

# Interpretation und Fehl-Interpretation bei der Auswertung grosser Routine- Datenbanken

Karl P. Pfeiffer  
Department für Med. Statistik,  
Informatik und Gesundheitsökonomie  
Med. Universität Innsbruck

9/30/2004

Interpretation  
Routinedatenbanken

1

## Inhalt

- Datenbasis
- Fehlerquellen bei der Berechnung von Kennzahlen
- Fehlende Informationen
- Das Gesetz der "grossen Zahl" und statistische Signifikanz
  - Sehr kleine Konfidenzintervalle
- Datenqualität
- Datenzweck

9/30/2004

Interpretation  
Routinedatenbanken

2



# Datenbasis

- MBDS der stationären Aufenthalte in Österreich
  - Jährlich ca. 2.200.000 Datensätze
- TISS- Daten bei Intensivaufenthalt, täglich
- SAPS-Daten bei Aufnahme in die Intensivstation
- Bevölkerungsstatistik Österreich

9/30/2004

Interpretation  
Routinedatenbanken

3



# Hypothese

- Der **Minimale Basis Datensatz (MBDS)** bildet das medizinische Leistungsgeschehen im stationären Bereich optimal ab
- Die kodierten Daten können auch für **medizinische Zwecke** genutzt werden
  - Es handelt sich um medizinische Inhalte
- Die Daten können für das **Controlling und für die Planung, Steuerung** und Versorgungsforschung unproblematisch verwendet werden
  - Es gibt keine Verzerrungen durch den Erhebungszweck

9/30/2004

Interpretation  
Routinedatenbanken

4

# Alternative



- **Dokumentiert wird was finanziert wird!**
  - Datenqualität ist unwesentlich – Geld zählt!
- **Kodierung ist administrative Dokumentation**
  - und keine medizinische!
    - Kodierung hat nur wenig mit dem Krankheitsbild zu tun
    - Krankheitsbilder können nur durch Freitext ausreichend beschrieben werden!
- **Die Verwendung der DRG-Daten für andere Zwecke als Finanzierung führt zu falschen Ergebnissen**
  - Epidemiologie, Leistungsplanung etc. brauchen andere Daten

9/30/2004

Interpretation  
Routinedatenbanken

5

# Verzerrung von Kennzahlen durch:

- Zweck der Dokumentation
- Qualität und Granularität der Kataloge
- Tagesklinische Aufnahmen
- Fallsplitting
- Transfer
- Wiederaufnahme
- Intensivaufenthalte
- Wochentag der Aufnahme und Entlassung
- Schweregrad
  - Versorgungsauftrag
- Angebot / Auslastung
  - Fehlende Informationen über nicht-stationären Bereich
- **Datenqualität**
  - **Überdokumentation**



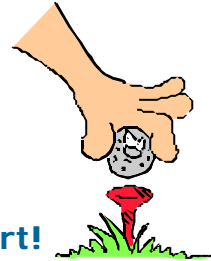
9/30/2004

Interpretation  
Routinedatenbanken

6

# Wie detailliert müssen Kennzahlen sein?

- Welche Einflussfaktoren müssen berücksichtigt werden:
  - Geschlecht
  - Alter
  - Diagnose, Leistung, DRG
  - **Krankenhaus**
    - Krankenhaustyp
  - **Region**
- **Globale Kennzahlen sind verzerrt!**
  - Asymmetrische und mehrgipfelige Verteilungen
  - Ausreisser
    - Der zentrale Grenzwertsatz ist nicht anwendbar!



9/30/2004

Interpretation  
Routinedatenbanken

7

# Statistische Kennzahlen

- **Robustheit der Kennzahlen?**
  - Welche Rolle spielen Ausreisser bei grossen Fallzahlen?
- Lageparameter
  - Mittelwert, Median, Modalwert
- Streuungsparameter
  - Standardabweichung, Quantile, Spannweite
- Korrelation
  - Nach Pearson, Spearman, Kendall
- Kontingenzkoeffizienten
- **Welche dieser Kennzahlen soll ein Dataminging-Tool berechnen?**
- **Grafische Darstellung**

