils für Universitätskliniken und nicht-Universitätskliniken dargestellt. Tabellen hierzu finden sich im Anhang D insbesondere in Tabelle D.4.

### Relativen Kostensummen je MT-Geräteart und Arbeitsbeschreibung

An Abbildung 4.27 wird deutlich, dass die Anteile der Kostensummen einer bestimmten Arbeitsbeschreibung in nicht erkennbarer Systematik von der MT-Geräteart abhängen.

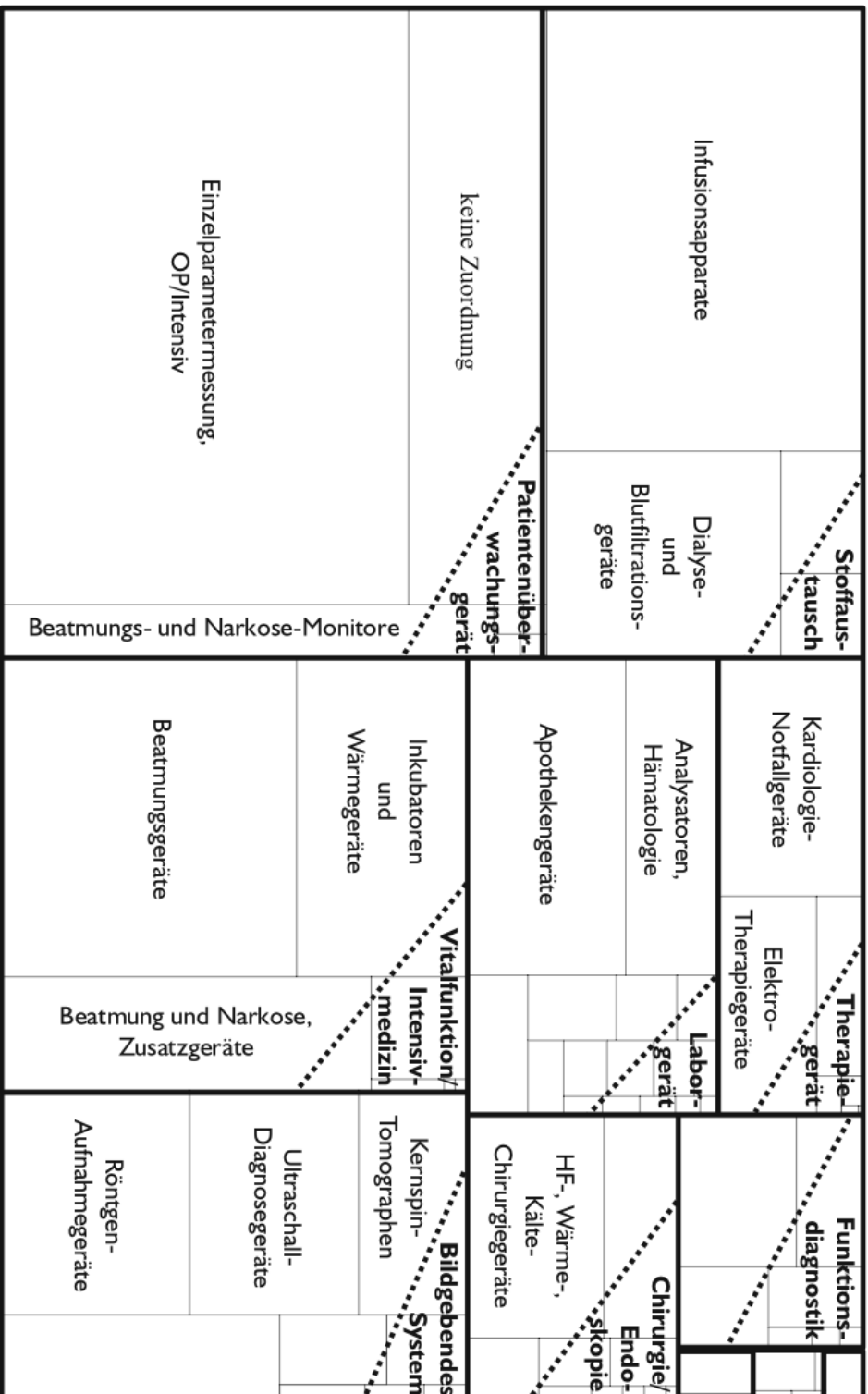


Abbildung 4.25: Squarified Treemap nach Algorithmus aus Abschnitt 2.5 Die Flächen entsprechen den Kostensummen an Geräten der jeweiligen MT-Geräteart und -Untergeräteart. Es sind die Anteile für Universitätskliniken angegeben. Deutung siehe Text.



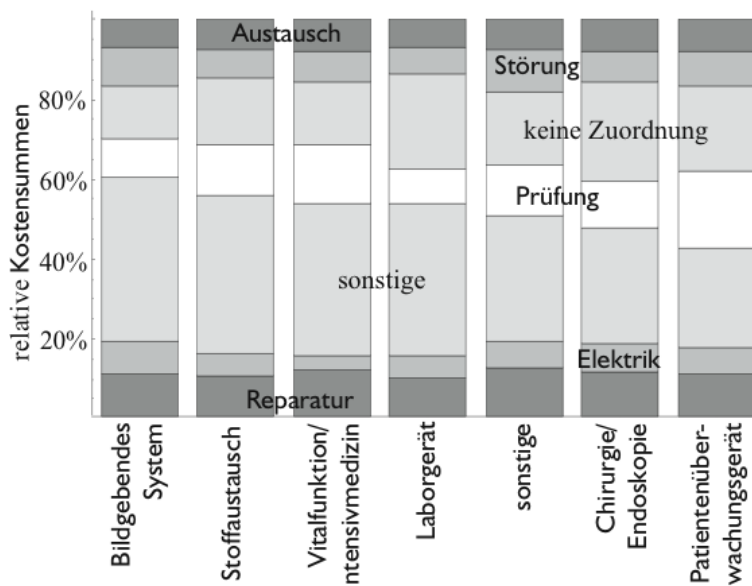


Abbildung 4.27: Balkendiagramm der relative Kostensummen für verschiedene MT-Gerätearten und Arbeitsbeschreibungen. Das Diagramm wurde durch den Algorithmus aus Abschnitt 2.5.2 sortiert – damit sind auch ab der siebthäufigsten Kategorie MT-Gerätearten und Arbeitsbeschreibungen zusammengefasst. Deutung siehe Text.

In Hinblick auf die in Abschnitt 2.4.2 vorgestellten Kennzahlen liefert diese Abbildung am ehesten Informationen zur Wartungs- / Instandhaltungsquote. Zwar ist eine Einteilung der Maßnahmen anhand der Arbeitsbeschreibungen nicht immer eindeutig in Wartungs- oder Instandsetzungsmaßnahmen möglich, dennoch bietet die Abbildung einen Anhaltspunkte für den Wert der Wartungs- / Instandhaltungsquote. Der Kostenanteil für Wartungen ist so gering, dass er unter „sonstige“ subsummiert wird. Dies bedeutet aber gleichzeitig, dass sein Anteil kleiner als der kleinste Anteil ist, der in der Abbildung dargestellt ist. Die Kosten für Wartungen machen demnach weniger als 10 % aller Kosten aus.

Für die Abbildung 4.27 ergibt sich eine Testgröße von  $T = 80.38 \cdot 10^6$ . Der kritische Wert für  $\alpha = 0.05$  und 36 Freiheitsgrade lautet  $k_{36;0.05} = 51.00$ . Damit wird die Nullhypothese mit überwältigender Deutlichkeit abgelehnt – die relativen Kostensummen der Arbeitsbeschreibung hängen also hoch signifikant von den MT-Gerätearten ab. Es ergibt sich ein  $C_{korr} = 19.26 \%$ .

### Relative Kostensummen je MT-Geräteart und Leistungsart

In Abbildung 4.28 werden die relativen Kostensummen für verschiedene MT-Gerätearten und Leistungsarten dargestellt. Aufgrund der hohen Anteile nicht zugeordneter Kostensummen sind die in Abbildung 4.28 erkennbaren Unterschiede in Abhängigkeit der MT-Geräteart vernachlässigbar. Auch erscheint eine statistische Analyse mit dem  $\chi^2$ -Test nicht angebracht.

Für die in Abschnitt 2.4.2 vorgestellten Kennzahlen liefert diese Abbildung am ehesten Informationen zur Eigenservicequote. Zwar ist eine Einteilung der Maßnahmen an-

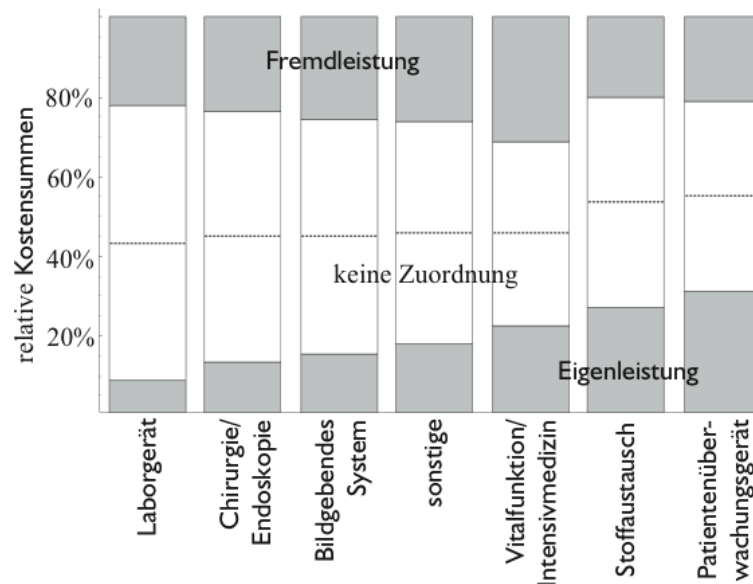


Abbildung 4.28: Balkendiagramm der relativen Kostensummen für verschiedene MT-Gerätearten und Leistungsarten. Das Diagramm wurde durch den Algorithmus aus Abschnitt 2.5.2 sortiert – damit sind auch ab der siebt-häufigsten Kategorie MT-Gerätearten zusammengefasst. Deutung siehe Text.

hand der Leistungsarten nur sehr lückenhaft vorhanden, dennoch bietet die Abbildung Anhaltspunkte für den Wert der Eigenservicequote. Der Kostenanteil der Eigenleistungen beträgt für verschiedene MT-Gerätearten ca. 50 %.

### Relative Kostensummen je Krankenhausart und Leistungsart

Auch in Abbildung 4.29 sind auf Grund der hohen Anteile nicht zugeordneter Kostensummen nur eingeschränkt Schlussfolgerungen möglich. Jedoch fällt bei öffentlichen Kliniken der geringe Eigenservicequote auf. Dies ist insbesondere deshalb bemerkenswert, weil gerade bei dieser Krankenhausart der Anteil der nicht zugeordneten Leistungsarten recht gering ist.

Die Eigenservicequote beträgt im Mittel ca. 60 %.

### 4.4.3 Verteilung der Leistungskosten

In diesem Abschnitt werden nicht die Verteilung der Kostensummen wie in Abschnitt 4.4.2 untersucht sondern die Verteilungen der Kosten je Maßnahme.

#### Darstellung der Verteilungsfunktion

In Abbildung 4.30 ist zu erkennen, dass die Kosten je Maßnahme noch stärker ungleich verteilt sind als die Kaufpreise von Geräten aus Abbildung 4.11. Auch die Unterschiede in den Ungleichverteilungen sind hier stärker. In privaten Kliniken sind die Kosten je Maßnahme wesentlich gleichmäßiger verteilt als in Universitätskliniken. Da bei diesen 80 % der Kostensummen nur 1 % der Maßnahmehzahl hervorgerufen werden, liegt die

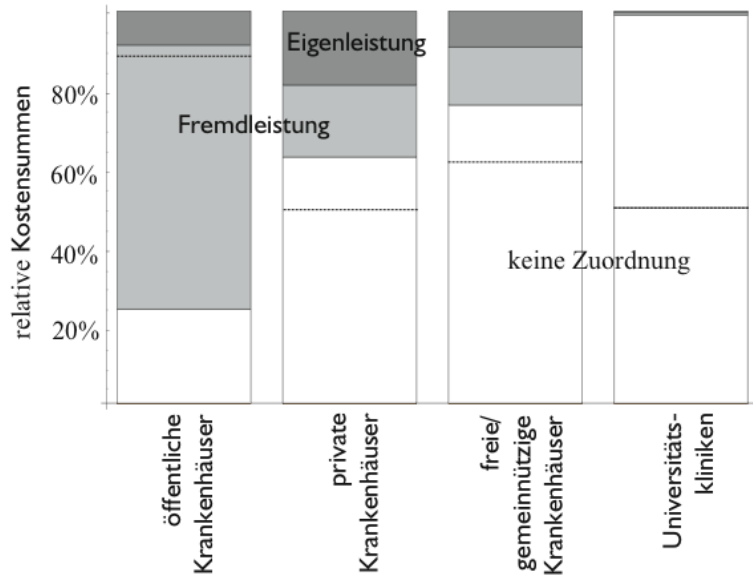


Abbildung 4.29: Balkendiagramm der relativen Kostensummen für verschiedene Krankenhausarten und Leistungsarten. Das Diagramm wurde durch den Algorithmus aus Abschnitt 2.5.2 sortiert. Deutung siehe Text.

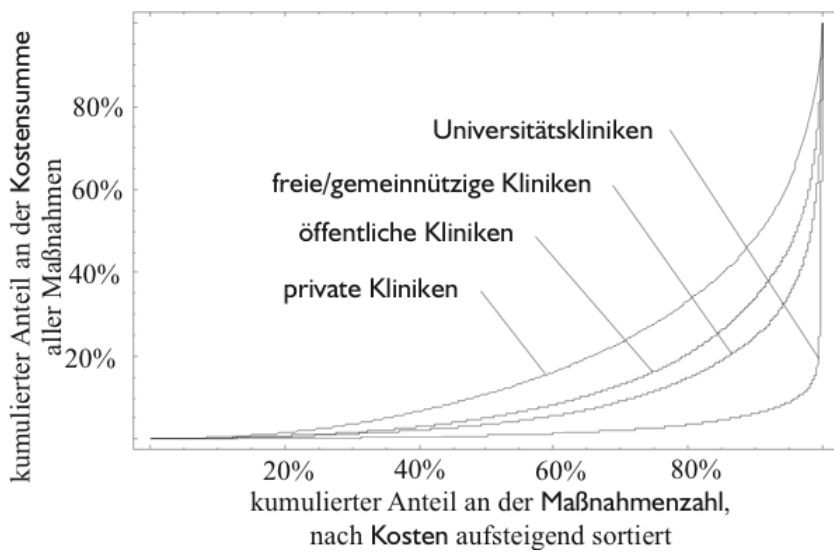


Abbildung 4.30: In dieser Abbildung die kumulierten relativen Kosten über der kumulierten relativen Maßnahmenzahl aufgetragen. Die Maßnahmen sind dabei aufsteigend nach Kosten sortiert. Eine solche Kurve nennt man auch Lorenzkurve. Die Fläche unter der gedachten (0,0)-(1,1)-Linie ist ein Maß für die Ungleichverteilung in diesem Fall der Kosten. Deutung siehe Text.

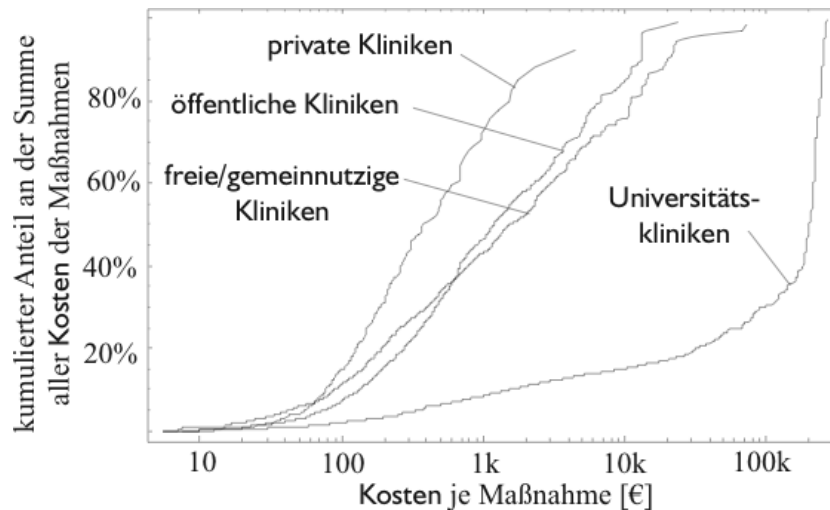


Abbildung 4.31: In dieser Abbildung sind die kumulierten normierten Kosten statt über Maßnahmenzahl (vgl. Abbildung 4.30) über den Kosten logarithmisch aufgetragen. Wieder sind die Maßnahmen nach Kosten aufsteigend sortiert. Man kann erkennen in welchen Kostenregionen unterschiedliche Anteile der Kostensummen der einzelnen Krankenhäuser vorliegen (siehe Text).

Vermutung nahe, dass hier mehrere Maßnahmen zu einer zusammengefasst wurden – wie es sich z.B. bei Wartungsverträgen anbietet.

Auch in Abbildung 4.31 ist deutlich zu erkennen, dass bei Universitätskliniken sehr hohe Kosten vorhanden sind – etwa 70 % der Kostensummen fallen für Maßnahmen an, die höher als 100 000 € sind. Für private Kliniken entfallen 80 % der Kostensummen auf Maßnahmen im Bereich von 100 € bis 2000 €. Wie in Abbildung 4.30 liegen freie / gemeinnützige und öffentliche Kliniken dicht beieinander.

Ein Grund für die Ungleichverteilung der Kosten kann – wie bei den Kaufpreisen die Gruppierung von Haupt- und Untergeräten – hier eine Gruppierung von Maßnahmen durch Wartungsverträge sein.

## Regression der Kosten

Wie in Abschnitt 4.3.3 sollen hier die Verteilungen der Kosten durch log-Normalverteilungen angenähert werden. In Abbildung 4.32 ist beispielhaft die Verteilung von Kosten aus Universitätskliniken angegeben. Eingezeichnet ist eine log-Normalverteilung, die durch Regression ermittelt wurde, sowie die Irrtumswahrscheinlichkeit für  $\alpha = 0.05$ , die sich aus dem Kolmogoroff-Smirnow-Test ergibt (siehe Abschnitt 2.3.3).

Man erkennt, dass für geringe Kosten die Kosten nicht signifikant log-normalverteilt sind. Dennoch soll im Folgenden die log-Normalverteilung zum Nähern der Daten verwendet werden, da der dadurch begangene Fehler als vernachlässigbar angesehen wird, bzw. der nötige Aufwand einer besseren Regression in keinem Verhältnis zum zu erwartenden Gewinn steht (vgl. auch Argumentation in Abschnitt 4.3.3).

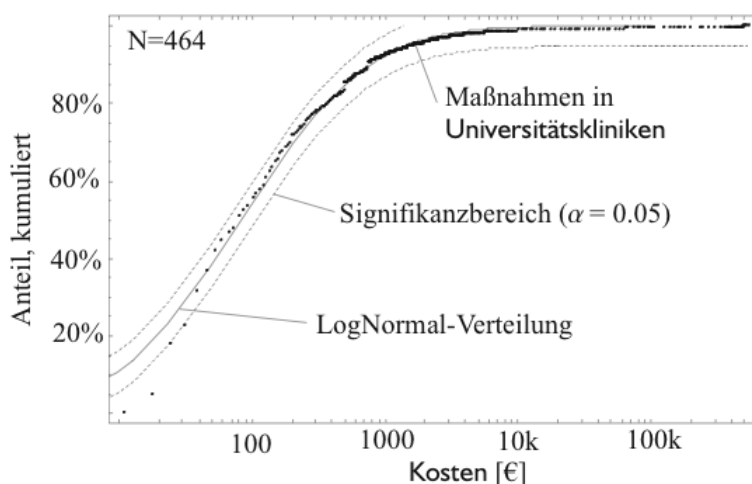


Abbildung 4.32: Beispielhafte Regression einer log-Normalverteilung an die empirische Verteilung der Kosten von Maßnahmen an Geräten aus Universitätskliniken. Eingezeichnet ist auch das Konfidenzintervall für  $\alpha=0.05$  nach Kolmogoroff-Smirnow (siehe Abschnitt 2.3.3). Zur Diskussion siehe Text.

### Darstellung der ermittelten Regressionsparameter

In Abbildung 4.33 sind die Regressionsparameter für Kosten für verschiedene Krankenhausarten angegeben. Es ist zu erkennen, dass die Verteilung der Kosten bei öffentlichen Kliniken (mit Universitätskliniken) stark schiefverteilt sind und einen niedrigen Häufungspunkt aufweisen. Bei privaten Kliniken ist es umgekehrt: die Verteilung weist einen hohen Häufungspunkt aus, besitzt aber eine geringere Schiefe. Bemerkenswert ist, dass es dadurch für beide Krankenhausarten zu einem ähnlich hohen Kostenerwartungswert kommt. Da die Verteilung der freien / gemeinnützigen Kliniken sowohl einen hohen Häufungspunkt als auch eine große Schiefe besitzt, ist hier der Kostenerwartungswert höher.

Da es nur vier verschiedene Krankenhausarten gibt, ist eine Untersuchung der Abhängigkeit der ermittelten log-Normalverteilungsparameter  $\alpha_L$  und  $\beta_L$  für die verschiedenen Krankenhausarten kaum sinnvoll. Insbesondere ist der in Abschnitt 2.3.3 vorgestellte Spearman'sche Rang-Korrelationskoeffizient erst ab einer Anzahl von 6 Wertepaaren einsetzbar. Aus diesen Gründen entfällt hier eine Untersuchung der ermittelten Parameterwerte  $\alpha_L$  und  $\beta_L$  auf eine gegenseitige Abhängigkeit.

In Abbildung 4.34 sind die Regressionsparameter sowohl für Maßnahmen mit verschiedenen Arbeitsbeschreibungen als auch verschiedener Leistungsarten angegeben. Zunächst kann man erkennen, dass die beiden Regressionsparameter leicht miteinander korreliert sind: mit steigendem Häufungspunkt nimmt auch die Schiefe zu. Normalerweise ist eine Korrelation zwischen Regressionsparameter nicht erwünscht, da dies ein Zeichen für eine zu komplexe Regressionsfunktion ist. In diesem Fall wäre die zweiparametrische Regressionsfunktion nur noch von einem echten Parameter abhängig, da sich der andere durch die Korrelation ermitteln ließe. Die Korrelation scheint aber zu schwach, als dass hier Maßnahmen erforderlich wären.

Da es nur zwei verschiedene Leistungsarten gibt, wird eine Untersuchung der Abhängigkeit der ermittelten log-Normalverteilungsparameter  $\alpha_L$  und  $\beta_L$  für die verschiedenen

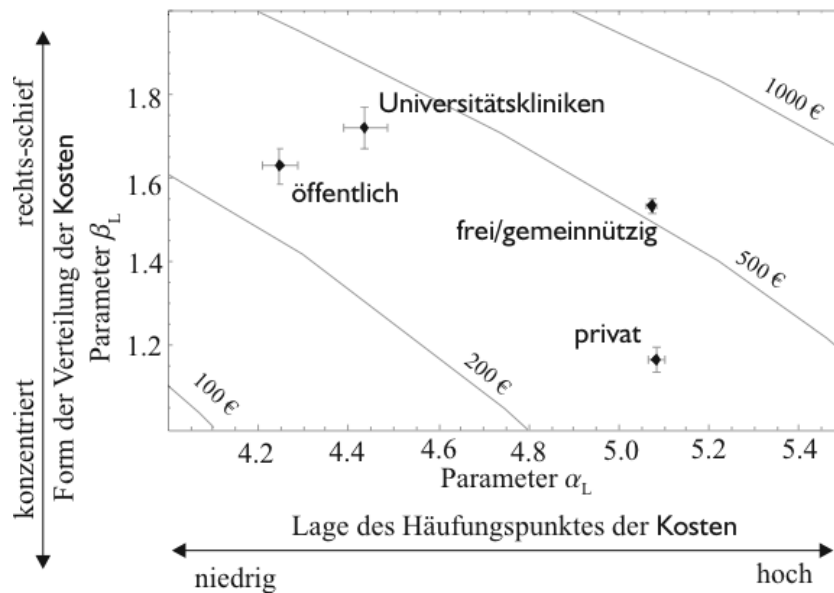


Abbildung 4.33: Aufgetragen sind hier die beiden Parameter der log-Normalverteilung für die Verteilungen der Kosten für verschiedene Krankenhausarten. Parameter  $\alpha$  gibt die Lage des Häufungspunktes an, Parameter  $\beta$  die Streuung der logarithmierten Kosten und damit die Schiefe der Verteilung an. Durch gestrichelte Linien sind beispielhaft Kostenerwartungswerte einiger Verteilungen angegeben. Deutung siehe Text.

Leistungsarten nicht durchgeführt.

Zur Untersuchung einer Abhängigkeit der ermittelten Parameter  $\alpha_L$  und  $\beta_L$  für verschiedene Arbeitsbeschreibungen kann der in Abschnitt 2.3.3 vorgestellte Spearmansche Rang-Korrelationskoeffizient  $r_S$  verwendet werden. Es ergibt sich  $r_S = 0.495$ . Der kritische Wert lautet  $k_{22;0.05} = 0.425$ . Die Nullhypothese kann also auf dem Signifikanzniveau  $\alpha = 0.05$  abgelehnt werden – d.h. es besteht ein signifikanter Zusammenhang zwischen den ermittelten Regressionsparametern. Dies ist ein Hinweis darauf, dass ein Parameter reicht, die Verteilung der Kosten für verschiedene Arbeitsbeschreibungen zu beschreiben – die Form der Verteilung ändert sich also abhängig von der Lage der Verteilung.

In dem durch die Leistungsarten markierten Bereich befinden sich die meisten Parameter für die verschiedenen Arbeitsbeschreibungen. Die Nähe der nicht zugeordneten Leistungsarten zu Fremdleistungen lässt vermuten, dass hier ein großer Teil von Fremdleistungen versteckt ist. Der mit versteckte Teil von Eigenleistungen macht sich durch die stärkere Streuung / Schiefe bemerkbar. Die Arbeitsbeschreibungen mit dem höchsten Kostenerwartungswert sind leider die, die nicht zugeordnet werden konnten. Dies ist mit darauf zurückzuführen, dass bei der Normierung der Arbeitsbeschreibungen nur auf die Anzahl der Nennungen geachtet wurde und nicht auf die mit ihnen verbundenen Kosten – ein weiteres Indiz dafür, wie wichtig eine gründliche Dokumentation ist.

In den Abbildungen 4.35 und 4.36 sind die ermittelten Parameter für Maßnahmen an Geräten gruppiert nach MT-Geräteart und -Untergeräteart dargestellt. In Abbildung 4.35 ergibt sich für den Spearmanschen Rang-Korrelationskoeffizient  $r_S = 0.236$ . Der kritische Wert lautet  $k_{21;0.05} = 0.435$ . Die Nullhypothese kann also nicht auf dem Si-

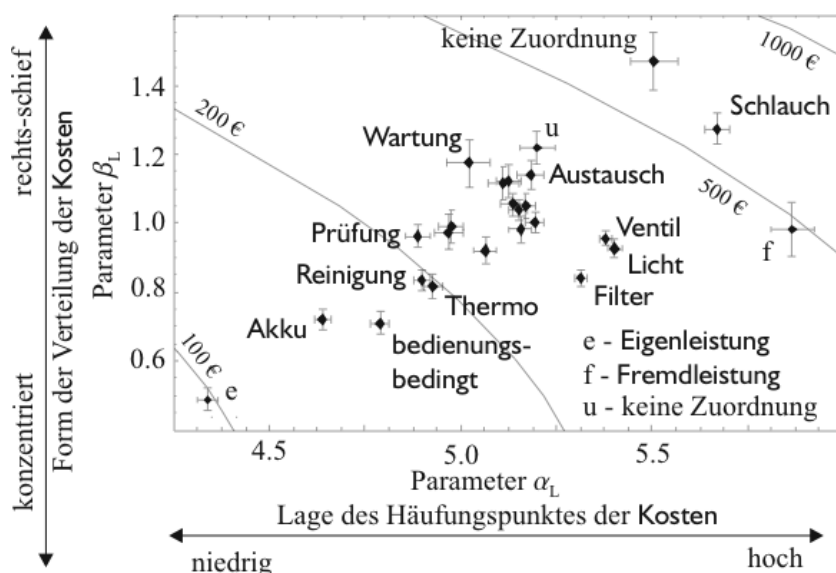


Abbildung 4.34: Aufgetragen sind hier die beiden Parameter der log-Normalverteilung für die Verteilungen der Kosten von Maßnahmen mit verschiedenen Arbeitsbeschreibungen und Leistungsarten. Parameter  $\alpha$  gibt die Lage des Häufungspunktes an, Parameter  $\beta$  die Streuung der logarithmierten Kosten und damit die Schiefe der Verteilung an. Durch gestrichelte Linien sind beispielhaft Kostenwertungswerte einiger Verteilungen angegeben. Deutung siehe Text.

gnifikanzniveau  $\alpha = 0.05$  abgelehnt werden – d.h. es besteht kein signifikanter Zusammenhang zwischen den ermittelten Regressionsparameter. Dies kann man als Hinweis werten, dass die verwendete log-Normalverteilung nicht zu viele Parameter benötigt, um die Daten zu beschreiben.

In Abbildung 4.36 ergibt sich für den Spearmanschen Rang-Korrelationskoeffizient  $r_S = 0.206$ . Der kritische Wert lautet  $k_{76;0.05} = 0.226$ . Die Nullhypothese kann also nicht auf dem Signifikanzniveau  $\alpha = 0.05$  abgelehnt werden – d.h. es besteht kein signifikanter Zusammenhang zwischen den ermittelten Regressionsparameter. Insgesamt fällt die geringere Korrelation der Parameter als bei Abbildung 4.34 auf.

In Abbildungen 4.37 sind die Regressionsparameter von Maßnahmen gruppiert nach Kaufpreis des betroffenen Gerätes aufgetragen. Hier ist nicht nur deutlich eine Korrelation der Regressionsparameter festzustellen, sondern auch ein Zusammenhang mit dem Kaufpreis: je größer der Kaufpreis desto höher der Häufungspunkt und damit auch der Streuung / Schiefe der Kosten.

Zur Untersuchung einer Abhängigkeit der ermittelten Parameter  $\alpha_L$  und  $\beta_L$  wird nach Abschnitt 2.3.3 der Spearmansche Rang-Korrelationskoeffizient  $r_S$  ermittelt. Es ergibt sich  $r_S = 0.879$ . Der kritische Wert lautet  $k_{13;0.05} = 0.560$ . Die Nullhypothese kann also auf dem Signifikanzniveau  $\alpha = 0.05$  abgelehnt werden – d.h. es besteht ein signifikanter Zusammenhang zwischen den ermittelten Regressionsparametern.

### Histogramm der Kosten und relative Kostensummen nach Leistungsart

In Abbildung 4.38 ist ein Histogramm der Kosten dargestellt. Der überwiegende Teil der Maßnahmen verursachen Kosten bis 5000 €.

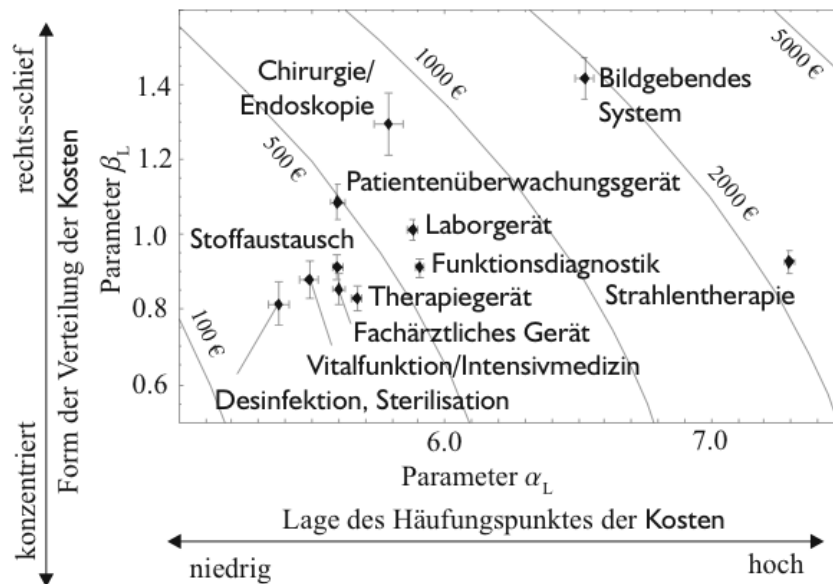


Abbildung 4.35: Aufgetragen sind hier die beiden Parameter der log-Normalverteilung für die Verteilungen der Kosten von Maßnahmen an verschiedenen MT-Gerätearten. Parameter  $\alpha$  gibt die Lage des Häufungspunktes an, Parameter  $\beta$  die Streuung der logarithmierten Kosten und damit die Schiefe der Verteilung an. Durch gestrichelte Linien sind beispielhaft Kostenerwartungswerte einiger Verteilungen angegeben. Deutung siehe Text.