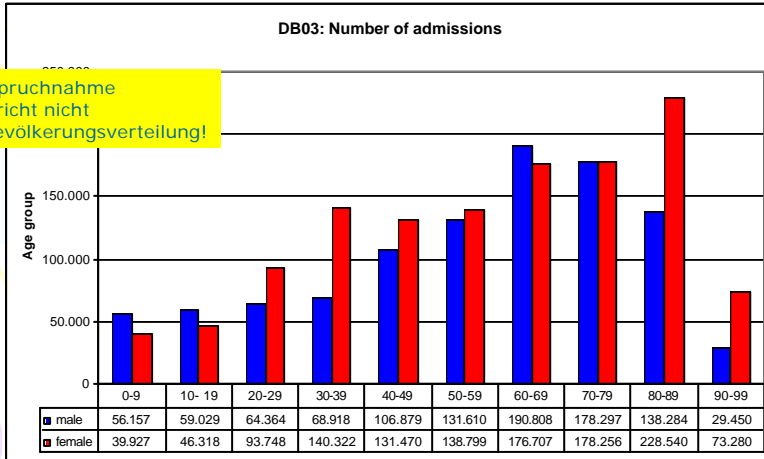
9/30/2004

Interpretation  
Routinedatenbanken

8

# Häufigkeit der Aufnahmen nach Alter und Geschlecht

Inanspruchnahme  
Entspricht nicht  
der Bevölkerungsverteilung!

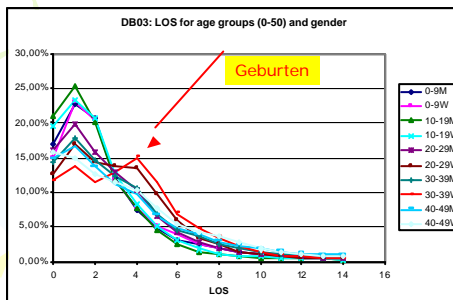


9/30/2004

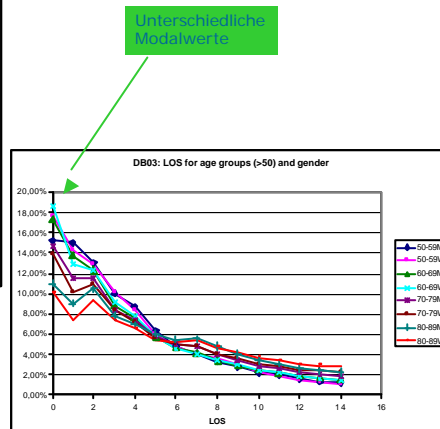
Interpretation  
Routinedatenbanken

9

# Verteilung der Belagsdauer nach Altersgruppen



Einflussfaktoren:  
• Tagesklinik  
• Fallsplitting  
• Entlassungstag  
• ...

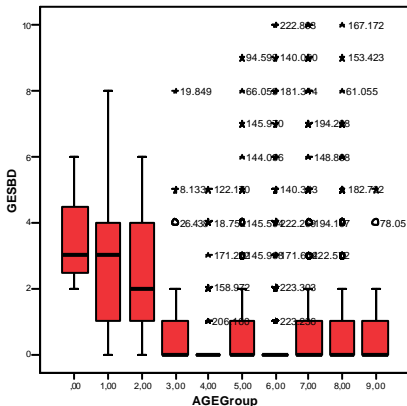


9/30/2004

Interpretation  
Routinedatenbanken

10

# DB03/Tirol: MEL22.02 LOS nach Altersgruppen



Sind Boxplots die bessere Grafik?

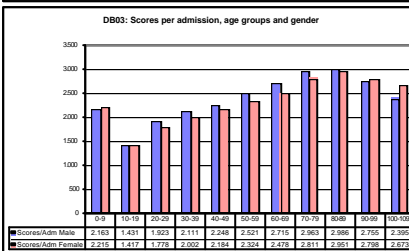
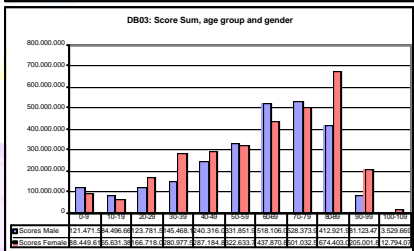
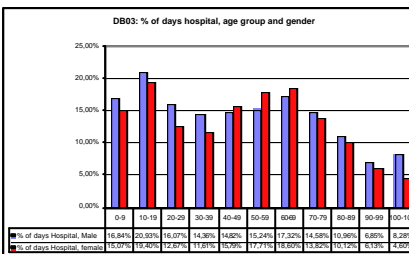
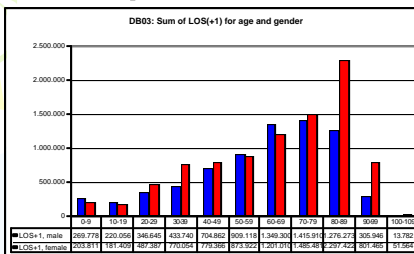
Wie homogen sind die LDF's  
Bezüglich Alter oder anderer  
Merkmale?

9/30/2004

Interpretation  
Routinedatenbanken

11

# Alters- und geschlechts- spezifische Kennzahlen



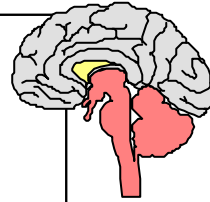
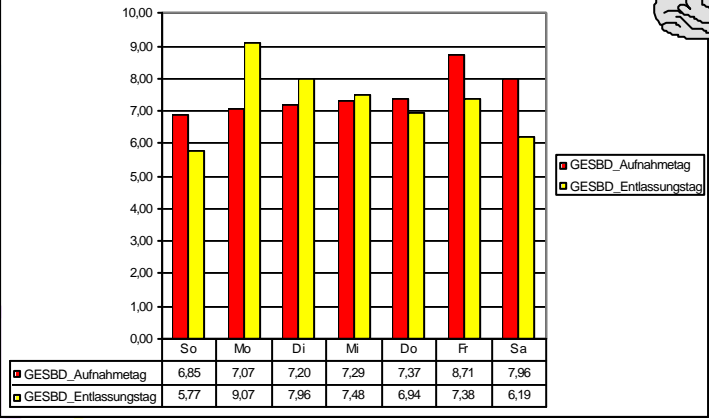
9/30/2004

Interpretation  
Routinedatenbanken

12

# Aufnahme- und Entlassungs-Wochentag

DB 2002: Belagsdauer (ohne 0-Tage) nach Aufnahme- und Entlassungstag

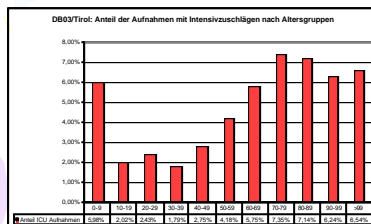
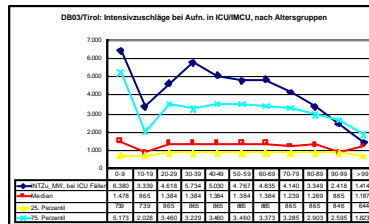
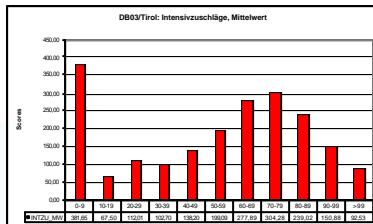


9/30/2004

Interpretation  
Routinedatenbanken

13

# Intensivzuschläge nach Altersgruppen



9/30/2004

Interpretation  
Routinedatenbanken

14

# Bedeutung der Standardisierung

- Nach Alter und Geschlecht
  - Basierend auf einer Standardbevölkerung
  - Um Verzerrungen durch unterschiedliche Bevölkerungsstrukturen auszugleichen
    - Rolle der „Standardbevölkerung“
      - » Europäer
- Probleme bei Kleinregionen mit kleinen Fallzahlen
  - Bei seltenen Erkrankungen

9/30/2004

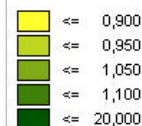
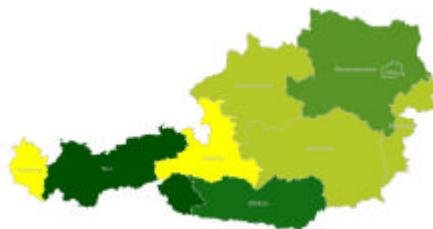
Interpretation  
Routinedatenbanken

15

## J03 – Akute Tonsillitis, 2001: Rohe Raten und Standardisierung

BLND	HR	HRst	HRew	CHF	SHR
Burgenland	40,71	43,20	43,78	0,92	0,93
Kärnten	50,77	51,50	46,51	1,10	1,09
Niederösterreich	48,07	48,62	46,56	1,03	1,03
Oberösterreich	45,25	43,38	49,30	0,92	0,92
Salzburg	40,36	38,68	49,40	0,82	0,82
Steiermark	42,68	43,70	45,82	0,93	0,93
Tirol	70,53	66,16	50,25	1,41	1,40
Vorarlberg	41,01	37,00	52,22	0,79	0,79
Wien	43,87	47,38	43,64	1,01	1,01
Gesamt	46,99	46,99	46,99	1,00	1,00