Darunter sind die relativen Kostensummen für die verschiedenen Leistungsarten angegeben. Für Kosten bis 500 € ist der Anteil von **Eigenleistungen** relativ hoch, für Kosten zwischen 500 € und 50 000 € überwiegen **Fremdleistungen**. Der Anteil ohne Zuordnung der Leistungsart ist für Maßnahmen mit Kosten über 50 000 € auffällig erhöht. Diese könnten durch Verträge verursacht sein. Dafür spricht auch, dass in Abbildung 4.34 die Kosten der einzelnen Maßnahmen, die keiner Leistungsart zugeordnet werden konnten, recht ähnlich zu **Fremdleistungen** verteilt sind. Die nicht zugeordneten Maßnahmen bestehen also wahrscheinlich zum überwiegenden Teil aus **Fremdleistungen**, die wiederum durch Verträge zusammengefasst sind.

Die Eigenservicequote aus Abschnitt 2.4.2 scheint also stark von den Kosten der betrachteten Maßnahmen abzuhängen.

#### 4.4.4 Verteilung der Durchführungszeitpunkte

In diesem Abschnitt werden Verteilungen von **Maßnahmenzahlen** in Abhängigkeit von der Zeit analysiert. Um die Durchführungszeitpunkte von Maßnahmen an verschiedenen Geräten vergleichen zu können, wird hier meist nicht der absolute Durchführungszeitpunkt angegeben sondern die Zeit, die das Gerät zum Durchführungszeitpunkt schon in Betrieb ist. Dieser so normierte Durchführungszeitpunkt wird **Betriebszeit** genannt (siehe Tabelle 4.1).

Um die Analysen zu ordnen, sind diese – wie die anderen Analysen auch – in zwei Gruppen eingeteilt: Verteilung von Häufigkeiten (**Gerätezahl** und **Maßnahmenzahl**) und Verteilung von Ausgaben (**Kaufpreissummen** und **Kostensummen**). Im Anschluss daran

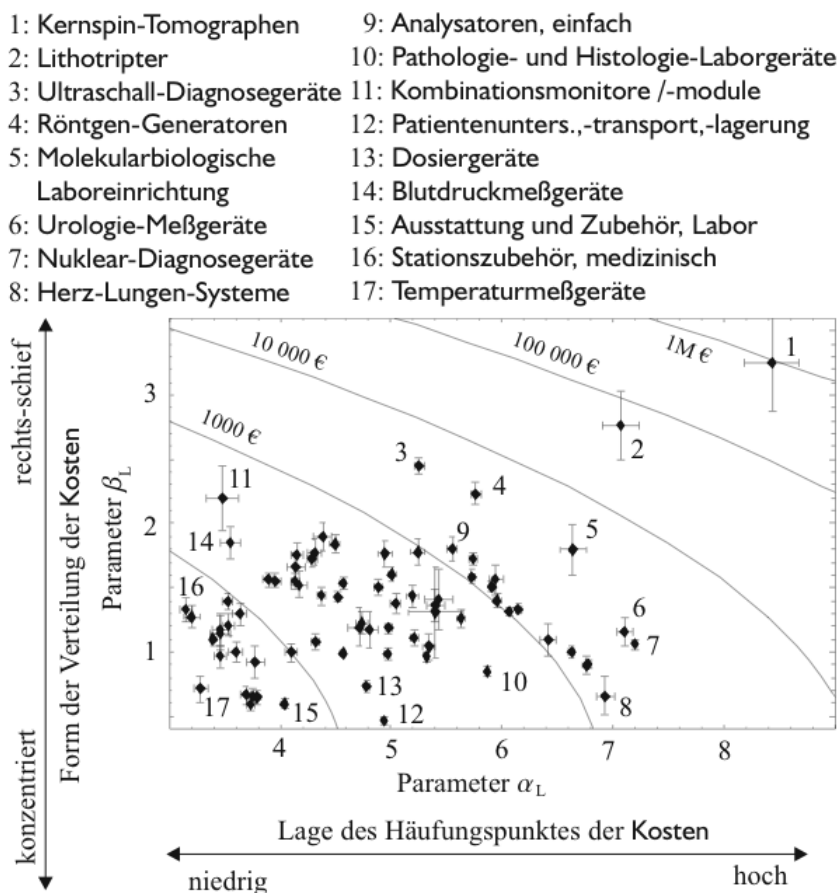


Abbildung 4.36: Aufgetragen sind hier die beiden Parameter der log-Normalverteilung für die Verteilungen der Kosten von Maßnahmen an verschiedenen MT-Untergerätearten. Parameter  $\alpha$  gibt die Lage des Häufungspunktes an, Parameter  $\beta$  die Streuung der logarithmierten Kosten und damit die Schiefe der Verteilung an. Durch gestrichelte Linien sind beispielhaft Kostenerwartungswerte einiger Verteilungen angegeben. Deutung siehe Text.

findet sich noch ein Teil, in dem die Verteilung der Zeiten zwischen aufeinanderfolgenden Maßnahmen (MTBF) durch Weibull-Verteilungen beschrieben werden.

### Verteilung der Maßnahmenzahlen

Abbildung 4.39 hat Ähnlichkeit mit Abbildung 4.17 – es können also die selben Überlegungen aus Abschnitt 4.3.4 herangezogen werden (zeitliche Datenfehler, siehe Abschnitt 4.1.2). Insgesamt ergibt sich auch hier, dass bei vielen Datenbanken der Krankenhäuser die Erfassung der Maßnahmen zeitlich nicht konstant erfolgt ist.

In Abbildung 4.40 ist ein Histogramm der Maßnahmenzahlen über der Betriebszeit dargestellt. Der Häufungspunkt der Maßnahmen liegt im Bereich von 3 bis 5 Jahren Betriebszeit. Anhand der eingezeichneten kumulierten Anteile kann man zum Beispiel ablesen, dass mindestens 20 % der Maßnahmen an Geräten vorgenommen werden, die älter als 10 Jahre sind. In Abbildung 4.44 sind statt der Maßnahmenzahl die Kosten-summe dargestellt.

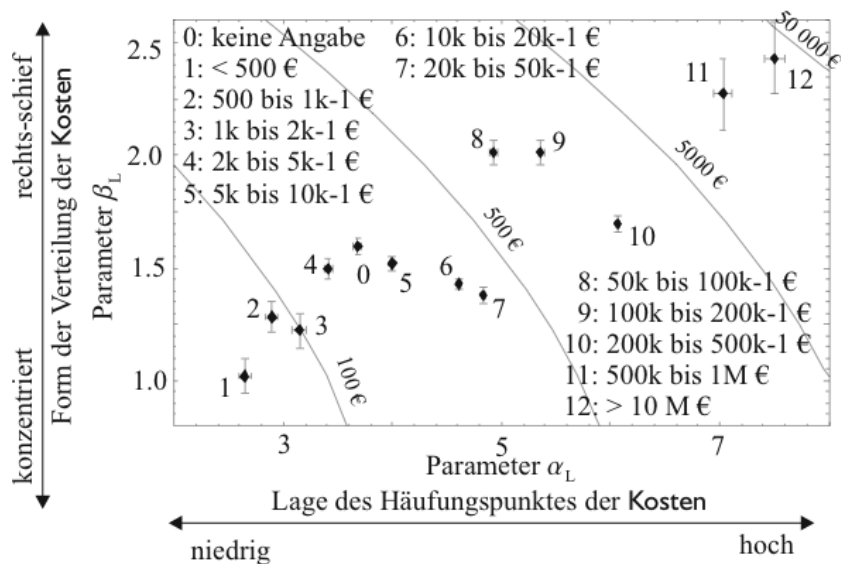


Abbildung 4.37: Aufgetragen sind hier die beiden Parameter der log-Normalverteilung für die Verteilungen der Kosten von Maßnahmen an Geräten mit verschiedenen Kaufpreisen. Parameter  $\alpha$  gibt die Lage des Häufungspunktes an, Parameter  $\beta$  die Streuung der logarithmierten Kosten und damit die Schiefe der Verteilung an. Durch gestrichelte Linien sind beispielhaft Kostenerwartungswerte einiger Verteilungen angegeben. Deutung siehe Text.

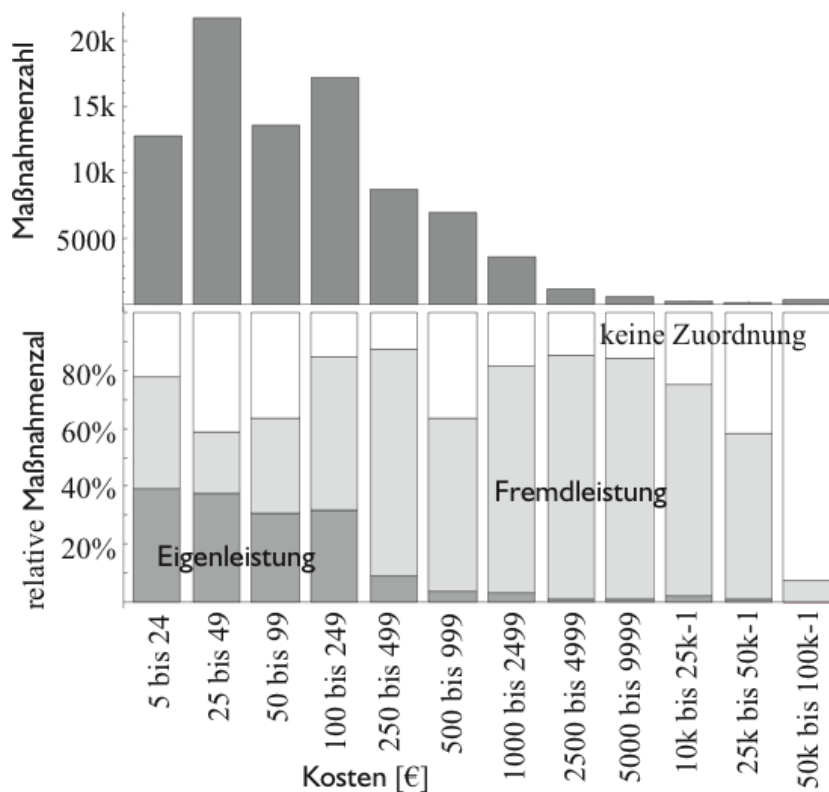


Abbildung 4.38: Kombination eines Histogramms der Kosten und eines Balkendiagramms, das die Verteilung der relativen Maßnahmenzahlen je Leistungsart und Kosten darstellt. Deutung siehe Text.

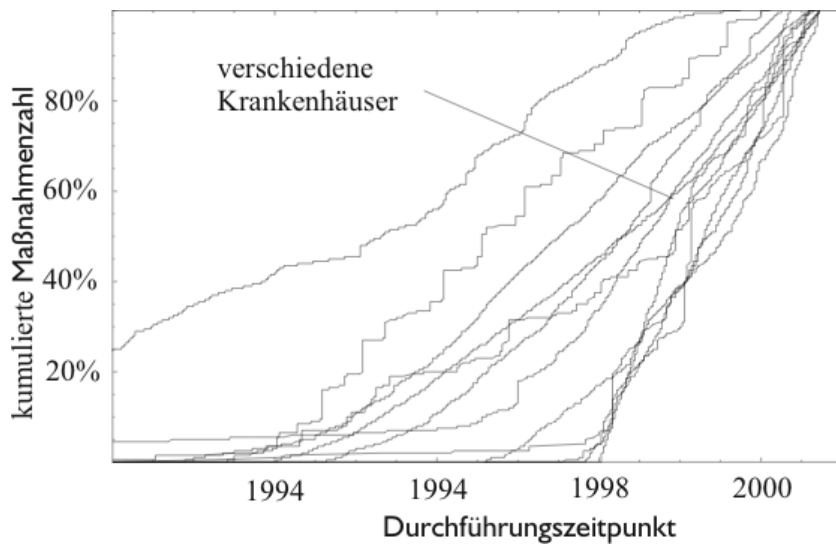


Abbildung 4.39: Kumulative Darstellung relativer Maßnahmenezahlen über Durchführungszeitpunkt für verschiedene Krankenhäuser. Es ist deutlich zu erkennen, wie heterogen die Erfassung des Gerätebestandes in den Krankenhäusern hinsichtlich Beginn und Konstanz der Erfassung erfolgt. Details siehe Text.

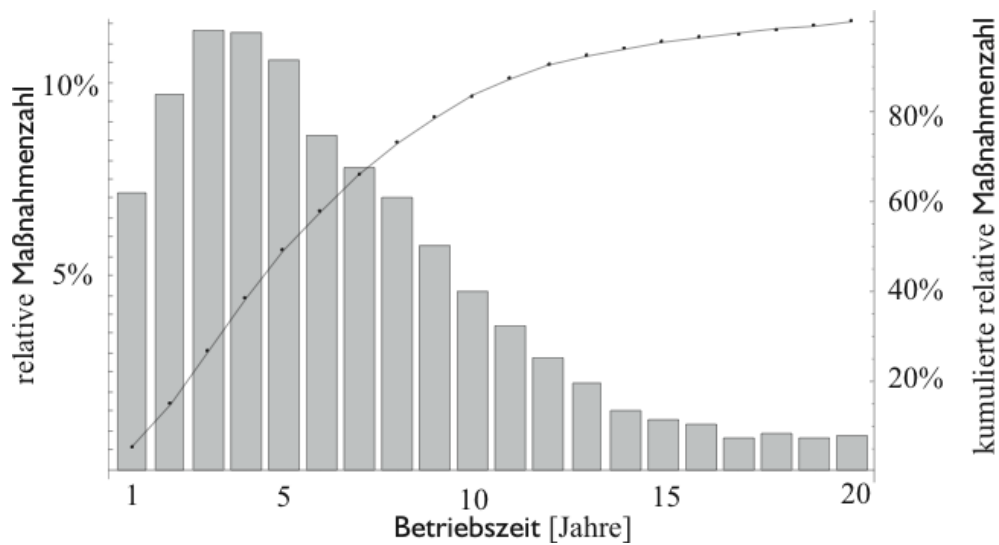


Abbildung 4.40: Histogramm und kumulative Darstellung der Maßnahmenezahlen in Abhängigkeit der Betriebszeit. Deutung siehe Text.

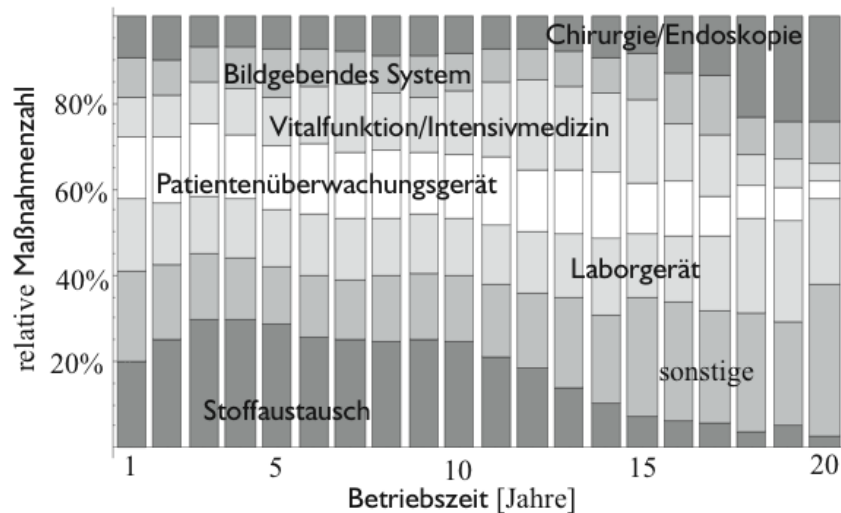


Abbildung 4.41: Balkendiagramm der relativen Maßnanzahlen für verschiedene MT-Gerätearten in Abhängigkeit der Betriebszeit. Deutung siehe Text.

Abbildung 4.41 zeigt die sich mit der Betriebszeit ändernden Anteile der Maßnanzahlen, die an verschiedenen MT-Geräteart durchgeführt werden. Der Anteil von Stoffaustauschgeräten ist im Zeitraum von 3 bis 10 Jahre deutlich erhöht. Ab dem 15. Betriebsjahr nehmen Anteile von Maßnahmen an Chirurgie- / Endoskopiegeräten sprunghaft zu. Dazwischen sind die Anteile von Vitalfunktions- / Intensivmedizingeräten leicht angehoben. Auch fällt auf, dass der Anteil der Stoffaustauschgeräte mit dem 10. Betriebsjahr abnimmt, bei Patientenüberwachungsgeräten nehmen die Anteile erst mit dem 15. Jahr ab. Die Anteile von Laborgeräten und Bildgebenden Systemen sind etwa konstant.

Für die gesamte Abbildung ergibt sich nach dem  $\chi^2$ -Test eine Testgröße  $T = 8396$ , der kritische Wert beträgt für  $(7 - 1)(20 - 1) = 114$  Freiheitsgrade  $k_{114;0.05} = 139.9$ . Damit wird die Nullhypothese mit überwältigender Deutlichkeit abgelehnt – die relative Maßnanzahlen je MT-Geräteart hängen also hoch signifikant von der Betriebszeit ab. Es ergibt sich ein Kontingenzkoeffizient von  $C_{korr} = 23.32\%$ .

Untersucht man nun die Anteile der einzelnen MT-Geräteart gegen den Rest, so ergeben sich für Stoffaustauschgeräte ein  $T = 3714$  und  $C_{korr} = 20.58\%$ , für Patientenüberwachungsgeräte ein  $T = 446.3$  und  $C_{korr} = 7.20\%$ , für Laborgeräte ein  $T = 380.3$  und  $C_{korr} = 6.65\%$ , für Vitalfunktions- / Intensivmedizingeräte ein  $T = 1720$  und  $C_{korr} = 14.08\%$ , für Bildgebende Systeme ein  $T = 283.0$  und  $C_{korr} = 5.74\%$  und für Chirurgie- / Endoskopiegeräte ein  $T = 1801$  und  $C_{korr} = 14.41\%$ . Der kritische Wert lautet für alle Testgrößen  $k_{19;0.05} = 30.14$ . Diese Zahlen unterstreichen die oben visuell gewonnen Erkenntnisse.

Abbildung 4.42 zeigt die sich mit der Betriebszeit ändernden Anteile der Maßnanzahlen für verschiedene Arbeitsbeschreibungen. Insgesamt scheinen die Anteile kaum von der Betriebszeit abzuhängen. Ab dem 15. Jahr nehmen jedoch Prüfungen zugunsten von Reparaturen ab. Die relativen Maßnanzahl von Reparaturen und Prüfungen bleibt in der Summe jedoch in etwa konstant.

Auch, wenn der Anteil der Wartungen nicht groß ist, kann man erkennen, dass ihr Anteil bei Geräten mit 5-jähriger Betriebszeit nur halb so groß ist wie bei Geräten mit

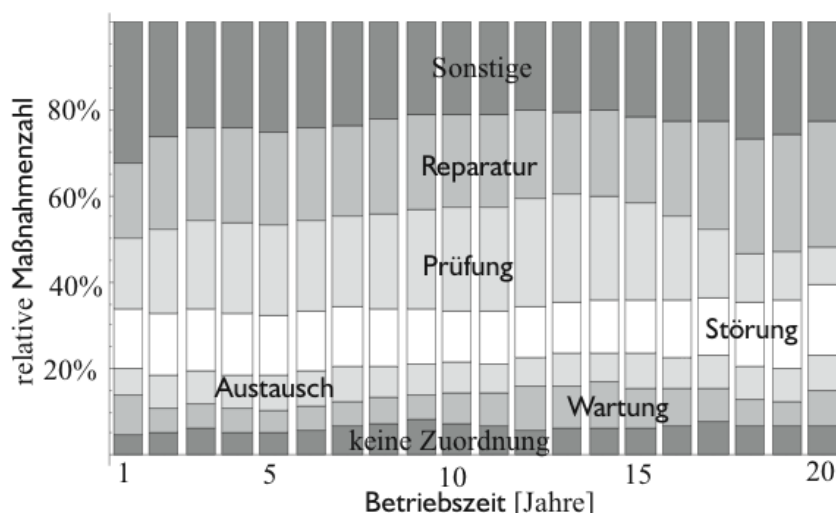


Abbildung 4.42: Balkendiagramm der relativen Maßnahmehzahlen für verschiedene Arbeitsbeschreibungen in Abhängigkeit der Betriebszeit. Deutung siehe Text.

15-jähriger Betriebszeit.

Die Wartungs- / Instandhaltungsquote (als Quotient der Maßnahmehzahlen statt als Quotient der Kostensummen wie in Abschnitt 2.4.2 vorgestellt) scheint insgesamt jedoch nur schwach von den Betriebszeiten der betroffenen Geräte abzuhängen. Es ergibt sich ein Wert von ca. 20 bis 30 %.

Für die gesamte Abbildung ergibt sich nach dem  $\chi^2$ -Test eine Testgröße  $T = 4313$ , der kritische Wert beträgt für  $(7 - 1)(20 - 1) = 114$  Freiheitsgrade  $k_{114;0.05} = 139.9$ . Damit wird die Nullhypothese mit überwältigender Deutlichkeit abgelehnt – die relativen Maßnahmehzahlen je Arbeitsbeschreibungen hängen also hoch signifikant von der Betriebsdauer ab. Es ergibt sich ein Kontingenzkoeffizient von  $C_{korr} = 12.51$  %.

Untersucht man nun die Anteile der einzelnen Arbeitsbeschreibungen gegen den Rest, so ergeben sich für Reparaturen ein  $T = 468.5$  und  $C_{korr} = 5.43$  %, für Prüfungen ein  $T = 1340$  und  $C_{korr} = 9.17$  %, für Störungen ein  $T = 227.8$  und  $C_{korr} = 3.79$  %, für Austauscharbeiten ein  $T = 105$  und  $C_{korr} = 2.57$  %, für Wartungen ein  $T = 1151$  und  $C_{korr} = 8.50$  % und für Arbeitsbeschreibungen, die nicht zugeordnet werden konnten, ein  $T = 503.7$  und  $C_{korr} = 5.63$  %. Der kritische Wert lautet für alle Testgrößen  $k_{19;0.05} = 30.14$ .

Für Reparaturen und Prüfungen gegen den Rest ergibt sich  $T = 964.5$ ,  $k_{19;0.05} = 30.14$  und  $C_{korr} = 7.78$  %. Für Prüfungen gegen Reparaturen ergibt sich  $T = 925.1$ ,  $k_{19;0.05} = 30.14$  und  $C_{korr} = 11.71$  %.

Abbildung 4.43 zeigt die relativen Maßnahmehzahlen für verschiedene Leistungsarten in Abhängigkeit der Betriebszeit. Als gestrichelte Linien sind die relativen Maßnahmehzahlen von Fremd- und Eigenleistungen angegeben unter Auslassung der Maßnahmen ohne Zuordnung.

Wieder lassen sich auf Grund des hohen Anteils nicht zugeordneter Maßnahmen keine eindeutigen Schlüsse ziehen. Die leichte Zunahme des Anteils an Fremdleistungen im Zeitraum von 11 bis 16 Jahren Betriebszeit wie auch die generelle Zunahme des Anteils von Fremdleistungen mit der Betriebszeit deutet auf noch nicht ausgelaufene Verträge hin.

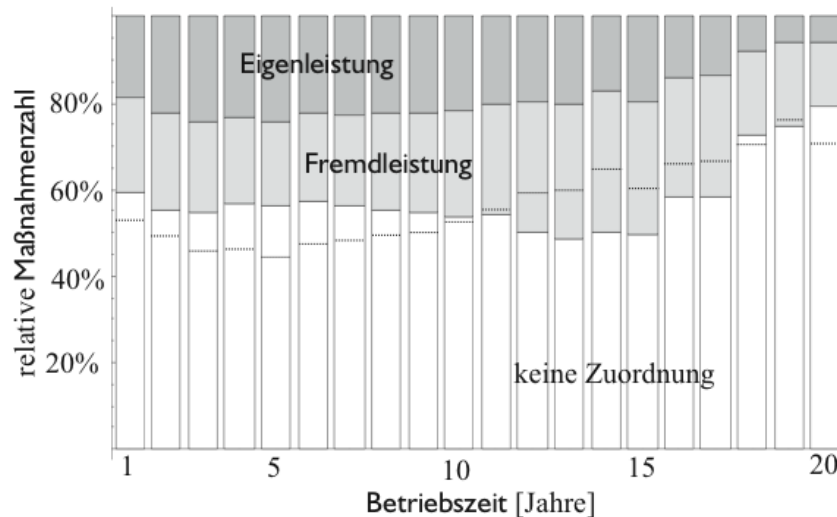


Abbildung 4.43: Balkendiagramm der relativen Maßnanzahlen für verschiedene Leistungsarten in Abhängigkeit der Betriebszeit. Deutung siehe Text.

Die Eigenservicequote (als Quotient der Maßnanzahlen statt als Quotient der Kostensummen wie in Abschnitt 2.4.2 vorgestellt) scheint also mit zunehmender Betriebszeiten der betroffenen Geräte abzunehmen. Diese Eigenservicequote beträgt in den ersten 10 Jahren ca. 50 % und wächst dann bis zum 20. Lebensjahr auf ca. 70 % an.

Aufgrund des vorliegenden Datenmaterials erscheint eine statistische Analyse wenig sinnvoll. Diese wäre jedoch möglich, wenn vollständig und regelmäßig dokumentiert werden wäre.

### Verteilung der Kostensummen

Abbildung 4.44 enthält ein Histogramm der Kostensummen über der Betriebszeit. Der Häufungspunkt der Kostensummen liegt im Bereich von 4 bis 6 Jahren Betriebszeit. Er liegt also etwas später als der Häufungspunkt der Maßnanzahl aus Abbildung 4.40. Anhand der eingezeichneten kumulierten Anteile kann man zum Beispiel ablesen, dass 20 % aller Kosten an Geräten anfallen werden, die älter als 10 Jahre sind. 50 % der Kosten fallen an Geräten zwischen dem 2. und 6. Lebensjahr an.

In Vergleich zu Abbildung 4.41 sieht man an Abbildung 4.45 eine deutliche Zunahme der Variation. Die zusätzlichen Schwankungen der Kosten lassen keine eindeutigen Tendenzen erkennen, verschieben jedoch gegenüber Abbildung 4.41 einige Schwerpunkte: Geräte, an denen viele Maßnahmen durchgeführt wurden, müssen nicht automatisch einen großen Anteil der Ausgaben verursacht haben.

Auffällig ist die Zunahme der relativen Kostensummen für Bildgebende Systeme zu den Zeitpunkten 6 und 17 Jahren. Eine Erklärung dafür kann nicht gegeben werden.

Für die gesamte Abbildung ergibt sich nach dem  $\chi^2$ -Test eine Testgröße  $T = 3.2610^6$ , der kritische Wert beträgt für  $(7 - 1)(20 - 1) = 114$  Freiheitsgrade  $k_{114;0.05} = 139.9$ . Damit wird die Nullhypothese mit überwältigender Deutlichkeit abgelehnt – die relativen Kostensummen je MT-Geräteart hängen also hoch signifikant von der Betriebsdauer ab. Es ergibt sich ein Kontingenzkoeffizient von  $C_{korr} = 24.69$  %.

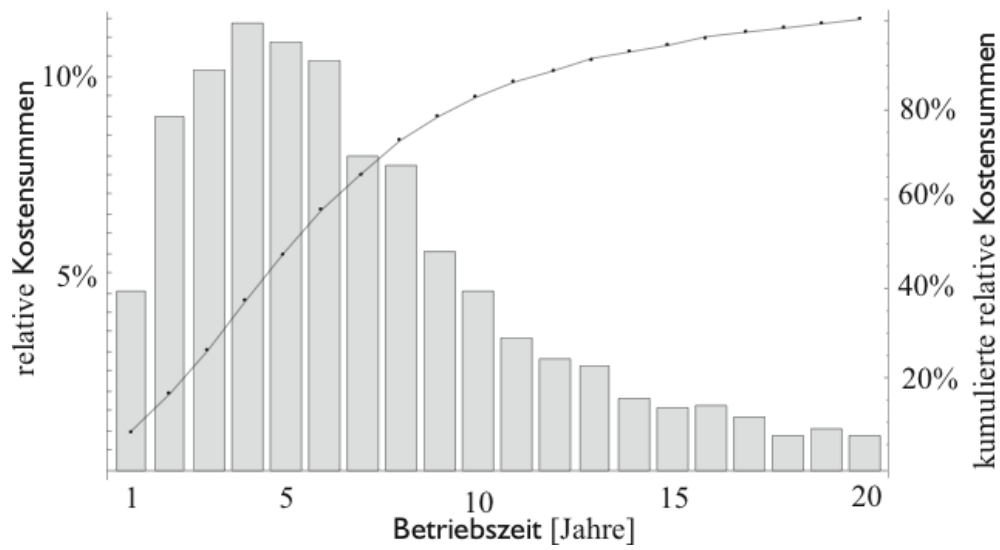


Abbildung 4.44: Histogramm und kumulative Darstellung der Kostensummen in Abhängigkeit der Betriebszeit. Deutung siehe Text.

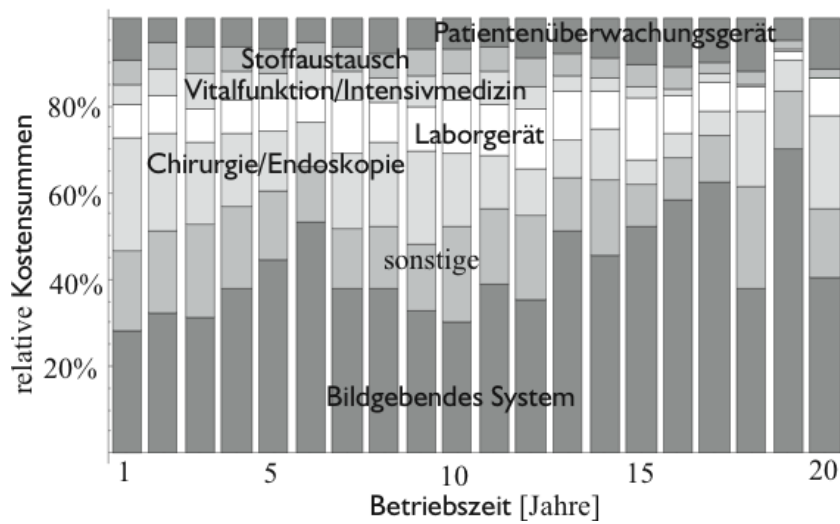


Abbildung 4.45: Balkendiagramm der relativen Kostensummen für verschiedene MT-Geräteart in Abhängigkeit der Betriebszeit. Deutung siehe Text.

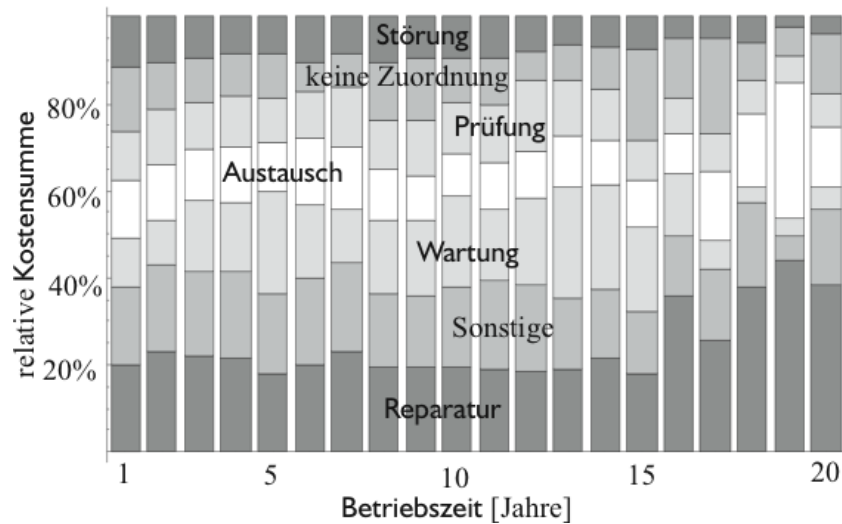


Abbildung 4.46: Balkendiagramm der relativen Kostensummen für verschiedene Arbeitsbeschreibungen in Abhängigkeit der Betriebszeit. Deutung siehe Text.

Abbildung 4.46 zeigt die sich mit der Betriebszeit ändernden Anteile der Kostensummen mit verschiedenen Arbeitsbeschreibungen. Die gegenüber Abbildung 4.42 zusätzlichen Schwankungen der relativen Kostensummen lassen keine eindeutigen Tendenzen erkennen. Ab dem 15. Betriebsjahr kann man jedoch deutliche eine Abnahme des Wartungsanteils und eine Zunahme des Reparaturanteils erkennen. Eine ähnliche Konstellation scheint es für Geräte zu geben, die 7 Jahre in Betrieb sind (siehe dazu auch Abbildung 5.1).

Die Wartungs- / Instandhaltungsquote aus Abschnitt 2.4.2 scheint also für Geräte, die 7 oder mehr als 15 Jahre in Betrieb sind, etwas niedriger zu sein.

Für die gesamte Abbildung ergibt sich nach dem  $\chi^2$ -Test eine Testgröße  $T = 4313$ , der kritische Wert beträgt für  $(7 - 1)(20 - 1) = 114$  Freiheitsgrade  $k_{114;0.05} = 139.9$ . Damit wird die Nullhypothese mit überwältigender Deutlichkeit abgelehnt – die relativen Kostensummen je Arbeitsbeschreibung hängen also hoch signifikant von der Betriebsdauer ab. Es ergibt sich ein Kontingenzkoeffizient von  $C_{korr} = 12.51\%$ .

Abbildung 4.47 zeigt die Anteile der Kostensummen je Leistungsart in Abhängigkeit der Betriebszeit. Bei Betrachtung der Kostensummen fällt im Gegensatz zur Betrachtung der Maßnahmenezahl in Abbildung 4.43 der kleinere Anteil nicht zugeordneter Maßnahmen auf. Für die Eigenservicequote aus Abschnitt 2.4.2 ergibt sich unter Auslassung der nicht zugeordneten Maßnahmen ein nicht von der Betriebszeit abhängiger Anteil von ca. 10 %.

### Regression von Weibull-Verteilungen der Durchführungszeitpunkte

Die in Abschnitt 2.4.3 vorgestellte Weibull-Verteilung soll nun durch Regression an die Durchführungszeitpunkte verschiedener Maßnahmen angepasst werden. Aus den Regressionsparameter kann man feststellen, ob Früh-, Zufalls- oder Spätausfälle vorherrschen. Außerdem kann aus den Parametern die Mean Time Between Failures MTBF bestimmt werden.

Es wird ausdrücklich darauf hingewiesen, dass hier die Zeiten untersucht werden,

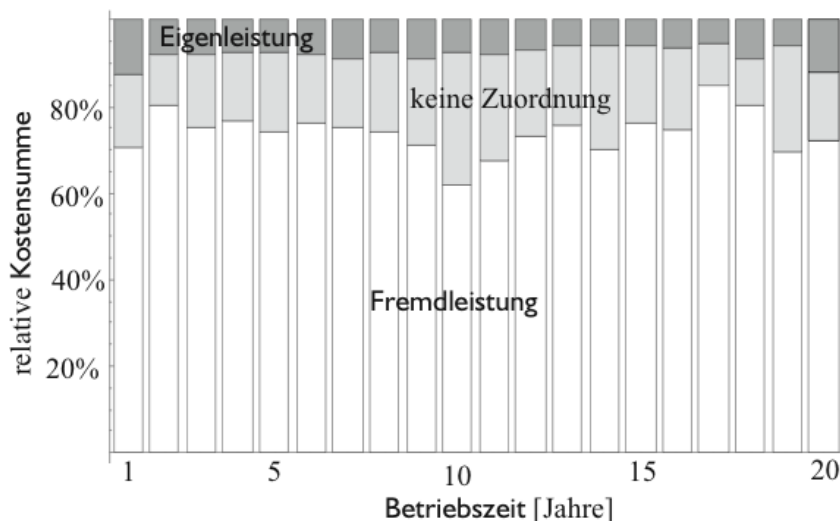


Abbildung 4.47: Balkendiagramm der relativen Kostensummen für verschiedene Leistungsart in Abhängigkeit der Betriebszeit. Deutung siehe Text.

die seit der letzten Maßnahme am betroffenen Gerät verstrichen sind. Das Alter des Gerätes spielt dabei keine Rolle (siehe dazu auch Abschnitt 2.4.3). Untersuchungen der Ausfallraten in Abhängigkeit der Betriebszeit finden sich am Ende dieses Abschnittes und in Anhang F.

In Abbildung 4.48 und 4.49 sind die ermittelten Parameter für Maßnahmen an verschiedenen MT-Gerätearten und -Untergerätearten dargestellt. Als gestrichelte Linien sind beispielhafte MTBF eingezeichnet.

Es fällt auf, dass insgesamt Frühausfälle vorherrschen. Dies kann teilweise auf den beschränkten Zeithorizont der vorliegenden Daten zurückzuführen sein.

Zur Untersuchung einer Abhängigkeit der ermittelten Parameter  $\alpha$  und  $\beta$  wird nach Abschnitt 2.3.3 der Spearmansche Rang-Korrelationskoeffizient  $r_S$  ermittelt. Es ergibt sich  $r_S = 0.839$ . Der kritische Wert lautet  $k_{12;0.05} = 0.587$ . Die Nullhypothese kann also auf dem Signifikanzniveau  $\alpha = 0.05$  abgelehnt werden – d.h. es besteht ein signifikanter Zusammenhang zwischen den ermittelten Regressionsparametern.

In Abbildung 4.48 fallen Geräte der Strahlentherapie, Desinfektion / Sterilisation und Bildgebende Systeme durch kleine MTBF und Frühausfälle auf. Für die restlichen Geräte sind MTBF und der Ausfallverlauf kaum korreliert.

In Abbildung 4.49 fallen Narkosegeräte und -Monitore durch Spätausfälle auf. Ansonsten herrschen auch hier Frühausfälle vor.

Für den Spearmansche Rang-Korrelationskoeffizient ergibt sich  $r_S = 0.636$ . Der kritische Wert lautet  $k_{78;0.05} = 0.223$ . Die Nullhypothese kann also auf dem Signifikanzniveau  $\alpha = 0.05$  abgelehnt werden – d.h. es besteht ein signifikanter Zusammenhang zwischen den ermittelten Regressionsparametern.

Betrachtet man nun in Abbildung 4.50 die Regressionsparameter für Geräte nach Kaufpreis gruppiert, so erkennt man, dass mit steigendem Kaufpreis die MTBF bei in etwa gleichbleibendem Ausfallverlauf abnimmt. D.h. teure Geräte werden häufiger geserviced.

Zur Untersuchung einer Abhängigkeit der ermittelten Parameter  $\alpha$  und  $\beta$  wird nach

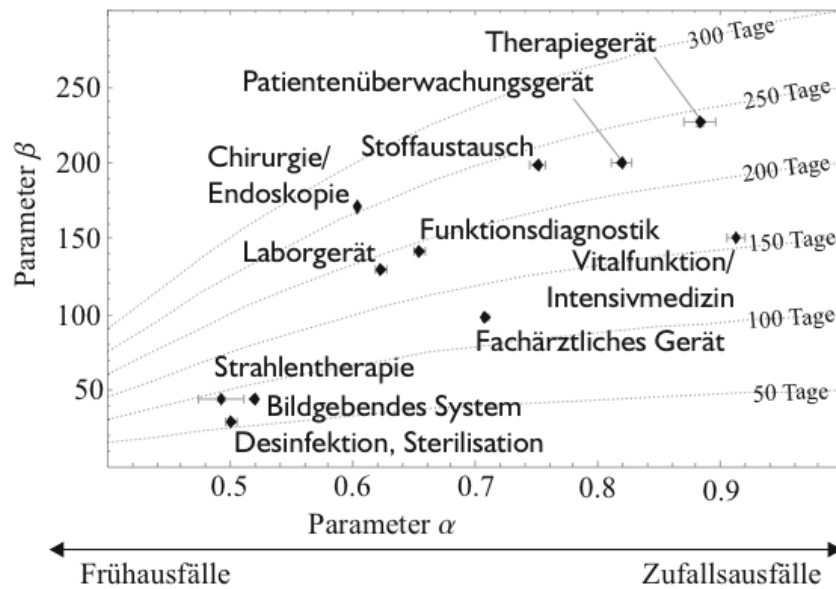


Abbildung 4.48: Darstellung der ermittelten Parameter  $\alpha$  und  $\beta$  der Weibull-Verteilung für Maßnahmen an Geräten verschiedener MT-Gerätearten. Der Parameter  $\alpha$  gibt an, ob Frühausfälle ( $\alpha < 0$ ), Zufallsausfälle ( $\alpha = 0$ ) oder Spätausfälle ( $\alpha > 0$ ) vorliegen. Durch gestrichelte Linien sind ausgewählte MTBF angegeben. Deutung siehe Text.

Abschnitt 2.3.3 der Spearmansche Rang-Korrelationskoeffizient  $r_S$  ermittelt. Es ergibt sich  $r_S = -0.119$ . Der kritische Wert lautet  $k_{12;0.05} = 0.587$ . Die Nullhypothese kann also nicht auf dem Signifikanzniveau  $\alpha = 0.05$  abgelehnt werden – d.h. es besteht kein signifikanter Zusammenhang zwischen den ermittelten Regressionsparameter.

In Abbildung 4.51 ist zu erkennen, dass in frei / gemeinnützigen Kliniken und Universitätskliniken Frühausfälle vorherrschen, in nicht-Universitätskliniken Zufallsausfälle und in privaten Kliniken Spätausfälle vorherrschen. Außerdem ist in privaten Kliniken die MTBF etwa doppelt so hoch wie bei den restlichen Krankenhäusern. Da es nur vier verschiedene Krankenhausarten gibt, entfällt hier eine statistische Analyse der Korrelation.

### Bestimmung der zeitabhängigen Ausfallrate

Statt der Ausfallrate aus Abschnitt 2.4.2, die die Wahrscheinlichkeit angibt, dass ein Gerät innerhalb einer folgenden Zeitspanne ausfällt, wird hier die Anzahl der Ausfälle in der folgenden Zeitspanne bestimmt – denn wir wissen ja, ein Gerät kann mehrmals ausfallen, da es wieder instandgesetzt werden kann.

In Abbildung 4.40 sind die Maßnahmenzahlen über der Betriebszeit des betroffenen Gerätes angegeben. Durch die zeitliche Gruppierung der Maßnahmenzahlen nach Jahren können die Maßnahmenzahlen gleich als Maßnahmenzahlen pro Jahr interpretiert werden.

Um nun die Maßnahmenzahl pro Gerät zu bestimmen, reicht es nicht, die Anzahl der in der Datenbank vorhandenen Geräte zu bestimmen und die Zahl der Maßnahmen durch diese Zahl zu teilen. Nicht alle Geräte in der Datenbank waren in den jeweiligen Zeiträumen in Betrieb bzw. wurden EDV-technisch erfasst, für die in Abbildung 4.40 die Anzahlen der durchgeführten Maßnahmen angegeben sind. Deswegen werden

- |   |                                     |
|---|-------------------------------------|
| 1: Kernspin-Tomographen                               | 12: Photometer                      |
| 2: Röntgenfilmentwicklungsmaschinen                   | 13: Ophthalmologische Geräte        |
| 3: Sterilisationsgeräte                               | 14: Lungenfunktions-Meßgeräte       |
| 4: Dosiergeräte                                       | 15: Inkubatoren und Wärmegeräte     |
| 5: Lithotripter                                       | 16: Meßgeräte, Einzelmessung, Labor |
| 6: Dialyse- und Blutfiltrationsgeräte                 | 17: Beatmungs- und Narkose-Monitore |
| 7: Röntgen-Generatoren                                | 18: Kardiologie-Notfallgeräte       |
| 8: Telemetrie-Anlagen                                 | 19: Dentaltechnische Geräte         |
| 9: Decken- & Wandversorg.-<br>und Befestigungssysteme | 20: Ergometer                       |
| 10: Perinatale Überwachung                            | 21: Personen-Waagen                 |
| 11: Inhalations-Narkosegeräte                         | 22: Sauerstoff-Therapiegeräte       |
|   | 23: Stationszubehör, medizinisch    |

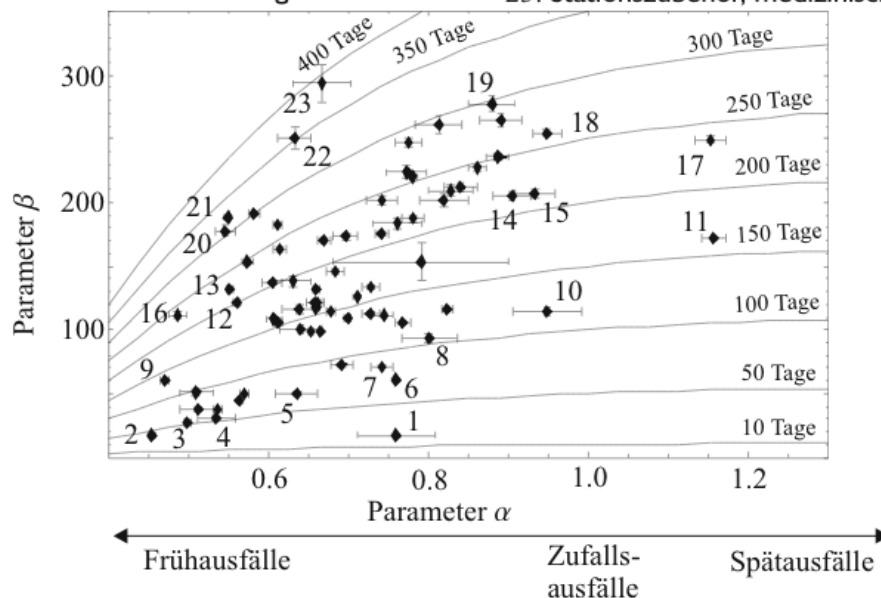


Abbildung 4.49: Darstellung der ermittelten Parameter  $\alpha$  und  $\beta$  der Weibull-Verteilung für Maßnahmen an Geräten verschiedener MT-Untergeräteart. Der Parameter  $\alpha$  gibt an, ob Frühausfälle ( $\alpha < 0$ ), Zufallsausfälle ( $\alpha = 0$ ) oder Spätausfälle ( $\alpha > 0$ ) vorliegen. Durch gestrichelte Linien sind ausgewählte MTBF angegeben. Deutung siehe Text.

nur Geräte gezählt, die vor und nach dem genannten Zeitraum mindestens eine Maßnahme vorweisen. Nur in diesem Fall kann davon ausgegangen werden, dass auch in der Zwischenzeit dieses Gerät eine Maßnahme zu den Werten in Abbildung 4.40 hätte beisteuern können.

Diese Geräte, die also vor und nach dem Betrachtungszeitraum mindestens eine erfasste Maßnahme zu verzeichnen haben, sollen als aktive Geräte betrachtet werden. Die Anzahl der aktiven Geräte ist in Abbildung 4.52 angegeben und wird im folgenden als **Aktivgerätezahl** genannt.

Nun kann man die **Maßnahmenzahl pro Jahr** aus Abbildung 4.40 durch die **Aktivgerätezahl** aus Abbildung 4.52 dividieren und erhält die Anzahl der Maßnahmen pro aktivem Gerät und Jahr – also die Anzahl der Maßnahmen pro Gerät und Jahr – in Abhängigkeit von der **Betriebszeit** des betroffenen Gerätes. Dies ist für alle Geräte in Abbildung 4.53 dargestellt. Die Werte für die einzelnen MT-Gerätearten sind im Anhang

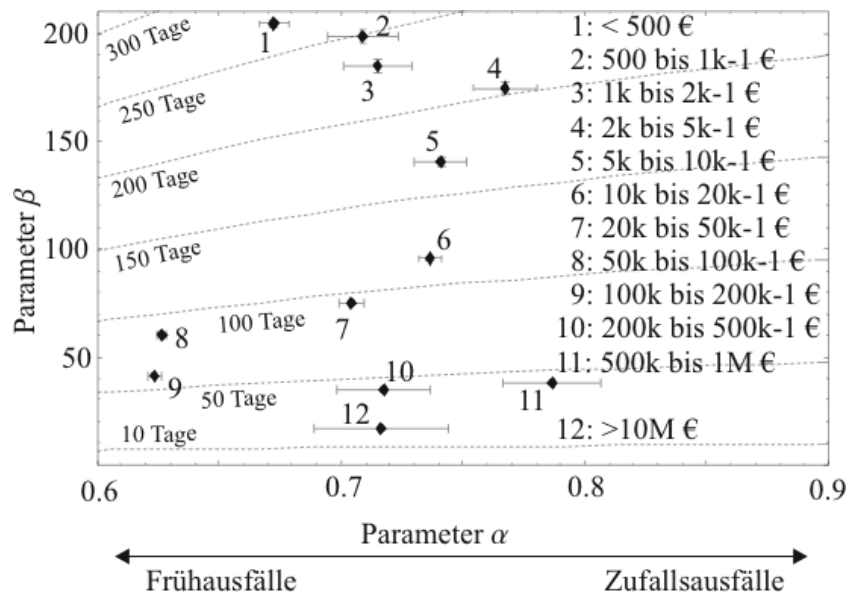


Abbildung 4.50: Darstellung der ermittelten Parameter  $\alpha$  und  $\beta$  der Weibull-Verteilung für Maßnahmen an Geräten verschiedener Kaufpreise. Der Parameter  $\alpha$  gibt an, ob Frühausfälle ( $\alpha < 0$ ), Zufallsausfälle ( $\alpha = 0$ ) oder Spätausfälle ( $\alpha > 0$ ) vorliegen. Durch gestrichelte Linien sind ausgewählte MTBF angegeben. Deutung siehe Text.

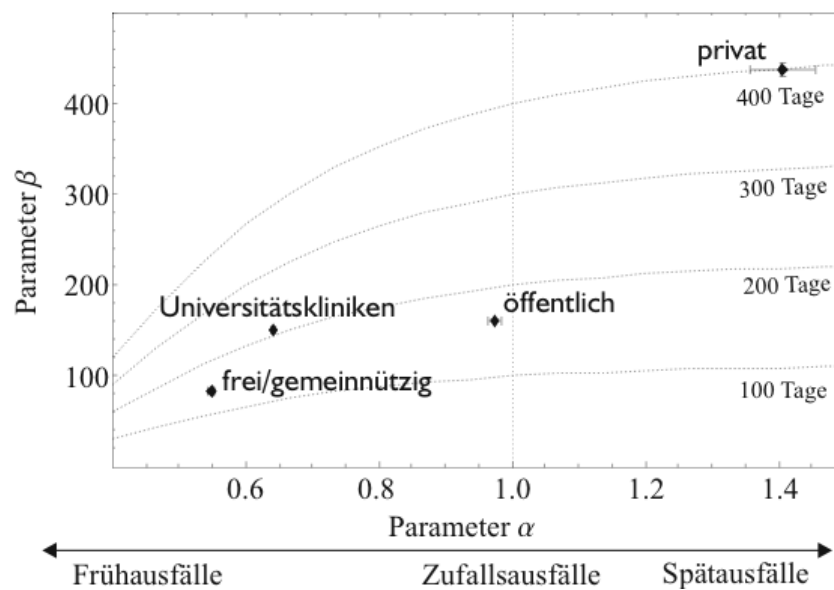


Abbildung 4.51: Darstellung der ermittelten Parameter  $\alpha$  und  $\beta$  der Weibull-Verteilung für Maßnahmen an Geräten verschiedener Krankenhausarten. Der Parameter  $\alpha$  gibt an, ob Frühausfälle ( $\alpha < 0$ ), Zufallsausfälle ( $\alpha = 0$ ) oder Spätausfälle ( $\alpha > 0$ ) vorliegen. Durch gestrichelte Linien sind ausgewählte MTBF angegeben. Deutung siehe Text.

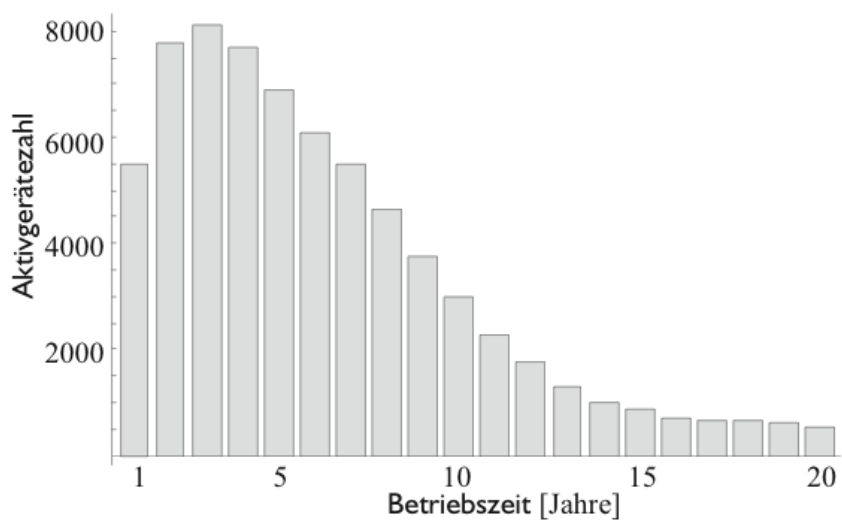


Abbildung 4.52: Darstellung der Aktivgerätezahl über der Betriebszeit für alle Geräte. Erklärung siehe Text.

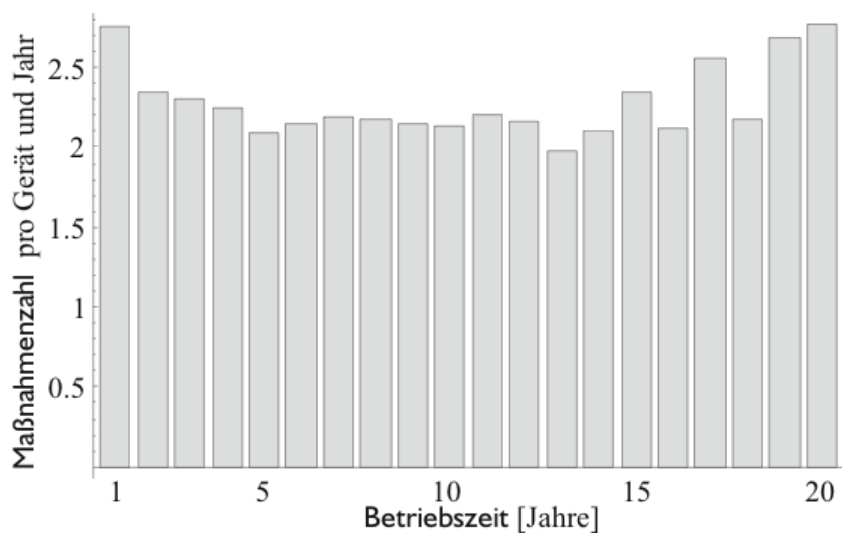


Abbildung 4.53: Darstellung der Maßnahmenzahl pro Gerät und Jahr über der Betriebszeit für alle Geräte. Erklärung siehe Text.

F abgebildet.

Insgesamt kann man erkennen, dass ausgeprägte „Badewannenkurve“ aus Abschnitt 2.4.3 nicht vorherrschen – die Werte sind etwa konstant. Die Zunahme der Schwankungen für größere Betriebszeiten ist auf die dann kleineren Aktivgerätezahlen zurückzuführen.

## 4.5 Analyse von Ungleichverteilungen

Auf der Suche nach interessanten Zusammenhängen in Datenbeständen erweisen sich Ungleichverteilungen jeder Art als guter Anhaltspunkt für weitere Analysen. Jeder Zusammenhang beruht letztlich auf einer Ungleichverteilung – damit gehört die Assoziationsanalyse im weiteren Sinne auch zu den Analysen von Ungleichverteilungen. Da die Assoziationsanalyse jedoch über das reine Auffinden von Ungleichverteilungen hinausgeht, ist ihr ein eigener Abschnitt gewidmet. Hier sollen Methoden vorgestellt werden, die sich auf das Darstellen von Ungleichverteilungen beschränken.

### 4.5.1 ABC-Analysen

Im Abschnitt 2.4.3 wurde die Mean Time Between Failures MTBF für verschiedene Maßnahmengruppen bestimmt. Anhand dieser Größe kann man zum Beispiel abschätzen, wie oft ein Gerät pro Jahr geserviced werden muss. Kennt man nun noch den Kostenerwartungswert, so kann man die pro Gerät und Jahr anfallenden Kosten nicht nur bestimmen, sondern auch noch ihr Zustandekommen nach Häufigkeit und Höhe darstellen.

Diese Analysen werden als ABC-Analysen bezeichnet ([16], Seite 16 ff), da man den Bereich der teuren und häufigen Maßnahmen als A-Bereich bezeichnet, den Bereich der teuren aber seltenen und den Bereich der häufigen aber günstigen Maßnahmen als B-Bereich und den Bereich der günstigen und seltenen Maßnahmen als C-Bereich bezeichnet.

In den Abbildungen 4.54 und 4.55 sind die ermittelten Kostenerwartungswerte und die Maßnahmenzahlen pro Gerät und Jahr (bestimmt aus der MTBF) für verschiedene MT-Gerätearten und -Untergerätearten dargestellt. Als gestrichelte Linien sind beispielhaft sich ergebende Kosten pro Jahr und Gerät eingezeichnet.

Global kann man erkennen, dass die Kosten pro Jahr und Gerät wesentlich stärker durch die Kosten pro Maßnahme beeinflusst werden als durch die Maßnahmenzahl pro Gerät und Jahr. Diese variieren nur im Bereich 0 bis 20, die Kosten pro Maßnahme variieren jedoch im Bereich von 50 € bis 1 000 000 €. Dies wird auch durch den fast vertikalen Verlauf der gestrichelten Linien deutlich.

Zur Untersuchung einer Abhängigkeit der ermittelten Kostenerwartungswerten und Maßnahmenzahl pro Gerät und Jahr wird nach Abschnitt 2.3.3 der Spearman'sche Rang-Korrelationskoeffizient  $r_S$  ermittelt. Es ergibt sich  $r_S = 0.291$ . Der kritische Wert lautet  $k_{11;0.05} = 0.618$ . Die Nullhypothese kann also nicht auf dem Signifikanzniveau  $\alpha = 0.05$  abgelehnt werden – d.h. es besteht kein Signifikanter Zusammenhang zwischen den ermittelten Kostenerwartungswerten und Maßnahmenzahl pro Gerät und Jahr.

In Abbildung 4.55 kann man ein Hauptfeld im Bereich von weniger als 2000 € Kosten pro Maßnahme und weniger als 4 Maßnahmen pro Jahr und Gerät erkennen.

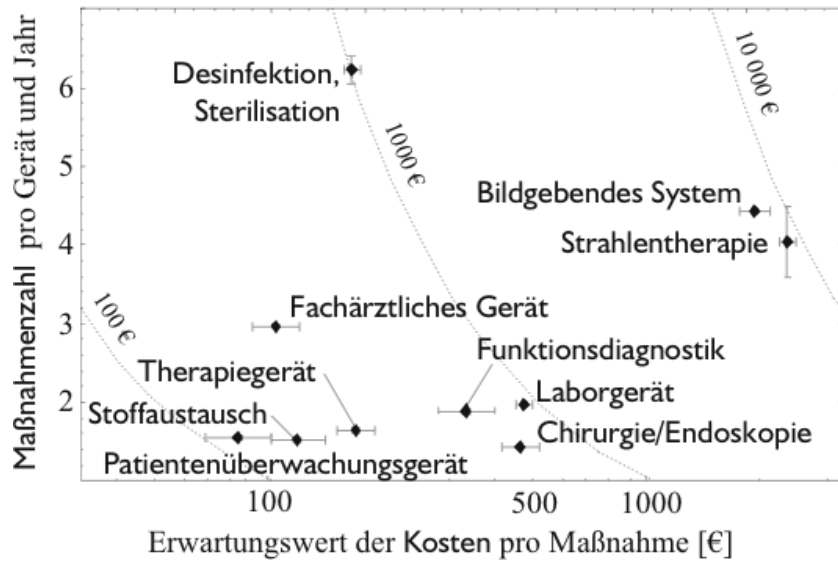


Abbildung 4.54: Darstellung der ermittelten Kostenerwartungswerte und der Maßnahmehzahl pro Gerät und Jahr je MT-Geräteart. Durch gestrichelte Linien sind ausgewählte Erwartungswerte für die anfallenden Kosten pro Gerät und Jahr angegeben. Deutung siehe Text.

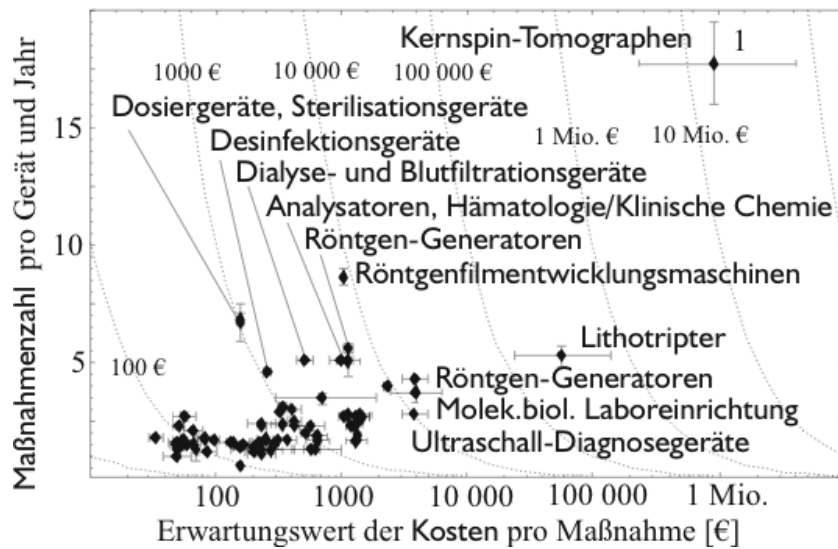


Abbildung 4.55: Darstellung der ermittelten Kostenerwartungswerte und der Maßnahmehzahl pro Gerät und Jahr je MT-Untergeräteart. Durch gestrichelte Linien sind ausgewählte Erwartungswerte für die anfallenden Kosten pro Gerät und Jahr angegeben. Deutung siehe Text.

MT-Gerätearten, die häufiger geserviert werden müssen, dabei jedoch nicht mehr Kosten verursachen, sind Dosiergeräte, Sterilisationsgeräte, Desinfektionsgeräte, Dialyse- und Blutfiltrationsgeräte, Analysatoren aus der klinischen Chemie und Hämatologie, Röntgen-Aufnahmegeräte und Röntgenfilmentwicklungsmaschinen. Geräte, die nicht häufiger bearbeitet werden müssen aber höhere Kosten pro Maßnahme verursachen, sind Lithotripter, Röntgen-Generatoren, molekularbiologische Laboreinrichtungen und Ultraschall-Diagnosegeräte. Der Ausreißer bei sowohl Maßnahmenzahl pro Gerät und Jahr als auch Kostenerwartungswerte sind Kernspintomographen. Aufgrund der sehr inhomogenen Verteilung der ermittelten Werte macht eine Analyse der Korrelation für diese Abbildung keinen Sinn.

Die in den Abbildungen 4.54 und 4.55 dargestellten Werte sind im Anhang G tabelliert angegeben.

In Abbildung 4.56 ist bei der Gruppierung der Maßnahmen nach Krankenhausart eine typische ABC-Anordnung zu finden. Den A-Bereich markieren die frei / gemeinnützigen Kliniken mit häufigen und teuren Maßnahmen. Den B-Bereich markieren die öffentlichen Kliniken, wobei Universitätskliniken eher teure Maßnahmen und die sonstigen öffentlichen Kliniken eher häufig Maßnahmen durchführen. Bei den Kosten pro Jahr und Gerät liegen die beiden Krankenhausarten dicht beieinander. Den C-Bereich markieren die privaten Kliniken mit seltenen und günstigen Maßnahmen.

Da es nur vier verschiedene Krankenhausarten gibt, ist eine Untersuchung der Abhängigkeit der ermittelten Kostenerwartungswerten und Maßnahmenzahl pro Gerät und Jahr für die verschiedenen Krankenhausarten kaum sinnvoll. Insbesondere ist der in Abschnitt 2.3.3 vorgestellte Spearman'sche Rang-Korrelationskoeffizient erst ab einer Anzahl von 6 Wertepaaren einsetzbar. Aus diesen Gründen entfällt hier eine Untersuchung der ermittelten Werte auf eine gegenseitige Abhängigkeit.

Abbildung 4.57 zeigt recht deutlich, dass mit steigendem Kaufpreis nicht nur der Kostenerwartungswert ansteigt sondern auch die zu erwartende Maßnahmenzahl pro Gerät und Jahr. Die Maßnahmenzahl pro Gerät und Jahr steigt erst bei Geräten mit einem Kaufpreis über 1000 € deutlich an.