Beachte die Spannweite!  
Trotz Standardisierung



9/30/2004

Interpretation  
Routinedatenbanken

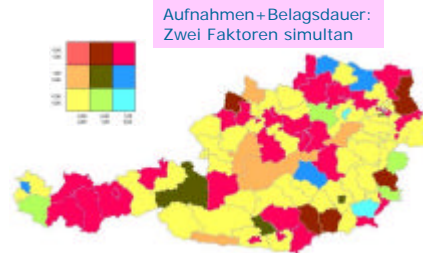
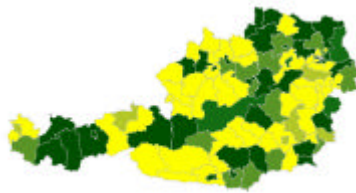
16



# Standardisierung von Aufnahmen und Belagsdauer

Bundesland	SHR	Konfidenzintervall für SHR Untere Grenze 5%	Konfidenzintervall für SHR Obere Grenze 5%	Signifikanz
Burgenland	0,93	0,93	0,93	sig. kleiner
Kärnten	1,09	1,09	1,09	sig. größer
Niederösterreich	1,03	1,03	1,03	sig. größer
Oberösterreich	0,92	0,92	0,92	sig. kleiner
Salzburg	0,82	0,82	0,82	sig. kleiner
Steiermark	0,93	0,93	0,93	sig. kleiner
Tirol	1,40	1,40	1,40	sig. größer
Vorarlberg	0,79	0,79	0,79	sig. kleiner
Wien	1,01	1,01	1,01	sig. größer

Bundesland	SHR	Konfidenzintervall für SHR untere Grenze 5%	Konfidenzintervall für SHR obere Grenze 5%	Signifikanz
Burgenland	0,96	0,875	1,043	kein sig. Unterschied
Kärnten	1,09	1,015	1,141	kein sig. Unterschied
Niederösterreich	1,03	0,991	1,069	kein sig. Unterschied
Oberösterreich	0,95	0,922	0,958	kein sig. Unterschied
Salzburg	0,83	0,777	0,883	sig. kleiner
Steiermark	0,90	0,855	0,936	sig. kleiner
Tirol	1,29	1,222	1,341	sig. größer
Vorarlberg	0,69	0,631	0,748	sig. kleiner
Wien	1,07	1,034	1,111	sig. größer



9/30/2004

Interpretation  
Routinedatenbanken

17

## Was sagt ein Case-Mix-Index?

- Wie gut geeignet ist der CMI für den Vergleich von Krankenanstalten?
  - Spiegelt er das Leistungsspektrum
    - Intensivzuschläge berücksichtigen ja/nein
    - Kinderabteilungen ja/nein
    - Sonderabteilungen ja/nein
    - Tagesklinische Patienten ja/nein