Zur Untersuchung einer Abhängigkeit zwischen den ermittelten Kostenerwartungswerten und Maßnahmenzahl pro Gerät und Jahr wird nach Abschnitt 2.3.3 der Spearman'sche Rang-Korrelationskoeffizient  $r_S$  ermittelt. Es ergibt sich  $r_S = 0.900$ . Der kritische Wert lautet  $k_{11;0.05} = 0.618$ . Die Nullhypothese kann also auf dem Signifikanzniveau  $\alpha = 0.05$  abgelehnt werden – d.h. es besteht ein signifikanter Zusammenhang zwischen den ermittelten Kostenerwartungswerten und Maßnahmenzahl pro Gerät und Jahr.

## 4.5.2 Lorenz-Kurven

In Abschnitt 2.5.3 wurden Lorenzkurven als geeignete Mittel dargestellt Ungleichverteilungen darzustellen. Hier sollen nun Ungleichverteilungen der Gerätezahl, Maßnahmenzahl, Kaufpreissumme und Kostensumme bezüglich der MT-Geräteart, MT-Untergeräteart, Abteilung und Hersteller untersucht werden. Dazu sind in Abbildung 4.58 Lorenzkurven verschiedener Kombinationen der oben Aufgeführten Kennzahlen und Merkmale dargestellt. Die Lorenzkurven mit der stärksten Wölbung gehören zu nicht gruppierten Geräten bzw. Maßnahmen. Es ist klar, dass wenn Geräte oder Maßnahmen z.B. nach Abteilungen gruppiert werden, die Ungleichverteilung nicht zunehmen kann.

Die Abbildung enthält vielfältige Informationen. Zunächst fällt auf, dass z.B. die Maßnahmenzahl bezüglich der Gerätezahl (oben links) weniger ungleich verteilt ist als

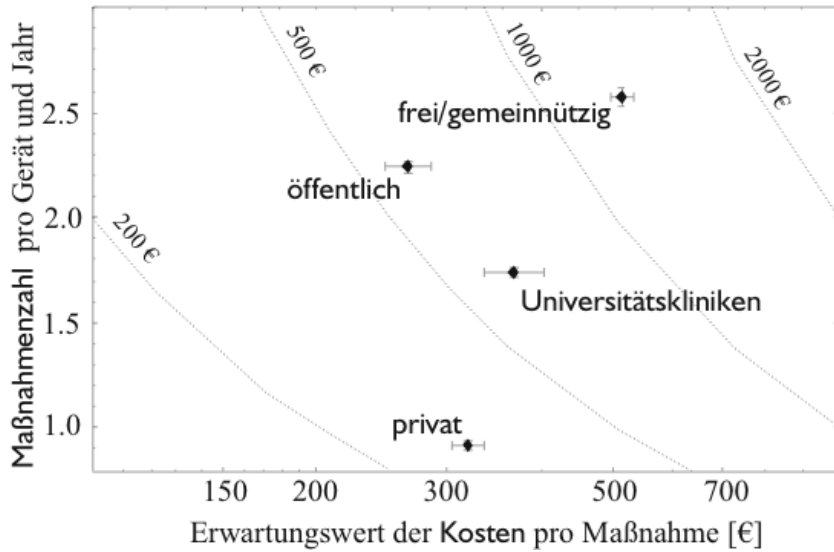


Abbildung 4.56: Darstellung der ermittelten Kostenerwartungswerte und der Maßnahmenzahl pro Gerät und Jahr je Krankenhausart. Durch gestrichelte Linien sind ausgewählte Erwartungswerte für die anfallenden Kosten pro Gerät und Jahr angegeben. Deutung siehe Text.

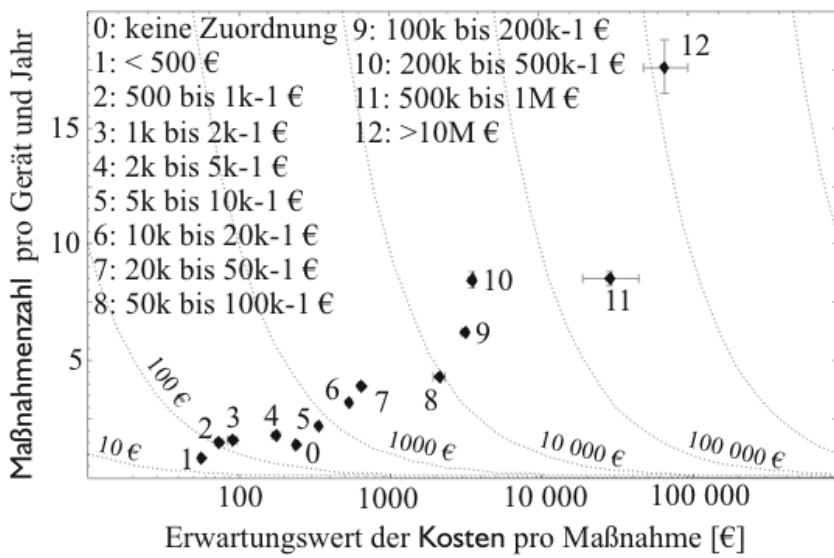


Abbildung 4.57: Darstellung der ermittelten Kostenerwartungswerte und der Maßnahmenzahl pro Gerät und Jahr je Kaufpreis. Durch gestrichelte Linien sind ausgewählte Erwartungswerte für die anfallenden Kosten pro Gerät und Jahr angegeben. Deutung siehe Text.

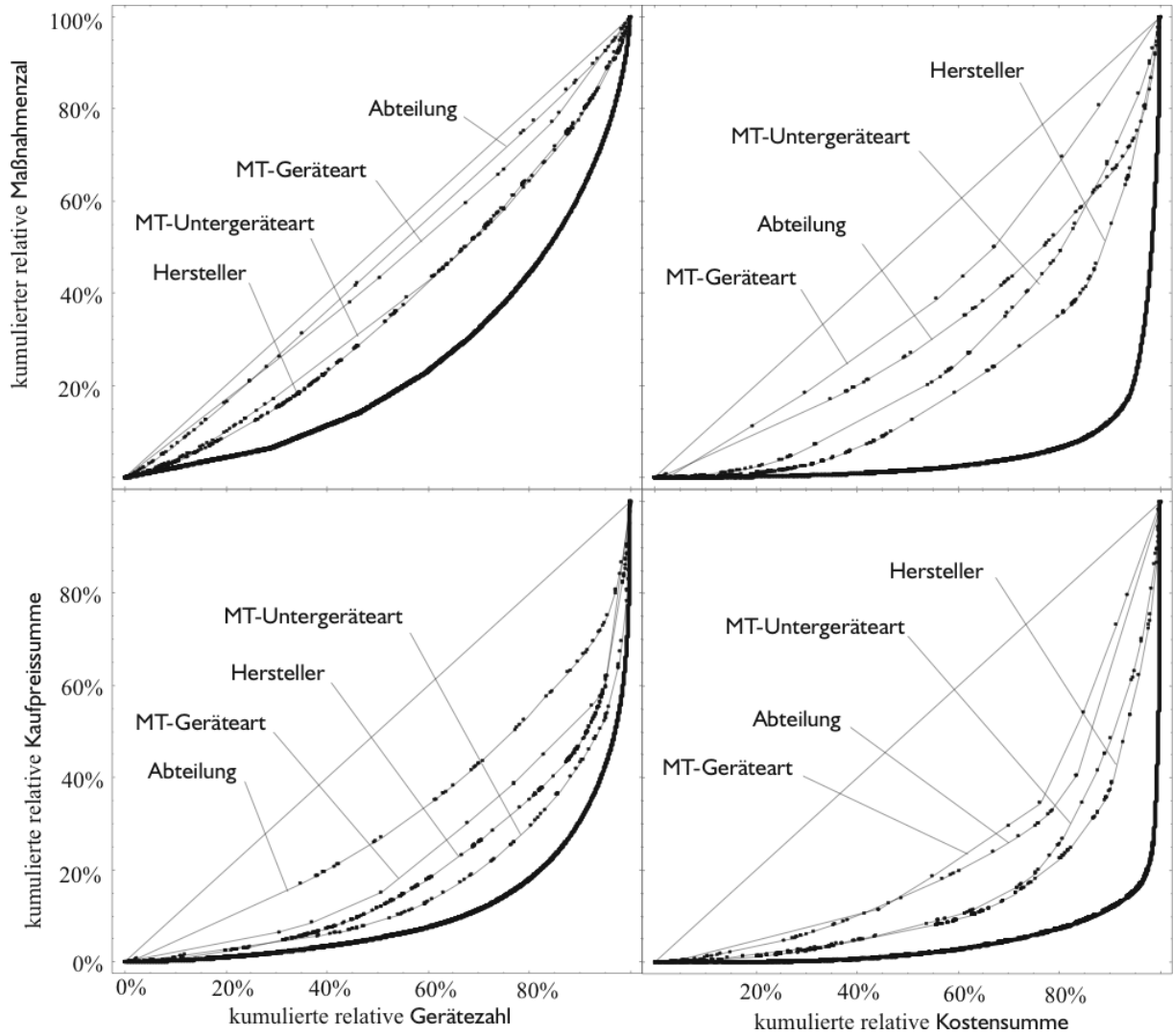


Abbildung 4.58: Lorenzdiagramme für verschiedene Merkmale. Erläuterungen siehe Text.

z.B. die Kaufpreissumme bezüglich der Kostensumme (unten rechts). Das bedeutet, dass die Anzahl der Maßnahmen pro Gerät für verschiedene Gerätegruppen weit weniger schwankt als die aufgewendeten Kosten pro Investition.

Weiter kann man ablesen, wie gut die verschiedenen Merkmale bezüglich der Kennzahlen die betrachteten Geräte und Maßnahmen in unterschiedliche Gruppen teilt. Zum Beispiel schwanken die Kosten pro Gerät je MT-Untergeräteart viel stärker als je Abteilung (unten links).

Die Abbildung verschafft einem also einen schnellen und informativen Überblick über viele Verteilungen. So kann zum Beispiel geprüft werden, ob die Erfassung eines Merkmales überhaupt lohnt: sollte ein Merkmal bezüglich aller interessanten Kennzahlen nur unwesentlich von der Gleichverteilung abweichen, so besitzt das Merkmal kaum Informationswert.

## 4.6 Assoziations-Analysen

Mit Hilfe der in Abschnitt 2.3.4 vorgestellten Methoden sollen nun die Daten untersucht werden. Dabei werden die Methoden im ersten Teil auf gemeinsam auftretende Ereignisse (Koinzidenzen) angewendet. Im zweiten Teil werden jedoch auch aufeinanderfolgende Ereignisse mittels Markoff-Ketten mit der selben Methode untersucht.

### 4.6.1 Analyse von Koinzidenzen

Bisher wurden einzelne Zusammenhänge untersucht, z.B. die Leistungsart in Abhängigkeit der Krankenhausart. Nunmehr soll nicht die Art der Abhängigkeit einiger Merkmale untersucht werden, sondern, ob überhaupt ein Zusammenhang zwischen beliebigen Merkmalen bzw. deren Ausprägung besteht. Die Analysen sind also weniger genau aber dafür breiter.

Ob ein Zusammenhang besteht, wird mit Hilfe des in Abschnitt 2.3.4 vorgestellten lifts entschieden. In den folgenden Abbildungen sind nur Zusammenhänge dargestellt, die einen lift größer 2 besitzen. Ein lift größer 2 bedeutet anschaulich, dass 2 Ausprägungen verschiedener Merkmale mehr als doppelt so häufig gemeinsam an einer Entität auftreten (Koinzidenz), als dies statistisch bei Gleichverteilung zu erwarten wäre. Der lift ist somit ein Maß für die Abweichung von der Gleichverteilung. Diese gehäuften Koinzidenzen sind jedoch nur ein notwendiges Kriterium für Korrelationen und kein hinreichendes. Zur Aufspüren von Zusammenhängen liefert diese Analyse aber wertvolle Hinweise.

Um die Darstellungen übersichtlich zu halten, muss nicht nur der lift eingezeichneter Zusammenhänge größer zwei sein, sondern auch die eingezeichneten Ausprägungen müssen mit einer „zentralen“ Ausprägung in Verbindung stehen. Wenn selbst diese Auswahl nicht zu einem übersichtlichen Grafen führt, werden im Einzelfall die schwächsten Zusammenhänge und die dann nicht mehr verbundenen Ausprägungen aus dem Graphen sukzessiv entfernt. So kann es zu Ausprägungen kommen, die nur noch indirekt mit der zentralen Ausprägung stehen. Die Bezeichnung der Krankenhäuser erfolgt wie in Abbildungen 4.1 und 4.2.

Abbildung 4.59 zeigt die Zusammenhänge und Ausprägungen, die mit einem lift größer 2 mit der zentralen Ausprägung Arbeitsbeschreibung = Akku und untereinander

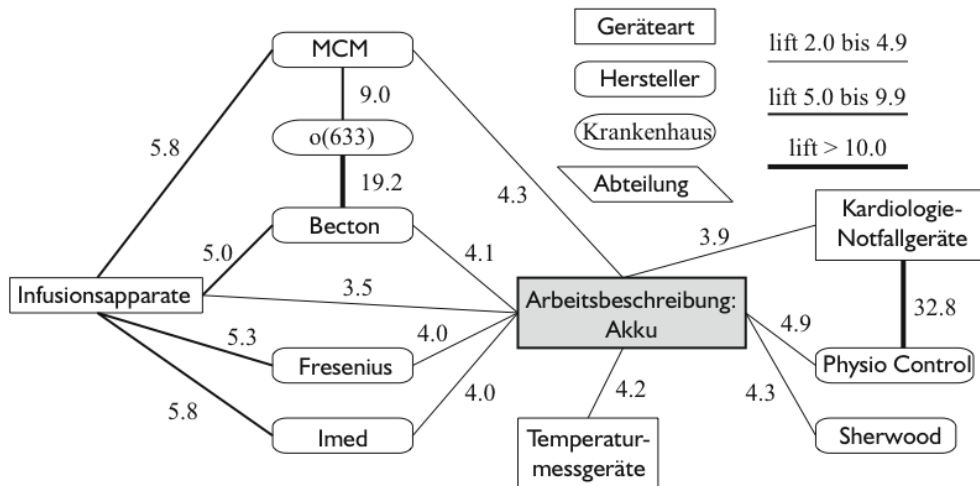


Abbildung 4.59: Darstellung der Zusammenhänge mit lift größer 2 für zentrale Ausprägung Arbeitsbeschreibung = Akku. Bei den Krankenhäusern sind Krankenhausart als Buchstabe und die Bettenzahl dahinter in Klammern angegeben. Die Kodierung der Krankenhausarten durch Buchstaben ist in Abbildung 4.1 angegeben. Erläuterungen siehe Text.

in Verbindung stehen. Es ist anhand dieser Grafen zu erkennen, warum zum Beispiel an Geräten bestimmter Hersteller viele Maßnahmen mit der Arbeitsbeschreibung Akku vorgenommen wurden: die Hersteller stellen viele Geräte einer Geräteart her, die viele Maßnahmen mit Arbeitsbeschreibung Akku erfordern. Über diese Wirkzusammenhänge muss nicht mehr spekuliert werden, sondern diese sind anhand der Darstellungen leicht zu erkennen.

Die Abbildung 4.60 hat als zentrale Ausprägung Arbeitsbeschreibung = bedienungsbedingt. Es ist zu erkennen, dass Maßnahmen mit dieser Ausprägung häufig als Eigenleistung und mit Kosten kleiner 50 € an eher günstigen Geräten abgewickelt werden. Außerdem ist wieder zu erkennen, wie die Hintergrundzusammenhänge dieser Ausprägungen strukturiert sind.

In den Abbildungen 4.61 und 4.62 sind die Ausprägungen Eigenleistung und Fremdleistung des Merkmales Leistungsart als zentrale Ausprägung dargestellt. In den Abbildungen 4.63 und 4.64 sind die Ausprägungen Elektrik und Mechanik des Merkmales Arbeitsbeschreibung als zentrale Ausprägung dargestellt. In Abbildung 4.65 folgt schließlich noch die Arbeitsbeschreibung Prüfung.

Bei den Krankenhäusern sind Krankenhausart als Buchstabe und die Bettenzahl dahinter in Klammern angegeben. Die Kodierung der Krankenhausarten durch Buchstaben ist in Abbildung 4.1 angegeben.

#### 4.6.2 Analyse von Sequenzen

Wurden im ersten Teil gemeinsam auftretende Ereignisse (Koinzidenzen) untersucht, so sollen hier aufeinanderfolgende Ereignisse (Sequenzen) untersucht werden. Die Abfolge von Ereignissen wird durch Markoff-Ketten dargestellt (siehe Abschnitt 2.3.5). Statt die Übergangswahrscheinlichkeiten anzugeben wird hier die Übergangswahrscheinlichkeit von Ereignis  $A$  nach  $B$  dividiert durch das Produkt der Einzelwahrscheinlichkeiten  $p(A)$



### 4.7.1 Zipf's Gesetz

In den Abbildungen 4.67 und 4.68 sind die relativen Worthäufigkeiten  $p_i$  logarithmiert über den zugehörigen Rängen  $i$  der nach Häufigkeit absteigend sortierten Worten logarithmisch für verschiedene Krankenhäuser aufgetragen. In Abbildung 4.67 sind Worte der originalen Gerätebezeichnung und in Abbildung 4.67 die Worte der originalen Arbeitsbeschreibungen dargestellt. Nach dem in Abschnitt 2.3.6 vorgestellten Zipf'schen Gesetz sollte sich in dieser Darstellungsart eine Gerade ergeben. Man erkennt, dass für niedrige Ränge sich  $a \approx 1$  (siehe Formel 2.26) ergibt. Für größere Ränge scheint  $a$  größer zu sein. In [17] wird für große (mehr als 5000 bis 10 000 Worte) Texte ein ähnliches Verhalten beobachtet. Es wird vermutet, dass die beiden unterschiedlichen Exponenten auf zwei verschiedene Wortschätze hindeuten: einen allgemeinen Wortschatz ( $a \approx 1$ ) und einen speziellen Wortschatz ( $a \approx 2$ ). Für Gerätebezeichnungen kann man so auf einen allgemeinen Wortschatz von ca. 100 Worten schließen. Für Arbeitsbeschreibungen ist ein Übergang nur schwach ausgeprägt und liegt etwa im Bereich von 500 bis 1000 Worten.

### 4.7.2 Wortentropie

Neben Prüfung der absoluten Worthäufigkeiten  $P_i$  bzw. der relativen Worthäufigkeiten  $p_i$  aus dem letzten Abschnitt auf Zipf-Verteilung kann auch die zugehörige Wortentropie nach Abschnitt 4.7.2 berechnet werden.

In Abbildung 4.69 sind neben der Wortentropie  $h$  die relative Wortentropie  $h_R$  und

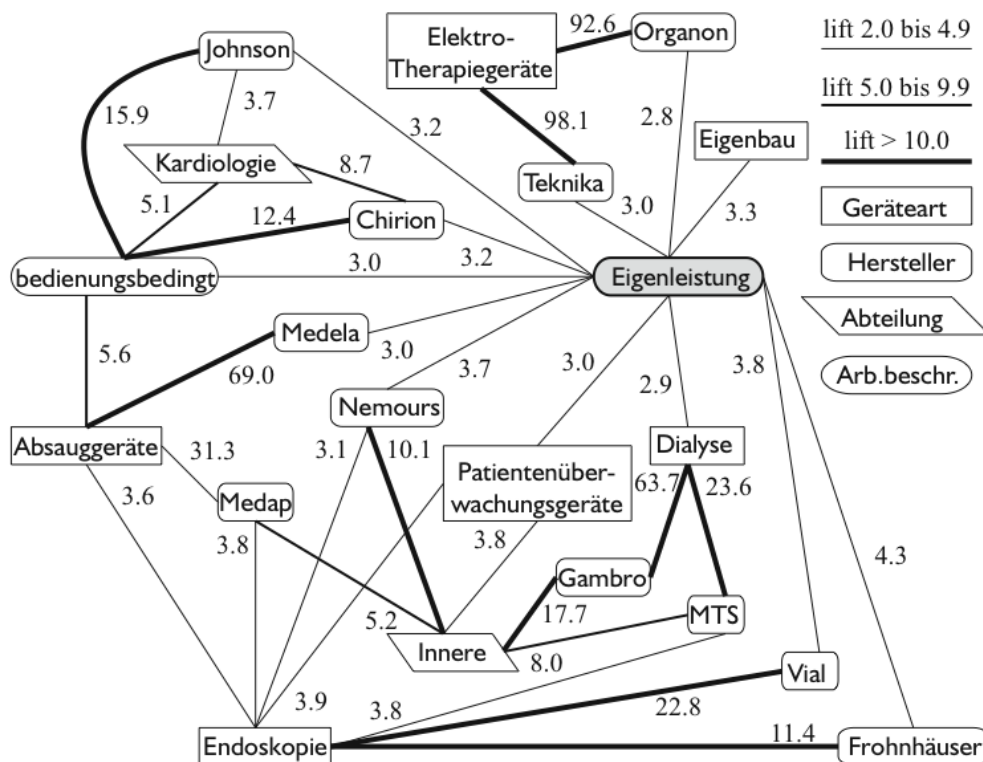


Abbildung 4.61: Darstellung der Zusammenhänge mit lift größer 2 für zentrale Ausprägung Leistungsart = Eigenleistung. Erläuterungen siehe Text.



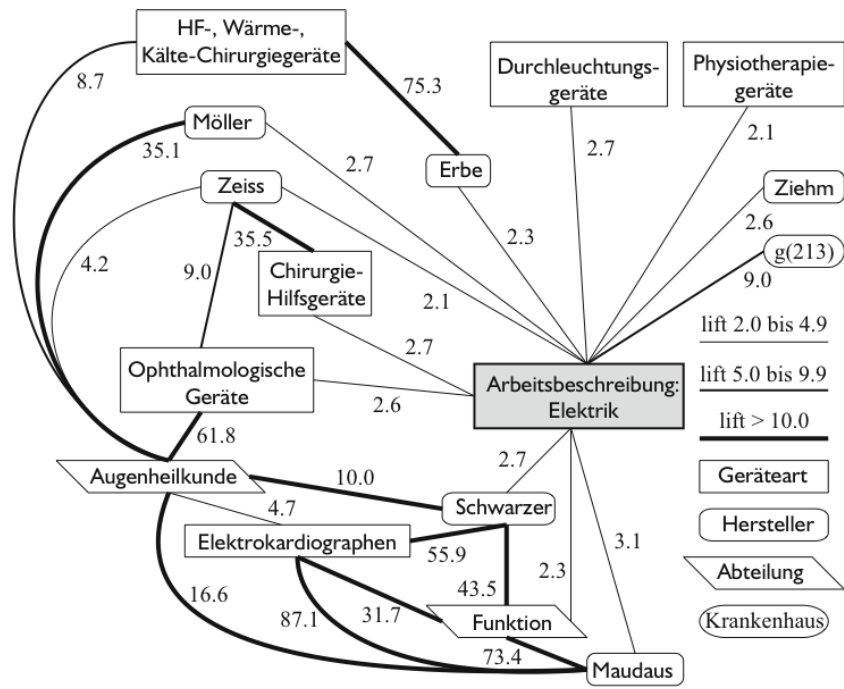


Abbildung 4.63: Darstellung der Zusammenhänge mit lift größer 2 für zentrale Ausprägung Arbeitsbeschreibung = Elektrik. Die Kodierung der Krankenhausarten ist in Abbildung 4.1 angegeben. Erläuterungen siehe Text.

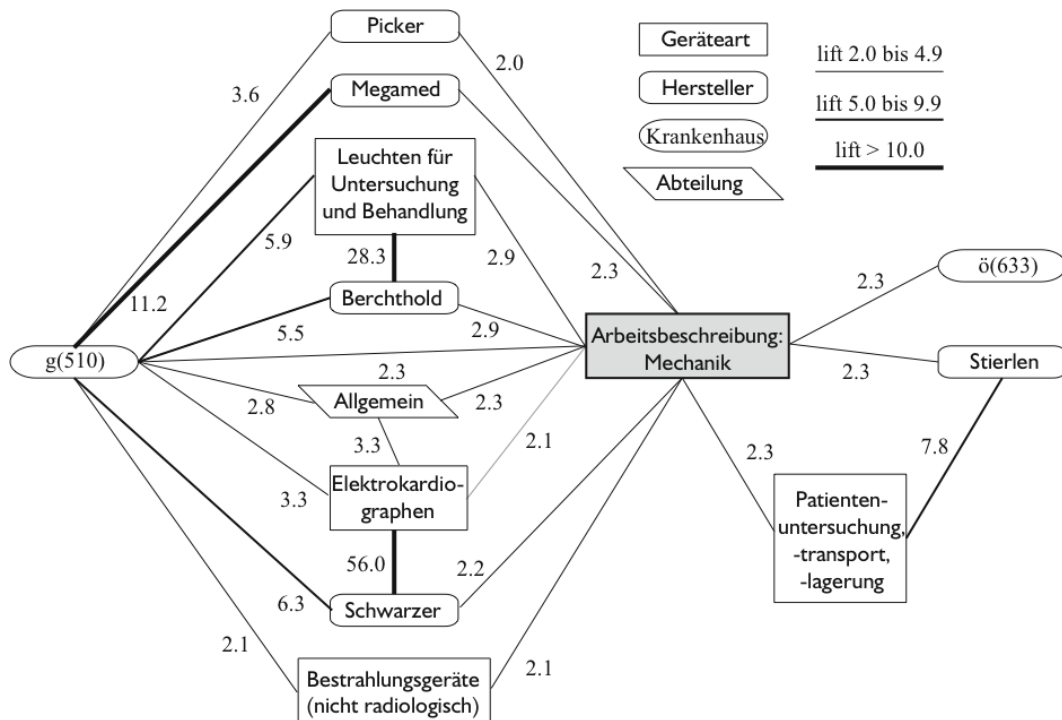


Abbildung 4.64: Darstellung der Zusammenhänge mit lift größer 2 für zentrale Ausprägung 'Arbeitsbeschreibung = Mechanik'. Die Kodierung der Krankenhausarten ist in Abbildung 4.1 angegeben. Erläuterungen siehe Text.

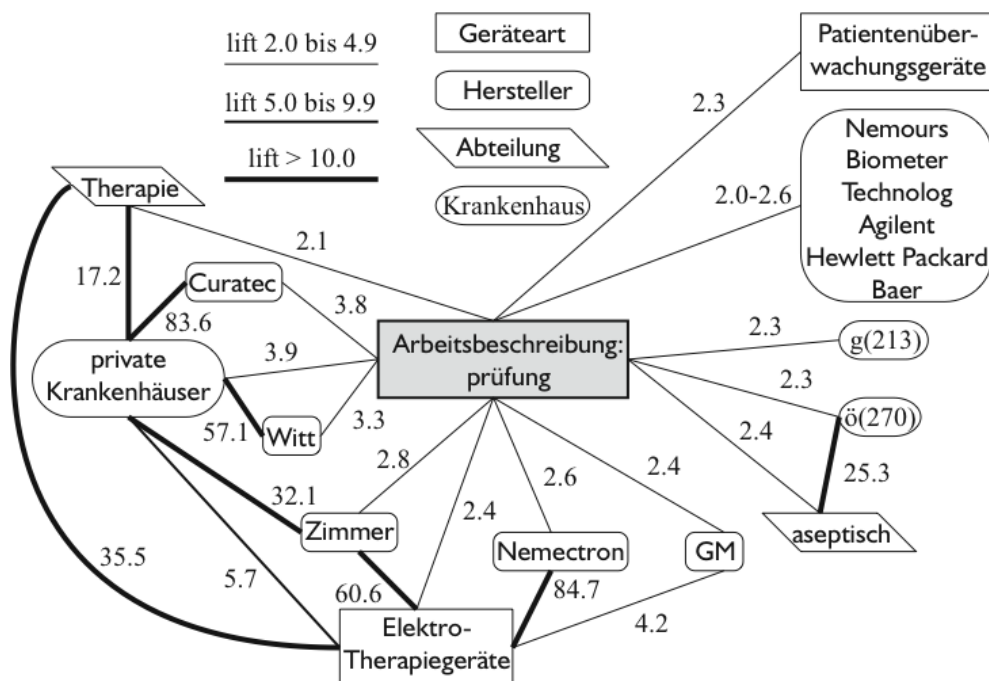


Abbildung 4.65: Darstellung der Zusammenhänge mit lift größer 2 für zentrale Ausprägung Arbeitsbeschreibung = Prüfung. Bei den Krankenhäusern sind Krankenhausart als Buchstabe und die Bettenzahl dahinter in Klammern angegeben. Die Kodierung der Krankenhausarten durch Buchstaben ist in Abbildung 4.1 angegeben. Erläuterungen siehe Text.

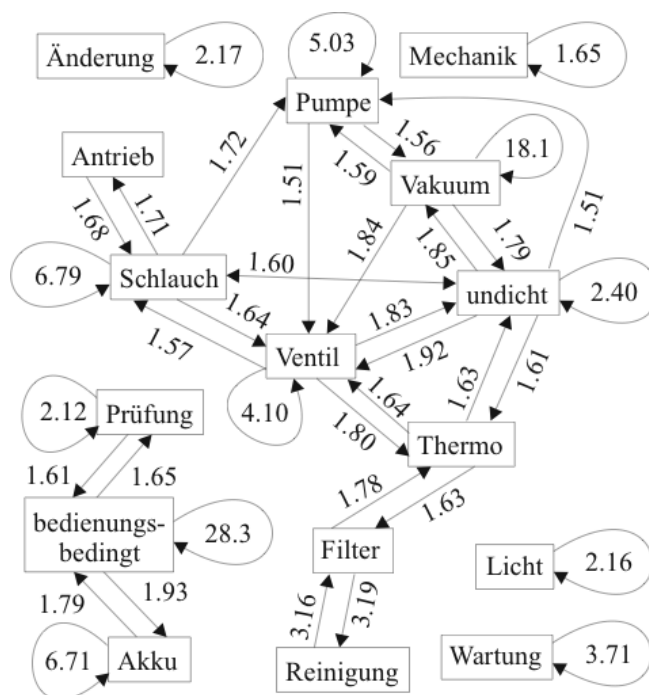


Abbildung 4.66: Darstellung der Zusammenhänge mit lift größer 1.5 für aufeinanderfolgende Arbeitsbeschreibungen. Erklärung siehe Text.

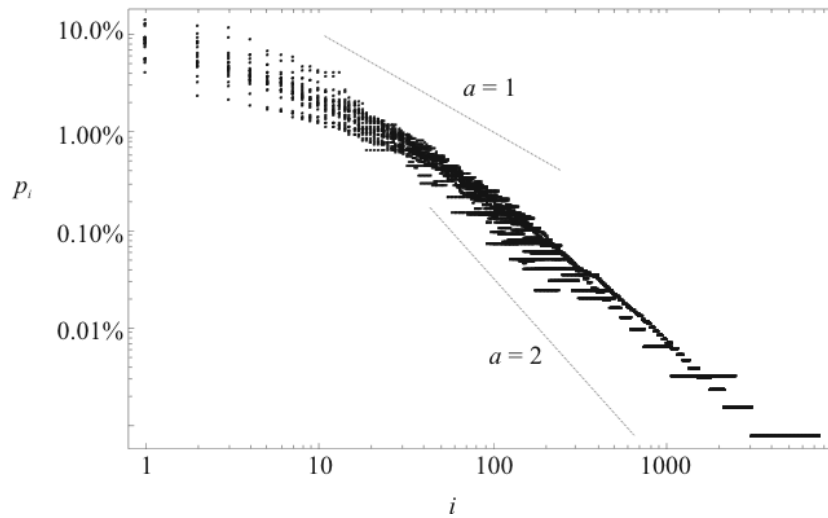


Abbildung 4.67: Darstellung der Zipf-Verteilung der relativen Worthäufigkeiten  $p_i$  der originalen Gerätebezeichnung in verschiedenen Krankenhäusern. Beispielhaft sind Geraden mit  $a = 1$  und  $a = 2$  eingezeichnet. Erläuterungen siehe Text.

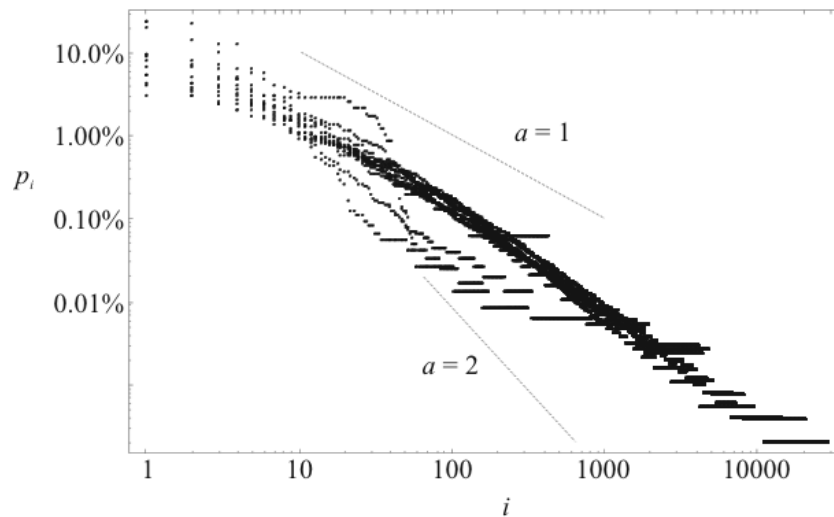


Abbildung 4.68: Darstellung der Zipf-Verteilung der relativen Worthäufigkeiten  $p_i$  der originalen Arbeitsbeschreibungen in verschiedenen Krankenhäusern. Beispielhaft sind Geraden mit  $a = 1$  und  $a = 2$  eingezeichnet. Erläuterungen siehe Text.

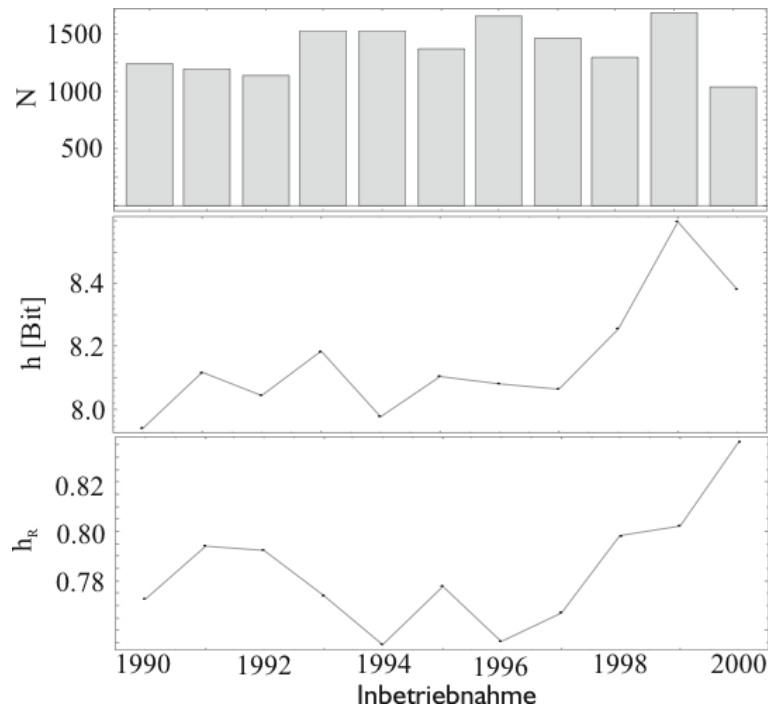


Abbildung 4.69: Darstellung der Anzahl  $N$  der unterschiedlichen Worte, der Wortentropie  $h$  – gemessen in Bit – und der relativen Wortentropie  $h_R$  der Gerätebezeichnung in Abhängigkeit der Inbetriebnahme. Erläuterungen siehe Text.

die Anzahl  $N$  der unterschiedlichen Worte der Gerätebezeichnungen in Abhängigkeit der Inbetriebnahme aufgetragen. Man erkennt, dass die Anzahl der unterschiedlichen Worte keinen Trend bezüglich der Inbetriebnahme besitzt – also etwa konstant bleibt. Die Wortentropie nimmt jedoch ab 1997 deutlich von 8 auf 8.5 Bit zu. Dies bedeutet, dass im Mittel pro Wort der Gerätebezeichnung zwischen  $2^8 = 256$  bzw.  $2^{8.5} \approx 362$  verschiedenen „Gerätearten“ unterschieden wird.

Eine Zunahme der Wortentropie kann zwei Ursachen haben: zum einen kann die Anzahl der unterschiedlichen Worte  $N$  wachsen und damit auch die Menge der pro Wort übertragenen Information. Es gibt aber auch noch eine andere Möglichkeit, die jedoch etwas schwieriger zu verstehen ist. Angenommen man führt ein neues Wort ein, so nimmt zwar die Information pro Wort zu, jedoch hängt die Zunahme an Information von der Häufigkeit des neuen Wortes ab. Verwendet man das neue Wort z.B. nur einmal, so nimmt der Informationsgehalt der anderen Worte nur minimal zu. Die maximale Information pro Wort wird dann übertragen, wenn die Häufigkeiten der verschiedenen Worte gleich hoch sind – die Entropie wird also bei Gleichverteilung maximal.

Damit kann also die Wortentropie auch dadurch zunehmen, dass die Verteilung der Worthäufigkeiten sich in Richtung Gleichverteilung ändert. Als Maß für die Gleichverteilung kann die relative Wortentropie  $h_R$  herangezogen werden, da sie unabhängig von verschiedenen  $N$  bei Gleichverteilung den Wert 1 annimmt.

In [Abbildung 4.69](#) ist zu erkennen, dass die Gleichverteilung der Worte zunimmt. Man kann also schließen, dass die Zunahme der Entropie  $h$  nicht durch die Zunahme des Wortschatzes  $N$  entstanden ist, sondern durch die Zunahme der Gleichverteilung  $h_R$ .

Eine Zunahme der Gleichverteilung ist aber auf eine Abnahme von vielen seltenen Worten zugunsten von wenigen häufigen Worten zurückzuführen. Diese Entwicklung ist zu begrüßen, denn es sind die vielen seltenen Worte, die eine Analyse von Datenbeständen sehr erschweren können. Qualitativ ist also die Entwicklung der Gerätebezeichnungen auf einem guten Weg. Ob diese Entwicklung aber quantitativ ausreicht, kann mit diesen Methoden nicht entschieden werden. Für diese Arbeit war die Normierung der Gerätebezeichnungen noch nicht ausreichend.

In Abbildung 4.70 sind neben der Wortentropie  $h$  die relative Wortentropie  $h_R$  und die Anzahl der unterschiedlichen Worte der Arbeitsbeschreibungen in Abhängigkeit des Durchführungszeitpunkt aufgetragen. Es können die selben Überlegungen wie bei Abbildung 4.69 angestellt werden. Das bedeutet, dass die Zunahme der Entropie  $h$  hauptsächlich auf einer Zunahme des Wortschatzes  $N$  beruht – mehr noch: es ist sogar eine leichte Abnahme der Gleichverteilung  $h_R$  zu beobachten. Das bedeutet, dass mit den neuen Worten auch viele sehr selten verwendete Worte eingeführt werden, die eine Analyse der Daten erschweren. Die Arbeitsbeschreibungen sind damit auf einem nicht so guten Weg wie die Gerätebezeichnungen.

In Abbildung 4.71 sind neben der Wortentropie  $h$  die relative Wortentropie  $h_R$  und die Anzahl der unterschiedlichen Worte der Arbeitsbeschreibungen in Abhängigkeit der Betriebszeit aufgetragen. Man erkennt, dass der Wortschatz  $N$  für ältere Geräte abnimmt und damit auch der Informationsgehalt  $h$  pro Wort. Jedoch erkennt man gleichzeitig eine Zunahme der Gleichverteilung  $h_R$ . Daraus kann man folgern, dass junge Geräte zwar viele verschiedene Ausfälle zeigen, diese jedoch recht ungleich verteilt sind – d.h. es gibt wenig häufige und viele seltene – man darf also eine gewisse Regularität der unterstellen. Mit zunehmendem Alter nimmt zwar die Anzahl der möglichen Ausfälle ab, aber die Gleichverteilung nimmt zu – d.h. wenige häufige Ausfälle nehmen zugunsten von vielen seltenen Ausfällen ab – die Zufälligkeit der auftretenden Fehler nimmt zu.

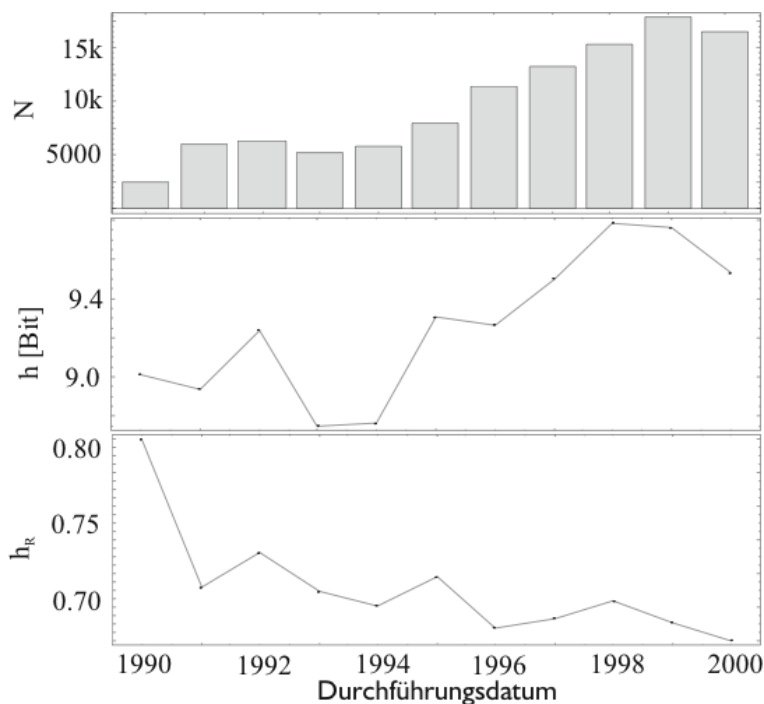


Abbildung 4.70: Darstellung der Anzahl  $N$  der unterschiedlichen Worte, der Wortentropie  $h$  – gemessen in Bit – und der relativen Wortentropie  $h_R$  der Arbeitsbeschreibung in Abhängigkeit des Durchführungsdatums. Erläuterungen siehe Text.

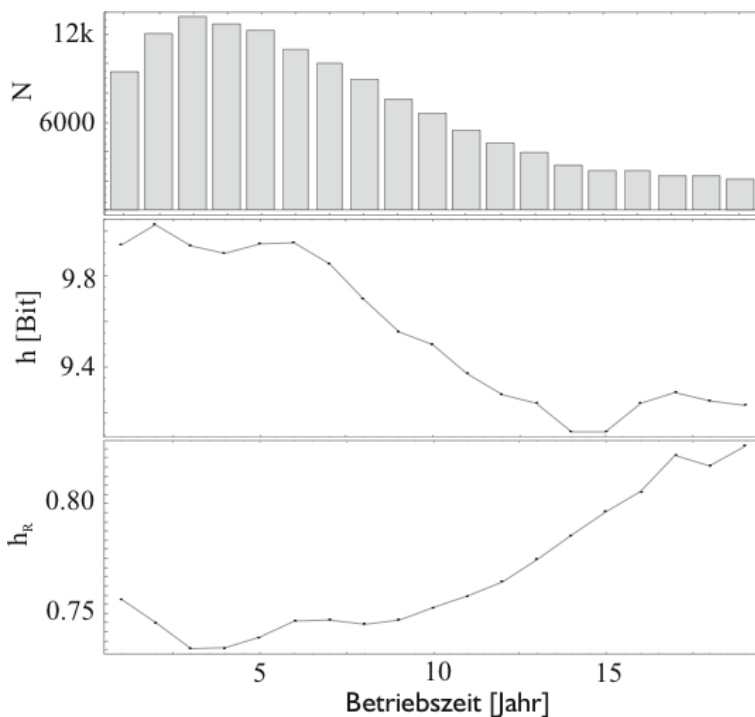


Abbildung 4.71: Darstellung der Anzahl  $N$  der unterschiedlichen Worte, der Wortentropie  $h$  – gemessen in Bit – und der relativen Wortentropie  $h_R$  der Arbeitsbeschreibung in Abhängigkeit der Betriebszeit. Erläuterungen siehe Text.

# 5 Schlussteil

## 5.1 Leiterkenntnisse und Verbesserungsvorschläge

Hier sollen die wichtigsten Erkenntnisse aus der bisherigen Arbeit zusammengefasst werden. Insbesondere im Analyseteil zeigt sich oft, dass viele Abbildungen keine „Hauptkenntnis“ besitzen, sondern je nach Standpunkt und Interessenlage eine Vielzahl von Detailinformationen liefern. Dies ist ausdrücklich kein Mangel sondern durchaus gewollt. Es macht kaum Sinn diese Details durchzugehen – diese sollten ja gerade durch die Abbildungen akzeptabel dargestellt werden. Das bedeutet, dass Erkenntnisse im Einzelfall auch dann bedeutsam sein können, wenn sie im Schlussteil nicht erwähnt werden.

### 5.1.1 Methodenteil

#### Software und Bedienung

Effektive Software und Hardware zur Verwaltung von Daten, wie sie bei der Gerätebewirtschaftung anfallen, steht kostengünstig zur Verfügung. Die leichte Verfügbarkeit sollte jedoch nicht mit einer leichten Bedienbarkeit verwechselt werden.

Zum Verwalten von Tabellen ist das Programm Microsoft Excel Standard. Mittlerweile sind auch vergleichbare Programme (z.B. open office) kostenlos verfügbar. Excel besitzt eine intuitive Bedienung und ist weit verbreitet. Jedoch lassen sich viele bei der Verwaltung von mittelgroßen Datenbeständen notwendigen Tätigkeiten damit kaum durchführen. So ist das notwendige Verknüpfen von Tabellen nur bedingt möglich.

Man kommt also um ein Datenbanksystem nicht herum. In dieser Arbeit wurde hauptsächlich mit Microsoft Access gearbeitet. Die intuitive Bedienung von Access kann jedoch auf Dauer keinen Sachverstand ersetzen – im Gegenteil: wenn die vielen in Access eingebauten Automatismen nicht das gewünschte Ergebnis liefern, ist eine Fehleranalyse oft schwieriger als bei manueller Verarbeitung. So sind in dieser Arbeit die Importassistenten von Access negativ aufgefallen.

Man kommt also ohne Fachpersonal nicht aus. Im Rahmen dieser Arbeit zeigte sich jedoch, dass mit Sachverstand und geeigneter Software Datenbestände sehr effektiv bearbeitet und verwaltet werden können. Fachpersonal ist also nicht bei jeder Standardeingabe erforderlich. Bei Beratungen der Medizinischen Hochschule Hannover, die Geräteinventarisierung betreffend, zeigte sich, dass ein mehrstündiger Test der eingegebenen Daten am Wochenbeginn schließlich ausreicht, um die Daten aus EDV-technischer Sicht zu pflegen. Durch ggf. sogar verschlüsselte Datenkommunikation muss das Fachpersonal nicht vor Ort sein. Eher erfordert eine effektive Kommunikation mit den Bedienern vor Ort eine Anwesenheit des Fachpersonals.

Breite Schulungen können die Abhängigkeit von einer Softwarelösung bedenklich erhöhen. Außerdem ist fraglich, ob das durch Schulungen erzielte Fachwissen reicht,

um seltene aber schwierige Probleme zu bewältigen. Auch ist zu bedenken, dass die Mitarbeiter, die Daten eingeben, kein Fachwissen benötigen, und die die Fachwissen haben bei Dateneingaben nicht effizient eingesetzt sind.

**Fazit:** (Software-)Investitionen können (Fach-)Personalkosten nur eingeschränkt ersetzen.

### Auswahl von Analyseverfahren

Bei der Analyse der Daten herrscht kein Mangel an Analyseverfahren und Möglichkeiten diese durchzuführen sondern eher ein Überangebot. Dies führt dazu, dass nicht das Durchführen einer Analyse schwierig ist sondern die Auswahl des Analyseverfahrens.

In dieser Arbeit stellte sich heraus, dass dem Versuch, durch aufwendige Analysen sichere Schlussfolgerungen auch bei fehlerbehafteter Daten zu ziehen, Grenzen gesetzt sind. Im Bereich von Gerätebewirtschaftungsdaten ist die Wahrscheinlichkeit systematischer Fehler recht hoch – und diese Fehler sind im Nachhinein durch Analysen kaum zu kompensieren.

Statt auf komplexe Analysen zu setzen, sollten Fehler beim Datengenerierungs- und -verarbeitungsprozess vermieden werden – hier zahlt sich ein erhöhter Aufwand in jedem Fall aus.

**Fazit:** Sorgsame Datenerhebung ist wichtiger als komplexe Datenanalyse.

### Vorbereitung für die Daten-Analyse

Die Methoden des Data Mining zeigen recht deutlich, dass beim Gewinnen von Informationen aus großen Datenbeständen die Hauptarbeit nicht in Auswahl und Anwenden der Analysen besteht sondern bei der Datenvorbereitung. Mehr noch: auch eine Prüfung der Ergebnisse und sogar der Aufgabenstellung ist wesentlicher Bestandteil des Ablaufes.

Ist die Aufgabenstellung nicht schon beim Datengenerierungsprozess vor Augen, so läuft man Gefahr, dass die gewonnen Daten am Ende nicht die Fragestellung beantworten können. Zwar ist man auch so nicht grundsätzlich vor Problemen geschützt, jedoch können ärgerliche Verfahrensfehler minimiert werden.

Auch die erforderliche Qualität der Ergebnisse sollte möglichst früh festgelegt werden: genügen wage Hinweise oder sind genaue Aussage nötig, welche Irrtumswahrscheinlichkeiten sind akzeptabel.

Oft ist es gar nicht möglich, die Daten mit der gewünschten Genauigkeit zu erfassen. Dann kann es durchaus sinnvoll sein, einen ganz anderen Weg zu gehen, als an einer Datenerfassung festzuhalten, die nicht in der Lage ist, die gestellten Fragen zu beantworten. Dies können z.B. Bereiche sein, in denen Informationen schlecht zugänglich sind (z.B. Daten über den Geräteinsatz oder Daten externer Dienstleister).

Es ist auch denkbar, dass Entscheidungen nur zum Teil auf Basis der vorliegenden Daten getroffen werden und eher von externen, wie z.B. politischen oder juristischen, Sachverhalten beeinflusst werden. In solchen Fällen sollten ungenaue Informationen ausreichen, weil die Entscheidungen hauptsächlich durch andere Faktoren bestimmt sind.

**Fazit:** Intensivierung EDV-gestützter Datenverarbeitung ist oft aber nicht immer effektiv.

### Genauigkeit von MTBF und Verlauf der Ausfallrate (Badewannenkurve)

Bei Geräten, die mehrmals in ihrem Leben ausfallen können (weil sie danach wieder instandgesetzt werden), sind die Ermittlungen von MTBF und der zeitlichen Ausfallrate nur schwer einheitlich durchführbar. Für die Ermittlung der MTBF muss die Betriebszeit des Gerätes nach einem Ausfall wieder auf Null zurückgesetzt werden. Für die Darstellung der Ausfallrate (die ggf. als Badewannenkurve darzustellen ist) zählt die Betriebszeit jedoch von Inbetriebnahme des Gerätes an (siehe 2.4.3).

Genau genommen treffen auch diese Extreme nicht die Wirklichkeit – es liegt eine Mischung von beidem vor. Auch nach einer Instandsetzung ist davon auszugehen, dass das Gerät nicht wieder in einen Neuzustand versetzt wurde – das Gerät hat also ein Gedächtnis. Das zurückstellen der Uhr auf Null nach einer Instandsetzung für die Berechnung der MTBF ist also nicht ganz korrekt, denn das bedeutet, das Gerät startet wieder fabrikneu.

Die einschlägigen Darstellungen (z.B. [7]) gehen nicht von einem so komplexen Fall von Geräten mit Gedächtnis aus, die mehrmals ausfallen können. Modelle, die dies berücksichtigen ([18] Seite 109ff) werden sehr schnell so komplex, dass sie in dieser Arbeit – auch wegen der fraglichen Datenqualität – nicht zur Anwendung kommen.

Analysen des Gerätelebenslaufes von instandsetzbaren Geräten mit Gedächtnis können bei nicht exzellenter Datenqualität kaum über die reine Beschreibung des Verhaltens hinausgehen. Modelle, die auch Aussagen über Wirkzusammenhänge zulassen stellen aufgrund ihrer Komplexität sehr hohe Anforderungen an die Daten.

**Fazit:** Analysen von Lebensläufen instandsetzbarer Geräte stellen hohe Anforderungen an die Daten.

### Wahl der Darstellung der Daten

Gerade bei der Darstellung von Kreuztabellen als Balkendiagramme sind die Variationsmöglichkeiten immens (siehe 2.5.2). Eine ungünstige Wahl der Darstellung muss nicht gleich zu einer Fehlinterpretation führen – jedoch kann das Erkennen wertvoller Informationen sehr erschwert werden. Auch eine übersichtliche Darstellung hierarchisch geordneter Kategorien, die sehr unterschiedlich besetzt sind, wie bei Treemaps, ergibt sich nicht von alleine.

Doch auch hier gilt wie bei den Analysen: auch noch so viel Aufwand vermag nicht mehr Informationen aus den Zahlen herauszuholen, als in ihnen steckt.

Die in dieser Arbeit verwendeten Algorithmen zur Darstellungsverbesserung (siehe 2.5) können ohne weitere Anforderungen an die Daten in leicht nachvollziehbarer Art und Weise die Aussagekraft der Abbildungen erhöhen. Leider sind bis jetzt keine Möglichkeiten realisiert, dies mit Standardanwendungen, wie z.B. Excel, automatisiert durchzuführen. Viele Anwendungen bieten aber durchaus die Möglichkeit zu solchen Erweiterungen.

**Fazit:** Gute Darstellungen machen das Auge zu einem guten Datenanalysewerkzeug.

## Denkbare Ergebnisanforderungen

Im Abschnitt 2.2.6 wurden einige Anforderungen an Data Mining Ergebnisse aufgezählt. Nach den nun durchgeführten Analysen ergibt sich in Hinblick auf die Gerätebewirtschaftung ein etwas detailliertes Bild.

Es wurden Validität und Neuheitsgrad als notwendige Kriterien für Ergebnisse aufgeführt. Bei den Analysen dieser Arbeit zeigte sich jedoch, dass diese Begriffe oft miteinander verknüpft sind. Neue Erkenntnisse besitzen nur selten sofort eine hohe Validität. Andererseits können auch schon bekannte Ergebnisse die Validität älterer Analysen erhöhen und sind damit nicht ganz überflüssig.

Als weiteres Kriterium wurde Kompaktheit / Verständlichkeit genannt. Nach den hier gemachten Erfahrungen gewinnt diese Bedingung erst an Bedeutung, wenn sie über die triviale Forderung, keine unverständlichen Ergebnisse zu erzeugen, hinausgeht, indem diese Forderung auch für die Analyse selbst erhoben wird. Nicht nur das Ergebnis selbst muss verständlich sein, auch der Weg zu diesem muss eine Minimum an Transparenz besitzen, sonst ist der Wert der Ergebnisse sehr eingeschränkt. Diese Einschränkung ist um so größer, je unsicherer die analysierten Daten sind.

Diese Forderung besitzt eine nicht zu unterschätzende Brisanz: sehr komplizierte Analysen scheiden damit praktisch aus – zumindest teilweise. An einem Gegenbeispiel sei dies erläutert: ein Taschenrechner liefert sehr wertvolle Ergebnisse auch für jemanden, der die Rechenoperationen, die er im Hintergrund ausführt, nicht kennt. Der numerische Wert von  $\sqrt{10}$  kann auch für jemanden sehr hilfreich sein, der nicht weiß, wie der Taschenrechner den Wert bestimmt. Solange es nur um Zahlen geht, ist die Sache unproblematisch. Sobald man aber von den rein numerischen Quantitäten zu Qualitäten übergeht, z.B. von  $\sqrt{10}$  zu Seitenlänge einer Fläche von 10 Flächeneinheiten, muss man wissen, was man rechnet. Der Taschenrechner kann eine Black Box sein, die Rechnung, die auf ihm durchgeführt wird, muss jedoch transparent sein. So besteht zumindest die Chance, einen etwaigen Fehler aufzudecken. Übertragen auf das Data Mining in dieser Arbeit heißt das, dass das Datenbanksystem (der Taschenrechner) im Detail unverstanden sein kann und damit sehr kompliziert, die darauf durchgeführten Analysen (die Rechnungen) jedoch sehr wohl verstanden werden müssen. Die Meinung, dass es reicht, nur die Ergebnisse zu verstehen und nicht die Analyse, wird hier nicht geteilt.

Schließlich sei noch die Nützlichkeit als Kriterium für Erkenntnisse aufgeführt. Nützlichkeit kann nur mit Rücksicht auf den Nutzer bewertet werden. In der Gerätebewirtschaftung von Medizintechnik sind drei Erkenntnisverwerter denkbar:

**Der Praktiker** denkt eher kurzfristig und braucht schnelle einfache Informationen (am besten ja/nein) ohne Herleitungen und Abwägungsmöglichkeiten. Typische Informationsform: Funktionswert  $f(x = x_1)$ . Typische Fragestellung: Ausmustern oder noch mal instandsetzen?

**Der Taktiker** denkt eher mittelfristig und braucht keine ja/nein-Entscheidungen – die trifft er selber. Er braucht abwägbare Informationen in einer Form, die es ermöglicht Informationen aus anderen Quellen zu integrieren. Im Bedarfsfall sollten Herleitungen der Ergebnisse verfügbar sein. Typische Informationsform: Funktion  $f(x)$ . Typische Fragestellung: Welche mittlere Nutzungszeit ist in Hinblick auf anfallende Kosten optimal?

**Der Strategie** denkt eher langfristig und ist nicht an einzelnen Zusammenhängen interessiert. Er benötigt einen Überblick über alle möglichen Zusammenhänge. Dabei geht es weniger um die Form des Zusammenhangs sondern eher um die Relevanz. Typische Informationsform: Funktional  $g(f_1, f_2, \dots)$ . Typische Fragestellung: Ist eine Änderung der Nutzungszeit zur Kostenreduzierung im Vergleich zu anderen Änderungen überhaupt relevant?

Der Schwerpunkt der Analysen in dieser Arbeit dienen dem Taktiker. Die Assoziationsanalysen von Koinzidenzen (siehe 4.6.1) sind eher für den Strategen von Bedeutung.

### Anpassung der Daten oder der Fragestellung

Das Ziel dieser Arbeit könnte man damit umschreiben, dass zwischen verschiedenen Bereichen versucht wird, eine Brücke zu schlagen. Der eine Bereich ist die Gerätebewirtschaftung mit ihren Fragestellungen in Hinblick auf Effizienz (siehe 1.3). Auf der anderen Seite sind die Datenbestände und ihre Verarbeitungsmethoden (siehe 2.2).

Um eine Brücke zu schlagen, kann man nun einerseits die Daten an die Fragestellungen der Gerätebewirtschaftung anpassen oder aber man passt die Fragestellungen an die Daten an. Beide Varianten haben Vor- und Nachteile.

Aus Sicht des Betreibers ist die Marschrichtung klar: er hat akut Entscheidungen zu treffen und erwartet für seine Fragestellungen zumindest Entscheidungshilfen. Es geht um relevante Informationen. Häufig jedoch erkaufte man sich Relevanz durch einen Verlust an Validität. Die meisten relevanten Fragestellungen können nicht direkt aus den Daten ermittelt werden (z.B. sind die Daten unvollständig oder unzureichend normiert). Aus diesem Grund müssen die Daten oft „interpretiert“ werden, was die Wahrscheinlichkeit von insbesondere systematischen Fehlern erhöht und damit die Validität der gefundenen Informationen vermindert.

Aus Sicht des „Data Miners“ geht der Weg von den Daten aus und soll zu validen Informationen führen – die Relevanz für etwaige Fragestellungen spielt prinzipiell keine so zentrale Rolle. Dabei kann es durchaus passieren das auf Grund der Daten relevante Fragestellungen – insbesondere für den Betreiber von medizin-technischen Geräten – nicht beantwortet werden können.

Eine Überbewertung des ersten Weges, beruht oft darauf, dass die fehlende Validität relevanter Ergebnisse viel schwerer zu entdecken ist als die fehlende Relevanz valider Ergebnisse. Außerdem ist die Wahrscheinlichkeit, dass irrelevante aber valide Informationen vielleicht in Zukunft relevant werden, wesentlich höher, als dass relevante aber nicht valide Informationen in Zukunft valide werden. Schließlich stellt noch das Wissen um fehlende Informationen eine deutlich bessere Ausgangsposition in Hinblick auf eine Verbesserung der Situation da, als das Vertrauen auf zweifelhafte Informationen.

**Fazit:** Die Relevanz von Informationen wird wegen ihrer leichteren Wahrnehmung oft gegenüber ihrer Validität überbewertet.

### 5.1.2 Vorbereitungsteil

#### Vorhandene Daten in elektronischer Form

Eingeschickt wurden Daten von 5% der angeschriebenen Krankenhäusern. Das sind 38% der eingegangenen Antworten. Nicht in elektronischer Form verfügbar waren Daten von

2.3% aller Krankenhäuser und von 16% der eingegangenen Antworten (siehe 3.1.3).

**Fazit:** Es gibt Daten, und sie sind verfügbar.

### Datenstruktur

Die Struktur der Daten ist sehr homogen. Die Daten liegen fast ausschließlich in Form von Tabellen vor und können nahezu verlustfrei in Textdateien umgewandelt werden. In Zeiten von allgegenwärtigen Multimedia- und XML-Metaformaten ist das durchaus bemerkenswert (siehe 3.1.4).

Auch wenn es nur ein technisches Detail ist: Sonderzeichen in den Textdateien erschweren die Verarbeitung (siehe 3.2.1). Alle nötigen Informationen können ohne Sonderzeichen transportiert werden. Sonderzeichen sind somit überflüssig – damit auch insbesondere Formatierungen in den Datenfeldern (siehe 3.2.2 und 3.4.2).

**Fazit:** Die Daten liegen vollständig als tabellierte Textdateien vor.

### Datensyntax

Der Syntax der Daten ist sehr inhomogen. Oft ist der Syntax der Daten vielfältiger als die dahinter liegende Semantik es erfordert. Zum Beispiel gibt es wesentlich mehr Gerätebeschreibungen in den Datenbanken als verschiedene Gerätetypen, die mit den Datenbanken verwaltet werden (siehe 3.3.2, 4.7.1 und 4.7.2).

Zellen, die gemischte Informationen enthalten (wie z.B. Gerätebezeichnung und Seriennummer) erschweren die Analyse und sogar die Verwaltung der Daten. Solche Zellen lassen sich leicht aus Zellen mit „atomaren“ Informationen gewinnen – umgekehrt ist die automatische Trennung der Information meist schwieriger (siehe Tabelle 3.2).

**Fazit:** Im Gegensatz zur homogenen Struktur der Daten ist der Syntax sehr inhomogen.

### Normierung der Daten

Die notwendige Homogenisierung des Syntax – die Normierung – kostet sehr viel Zeit. Selbst, wenn die Möglichkeit einer teilweisen Automatisierung dieses Prozesses bestehen sollte, so werden der dadurch zu erwartenden Zeitgewinn durch die erforderliche Entwicklungs- und Implementationszeit geschmälert.

Zum Identifizieren von wichtigen Zeichenketten in großen Textmengen hat sich ein Algorithmus mit Häufigkeitsdifferenzen von Teilzeichenketten bewährt (siehe 3.5.2). Suchalgorithmen auf Basis von Trigrammen eignen sich sehr gut, um unscharf nach Teilzeichenketten zu suchen. Dies wird z.B. erforderlich, wenn man bestehende Begriffskataloge auf nicht normierte Daten anwenden möchte.

Bestehende Begriffskataloge haben meist den Vorteil, dass sie vollständig und gut strukturiert sind. Beim Anwenden dieser Kataloge auf die nicht normierten Daten steht man aber oft vor dem Problem, dass diese externen Kataloge nicht die vorhandenen Daten effektiv abbilden können. Das bedeutet, dass viele Begriffe sehr selten benutzt werden und wenige sehr häufig.

Begriffskataloge, die man durch Finden von interessanten Teilzeichenketten gebildet hat, weisen dieses Problem meist nicht so deutlich auf. Dafür sind die so gewonnen Kataloge meist weder vollständig noch gut strukturierbar.

Eine vollständige Normierung der Daten scheitert oft an der verfügbaren Zeit aber auch nicht zuletzt an dem Willen und der Möglichkeit für jeden Einzelfall eine Zuordnung zu finden oder zu schaffen. Man wird im Normalfall also nur einen Teil der Daten normalisieren. Welchen Teil soll man wählen? Wer z.B. nur nach der Häufigkeit von Geräten seine Schwerpunkte auswählt, kann bei einer Betrachtung von Kaufpreisen sehr ineffektiv vorgegangen sein. Die Wahl der Beschränkung sollte also mit Bedacht erfolgen und ist manchmal nur mit großen Einbußen an Normierungsqualität möglich.

Schon an diesen Beispielen erkennt man kann, dass eine Normierung immer nur ein Kompromiss zwischen vielfältigen Anforderungen sein kann. Sie ist selten ein Ergebnis einer gradlinigen Entwicklung. Darüber hinaus hängt eine gute Normierung von den Daten ab, d.h. wenn die Daten sich ändern, muss sich ggf. auch die Normierung ändern.