9/30/2004

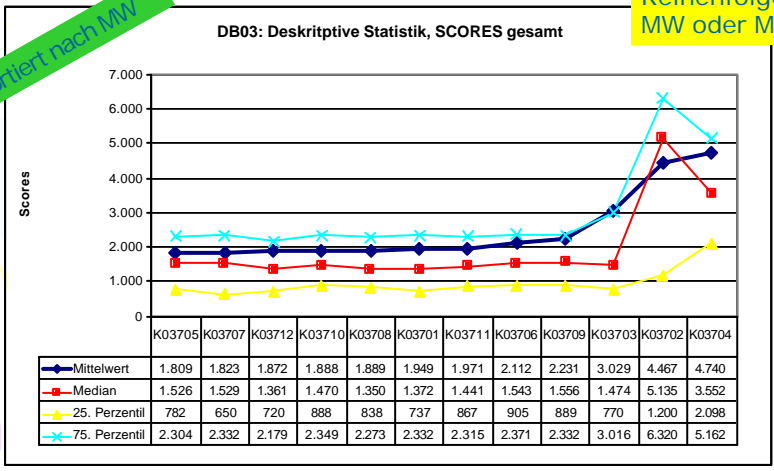
Interpretation  
Routinedatenbanken

18

# DB03, 12KA´s: Score gesamt

Sortiert nach MW

Reihenfolge: MW oder Median

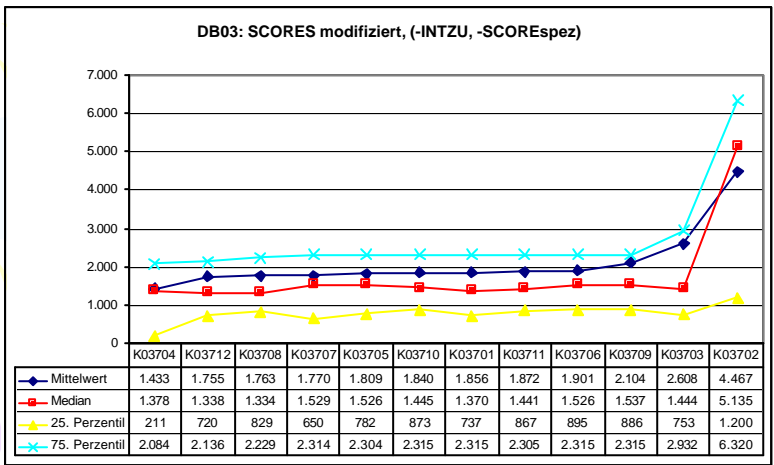


9/30/2004

Interpretation  
Routinedatenbanken

19

# DB03, 12 KA´s: Scores, modifiziert

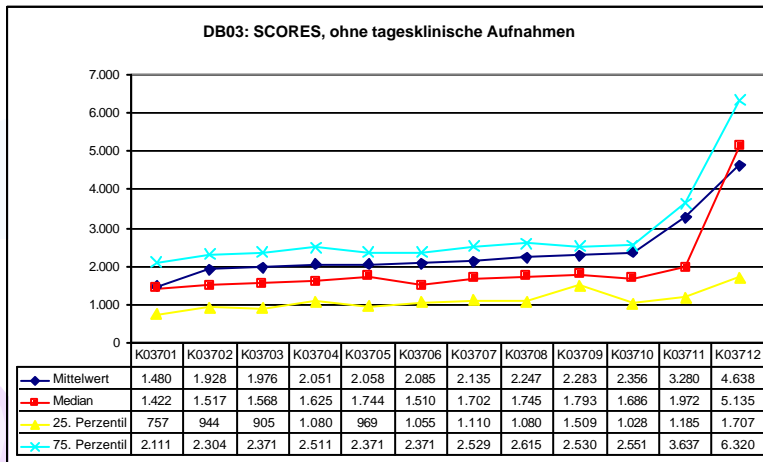


9/30/2004

Interpretation  
Routinedatenbanken

20

# DB03, 12 KA's: Scores modifiziert, ohne tageskl. Aufnahmen

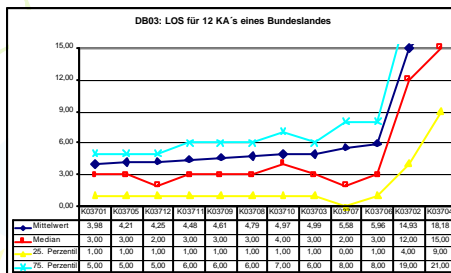


9/30/2004

Interpretation  
Routinedatenbanken

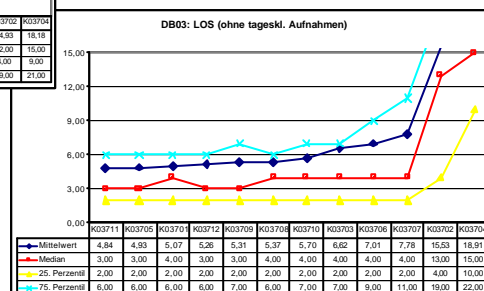
21

# DB03, 12 KA's: LOS, mit/ohne tagesklinische Aufnahmen



Welche Kennzahl ist für die LOS repräsentativ?  
Mittelwert  
Mittelwert (ohne Tagesklinik)  
Median

Beachte die Änderungen der Reihenfolge bei Sortierung nach:  
Mittelwert  
Median