**Fazit:** Auch mit viel Aufwand erstellte Normierungen sind bestenfalls erträgliche Kompromisse.

### Erfasste Merkmale

Wichtige Größen sind nicht erfasst (siehe 3.3.3 und 4.1.1). Die Datenbanken spiegeln hauptsächlich nur wieder, was die Geräte dem Betreiber kosten und nicht, was sie ihm nützen. Ohne diese Informationen ist die Bewirtschaftung nur sehr eingeschränkt an Hand der Daten optimierbar.

Um den Eingabeaufwand nicht zu erhöhen, wäre zu überlegen, ob schlecht erfasste Merkmale (z.B. „ortsfest“) zugunsten einer besseren Erfassung wichtiger Merkmale (z.B. „Betriebsstunden“) ganz aus der Datenbank entfernt werden sollen.

Sehr viele der Datenbanken bieten für jeden Geräte- oder Maßnahmedatensatz ein Freitextfeld „Bemerkung“ an. Obwohl ein solches Feld bei den Nutzern, die etwas in die Datenbank eingeben, beliebter sein dürfte, als bei denen, die etwas aus der Datenbank herausholen möchten, ist ein solches Feld wichtig. Hier ist ein Auffangbecken für Informationen, die sonst nicht abgelegt werden konnten. Zwar sind die Informationen in den Bemerkungsfeldern nur sehr schwach strukturiert, aber die Alternative, dass diese Informationen gar nicht in der Datenbank abgelegt werden, scheint nicht akzeptabel. Immerhin kann dieses Feld bei gelegentlicher Prüfung Hinweise geben, welches Merkmal neu hinzugenommen werden könnte oder welches Merkmal überarbeitungsbedürftig ist.

**Fazit:** Schwerpunkt der Daten liegt bei den Kosten der Geräte – nicht bei dem Nutzen. Eine Spalte für Daten, die Zusatzinformationen liefern, ist unverzichtbar.

### Erzwungene und automatisch generierte Eingaben

Keine Eingabe ist besser als eine automatisch erzeugte aber mit hoher Wahrscheinlichkeit falsche Eingabe. Viele Kaufpreise lauten Null Euro. Die wenigen Fälle, in denen diese Angabe Sinn macht, sind leider im Nachhinein nicht mehr von denen zu unterscheiden, bei denen dies nicht der Fall ist. Abfragen, die unerwünschte Eingaben zurückweisen und damit den Nutzer zwingen irgendetwas einzugeben, scheinen die Situation nur zu verschlimmern.

**Fazit:** Falsche Daten sind schlimmer als keine Daten.

### Weitere Datenquellen

Hochwertige und nützliche Daten liegen oft extern vor. Geräteinformationen müssen beim Aufbau einer Datenbank nicht unbedingt mühevoll eingepflegt werden. Denn mit Sicherheit sind diese Daten beim Hersteller abrufbar. Weitere Datenquellen sind für Informationen externer Dienstleister über durchgeführte Maßnahmen und Normungsinstitutionen für Begriffskataloge (z.B. DIMDI).

Auch interne Datenquellen, wie z.B. die Personalabteilung für Informationen über die Arbeitszeit oder Belegungspläne der Stationen über die Auslastung der Geräte, könnten gewinnbringend in die Datenbanken integriert werden. Informationen, die oft übergangen werden, sind Daten in Papierform und implizites Wissen der Mitarbeiter.

Nicht immer lohnt der hohe, zu erwartende Aufwand einer Integration solcher Daten. Vergessen werden darf allerdings nicht, dass ggf. auch andere Abteilungen von der EDV-technischen Erfassung solcher Informationen profitieren würden und so auf lange Sicht Synergieeffekte zu erwarten sind. Diese Überlegungen werden in Zukunft eher an Bedeutung gewinnen.

**Fazit:** Auch aufwendige Datenimporte können effektiver sein als eigene Datenerhebung.

### Import- und Interpretationsfehler

Hier soll betont werden, dass die bei den Analysen immer wieder in Erscheinung getretenen Datenfehler und Datenlücken nicht ausschließlich bei der Datengenerierung entstanden sind. Es muss vielmehr davon ausgegangen werden, dass viele Informationen der Daten in einem nicht ohne weiteres entschlüsselbaren Format abgelegt wurden und damit nicht oder nur fehlerhaft ausgewertet werden konnten. Ein Teil der Fehler ist also auch auf die in dieser Arbeit durchgeführten Import- und Vorbereitungsprozesse der Daten zurückzuführen.

Zum Beispiel ist es denkbar, dass Informationen über Haupt- und Untergeräte in den Identifikationsnummern hinterlegt sind. Um solche undokumentierten Informationen zu nutzen, müssen entweder auf anderen Kanälen Informationen beschafft werden oder die Daten müssen sehr genau analysiert werden. Die letzte Möglichkeit fällt aus zeitlichen Gründen weg. Die erste Möglichkeit wurde auch nicht gewählt, da sie sich schnell als „Fass ohne Boden“ herausstellen würde. Stichproben haben gezeigt, dass die Betreiber zusätzlich zu den Datenbanken über eine Menge relevanter Informationen verfügen, deren Integration auch den Rahmen dieser Arbeit sprengen würde. Das sind dann Informationen wie z.B.: „Fremdleistungen an Geräten des Herstellers XYZ erkennt man bis zum Jahr 1993 immer an der fehlenden Kostenstellenummer...“.

Durch Weglassen dieser Informationen werden zwar teilweise wichtige Informationen übergangen, aber dieses Vorgehen besitzt auch unübersehbare Vorteile.

1. Die zu untersuchenden Daten sind in Form von Datenbanken und Textdateien gut dokumentiert. Sie werden nicht durch nur schlecht dokumentierbare Informationen angereichert.

2. Durch die gute Dokumentation der zu analysierenden Daten ist eine bessere und wichtige Abschätzung des nötigen Arbeitsaufwandes möglich.
3. Das „Nachreichen“ von Informationen wird prinzipiell ausgeschlossen, da sich die Untersuchungen ausschließlich auf statische Datenbestände stützen.
4. Die Feststellung, dass Zusatzinformationen nötig sind, um die Daten richtig zu interpretieren, ist auch schon eine wichtige Information an sich. Der Großteil der abgelegten Daten zeigt, dass eine verschlüsselte Kodierung von Informationen heutzutage nicht mehr notwendig ist.

**Fazit:** Verschlüsselt kodierte Informationen sollten vermieden werden.

### 5.1.3 Analyseteil

#### Unstetige Erfassung

Die Daten sind zeitlich nicht konstant erfasst (siehe Abbildungen 4.17 und 4.39). Damit verlieren Untersuchungen von absoluten Größen über der Zeit ihren Sinn: untersucht man die Veränderung der erfassten Größen oder die Änderung der Erfassung dieser Größen?

Durch Verwendung von relativen Größen ist nur ansatzweise möglich, diesen Fehler zu beseitigen (siehe Abschnitt 4.1.2).

#### Statistische Tests und systematische Datenfehler

Bei den statistischen Tests wird die Nullhypothese (Gleichverteilung) oft mit überwältigender Deutlichkeit abgelehnt. Gewöhnlich ist das ein erfreuliches Ergebnis, denn dies zeigt, dass die Daten etwas mehr als nur Rauschen zeigen. Bei Betrachtung vieler Darstellungen kommt aber dennoch der Eindruck auf, sie machen eine sehr irregulären, verrauschten Eindruck.

Dabei fällt auf, dass im Allgemeinen die Irregularitäten von den Gerätezahlen über den Maßnahmenzahlen bis zu den Kostensummen immer mehr zunehmen (siehe Abbildungen 4.4, 4.18 und 4.24). Das stärkere Rauschen der Maßnahmenzahlen gegenüber den Gerätezahlen kann man sich folgendermaßen erklären. Die Gerätezahlen sind gleich den Maßnahmenzahlen, wenn für jedes Gerät eine Maßnahme vorliegt. Die Anzahl der Maßnahmen pro Gerät variiert jedoch, und damit kommt zur Variation der Maßnahmenzahlen noch die Variation der Gerätezahl hinzu.

Genauso kann man auch bei der Zunahme des Rauschens bei Kostensummen gegenüber Maßnahmenzahlen argumentieren: die Kostensummen sind gleich den Maßnahmenzahlen, wenn jede Maßnahme 1 € kostet. Wieder sind die Kosten pro Maßnahme aber nicht konstant, und diese Variation der Kosten tritt noch zur Variation der Maßnahmenzahl hinzu.

Gewöhnlich ist Rauschen auf zufällige Störeinflüsse zurückzuführen. Statistische Tests können ihre Leistungsfähigkeit nur in Gegenwart von zufälligen Störeinflüssen entfalten. Nicht zufällige Störeinflüsse – so genannte systematische Fehler – werden von den Tests als vermeintlich interessante Zusammenhänge identifiziert.

Die deutliche Ablehnung der Nullhypothese bei gleichzeitigem Fehlen von deutbaren Zusammenhängen ist mit einer gewissen Wahrscheinlichkeit auf systematische Fehler zurückzuführen. Natürlich kann dieser Nachweis nur schwer erbracht werden, denn durch eine unpassende Darstellung der Daten können Zusammenhänge verschleiert werden, die ein statistischer Test aufgedeckt hat. Die teilweise aufwendigen Darstellungsverbesserungen in dieser Arbeit sollen dieses Argument entkräften. Auf der anderen Seite ist das Vorhandensein von systematischen Fehlern bei der Verarbeitung (Import und Normierung) aber auch nicht zuletzt bei der Erfassung nicht auszuschließen.

Was bleibt, ist die Erkenntnis, dass aus fehlerhaften Daten fehlerbehaftete Schlussfolgerungen gezogen werden. Dass sich diese Fehler bei Untersuchung einer ausreichend großen Zahl von Betrachtungseinheiten zumindest teilweise herausmitteln, hat sich nicht in dem erhofften Maße gezeigt.

### **Umgang mit Haupt-, Untergeräten und Wartungsverträgen**

Die schwache Korrelation von Geräte und Bettenzahl der Krankenhäuser deutet auf neben einer ungleichmäßigen Erfassung auch auf eine nicht einheitliche Behandlung von Haupt- und Untergeräten hin (siehe Abbildung 4.1). Versuche, dies zu vereinheitlichen, werden immer mit Willkür behaftet sein. Auf diesem Weg ist das Problem also nur schwer lösbar. Auf Möglichkeiten, das Problem abzuschwächen, wird im Folgenden noch eingegangen.

Ähnlich dem Problem bei Haupt- und Untergeräten können auch Maßnahmen z.B. durch Wartungsverträge zusammengefasst werden. Eine solche Zusammenfassung führt natürlich zu sehr hohen Kosten je vermeintlicher Maßnahme und Verlust differenzierter Informationen.

### **Robuste Kennzahlen**

Kennzahlen sollen verteilte Informationen prägnant zusammenfassen (siehe 4.1.3). Gegenüber von Datenfehlern soll diese Zusammenfassung weitestgehend unempfindlich (robust) sein.

Datenfehler ergeben sich z.B. aus der unstetigen Erfassung oder auch durch nicht einheitliche Zusammenfassung von Geräten und Maßnahmen. Unter gewissen Bedingungen sind relative Größen gegenüber diesen Datenfehlern robust. Aus diesem Grund bieten sich relative Größen an, wie zum Beispiel relative Geräte- und Maßnahmenzahl sowie die relative Kaufpreis- und Kostensumme.

Weiter wichtige Kennzahlen – jedoch nicht so robust – sind durchschnittlicher Kaufpreis pro Gerät und durchschnittliche Kosten pro Maßnahme sowie MTBF und Maßnahmen pro Jahr und Gerät.

### **Die „dicken Brocken“ der Gerätebewirtschaftung**

Um mit möglichst wenig Aufwand Einsparungen zu erzielen, kann es sinnvoll sein, sich zunächst auf die Gerätegruppen zu konzentrieren, die

- am häufigsten sind
- den größten Teil der Investitionen ausmachen

- die meisten Maßnahmen verursachen
- den größten Teil der laufenden Kosten ausmachen.

Einen optischen Überblick über die möglichen Schwerpunkte bieten die Abbildungen 4.5, 4.6, 4.9, 4.10, 4.19, 4.20, 4.25 und 4.26. Tabellen finden sich dazu in Anhang D.

### Ungleichverteilung der Kaufpreise und Kosten

Unabhängig von der Krankenhausart zeigt sich, dass etwa die teuersten 20% der MT-Geräte 80% aller Kaufpreise ausmachen. Damit gilt auch umgekehrt: die günstigsten 80% der Geräte machen 20% der Investitionen aus. Insbesondere bei Universitätskliniken schwanken die Kosten je Maßnahme sehr stark (siehe Abbildung 4.11) .

Insgesamt kann die Verteilung der Kaufpreise und Kosten überraschend gut durch log-Normalverteilungen beschrieben werden (siehe Abbildungen 4.13 und 4.32, aber auch E.2) – insbesondere, wenn man bedenkt, dass die Kaufpreise und Kosten keiner inneren Logik unterworfen oder einem Naturgesetz unterworfen sind.

Die Erwartungswerte der ermittelten log-Normalverteilungen sind zwar ungenauer als die auf gewöhnlichem Weg ermittelten Mittelwerte, aber dafür bieten die ermittelten Parameter weitere Informationen über das Zustandekommen des Mittelwertes: wird er durch den Häufungspunkt der Verteilung bestimmt oder ist er aufgrund der Schiefe der Verteilung verschoben (siehe Abschnitt 2.3.1).

Bei den Analysen zeigte sich darüber hinaus die deutliche Tendenz, dass Geräte mit hohem Kaufpreis nicht nur teure Maßnahmen nach sich ziehen, sondern diese auch auffallend häufig sind (siehe Abbildung 4.57). Zu beachten ist dabei jedoch, dass die Häufigkeiten je nach Geräteart um den Faktor 10 schwanken können, die Kosten je Maßnahme jedoch um den Faktor 1000 (siehe z.B. Abbildung 4.55).

### Private Krankenhäuser

Bei vielen Analysen nahmen private Krankenhäuser eine Sonderstellung ein. So werden in privaten Krankenhäusern generell weniger Maßnahmen durchgeführt (siehe Abbildung 4.56). Die Geräte neigen am ehesten zu Spätausfällen (siehe Abbildung 4.51).

Der Häufungspunkt der Kosten je Maßnahme liegt zwar relativ hoch, jedoch ist der Erwartungswert der Kosten durch die geringe Schiefe der Kostenverteilung im Vergleich niedrig (siehe Abbildung 4.33). Das bedeutet, im Vergleich zu anderen Krankenhausarten gibt es bei privaten Krankenhäusern zwar nicht so viele sehr günstige Maßnahmen, aber eben auch nicht so viele sehr teure Maßnahmen. Diese im Vergleich zu den sehr günstigen Maßnahmen sehr seltenen Maßnahmen ziehen jedoch den Mittelwert der Kosten stark nach oben.

### Ausfallraten und „Badewannenkurve“

Wie schon weiter oben geschrieben wurde, werden in dieser Arbeit Ausfälle in Abhängigkeit der Zeit seit der letzten Maßnahme und in Abhängigkeit der Zeit seit Inbetriebnahme getrennt untersucht.

Untersucht man die Verteilung der Zeiten zwischen Maßnahmen, so überwiegen Frühausfälle, d.h. die Wahrscheinlichkeit, dass es zu einem Ausfall kommt, nimmt mit zunehmender Zeit seit der letzten Maßnahme ab (siehe Abbildung 4.48). Ausnahmen bilden

hier die schon genannten Maßnahmen in privaten Kliniken (siehe Abbildung 4.51) und Maßnahmen an Beatmungs- und Narkosegeräten (siehe Abbildung 4.49).

Untersucht man die Verteilung der Zeiten seit Inbetriebnahme des betroffenen Gerätes, so sind die Maßnahmen pro Gerät und Jahr relativ konstant für verschiedene Gerätearten – eine in einschlägiger Literatur ([7] Seite 128) beschriebene „Badewannenkurve“ ist in diesem Zusammenhang nicht deutlich zu erkennen (siehe Abbildungen in Anhang F).

Ein Drittel der Maßnahmen fällt innerhalb des 3. bis 5. Betriebsjahres an. An Geräten, die länger als 10 Jahre in Betrieb sind, werden noch 20 % der Maßnahmen durchgeführt (siehe Abbildung 4.40). Der Anteil von Maßnahmen an Stoffaustauschgeräten ist bis zum 10. Betriebsjahr erhöht, der von Vital- / Intensivmedizingeräten in den Betriebsjahren 10 bis 15 und für Chirurgiegeräte ab dem 15. Betriebsjahr auffällig erhöht (siehe Abbildung 4.41).

Für die mit den Maßnahmen verbundenen Kosten gilt: 50 % der Kosten fallen in den Betriebsjahren 2 bis 6 an, und 20 % der Kosten verursachen Geräte, die länger als 10 Jahre in Betrieb sind (siehe Abbildung 4.44).

### Wartungs- / Instandhaltungs-, Eigenservice- und Jahresinstandhaltungsquote

In Abbildung 5.1 ist oben die Wartungs- / Instandhaltungsquote und unten die Eigenservicequote über der Betriebszeit dargestellt, die sich aus den Abbildungen 4.46 und 4.47 ergeben. Die Eigenservicequote gibt damit den Anteil der Kosten und nicht der Maßnahmenzahl an. Die Zuordnungen der Kosten zu Wartungen und Instandsetzungen bzw. zu Eigen- und Fremdleistungen sind natürlich mit Fehlern verbunden.

Die Wartungen sind im 7. und ab dem 15. Betriebsjahr gegenüber den Instandsetzungsmaßnahmen verringert. Dies kann darauf zurückzuführen sein, dass ein Teil der Geräte ab dem 7. Jahr nicht mehr gewartet, aber dann auch recht zügig ausgemustert werden. Die Geräte, die länger betrieben werden, werden generell ab dem 15. Betriebsjahr kaum noch gewartet.

Die Eigenservicequote schwankt vom 2. bis zum 17. Betriebsjahr zwischen 10 und 15 %, wobei sie ab dem 10. Jahr etwas abnimmt.

In den Tabellen 5.1 und 5.2 sind die Jahresinstandhaltungs- und Eigenservicequote je MT-Geräte- und MT-Untergeräteart angegeben. Die Eigenservicequoten sind jeweils bezüglich der Maßnahmenzahlen und der mit ihnen verknüpften Kosten aufgelistet. Ermittelt wurden diese Zahlen aus den Kaufpreisen aus Abbildungen 4.15 und 4.16, aus den Kosten je Maßnahme und der Maßnahmenzahl pro Jahr und Gerät aus Abbildungen 4.54 und 4.55. Anhand dieser Werte kann die Art der Instandhaltung schnell beurteilt werden. Es sei hier nochmal betont, dass die Zahlen mit Unsicherheiten behaftet sind.

In Abbildung 5.2 ist die Jahresinstandhaltungs- und Eigenservicequote (nach Kostenanteilen) für verschiedene Krankenhäuser angegeben. Man erkennt keinen Zusammenhang von Eigenservice- und Jahresinstandhaltungsquote. Ein Minimum bei ca. 65 % – wie in [8] angegeben – kann aufgrund der vorliegenden Zahlen nicht bestätigt werden.

### Aufdecken von Wirkzusammenhängen durch Assoziationsanalysen

Im Abschnitt 4.6.1 werden Assoziationen zwischen Ereignissen untersucht. Assoziationen sind zwar ein notwendiges aber kein hinreichendes Kriterium für kausale Zusam-

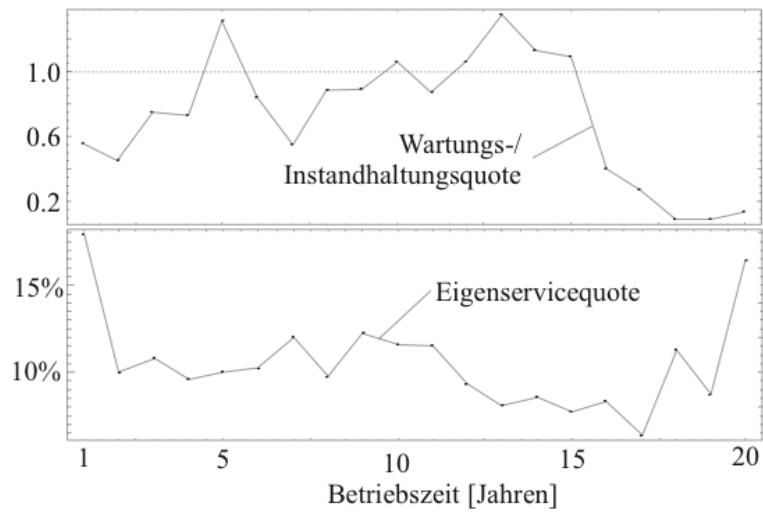


Abbildung 5.1: In der oberen Abbildung ist die Wartungs- / Instandhaltungsquote, in der unten die Eigenservicequote über der Betriebszeit dargestellt. Erläuterungen siehe Text.

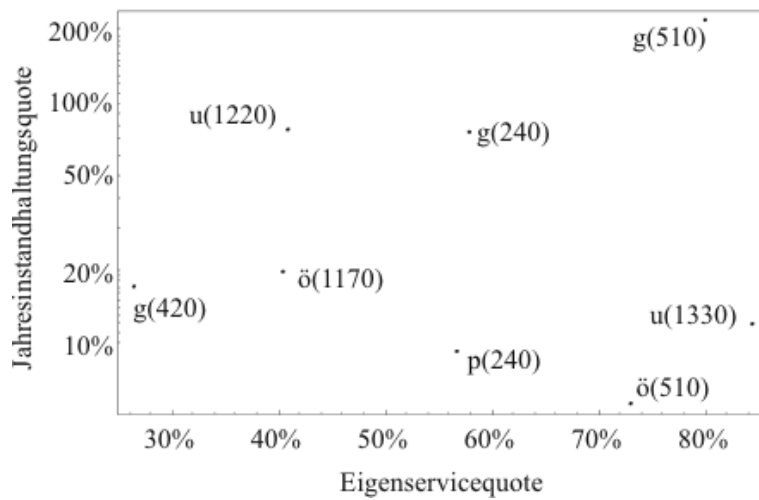


Abbildung 5.2: Für verschiedene Krankenhäuser sind hier die ermittelten Jahresinstandhaltungs- und Eigenservicequoten angegeben. Die Kennzeichnung der Krankenhäuser erfolgt wie in Abbildung 4.1

Tabelle 5.1: Jahresinstandhaltungs- und Eigenservicequote je MT-Geräte- und MT-Untergeräteart, Teil I. Die Eigenservicequoten sind jeweils bezüglich der Maßnahmenzahlen und der mit ihnen verknüpften Kosten angegeben.

EMTEC-Geräteart	Eigenservicequote		Jahresinstandhaltungsquote
	Anzahl	Kosten	
<b>Stoffaustausch</b>	<b>75.5 %</b>	<b>41.5 %</b>	<b>11.7 %</b>
Infusionsapparate	77.7 %	55.6 %	7.59 %
Absauggeräte	71.9 %	37.7 %	7.66 %
Bluttemperierungsgeräte	75.3 %	20.4 %	3.91 %
Dialyse- und Blutfiltrationsgeräte	63.2 %	15.9 %	17.8 %
Blutzuführungs / -entnahme / -aufarbeitungsgeräte	18.5 %	1.51 %	16.5 %
<b>Patientenüberwachungsgerät</b>	<b>60.6 %</b>	<b>27.3 %</b>	<b>5.91 %</b>
Einzelparametermessung, OP / Intensiv	64.9 %	38.1 %	9.78 %
Personen-Waagen	47.2 %	19.0 %	26.8 %
Beatmungs- und Narkose-Monitore	30.3 %	7.78 %	8.55 %
Patientenüberwachungsgeräte	91.0 %	42.5 %	3.34 %
Kombinationsmonitore / -module	68.9 %	9.75 %	13.3 %
Perinatale Überwachung	15.7 %	3.38 %	12.3 %
<b>Chirurgie / Endoskopie</b>	<b>60.1 %</b>	<b>11.9 %</b>	<b>17.0 %</b>
Patientenuntersuchung,-transport,-lagerung	92.8 %	83.6 %	4.66 %
Endoskope und Zubehör	30.6 %	3.07 %	64.6 %
Leuchten für Untersuchung und Behandlung	87.7 %	44.1 %	5.08 %
HF-, Wärme-, Kälte-Chirurgiegeräte	62.5 %	23.3 %	4.08 %
Operationstische	38.8 %	5.57 %	8.64 %
Decken-& Wandversorg.-und befestigungssysteme	38.0 %	1.28 %	85.4 %
Stühle, medizinisch	93.4 %	68.1 %	9.09 %
Chirurgie-Hilfsgeräte	78.1 %	28.6 %	0.93 %
Lithotripter	14.3 %	0.17 %	1615 %
<b>Vitalfunktion / Intensivmedizin</b>	<b>32.1 %</b>	<b>8.66 %</b>	<b>7.57 %</b>
Beatmung und Narkose, Zusatzgeräte	67.7 %	27.3 %	13.67 %
Inhalations-Narkosegeräte	11.6 %	1.69 %	1.79 %
Beatmungsgeräte	36.8 %	12.6 %	4.97 %
Inkubatoren und Wärmegeräte	39.1 %	9.35 %	10.4 %
<b>Funktionsdiagnostik</b>	<b>61.3 %</b>	<b>13.7 %</b>	<b>16.1 %</b>
Blutdruckmeßgeräte	52.5 %	20.4 %	4.62 %
Temperaturmeßgeräte	81.4 %	65.8 %	30.8 %
Elektrokardiographen	58.7 %	12.0 %	33.6 %
Neurologie-Meßgeräte	69.6 %	13.2 %	9.14 %
Ergometer	61.5 %	20.2 %	21.5 %
Lungenfunktions-Meßgeräte	51.9 %	6.11 %	10.2 %

Tabelle 5.2: Jahresinstandhaltungs- und Eigenservicequote je MT-Geräte- und MT-Untergeräteart, Teil II. Die Eigenservicequoten sind jeweils bezüglich der Maßnahmenzahlen und der mit ihnen verknüpften Kosten angegeben.

EMTEC-Geräteart	Eigenservicequote		Jahresinstandhaltungsquote
	Anzahl	Kosten	
<b>Bildgebendes System</b>	<b>38.4 %</b>	<b>2.24 %</b>	<b>27.1 %</b>
Bildbetrachtungseinrichtungen	69.1 %	12.3 %	88.3 %
Ultraschall-Diagnosegeräte	49.5 %	9.29 %	29.2 %
Röntgen-Aufnahmegeräte	17.9 %	0.96 %	21.4 %
Systemkomponenten, Röntgen	18.9 %	0.37 %	0.41 %
Röntgenfilmentwicklungsmaschinen	64.1 %	9.29 %	240 %
Röntgen-Generatoren	31.1 %	1.35 %	7.59 %
Nuklear-Diagnosegeräte	3.50 %	0.74 %	5.84 %
<b>Therapiegerät (ohne OP-Tech. &amp; Strahlenther.)</b>	<b>60.9 %</b>	<b>29.1 %</b>	<b>6.12 %</b>
Kardiologie-Notfallgeräte	52.8 %	24.7 %	4.51 %
Bestrahlungsgeräte (nicht radiologisch)	71.9 %	62.5 %	11.3 %
Physiotherapiegeräte	67.1 %	45.7 %	3.00 %
Stimulationsgeräte Nerven / Muskeln, Therapie	43.1 %	13.6 %	3.74 %
Elektro-Therapiegeräte	83.9 %	65.5 %	3.53 %
Wärmetherapiegeräte / Diathermiegeräte	65.7 %	22.6 %	49.9 %
<b>Laborgerät</b>	<b>40.3 %</b>	<b>8.07 %</b>	<b>59.9 %</b>
Probenvorbereitung / -aufbereitung / -transport	71.9 %	29.8 %	73.5 %
Mikroskope	36.6 %	4.17 %	19.4 %
Labor-Zentrifugen	24.1 %	6.35 %	9.37 %
Analysatoren, einfach	31.2 %	4.08 %	283 %
Analysatoren, Hämatologie	49.9 %	11.9 %	63.6 %
Apothekengeräte	42.2 %	14.9 %	29.0 %
Photometer	31.0 %	1.67 %	39.1 %
Dosiergeräte	9.85 %	2.52 %	80.6 %
Ausstattung und Zubehör, Labor	88.2 %	78.7 %	931 %
Pathologie- und Histologie-Laborgeräte	7.35 %	1.41 %	11.8 %
Meßgeräte, Einzelmessung, Labor	73.2 %	24.9 %	218 %
<b>Desinfektion, Sterilisation</b>	<b>85.3 %</b>	<b>59.3 %</b>	<b>36.4 %</b>
Sterilisationsgeräte	88.6 %	65.2 %	25.6 %
Desinfektionsgeräte	69.9 %	20.6 %	55.7 %
<b>Fachärztliches Gerät</b>	<b>65.9 %</b>	<b>26.8 %</b>	<b>1.86 %</b>
HNO-Geräte	29.1 %	5.02 %	10.3 %
Ophthalmologische Geräte	87.2 %	53.7 %	11.8 %
Dentaltechnische Geräte	28.2 %	4.77 %	40.7 %

menhänge. Dennoch liefern die Analysen Hinweise auf deutliche Wirkzusammenhänge.

Oft zeigt sich, dass Assoziationen zwischen Ereignissen durch mehrere Verbindungen über Umwege verknüpft sind – mehr noch: es zeichnen sich Gruppen von Ereignissen ab, die untereinander häufig verbunden sind und sog. „Cluster“ bilden. Eine Bildung von Clustern kann auf versteckte Merkmale von Betrachtungseinheiten hindeuten, die die Betrachtungseinheiten mit wertvollen Informationen ausstatten können. Zum Beispiel erkennt man in Abbildung 4.6.2 zwei Cluster von Arbeitsbeschreibungen. Hier könnte sich z.B. anbieten, eine organisatorische Trennung einer bisher auf andere Art getrennten Instandhaltung vorzunehmen, da hier Geräte, die zwischen den Einheiten hin und her pendeln, minimiert werden.

### Wortschätze und -verteilungen

Im Abschnitt 4.7.1 wurde für Gerätebezeichnungen ein Stammwortschatz von ca. 100 und für Arbeitsbeschreibungen von ca. 500 bis 1000 Worten ermittelt.

Für Analysen dieser Daten ist es günstig, wenn diese Worte etwa gleichhäufig verteilt sind – d.h. die vielen verschiedenen Worte, die nur selten auftauchen, erschweren die Analysen. In Abschnitt 4.7.2 wurden mit Hilfe der Entropie diese Worthäufigkeitsverteilungen untersucht. Es zeigt sich, dass die Gerätebezeichnungen sich tendenziell schon auf einem guten Weg befinden – d.h. also die Ungleichverteilung der verwendeten Worte abnimmt. Die Verteilung der Arbeitsbeschreibungen verschlechtert sich in dieser Hinsicht sogar leicht – hier ist also insbesondere Handlungsbedarf in Hinblick auf eine Normierung angezeigt.

Bemerkenswert sind auch die Veränderungen der Arbeitsbeschreibungen in Abhängigkeit der Betriebszeit des betroffenen Gerätes. So nimmt der Umfang des verwendeten Wortschatzes mit zunehmender Betriebszeit ab, die Zufälligkeit der verbleibenden Worte nimmt aber zu – und damit, so wird hier vermutet, auch die Zufälligkeit der zu erwartenden Fehler.

### Allgemeine Bemerkungen zu Gleich- und Ungleichverteilungen

Das Ziel aller in dieser Arbeit verwendeten Analysen kann man als Aufspüren und Darstellen von Ungleichverteilungen auffassen. Die Art der Darstellung und auch die Art der Ungleichverteilung, die gesucht wird, mögen sich unterscheiden. Jedoch ist dieses Prinzip meist dasselbe: fast jede in dieser Arbeit gewonnene Erkenntnis beruht auf einer Ungleichverteilung innerhalb der untersuchten Daten.

Auf der anderen Seite war es zum Durchführen der Analysen oft nötig, die Daten zu normieren. Allgemein hat eine Normierung das Ziel, Ungleichverteilungen in Gleichverteilungen umzuwandeln – viele seltene Bezeichnungen sollen in wenige Häufige verändert werden.

Um es deutlich zu sagen: das Vermeiden von Ungleichverteilungen bei der Eingabe erleichtert das Finden interessanter Ungleichverteilungen bei der Analyse. Sonst untersucht die Analyse eher die Art der Eingabe als die Dinge, die eingegeben wurden.

### 5.1.4 Unterstützung bei der Gerätebewirtschaftung der Medizinischen Hochschule Hannover

Insbesondere die hier im letzten Teil gewonnen Erkenntnisse sind nicht nur auf Basis der eingesandten Daten gewonnen worden. Einige Informationen sind auch aus der Zusammenarbeit mit der Gerätebewirtschaftung der Medizinischen Hochschule Hannover (MHH) in diese Arbeit eingeflossen.

Eine lehrreiche Aufgabe war, die Organisation einer Neuerfassung der Geräte zu unterstützen. Dabei ging es hauptsächlich um die Planung des Ablaufs der Speicherung, Verarbeitung und Integration der neuen Datensätze in die Gerätedatenbank.

Die Mitarbeiter haben die neu erstellten Datensätze jeden Tag in neue Textdateien gespeichert. Diese Tagesdateien wurden dann in die Gerätedatenbank integriert. Ein Vorschlag als dieser Zusammenarbeit von exotischen Geräten auch ein Foto zu machen, das mit den Datensätzen später verknüpft werden kann, wurde zwar interessiert aufgenommen, scheiterte aber am Fehlen von Kameras.

Die Eindeutigkeit der neuen Geräteummern wurde durch die Ausgabe von Geräteaufklebern mit eindeutiger Nummer sichergestellt. Beim Erstellen der Datensätze wurden nun jedoch teilweise die Identifikationsnummern falsch eingegeben. Da die Ausgabe der Geräteaufkleber nicht systematisch erfolgen sollte, konnte man solche Fehler nicht sofort ermitteln. Diese Fehler tauchten aber dann plötzlich wieder auf, als nun das Gerät mit der richtigen Nummer in die Datenbank importiert werden sollte. Zunächst hat man dies wohl noch in Kauf genommen, als die Fehler jedoch zunahm wurde um Hilfe gebeten.

Dadurch, dass die Tagesdateien sorgfältig erstellt und aufgehoben wurden, war es leicht möglich, die Fehler zurückzuverfolgen, die betroffenen Geräte zu identifizieren und die fehlgeschlagenen Importe nachzuholen, ohne dass Datensätze wieder neu aufgenommen werden mussten.

Andere Tätigkeiten bestanden darin, Tabellen aus verschiedenen Datenbeständen mit Hilfe von recht heterogen aufgebauten "Umschlüsselungs"-Tabellen miteinander in Beziehung zu setzen und zu gruppieren. Letztendlich ging es darum, aktuelle Datenbestände aus einer Abteilung in die „Sprache“ einer anderen Abteilung zu übersetzen, da hier die Daten nicht mehr aktuell waren.

Herausforderung bei dieser Zusammenarbeit war meistens die Verständigung über die nötigen Maßnahmen mit den Auftraggebern.

## 5.2 Zusammenfassung

In dieser Untersuchung werden auf eine Umfrage hin 22 eingesandte Datenbanken aus verschiedenen Krankenhäusern untersucht. Die Datenbanken enthalten alphanumerische Informationen über medizin-technische Geräte und die an ihnen durchgeführten Maßnahmen. Insgesamt stehen Daten von ca. 80.000 medizin-technischen Geräte und 220.000 Maßnahmen im Umfang von weit über 100 MB Textdateien für Analysen zur Verfügung. Explizit ausgeschlossen von der Analyse werden Daten, die nicht in elektronischer Form vorliegen.

Es wird der Zustand der Daten ermittelt und analysiert. Die Darstellung der Ergebnisse erfolgt insbesondere in Hinblick auf mögliche Empfehlungen zur Verbesserung der

Erfassung der Daten und zur Wahl von Schwerpunkten mit hohem Optimierungspotential bei der Gerätebewirtschaftung.

Aufgrund des hohen Anteils von Text in den Daten sind ähnliche Informationen häufig sehr unterschiedlich repräsentiert. Um sinnvoll Analysen durchführen zu können, sind aufwändige Homogenisierungen bzw. Begriffsnormierungen notwendig. Dazu werden die vorliegenden Daten mit bereits existierenden Begriffskatalogen als auch mit aus den Daten erstellten Begriffskatalogen verglichen und verknüpft. Der Umfang der Daten macht ein weitgehend automatisiertes Vorgehen notwendig. Die Algorithmen gehen auf unscharfe Suchalgorithmen, die auf Trigrammen basieren, zurück oder wurden teilweise selbst entwickelt.

Trotz der Normierungen ist eine übersichtliche Darstellung der vielschichtigen und häufig sehr inhomogen besetzten Daten meist nur mit Aufwand zu erreichen. So können z.B. durch automatisierte Umsortierungen auch umfangreiche Kreuztabellen als Balkendiagramme in optisch ansprechender Form dargestellt werden. Zur übersichtlichen Darstellung hierarchischer Merkmale, die sehr unterschiedlich besetzt sind werden sog. „squarified Treemaps“ eingesetzt.

Darüber hinaus wurden die Daten mit Assoziations- und Regressionsanalysen, und allgemein auf Ungleichverteilungen hin untersucht. Dabei zeigte sich immer wieder, dass vermutlich systematische Fehler in den Daten eine Analyse erschweren und die Validität der Ergebnisse geschmälert wird.

Eine Zusammenfassung der Daten durch bekannte Kennzahlen der Instandhaltung soll dem Betreiber medizin-technischer Geräte einen Vergleich mit eigenen Kennzahlen ermöglichen und damit wertvolle Hinweise zur Beurteilung des Zustandes der eigenen Gerätebewirtschaftung geben. Darüber hinaus bietet die ausführliche Darstellung der Datenaufbereitung (insbes. der Normierung) dem Betreiber die Möglichkeit, die dazu nötigen Maßnahmen vom Arbeitsaufwand, von der Methodik und von dem zu erwartenden Gewinnen abzuschätzen.

### 5.3 Schlusswort

Die Arbeit hat am ungünstigen Ende angesetzt. Große Mühen bei der Analyse können nur in günstigen Fällen Fehler bei der Datenerhebung rückgängig machen. Die Hoffnung, dass die Fehler sich bei Betrachtung großer Datenbestände von selber oder mit etwas Hilfe wegmitteln, hat sich bei dem nicht überraschenden Vorhandensein systematischer Fehler leider nicht erfüllt. Es kann zwar nicht ausgeschlossen werden, dass es nicht doch Möglichkeiten gibt, die gemachten Fehler zu korrigieren – aber hier werden solche Möglichkeiten als unwahrscheinlich angesehen.

Bleibt also die Möglichkeit die Fehler bei der Erhebung zu vermeiden. Die Forderung, alles vollständig und fehlerfrei zu erfassen, ist illusorisch. Akzeptabel ist nur eine Beschränkung auf die notwendigsten Informationen. Doch was ist notwendig?

Bevor diese Frage beantwortet wird, soll hier ausdrücklich auf eine jetzt schon mögliche Schlussfolgerung hingewiesen werden: *Richtigkeit geht vor Vollständigkeit – d.h. besser wenige Informationen richtig erfassen als viele fehlerhaft.*

Doch, was sind nun wichtige Informationen? Das hängt von der Fragestellung ab, die man anhand der Daten beantworten möchte. Von den vielen Fragestellungen, die hier von Interesse sein könnten, soll hier die in der Einleitung aufgeworfene Frage

aufgegriffen werden: Wann wird die Gerätebewirtschaftung optimal betrieben, was ist dazu zu tun?

Optimal wäre eine Bewirtschaftung, die mit minimalem Aufwand einen maximalen Erfolg erzielt. Egal wie man Aufwand oder auch Erfolg im Einzelfall definiert, ist klar, dass zur Lösung des Optimierungsproblems sowohl Daten über die Aufwendungen wie auch über die erzielten Erfolge vorliegen müssen. Es zeigt sich aber, dass die Daten fast ausschließlich Informationen über die Aufwendungen enthalten und nicht über die erzielten Erfolge. Darum hier die zweite Schlussfolgerung: *Die Daten sollten nicht nur wiedergeben, welchen Aufwand der Nutzer für die Geräte betreiben muss, sondern auch den Nutzen der Geräte für den Betreiber.*

Da diese Informationen hier nicht vorlagen, kann ein Optimum leider prinzipiell nicht ermittelt werden. Damit bleibt das Ziel der optimalen Bewirtschaftung im Rahmen dieser Arbeit im Dunkeln. Das bedeutet jedoch nicht, dass keine Empfehlungen ausgesprochen werden können, in welche Richtung man sich bewegen sollte, um sich diesem Ziel zu nähern. Als Hinweise, wo Optimierungspotentiale verborgen sind, können sämtliche Hinweise auf Ungleichverteilungen gedeutet werden, von denen einige gefunden und eingehend untersucht worden sind. Entdeckte und verstandene Ungleichverteilungen geben einem die Möglichkeit, z.B zunächst seltene aber teure Maßnahmen zu identifizieren, um über möglichst wenig Maßnahmen an möglichst viel Einsparpotential zu gelangen. Damit lautet die dritte Schlussfolgerung: *Ungleichverteilungen finden, verstehen und sich an ihnen orientieren beim Handeln.* Zum Glück lassen sich Ungleichverteilungen auch in Gegenwart von Datenfehlern oft recht sicher ermitteln.

Darüber hinaus wurde in dieser Arbeit versucht, über die rein deskriptiven Verteilungsanalysen hinaus eine Untersuchung von kausalen Zusammenhängen vorzunehmen. Dazu wurden Assoziationsanalysen eingesetzt. Es ist ganz klar, dass auch diese Analysen nicht die Unterschiede zwischen Koinzidenzen, Korrelationen und kausalen Zusammenhängen überbrücken können. Dennoch liefern die Analysen wertvolle Informationen für höchstwahrscheinlich kausale Zusammenhänge. In den Assoziationsdiagrammen können darüber hinaus Cluster gefunden werden, die auf syntethische Merkmale hinweisen, die durchaus wertvolle Informationen beinhalten können. Diese Untersuchungen sollen jedoch folgenden Arbeiten vorbehalten bleiben.

## 5.4 Ausblick

Die Zukunftsaussichten von EDV-gestützter Gerätebewirtschaftung werden sehr positiv bewertet. Sie steht grundsätzlich konkurrenzlos da und ihre Voraussetzungen sind entweder schon erfüllt, oder es besteht kaum Zweifel daran, dass sie in Zukunft erfüllt sein werden, und schließlich bietet sie noch großes Entwicklungspotential. Diese Ausgangskonstellation, gepaart mit dem derzeitigen Handlungsdruck im Gesundheitswesen, lässt schon fast eine zu stürmische Weiterentwicklung erwarten.

Während sich umfangreiche Datenverwaltung mit Aktenordnern vom Standard zur Notlösung entwickelt, reift die EDV vom Experiment zum Standard. Andere Alternativen sind nicht in Sicht.

Schon erfüllte Voraussetzungen sind die nötige Hardware, Software und Methodik, wobei am ehesten Entwicklungen im Bereich der Software in nächster Zeit Veränderungen erwarten lassen. Hardware und Methodik sind in einem Maße entwickelt, dass

zukünftige Entwicklungen für die Gerätebewirtschaftung kaum relevant sein werden.

Voraussetzungen, die in Kürze erfüllt sein werden, sind die grundsätzlich elektronische Erfassung von Daten und die breite Akzeptanz bei den Nutzern. Wenn in Zukunft Daten erfasst werden, dann nur noch in Ausnahmefällen nicht direkt in elektronischer Form. Es steht eine Generation von Anwendern an, die mit einem Computer aufgewachsen ist. Die Akzeptanzprobleme der Vergangenheit werden in Zukunft keine Rolle mehr spielen.

Entwicklungspotential entsteht auch hier durch die viel beschworene Vernetzung und der damit allgegenwärtige Zugriff auf andere Datenbestände. Die Vernetzung von Datenbeständen hat ein so großes Entwicklungspotential, dass hier sogar noch Entwicklungen von leistungsfähigerer Hardware und Methodik notwendig werden kann. Aber dies wird dann nicht mehr im Rahmen der Gerätebewirtschaftung stattfinden.

Der Motor, der diese Entwicklung wesentlich mit antreibt, ist der momentane Handlungsdruck, im gesamten Gesundheitswesen zu rationalisieren. Gerade die letzten beiden Aspekte können durchaus auch als bedenklich eingestuft werden. Aber wie so oft sind auch in diesem Fall Mahnungen eher ein notwendiges Kriterium für erfreuliche Veränderungen als ein hinreichendes Kriterium für negative Veränderungen.

# A Das Anschreiben im Wortlaut

2

Wir werden Sie auch über die erzielten Ergebnisse selbstverständlich gerne unterrichten.

Die Daten dienen der Untermauerung eines Promotionsvorhabens.

Wir würden uns über diese Unterstützung unserer wissenschaftlichen Arbeiten sehr freuen.

Für Rückfragen stehe ich Ihnen jederzeit gerne zur Verfügung.

Mit freundlichen Grüßen und bestem Dank

Prof. Dr. C. Hartung

Anlage: TK 2000 Programm

2

**Medizinische Hochschule Hannover**  
Biomedizinische Technik  
und Krankenhaustechnik  
Zentrum Biometrie, Medizinische Informatik und Medizintechnik  
Prof. Dr. C. Hartung

M.H.H. - Biomedizinische Technik, CE 4400, D-30625 Hannover

Bitte Zeichen, Ihre Nachnamen  
Hilf/PI

Telefon (0511) 532-2710

Fax 18.07.00

**• Medizintechnik, Gerätelebensläufe**

Sehr geehrter Herr Verwaltungsdirektor,

im Rahmen der übergeordneten, von allen Häusern zu bewältigenden Aufgabe der ökonomischen Sicherung und Verbesserung einer sozialgerechten und qualitativ hochwertigen medizinischen Versorgung, beschäftigen wir uns seit Jahrzehnten mit Fragen der Sicherheit, Verfügbarkeit und Wirtschaftlichkeit der Krankenhaustechnik. Regelmäßig berichten wir über unsere Erkenntnisse u. a. auf der Ihnen sicherlich bekannten Veranstaltungreihe TK – Technik im Krankenhaus in der Medizinischen Hochschule Hannover (s.H. Anlage).

Derzeit sammeln wir Servicedaten medizinisch-technischer Geräte deutscher Krankenhäuser und analysieren diese, mit dem Ziel, Empfehlungen zu erarbeiten, wie die Performance der Geräte und des Services erhalten und verbessert werden können.

Hierdurch bitte ich Sie herzlich, unsere wissenschaftlichen Arbeiten zu unterstützen. Sollen in Ihrem Haus diese Daten computergestützt dokumentiert sein, so bitte ich Sie, diese auf einem Datenträger in PC-kompatiblen Format uns zur Verfügung zu stellen.

Wenn verfügbar, benötigen wir: u. a. Beschaffungsdatum, Beschaffungspreis, Finanzierungsart, Gerätetyp, Anzahl, Hersteller, hauptsächlichiger Einsatzort im Haus, Instandhaltungsmaßnahmen, geordnet nach Inspektion, Wartung, Instandsetzung, STK, MTK, Instandhaltungs- und Betriebskosten, Betriebsstunden, etwaigen Ausmusterungszeitpunkt, geplanter Ersatz, eigen bzw. fremd geservict unterschieden.

Sämtlich Daten werden aus Datenschutzgründen ggf. durch uns anonymisiert. Auch über die Herkunft der Daten werden wir Stillschweigen bewahren.

Lernfortschritt: Carl-Neuberg-Str. 1, D-30625 Hannover  
Sekretariat: Telefon (0511) 532-2449, Fax (0511) 531-4143

Abbildung A.1: Anschreiben im Wortlaut

# B Vollständige Regelsätze für Normierungen

## B.1 Merkmal Arbeitsbeschreibung

Angegeben sind hier für alle normierten Arbeitsbeschreibungen die zugeordneten Teilzeichenketten jeweils nach Häufigkeit absteigend sortiert. Die Häufigkeiten der Teilzeichenketten sind in Klammern angegeben.

**Änderung** \_inv\_ (11173), montier (9530), \_demont (4739), \_mont\_ (3881), \_inventarisier (3614), montage\_ (3381), \_eingebaut\_ (2985), \_inbetriebnahme\_ (2941), \_montage\_ (2347), \_installier (2002), \_anbringen\_ (1915), \_ausgebaut\_ (1525), \_umbau (1272), \_verlegt\_ (1078), \_verlegen\_ (955), \_umrüstung\_ (940), \_abgebaut\_ (706), \_ausgemustert\_ (696), \_ausmustern\_ (694), \_einbau\_ (687), \_aendern\_ (663), \_angebaut\_ (607), \_gestrichen\_ (538)

**Akku** \_akku (8328), \_batterie (2916), batt\_ (1848), \_accu\_ (1193)

**Antrieb** motor (5329), \_antrieb (2161)

**Austausch** erneuer (33043), tausch (28365), wechsel (19258), gew\_ (9337), \_ersatz (9105), \_ersetzt (5299)

**bedienungsbedingt** \_bedienungsbedingt (2272), \_sturzschaeden\_ (947)

**Elektrik** kabel (13336), schalter (9858), steckd (9703), sicherung (6998), kontakt (6494), elektr (6310), schaltet\_ (5375), spannung (3737), platine\_ (3585), taste (3496), \_unterbr (3258), netzteil (2728), schalten\_ (2666), \_kurzschluss\_ (2082), elstoe-rung\_ (1169), \_flackert\_ (861), \_relais\_ (777)

**Filter** \_filter (7128)

**Justage** just (6458), \_eingestellt\_ (4514), gerichtet\_ (4362), abgleich\_ (3492), \_kalibr (3345), eichung\_ (2859), \_abgeglichen\_ (1523), \_einstellen\_ (776), \_nachgestellt\_ (611)

**Licht** \_bel\_ (16494), lampe (14275), licht (13489), leucht (12130)

**Mechanik** \_mech (4037), gebrochen\_ (3908), \_schloss (3664), \_gehaeuse (2898), \_verschleiss\_ (2830), bruch\_ (2722), klemmt\_ (2566), schraube (2459), geris-sen\_ (2287), \_keilriemen (1417), \_bremse (1261), riemen\_ (1252), geloest\_ (1078), welle\_ (940), \_geraeusche\_ (809), \_rahmen\_ (750), verbindung\_ (698), \_schloesser\_ (635), \_abdeckung\_ (624), \_feder\_ (564)

**Prüfung** prue (103507), stk\_ (49366), \_stk (37889), ueberp (36158), kontr (28724),  
probe (8455), test (7832), \_funktionspr (6320), \_gepr\_ (2489)

**Pumpe** pump (11771)

**Reinigung** reinig (21949), \_verstopf (13215), \_gespueht\_ (2439), \_verschmutz (867)

**Reparatur** \_rep (456129), reparatur (352226), \_instand (8226)

**Schlauch** schlauch (4620), schlaeuche\_ (1155)

**Störung** \_def (124564), fehler (46934), stoer (42661), defekt (34667), \_geraetebe-  
dingt (9092), \_keine\_funktion\_ (4304), ausfall\_ (2787), \_ausgefallen\_ (2132),  
\_setzt\_aus\_ (2012), \_beschaedig (1777)

**Thermo** \_thermo (2468), \_warm (1816), \_temperatur (1562), \_temp\_ (986), \_waer-  
me (921)

**undicht** dicht (16756), \_tropft\_ (921), \_tropfen (863)

**Vakuum** \_vakuum (1064), \_vacuum (754)

**Ventil** ventil (8100)

**Wartung** wartung (60533), inspektion\_ (14827), \_ueberhol (12606), \_wartstk\_ (11670),  
\_wartg\_ (1966), \_wart\_ (1526), \_gewartet\_ (798)

## B.2 Merkmal Abteilung

Angegeben sind hier für alle normierten Abteilungen die zugeordneten Teilzeichenketten jeweils nach Häufigkeit absteigend sortiert. Die Häufigkeiten der Teilzeichenketten sind in Klammern angegeben.

abdominal \_abdom (489); Akademie \_akadem (460); allgemein \_allg (2454); Ambulanz ambu (1158), \_amb (1111); Anästhesie \_anae (7714); Anatomie anatomie\_ (1037); Angiologie \_angio (202); Apotheke apotheke\_ (453); aseptisch \_asep (215); Aufnahme aufnahme (647); Aufwachraum \_aufwach (173); Augenheilkunde \_augen (1284); Biochemie \_biochem (721); Biologie biol (2077); Biophysik \_biophys (573); Chemie chemi (2183), \_chem\_ (121); Chirurgie chir (7453); Dermatologie haut (821), \_dermatol (396); Diätetik \_diaet (245); Diagnostik diagn (1214); Dialyse dialyse (361); Endokrinologie \_endokr (587); Endoskopie endoskop (1074), \_endo\_ (261); Epidemiologie \_epidemiolog (123); Forschungforsch (563); Funktion \_funkt (421); Gastro \_gastro (1696); Gefäß \_gefaess (310); Gynäkologie \_gyn (1684), \_frauen (465); Hämatologie haem (1517), blut (1397); Herzkatheter \_herzkathe (312); HNO \_hno\_ (2374); Hörsaal \_hoersaal (104); human-genetik \_humangenetik\_ (228); Immunologie \_immun (892); innere \_innere\_ (3015); Intensiv \_intensiv\_ (2059); Kardiologie kard (2817), \_herz (1683); Kernspin \_kernspin (162); Kiefer \_kiefer (371); Kinderheilkunde kind (2921), paedi (1046), paed\_ (778); Kreissaal \_kreiss (536); Labor labor (7589), lab\_ (440); Medizintechnik \_medizintechnik\_ (409), \_medizin\_technik\_ (41); Mikrobiologie \_mikrobio (770); molekular] \_molek (1442); Neurologie \_neuro (2757); Notfall \_not (780); Nuklearmedizin \_nukl (3205); Onkologie \_onkol (753); OP] \_op\_ (8287), \_zop\_ (2056); Orthopädie \_orth (378); Pathologie patholog (2046); Pflege pflege (672); Pharmakologie pharma (2054); Physiologie physio (1919); Pneumologie \_pneumol (577); Poliklinik \_poli (3182); Psychiatrie \_psychiatri (530), \_psych\_ (361); Radiologie radiolog (1109), \_strahlen (868); Rechenzentrum \_rz\_ (1350), rechenzentrum\_ (182); Röntgen roent (2530); Schule \_lehr (794), schule (397); Sonographie \_sono (176); Sport sport (587); Sterilisation steri (1129); Technik techn (1743); Therapie therapie\_ (1471); Tier tier (641); Transplantation \_transpl (787); Transport transport (304); Unfall \_unfall (1436); Urologie urol (2487); Virologie \_virolog (475); Werkstatt werkst (333); Zahn \_zahn (1687), \_zmk\_ (356);

# C Gerätelisten der Häufigkeitsanalysen

Tabelle C.1: Auflistung der in Abbildungen 4.15 und 4.16 dargestellten Kaufpreise und der in den Abbildungen 4.5 und 4.6 dargestellten Werte – jedoch bezogen auf die Anzahl der Betten pro Gerät, Teil I.

EMTEC-Geräteart	Betten pro Gerät		mittlerer Kaufpreis [€]
	Univ.	sonst.	
<b>Stoffaustausch</b>	<b>0.48</b>	<b>1.18</b>	<b>1521</b>
Infusionsapparate	0.58	1.53	1467
Absauggeräte	3.77	6.78	1176
Bluttemperierungsgeräte	17.6	48.6	3309
Dialyse- und Blutfiltrationsgeräte	39.3	57.5	20046
Blutzuführungs / -entnahme / -aufarbeitungsgeräte	39.9	162	30778
Dialyse-Hilfsgeräte	638	1585	–
<b>Patientenüberwachungsgerät</b>	<b>0.47</b>	<b>1.62</b>	<b>5723</b>
Einzelparametermessung, OP / Intensiv	1.32	5.03	2917
Personen-Waagen	2.82	12.1	1056
Beatmungs- und Narkose-Monitore	2.68	12.4	2130
Patientenüberwachungsgeräte	7.23	15.4	8692
Kombinationsmonitore / -module	16.5	31.1	2674
Perinatale Überwachung	82.4	84.4	9066
Telemetrie-Anlagen	73.0	304	–
<b>Chirurgie / Endoskopie</b>	<b>0.32</b>	<b>1.95</b>	<b>8126</b>
Patientenuntersuchung,-transport,-lagerung	0.84	6.84	3172
Endoskope und Zubehör	2.25	7.64	9487
Leuchten für Untersuchung und Behandlung	4.25	12.6	3603
HF-, Wärme-, Kälte-Chirurgiegeräte	6.41	17.0	6168
Operationstische	6.94	26.4	20923
Stationszubehör, medizinisch	2.19	32.9	715
Decken-& Wandversorg.-und befestigungssysteme	16.9	111	16408
Stühle, medizinisch	5.35	129	3137
Chirurgie-Hilfsgeräte	23.0	132	26007
Lithotripter	98.2	344	35714
Laser-Chirurgiegeräte	116	566	–

Tabelle C.2: Auflistung der in Abbildungen 4.15 und 4.16 dargestellten Kaufpreise und der in den Abbildungen 4.5 und 4.6 dargestellten Werte – jedoch bezogen auf die Anzahl der Betten pro Gerät, Teil II.

EMTEC-Geräteart	Betten pro Gerät		mittlerer Kaufpreis [€]
	Univ.	sonst.	
<b>Vitalfunktion / Intensivmedizin</b>	<b>0.97</b>	<b>2.53</b>	<b>8813</b>
Beatmung und Narkose, Zusatzgeräte	2.97	6.47	1452
Inhalations-Narkosegeräte	3.14	9.33	25842
Beatmungsgeräte	3.96	11.0	–
Inkubatoren und Wärmegeräte	15.6	55.1	4296
Kreislaufbeeinflussungsgeräte	283	93.3	2428
Notfallausrüstung	47.3	108	3414
Herz-Lungen-Systeme	39.9	1585	–
Druckkammer	511	3964.5	–
<b>Funktionsdiagnostik</b>	<b>0.86</b>	<b>2.93</b>	<b>6664</b>
Blutdruckmeßgeräte	6.35	6.36	–
Temperaturmeßgeräte	17.0	23.9	422
Elektrokardiographen	13.5	24.7	5121
Neurologie-Meßgeräte	7.81	38.3	12630
Gefäßdiagnostik-Meßgeräte	12.2	42.2	4203
Ergometer	22.8	60.5	3247
Lungenfunktions-Meßgeräte	21.8	96.7	5517
Funktionsdiagnostik-Aufzeichnungsgeräte	1.86	114	2890
Vitalitätsprüfgeräte	182	283	–
Urologie-Meßgeräte	127	495	6247
Stethoskope	319	2643	–
Haut-Meß- und Prüfgeräte	638	3964	–
<b>Bildgebendes System</b>	<b>0.82</b>	<b>4.12</b>	<b>105786</b>
Bildbetrachtungseinrichtungen	2.20	11.2	915
Ultraschall-Diagnosegeräte	5.00	11.7	63311
Röntgen-Aufnahmeggeräte	3.16	52.2	93045
Systemkomponenten, Röntgen	11.5	176	948330
Röntgenfilmentwicklungsmaschinen	21.6	184	21607
Durchleuchtungsgeräte	0.00	184	87592
Röntgen-Generatoren	19.9	293	307980
Nuklear-Diagnosegeräte	53.2	720	254699
Bildaufzeichnungs- und Dokumentationseinr.	159	1321	–
Kernspin-Tomographen	46.4	2643	–

Tabelle C.3: Aufzählung der in Abbildungen 4.15 und 4.16 dargestellten Kaufpreise und der in den Abbildungen 4.5 und 4.6 dargestellten Werte – jedoch bezogen auf die Anzahl der Betten pro Gerät, Teil III.

EMTEC-Geräteart	Betten pro Gerät		mittlerer Kaufpreis [€]
	Univ.	sonst.	
<b>Therapiegerät (ohne OP-Tech. &amp; Strahlenther.)</b>	<b>1.69</b>	<b>3.58</b>	<b>3282</b>
Kardiologie-Notfallgeräte	6.20	13.3	6395
Bestrahlungsgeräte (nicht radiologisch)	6.90	22.6	1261
Physiotherapiegeräte	12.6	23.4	2370
Sauerstoff-Therapiegeräte	23.0	27.2	923
Stimulationsgeräte Nerven / Muskeln, Therapie	15.6	30.5	2779
Elektro-Therapiegeräte	15.4	32.0	2666
Wärme-therapiegeräte / Diathermiegeräte	29.7	62.4	2439
<b>Laborgerät</b>	0.11	3.22	3384
Probenvorbereitung / -aufbereitung / -transport	<b>0.27</b>	<b>9.18</b>	<b>631</b>
Mikroskope	0.79	23.5	4961
Labor-Zentrifugen	0.99	25.6	5051
Analysatoren, einfach	1.56	38.1	2516
Analysatoren, Hämatologie	6.27	44.3	20782
Apothekengeräte	2.24	48.6	5571
Photometer	2.59	70.2	8069
Dosiergeräte	6.88	95.5	4028
Ausstattung und Zubehör, Labor	2.31	121	2256
Pathologie- und Histologie-Laborgeräte	5.17	220	14026
Meßgeräte, Einzelmessung, Labor	4.16	283	707
Analysatoren, Klinische Chemie	212	1585	–
Spektroskopiegeräte	28.7	1982	34404
Analysatoren, Nuklear-Labor	27.1	2643	25280
Molekularbiologische Laboreinrichtung	111	217444	434863
<b>Desinfektion, Sterilisation</b>		<b>25.3</b>	<b>11082</b>
Sterilisationsgeräte	3.92	36.7	16653
Desinfektionsgeräte	10.3	80.9	5625
<b>Fachärztliches Gerät</b>	<b>2.82</b>	<b>34.2</b>	<b>25416</b>
HNO-Geräte	27.4	55.1	7934
Ophthalmologische Geräte	7.32	99.1	9611
Dentaltechnische Geräte	22.0	7929	2119
<b>Medizinische Physik / Strahlenschutz</b>	<b>19.5</b>	<b>345</b>	<b>8365</b>

# D Hauptanteile je MT-Untergerätegruppen für verschiedene Merkmale

Tabelle D.1: Auflistung der Hauptanteile der Gerätezahlen je MT-Untergeräteart aus Abbildungen 4.5 und 4.6.

EMTEC-Geräteart	relative Gerätezahl	kumuliert
<b>Universitätskliniken</b>		
Probenvorbereitung / -aufbereitung / -transport	17.70 %	17.70 %
Infusionsapparate	8.15 %	25.86 %
Mikroskope	6.03 %	31.90 %
Patientenuntersuchung,-transport,-lagerung	5.65 %	37.56 %
Labor-Zentrifugen	4.80 %	42.37 %
Einzelparametermessung, OP / Intensiv	3.59 %	45.97 %
Analysatoren, einfach	3.05 %	49.02 %
Funktionsdiagnostik-Aufzeichnungsgeräte	2.56 %	51.59 %
⋮	⋮	⋮
<b>nicht Universitätskliniken</b>		
Infusionsapparate	17.61 %	17.61 %
Einzelparametermessung, OP / Intensiv	5.12 %	22.74 %
Inhalations-Narkosegeräte	4.52 %	27.26 %
Beatmungsgeräte	4.26 %	31.53 %
Labor-Zentrifugen	3.46 %	35.00 %
Sterilisationsgeräte	3.38 %	38.38 %
Endoskope und Zubehör	3.21 %	41.59 %
Röntgen-Aufnahmeggeräte	3.13 %	44.73 %
Patientenuntersuchung,-transport,-lagerung	2.95 %	47.68 %
Röntgenfilmentwicklungsmaschinen	2.32 %	50.00 %
⋮	⋮	⋮

Tabelle D.2: Aufistung der Hauptanteile der Kaufpreissummen je MT-Untergeräteart aus Abbildungen 4.9 und 4.10.

EMTEC-Geräteart	relative Kaufpreissummen	kumuliert
<b>Universitätskliniken</b>		
Kernspin-Tomographen	18.93 %	18.93 %
Ultraschall-Diagnosegeräte	7.66 %	26.60 %
Röntgen-Generatoren	5.28 %	31.88 %
Röntgen-Aufnahmeggeräte	4.86 %	36.74 %
Mikroskope	4.43 %	41.17 %
Probenvorbereitung / -aufbereitung / -transport	2.98 %	44.16 %
Labor-Zentrifugen	2.83 %	46.99 %
Analysatoren, einfach	2.55 %	49.55 %
Inhalations-Narkosegeräte	2.09 %	51.65 %
⋮	⋮	⋮
<b>nicht Universitätskliniken</b>		
Kernspin-Tomographen	7.73 %	7.73 %
Ultraschall-Diagnosegeräte	6.83 %	14.57 %
Röntgen-Generatoren	5.73 %	20.31 %
Röntgen-Aufnahmeggeräte	5.43 %	25.74 %
Mikroskope	4.01 %	29.76 %
Inhalations-Narkosegeräte	3.67 %	33.44 %
Photometer	2.40 %	35.84 %
Systemkomponenten, Röntgen	2.30 %	38.14 %
Beatmungsgeräte	2.29 %	40.44 %
Infusionsapparate	2.22 %	42.66 %
Labor-Zentrifugen	2.10 %	44.77 %
Endoskope und Zubehör	2.00 %	46.78 %
Operationstische	1.94 %	48.72 %
Patientenuntersuchung,-transport,-lagerung	1.89 %	50.61 %
⋮	⋮	⋮



Tabelle D.4: Auflistung der Hauptanteile der Kostensummen je MT-Untergeräteart aus Abbildungen 4.25 und 4.26.