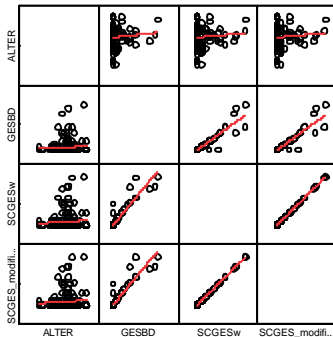
9/30/2004

Interpretation  
Routinedatenbanken

22

# Korrelation: HD3: =C50, LOS>0, MEL22.02

C50, LOS>0; MEL22.02 - Chemotherapy B



Die Beziehungen sind meist:  
Nicht-linear  
Es gibt Ausreisser und Hebepunkte

Korrelationen					
	ALTER	GESBD	SCGESw	AnzDag	SCGES_modifiziert
ALTER	1				
	Korrelation nach Pearson	,071	,076	,227	,076
	Signifikanz (2-seitig)	,366	,329	,003	,329
	N	165	165	165	165
GESBD		1			
	Korrelation nach Pearson	,071	,339	,469	,339
	Signifikanz (2-seitig)	,366	,000	,000	,000
	N	165	165	165	165
SCGESw			1		
	Korrelation nach Pearson	,076	,339	,471	1,000
	Signifikanz (2-seitig)	,329	,000	,000	,000
	N	165	165	165	165
AnzDag				1	
	Korrelation nach Pearson	,227	,469	,471	,471
	Signifikanz (2-seitig)	,003	,000	,000	,000
	N	165	165	165	165
SCGES_modifiziert					1
	Korrelation nach Pearson	,076	,339	1,000	,471
	Signifikanz (2-seitig)	,329	,000	,000	,000
	N	165	165	165	165

\*\*. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Korrelationen					
Spalten/Rei	ALTER	GESBD	SCGESw	AnzDag	SCGES_modifiziert
ALTER	1,000				
	Korrelationskoeffizient	-,021	,072	,209	-,012
	Sig. (2-seitig)	,791	,678	,001	,878
	N	165	165	165	165
GESBD		1,000			
	Korrelationskoeffizient	,021	,306	,311	,296
	Sig. (2-seitig)	,791	,000	,000	,000
	N	165	165	165	165
SCGESw			1,000		
	Korrelationskoeffizient	,072	,306	,295	1,000
	Sig. (2-seitig)	,678	,000	,000	,000
	N	165	165	165	165
AnzDag				1,000	
	Korrelationskoeffizient	,209	,311	,295	,295
	Sig. (2-seitig)	,001	,000	,000	,000
	N	165	165	165	165
SCGES_modifiziert					1,000
	Korrelationskoeffizient	-,012	,296	,295	,295
	Sig. (2-seitig)	,878	,000	,000	,000
	N	165	165	165	165

\*\*. Die Korrelation ist auf dem 0,01 Niveau signifikant (Zweiseitig).

9/30/2004

Interpretation  
Routinedatenbanken

23

## Signifikanztests und Routinedaten

- Aufgrund der hohen Fallzahlen wird sehr schnell ein „hoch signifikantes“ Ergebnis erreicht
  - Anreiz zur Überinterpretation
- Wesentliche Kennzahlen, z.B. LOS sind asymmetrisch verteilt
  - Daher kommen nur parameterfreie Methoden in Frage
  - Probleme bei der Berechnung von Konfidenzintervallen
    - Besser sind Quantile

9/30/2004

Interpretation  
Routinedatenbanken

24

# T-Tests und Signifikanz

Gruppenstatistiken

GESBD	GESCHL	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
ALTER	M	1025271	6,07	16,792	,017
	W	1252160	6,29	10,632	,010
SCGESw	M	1025271	50,94	23,890	,024
	W	1252160	53,28	24,138	,022
LK	M	1025271	627,5680	1706,96596	4,64860
	W	1252160	429,9702	9639,96194	3,25287
TK	M	1025271	663,4956	1622,95423	1,60283
	W	1252160	609,4793	1307,86013	1,16878
INGESBD	M	1025271	359,4187	114,52181	2,03830
	W	1252160	619,8104	1638,07813	1,46388
LNGESBD1	M	1025271	1,4286	,5	,00000
	W	1078291	1,5055	,5	,00000
LNCSGES	M	1025271	1,4583	,5	,00000
	W	1252160	1,5271	,5	,00000
LNCSGES	M	1024464	7,2932	1,1	,00000
	W	1250755	7,3413	,5	,00000