EMTEC-Geräteart	relative Kosten- summen	kumuliert
<b>Universitätskliniken</b>		
Einzelparametermessung, OP / Intensiv	20.10 %	20.10 %
Infusionsapparate	11.76 %	31.86 %
Beatmungsgeräte	7.77 %	39.64 %
Inkubatoren und Wärmegeräte	4.51 %	44.16 %
Apothekengeräte	4.15 %	48.31 %
Dialyse- und Blutfiltrationsgeräte	4.09 %	52.41 %
⋮	⋮	⋮
<b>nicht Universitätskliniken</b>		
Infusionsapparate	12.41 %	12.41 %
Einzelparametermessung, OP / Intensiv	11.95 %	24.36 %
Beatmungsgeräte	6.41 %	30.77 %
Dialyse- und Blutfiltrationsgeräte	5.33 %	36.11 %
Röntgen-Aufnahmeggeräte	4.09 %	40.20 %
Apothekengeräte	3.79 %	43.99 %
Ultraschall-Diagnosegeräte	3.78 %	47.78 %
Beatmung und Narkose, Zusatzgeräte	3.55 %	51.34 %
⋮	⋮	⋮

# E Bemerkungen zu Regression der Kaufpreise

## E.1 Johnson Transformationen

Oft liegen Daten vor, die durch klassische Verteilungsfunktionen nur mit unververtretbarem Fehler approximiert werden können. Um dennoch die Daten ausreichend gut durch eine analytisch darstellbare Funktion anzunähern, erweitert man die klassischen Verteilungsfunktionen durch ein Verteilungsfunktion-System. Hier wird das Johnson-System betrachtet ([15] Seite 164).

Dieses System entsteht durch Anwendung von drei Transformationen auf die Normalverteilung  $N(\mu, \sigma)$ , um diese an verschiedene Anforderungen anzupassen. Sei  $Z \sim N(\mu', \sigma)$  eine Normalverteilte Zufallsvariable, dann schreiben sich die Transformationen als

- $S_L$  (lognormal) Transformation:  $Y = e^{\frac{Z-\gamma}{\delta}}$
- $S_U$  (unbegrenzt) Transformation:  $Y = \sinh\left(\frac{Z-\gamma}{\delta}\right)$
- $S_B$  (begrenzt) Transformation:  $Y = \frac{1}{1+e^{\frac{Z-\gamma}{\delta}}}$

Die Wirkungen dieser Transformationen lassen sich gut im  $(\beta_1, \beta_2)$ -Raum darstellen. Für die Parameter  $\beta_1$  und  $\beta_2$  gilt

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \quad \text{und} \quad \beta_2 = \frac{\mu_4}{\mu_2^2} \quad (\text{E.1})$$

Dabei sind die  $\mu_i$  die i. zentralen Momente einer Zufallsvariablen  $X$ , für die gelten

$$\mu_i = E((X - \mu)^i) \quad \text{mit} \quad \mu = E(X) \quad (\text{E.2})$$

$E$  ist hierbei der Erwartungswert (siehe Abschnitt 2.3.1) und  $\mu$  wird als Mittelwert bezeichnet.  $\sqrt{\beta_1}$  werden als Schiefe (Skewness) und  $\beta_2$  als Wölbung (Kurtosis) bezeichnet.

Die Parameter  $\gamma$  und  $\delta$  der Johnson Transformationen sowie die Parameter  $\mu'$  und  $\sigma$  der Normalverteilung lassen sich aus den Größen  $\mu$ ,  $\mu_2$ ,  $\mu_3$  und  $\mu_4$  bestimmen. Für eine genauere Darstellung sei hier auf [15] Seite 164 verwiesen.

In Abbildung E.1 sind die Gebiete, die sich mit den Transformationen  $S_B$ ,  $S_L$  und  $S_U$  erreichen lassen im  $(\beta_1, \beta_2)$ -Raum dargestellt. Beispielhaft sind auch die empirischen Verteilungen der Kaufpreise von Geräten aus verschiedenen Krankenhäusern als Punkte eingezeichnet. Es ist deutlich zu erkennen, dass die empirischen Kaufpreis-Verteilungen im  $S_B$ -Bereich zu finden sind.

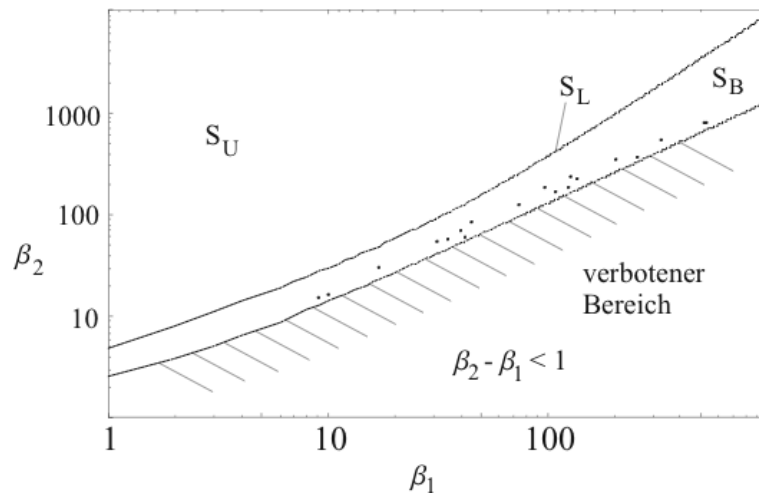


Abbildung E.1: Darstellung im  $(\beta_1, \beta_2)$ -Raum der Bereiche, die sich durch Johnson Transformationen  $S_B$ ,  $S_L$  und  $S_U$  erreichen lassen. Beispielhaft sind Kaufpreis-Verteilungen von verschiedenen Krankenhäusern als Punkte eingezeichnet. Deutung siehe Text.

## E.2 Problematische Kaufpreisverteilungen

In Abbildung E.2 ist beispielhaft die empirische Kaufpreis-Verteilung von Blutdruckmessgeräten dargestellt. Man kann deutlich erkennen, dass eine Regression durch eine log-Normalverteilung – wenn überhaupt numerisch möglich – mit sehr hohen Fehlern behaftet sein wird. Die globale Abweichung von der log-Normalverteilung kann dadurch erklärt werden, dass die MT-Untergeräteart Blutdruckmessgerät in Hinblick auf den Kaufpreis eine Mischung von zwei unterschiedlichen Gerätegruppen ist. An den gestrichelten Hilfslinien lässt sich ablesen, dass die Grenze zwischen den Gerätegruppen etwa bei 1000 € verläuft. Je Gruppe liegt dann wieder eine log-Normalverteilung der Kaufpreise vor. Außerdem sind die Häufungspunkte und Anteile der beiden Gerätegruppen eingezeichnet.

Eine ähnliche Konstellation lässt sich auch bei der MT-Untergeräteart Beatmungsgeräte finden. Hier liegt die Grenze zwischen den Gruppen bei 10 000 €. Der Bestand wird auf beide Gruppen etwa zur Hälfte aufgeteilt. Die Häufungspunkte liegen bei 3000 € und 60 000 €.

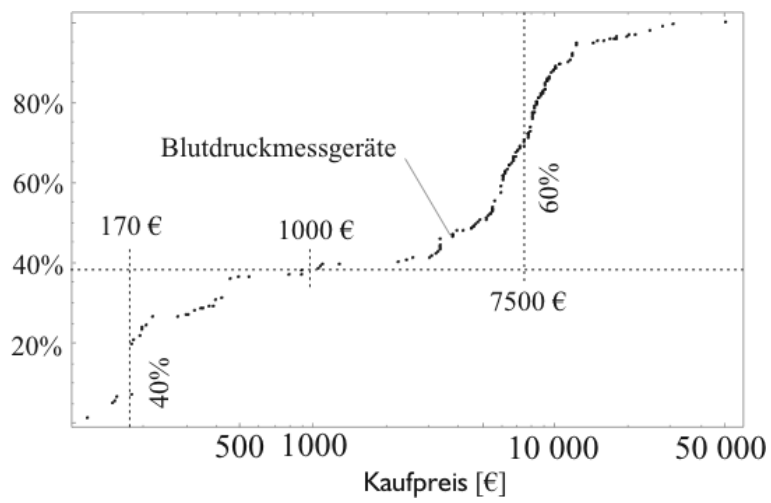


Abbildung E.2: Dargestellt ist beispielhaft die empirische Verteilungsfunktion der Kaufpreise von Blutdruckmessgeräten. Deutung siehe Text.

## F Maßnahmen in Abhängigkeit der Betriebszeit je Geräteart

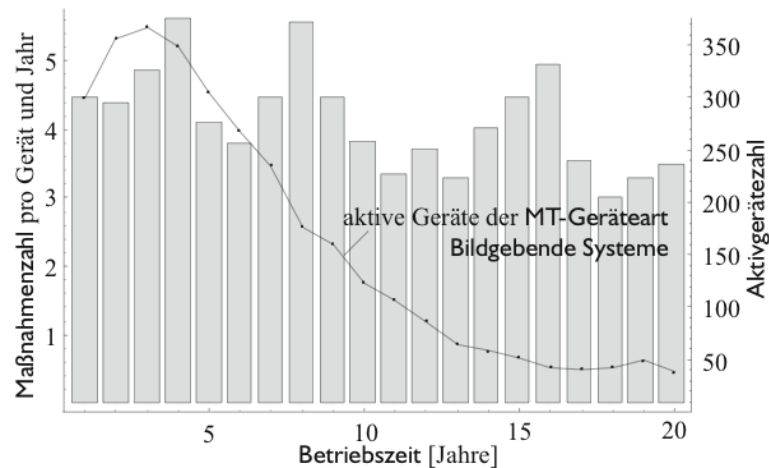


Abbildung F.1: Darstellung der Aktivgerätezahl der Geräteart Bildgebende Systeme in Abhängigkeit der Betriebszeit als Linie und Maßnahmenzahl pro Gerät und Jahr auch in Abhängigkeit der Betriebszeit als Balkendiagramm. Erklärung siehe Abschnitt 4.4.4.

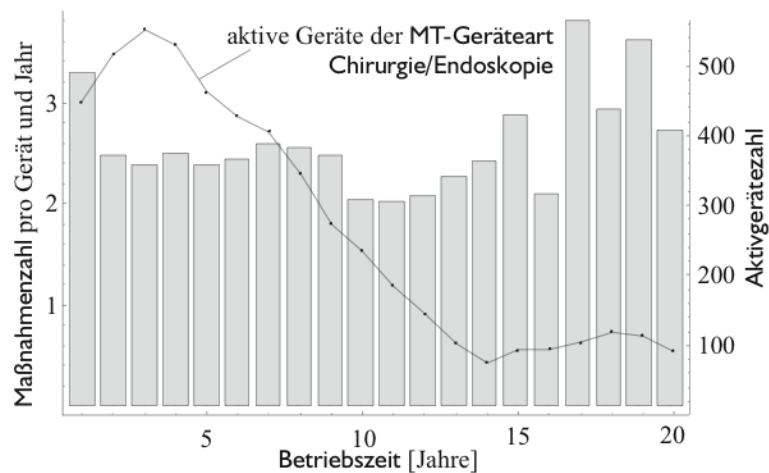


Abbildung F.2: Darstellung der Aktivgerätezahl der Geräteart Chirurgie / Endoskopie in Abhängigkeit der Betriebszeit als Linie und Maßnahmenzahl pro Gerät und Jahr auch in Abhängigkeit der Betriebszeit als Balkendiagramm. Erklärung siehe Abschnitt 4.4.4.

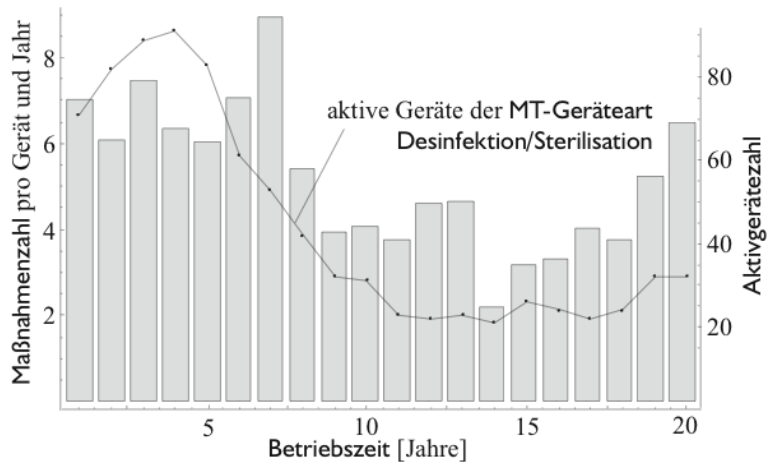


Abbildung F.3: Darstellung der Aktivgerätezahl der Geräteart Desinfektion / Sterilisation in Abhängigkeit der Betriebszeit als Linie und Maßnahmenzahl pro Gerät und Jahr auch in Abhängigkeit der Betriebszeit als Balkendiagramm. Erklärung siehe Abschnitt 4.4.4.

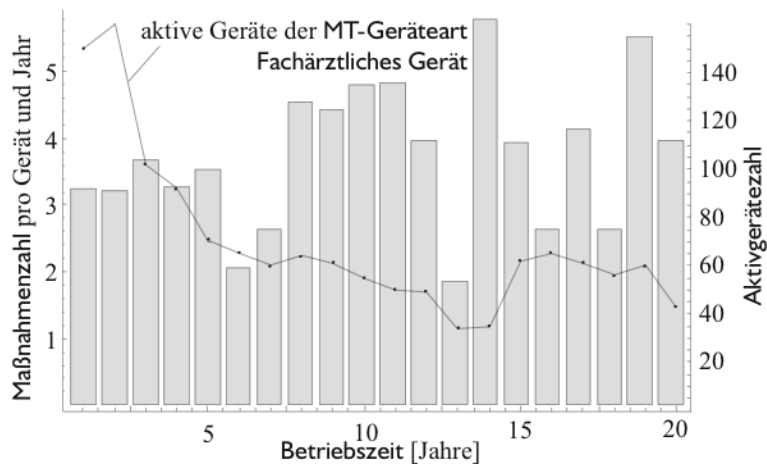


Abbildung F.4: Darstellung der Aktivgerätezahl der Geräteart Fachärztliches Gerät in Abhängigkeit der Betriebszeit als Linie und Maßnahmenzahl pro Gerät und Jahr auch in Abhängigkeit der Betriebszeit als Balkendiagramm. Erklärung siehe Abschnitt 4.4.4.

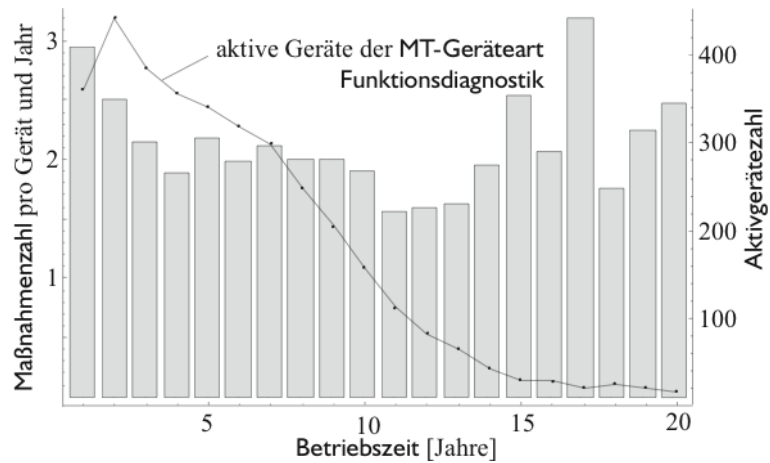


Abbildung F.5: Darstellung der Aktivgerätezahl der Geräteart Funktionsdiagnostik in Abhängigkeit der Betriebszeit als Linie und Maßnahmenzahl pro Gerät und Jahr auch in Abhängigkeit der Betriebszeit als Balkendiagramm. Erklärung siehe Abschnitt 4.4.4.

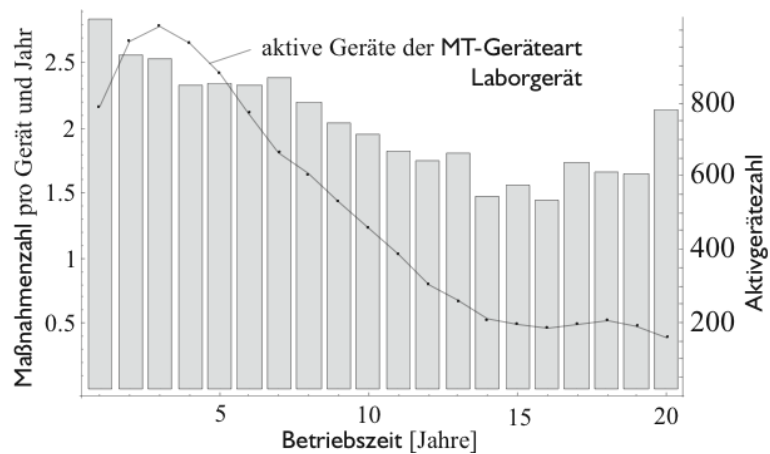


Abbildung F.6: Darstellung der Aktivgerätezahl der Geräteart Laborgerät in Abhängigkeit der Betriebszeit als Linie und Maßnahmenzahl pro Gerät und Jahr auch in Abhängigkeit der Betriebszeit als Balkendiagramm. Erklärung siehe Abschnitt 4.4.4.

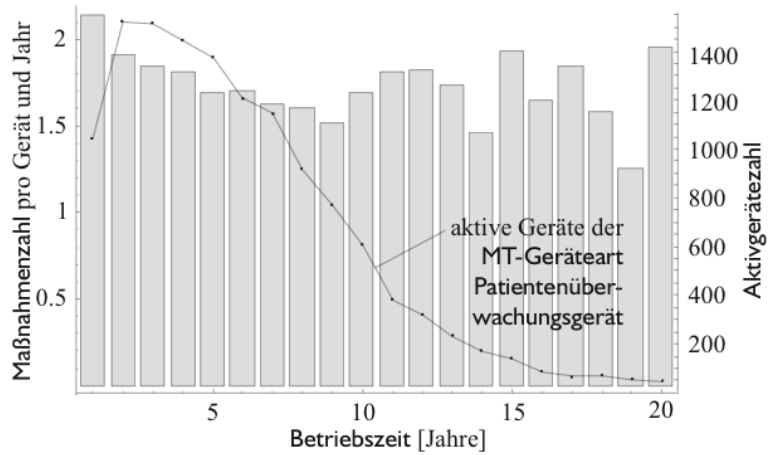


Abbildung F.7: Darstellung der Aktivgerätezahl der Geräteart Patientenüberwachungsgerät in Abhängigkeit der Betriebszeit als Linie und Maßnahmenzahl pro Gerät und Jahr auch in Abhängigkeit der Betriebszeit als Balkendiagramm. Erklärung siehe Abschnitt 4.4.4.

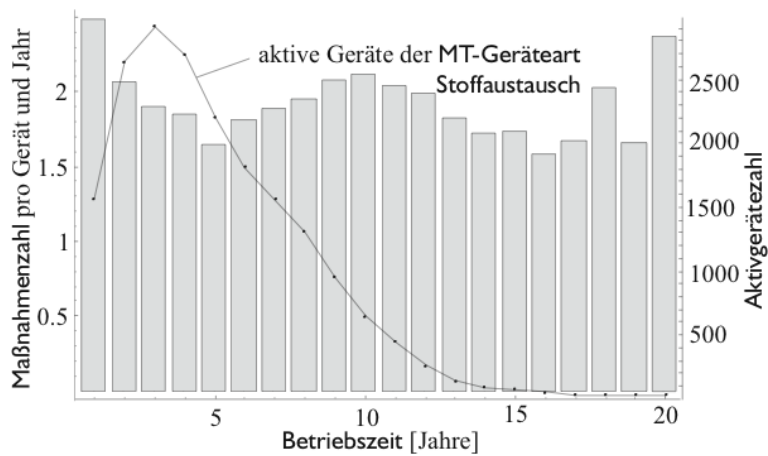


Abbildung F.8: Darstellung der Aktivgerätezahl der Geräteart Stoffaustausch in Abhängigkeit der Betriebszeit als Linie und Maßnahmenzahl pro Gerät und Jahr auch in Abhängigkeit der Betriebszeit als Balkendiagramm. Erklärung siehe Abschnitt 4.4.4.

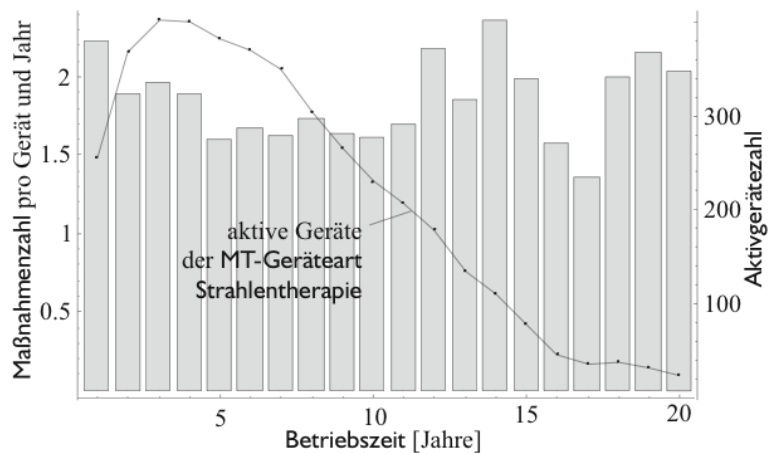


Abbildung F.9: Darstellung der Aktivgerätezahl der Geräteart Strahlentherapie in Abhängigkeit der Betriebszeit als Linie und Maßnahmenzahl pro Gerät und Jahr auch in Abhängigkeit der Betriebszeit als Balkendiagramm. Erklärung siehe Abschnitt 4.4.4.

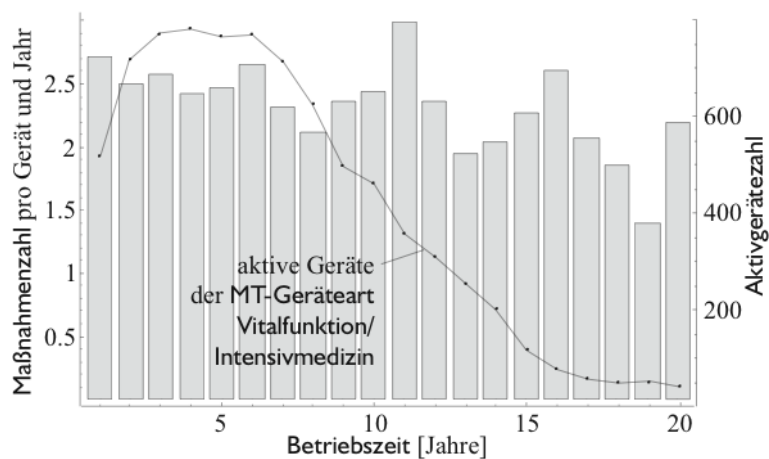


Abbildung F.10: Darstellung der Aktivgerätezahl der Geräteart Vitalfunktion / Intensivmedizin in Abhängigkeit der Betriebszeit als Linie und Maßnahmenzahl pro Gerät und Jahr auch in Abhängigkeit der Betriebszeit als Balkendiagramm. Erklärung siehe Abschnitt 4.4.4.

# G Vollständige Gerätelisten der ABC-Analyse

Zur übersichtlicheren Darstellung der in Abbildungen 4.54 und 4.55 dargestellten Werte sind diese in den Tabellen G.2, G.2 und G.2 aufgelistet.

Tabelle G.1: Auflistung der in Abbildungen 4.54 und 4.55 dargestellten Werte, Teil I.

EMTEC-Geräteart	Kosten pro Maßnahme [€]	Maßnahmen pro Jahr
<b>Bildgebendes System</b>	<b>1855</b>	<b>4.42</b>
Röntgen-Aufnahmegерäte	1121	5.58
Durchleuchtungsgeräte	1104	2.72
Systemkomponenten, Röntgen	1167	2.23
Röntgen-Generatoren	3803	4.26
Röntgenfilmentwicklungsmaschinen	1022	8.62
Ultraschall-Diagnosegeräte	3746	2.77
Nuklear-Diagnosegeräte	2288	3.96
Kernspin-Tomographen	922205	17.74
Bilddarstellungseinrichtungen	269	1.22
<b>Chirurgie / Endoskopie</b>	<b>450</b>	<b>1.43</b>
HF-, Wärme-, Kälte-Chirurgiegeräte	95	1.65
Laser-Chirurgiegeräte	563	2.24
Chirurgie-Hilfsgeräte	65	2.06
Lithotripter	55880	5.26
Operationstische	644	1.61
Decken- & Wandversorg.- und befestigungssysteme	1040	2.67
Stühle, medizinisch	50	2.27
Leuchten für Untersuchung und Behandlung	48	1.49
Endoskope und Zubehör	1451	2.63
Patientenuntersuchung,-transport,-lagerung	155	0.55
Stationszubehör, medizinisch	49	0.94
<b>Desinfektion, Sterilisation</b>	<b>163</b>	<b>6.23</b>
Desinfektionsgeräte	253	4.60
Sterilisationsgeräte	155	6.83

Tabelle G.2: Auflistung der in Abbildungen 4.54 und 4.55 dargestellten Werte, Teil II.

EMTEC-Geräteart	Kosten pro Maßnahme [€]	Maßnahmen pro Jahr
<b>Fachärztliches Gerät</b>	<b>102</b>	<b>2.97</b>
Ophthalmologische Geräte	53	1.63
HNO-Geräte	364	1.69
Dentalmedizinische Geräte	248	%
Dentaltechnische Geräte	617	1.23
<b>Funktionsdiagnostik</b>	<b>324</b>	<b>1.89</b>
Neurologie-Meßgeräte	228	2.36
Elektrokardiographen	415	2.43
Blutdruckmeßgeräte	137	1.53
Gefäßdiagnostik-Meßgeräte	473	%
Temperaturmeßgeräte	33	1.79
Ergometer	198	1.19
Lungenfunktions-Meßgeräte	307	1.69
Urologie-Meßgeräte	2380	%
Funktionsdiagnostik-Aufzeichnungsgeräte	316	%
<b>Laborgerät</b>	<b>461</b>	<b>1.97</b>
Labor-Zentrifugen	287	1.46
Probenvorbereitung / -aufbereitung / -transport	153	1.36
Dosiergeräte	154	6.66
Ausstattung und Zubehör, Labor	69	1.27
Apothekengeräte	398	2.96
Analysatoren, einfach	1372	2.72
Chromatographiegeräte	1245	2.23
Analysatoren, Hämatologie	973	5.06
Mikroskope	567	1.27
Photometer	633	1.82
Analysatoren, Klinische Chemie	1112	5.03
Meßgeräte, Einzelmessung, Labor	222	1.55
Molekularbiologische Laboreinrichtung	3832	3.72
Spektroskopiegeräte	1295	2.32
Analysatoren, Nuklear-Labor	1323	1.87
Pathologie- und Histologie-Laborgeräte	510	1.93

Tabelle G.3: Auflistung der in Abbildungen 4.54 und 4.55 dargestellten Werte, Teil III.

EMTEC-Geräteart	Kosten pro Maßnahme [€]	Maßnahmen pro Jahr
<b>Patientenüberwachungsgerät</b>	<b>166</b>	<b>1.63</b>
Patientenüberwachungsgeräte	56	2.63
Einzelparametermessung, OP / Intensiv	164	1.49
Beatmungs- und Narkose-Monitore	129	1.54
Kombinationsmonitore / -module	227	1.25
Telemetrie-Anlagen	691	3.43
Perinatale Überwachung	341	3.10
Personen-Waagen	84	1.13
<b>Stoffaustausch</b>	<b>82</b>	<b>1.55</b>
Dialyse- und Blutfiltrationsgeräte	505	5.08
Infusionsapparate	58	1.44
Absauggeräte	48	1.57
Blutzuführungs / -entnahme / -aufarbeitungsgeräte	1287	2.64
Bluttemperierungsgeräte	55	1.69
<b>Therapiegerät (ohne OP-Tech. &amp; Strahlenther.)</b>	<b>117</b>	<b>4.03</b>
Kardiologie-Notfallgeräte	196	1.40
Elektro-Therapiegeräte	50	1.40
Wärmetherapiegeräte / Diathermiegeräte	336	2.32
Bestrahlungsgeräte (nicht radiologisch)	66	1.50
Physiotherapiegeräte	49	1.30
Stimulationsgeräte Nerven / Muskeln, Therapie	54	1.57
Sauerstoff-Therapiegeräte	227	1.04
<b>Vitalfunktion / Intensivmedizin</b>	<b>262</b>	<b>1.52</b>
Beatmungsgeräte	317	2.84
Inhalations-Narkosegeräte	229	2.23
Beatmung und Narkose, Zusatzgeräte	80	1.73
Inkubatoren und Wärmegeräte	245	1.71
Herz-Lungen-Systeme	1293	1.62

# Literaturverzeichnis

- [1] Vossen G; Datenmodelle, Datenbanksprachen und Datenbankmanagementsysteme; 4. Auflage; München, Wien; Oldenbourg 2000 [2.1](#), [2.1.1](#)
- [2] Hippner, Küsters, Meyer, Wilde (Hrsg.); Handbuch Data Mining im Marketing; 1. Auflage; Braunschweig, Wiesbaden; Vieweg 2001 [2.2](#), [2.2](#), [2.3.4](#)
- [3] Sachs L; Angewandte Statistik; 9. Auflage; Berlin, Heidelberg, New York; Springer 1999 [2.3.1](#), [2.1](#), [2.3.1](#), [2.3.1](#), [2.3.2](#), [2.3.2](#), [2.3.3](#), [2.3.3](#), [2.3.3](#)
- [4] Hartung J; Multivariate Statistik; 6. Auflage; München, Wien; Oldenbourg 1999 [2.3.3](#), [3.6.2](#)
- [5] Bronstein, I N, Semendjajew, K A, Zeidler (Hrsg.); Teubner-Taschenbuch der Mathematik; Teil 1; Stuttgart, Leipzig; Teubner 1996 [2.3.2](#), [2.3.5](#)
- [6] Zipf GK; Human behavior and the principle of least effort : an introduction to human ecology; Repr., facs. of 1949 ed.; New York (u.a.); Hafner 1965 [2.3.6](#)
- [7] Warnecke H J; Instandhaltung, Grundlagen; Köln; TÜV Rheinland GmbH; 1981 [2.4.1](#), [2.4](#), [2.4.3](#), [2.4.3](#), [5.1.1](#), [5.1.3](#)
- [8] Nippa J; Praktische Kennzahlen für das Instandhaltungsmanagement; mt-Medizintechnik 121 (2001), Nr. 4, S. 139ff [2.4.2](#), [5.1.3](#)
- [9] Adler N; Instandhaltung versus investiver Maßnahmen, aus: Hartung C (Hrsg.): Sanierung und Erneuerung technischer Anlagen im Krankenhaus; Tagungsband; Hannover; Fachverlag Krankenhaustechnik; 1985 [2.4.2](#)
- [10] Knicker N; Instandhaltung versus investiver Maßnahmen, aus: Hartung C (Hrsg.): Sicherung und Verbesserung der technischen Qualität im Krankenhaus; Tagungsband; Hannover; Fachverlag Krankenhaustechnik; 1985 [2.4.2](#)
- [11] Bruls M, Huizing K, Wijk J J v; Squarified Treemaps. Proc. of VisSym 20000 (May 2931, Amsterdam, The Netherlands; Wien, New York; Springer 2000 [2.5.1](#), [2.5.1](#)
- [12] Hartung J; Statistik; 6. Auflage; München, Wien; Oldenbourg 2002 [2.5.3](#)
- [13] Schaich E; Lorenzkurve und Gini-Koeffizient in kritischer Betrachtung; Jahrbücher für Nationalökonomie und Statistik; 185 (1971); S.193-298 [2.5.3](#)
- [14] Rapp R; Text-Detektor, Fehlertolerantes Retrieval ganz einfach; c't (Computer Zeitschrift); Ausgabe 4/1997; Seite 386 [3.5.3](#)

- [15] Rose C, Smith M D; Mathematical statistics with Mathematica; New York, Berlin, Heidelberg; Springer Verlag 2001 [E.1](#), [E.1](#)
- [16] Vollmuth H J; Controlling-Instrumente von A-Z; 5. erw. Auflage; Planegg / München; WRS Verlag; 2000 [4.5.1](#)
- [17] Montemurro M A; Beyond the Zipf-Mandelbrot law in quantitative linguistics; <http://arXiv.org/abs/cond-mat/0104066> [4.7.1](#)
- [18] Briolini A; Zuverlässigkeit von Geräten und Systemen; 4. Auflage; Berlin, Heidelberg, New York (u.a.); Springer Verlag; 1997 [5.1.1](#)