Hohe Streuungen

Alles ist signifikant!\*

Kleine Konfidenzintervalle

Test bei unabhängigen Stichproben

	Levene-Test der Varianzhomogenität	t-Test für die Mittelwertgleichheit								
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
								Untere	Oberer	
GESBD	Varianzen sind gleich	32,096	,000	-12,930	2277429	,000	,23	,018	-,263	-,191
	Varianzen sind nicht gleich			-11,673	1962149	,000	,23	,019	-,264	-,189
ALTER	Varianzen sind gleich	1728,662	,000	-73,277	2277429	,000	-2,34	,032	-2,408	-2,282
	Varianzen sind nicht gleich			-73,353	2198116	,000	-2,34	,032	-2,408	-2,282
SCGESw	Varianzen sind gleich	3080,000	,000	17,638	2277429	,000	97,5899	5,53323	86,75294	108,44279
	Varianzen sind nicht gleich			17,202	1901753	,000	97,5899	5,67368	86,47768	108,71807
LK	Varianzen sind gleich	2400,471	,000	27,613	2277429	,000	54,0163	1,94213	50,20979	57,82282
	Varianzen sind nicht gleich			27,230	1953285	,000	54,0163	1,98371	50,12831	57,90430
TK	Varianzen sind gleich	176,940	,000	-24,277	2277429	,000	-60,3917	2,48756	-65,26766	-55,51620
	Varianzen sind nicht gleich			-23,680	1904003	,000	-60,3917	2,55028	-65,30019	-55,38027
INGESBD	Varianzen sind gleich	1147,252	,000	-54,691	1946276	,000	-,0768	,00145	-,07899	-,07408
	Varianzen sind nicht gleich			-54,519	1940415	,000	-,0768	,00141	-,07860	-,07407
LNGESBD1	Varianzen sind gleich	203,219	,000	-54,454	2277429	,000	-,0688	,00126	-,07128	-,06632
	Varianzen sind nicht gleich			-54,394	2180111	,000	-,0688	,00126	-,07128	-,06632
LNCSGES	Varianzen sind gleich	1407,396	,000	-36,302	2275217	,000	-,0481	,00132	-,05068	-,04547
	Varianzen sind nicht gleich			-36,138	2148369	,000	-,0481	,00133	-,05068	-,04546

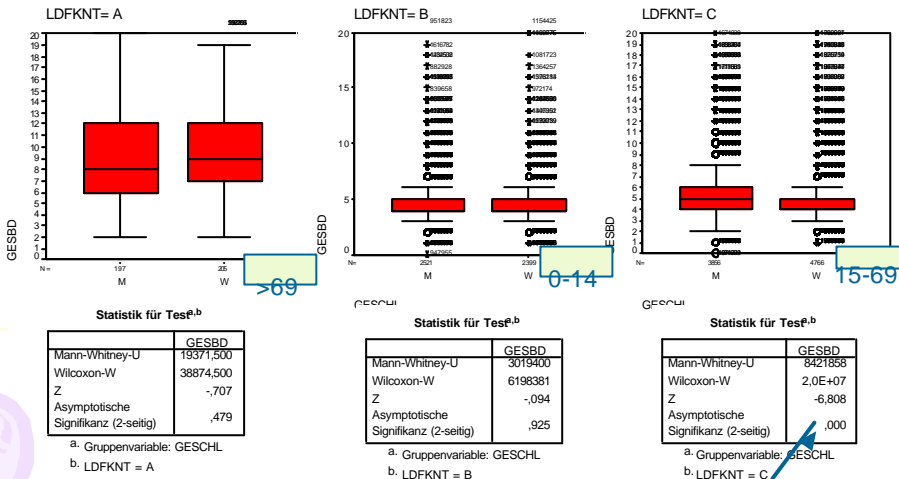


9/30/2004

Interpretation  
Routinedatenbanken

25

# MEL06.01: Appendektomie



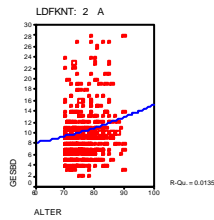
Signifikant, da Grosse Fallzahl

9/30/2004

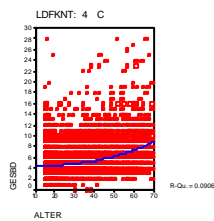
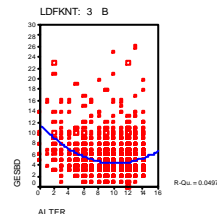
Interpretation  
Routinedatenbanken

26

# MEL06.01: Korrelation zwischen Alter und Belagsdauer



Linearität der Beziehung!?



Korrelationen

	ALTER	GEBSD	SCGESw
Kardial-Traub	Korrelationskoeffizient	1,000	1,113
	Sig. (2-seitig)	,002	,034
	N	402	402
GEBSD	Korrelationskoeffizient	,113	1,000
	Sig. (2-seitig)	,002	,000
	N	402	402
SCGESw	Korrelationskoeffizient	,081	,034
	Sig. (2-seitig)	,000	,000
	N	402	402
Spätmann-Rho	Korrelationskoeffizient	1,000	1,159
	Sig. (2-seitig)	,001	,036
	N	402	402
GEBSD	Korrelationskoeffizient	,159	1,000
	Sig. (2-seitig)	,001	,000
	N	402	402
SCGESw	Korrelationskoeffizient	,109	,039
	Sig. (2-seitig)	,000	,000
	N	402	402

<sup>a</sup> Die Korrelation ist auf dem 0.01 Niveau signifikant (zweiseitig).  
<sup>b</sup> Die Korrelation ist auf dem 0.05 Niveau signifikant (zweiseitig).  
<sup>c</sup> LDFKNT = A

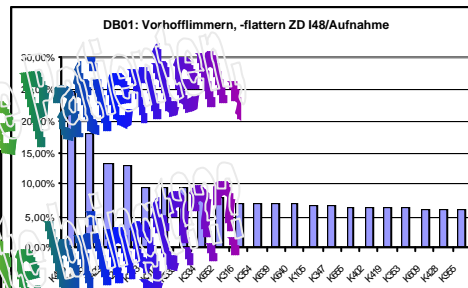
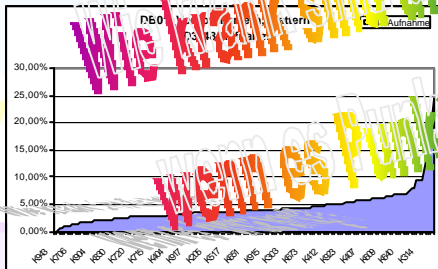
9/30/2004

Interpretation  
Routinedatenbanken

27

# Dokumentation von Zusatzdiagnosen

ZD: I48  
N=73.776



9/30/2004

Interpretation  
Routinedatenbanken

28

# Evaluierung eines MBDS

$$SumScore = \sum_{j=1}^{m_w} g_w(j)W(j) + \sum_{j=1}^{m_v} g_v(j)V(j)$$

$m_w$ ...number of discret variables

$m_v$ ...number of continuous variables

$g_v(j), g_w(j)$ ...weights of the scores of variable X(j)

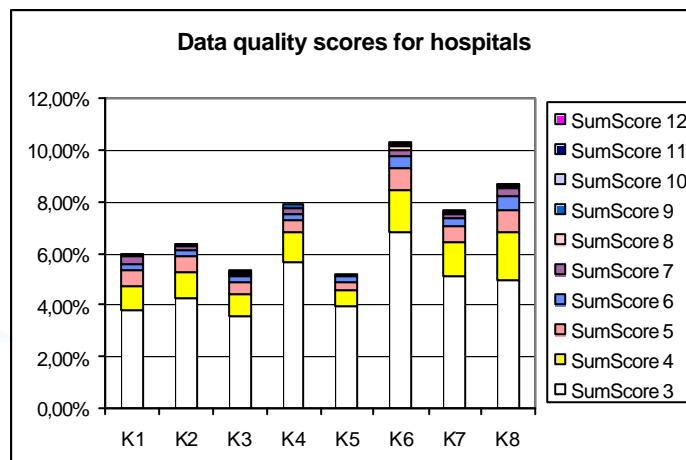
- Evaluierung der Datenqualität einer Krankenanstalt
  - Verteilung der Scores

9/30/2004

Interpration  
Routinedatenbanken

29

# Gesamtbeurteilung der Datenqualität von 8 Krankenanstalten



9/30/2004

Interpration  
Routinedatenbanken

30



# Zusammenfassung

- **Der Zweck verzerrt!**
- Grosse Fallzahlen sind keine Garantie für unverzerrte Kennzahlen
- Beachte die hohe Variabilität
  - Beschreibe die Variabilität
- Was bedeutet „statistisch signifikant“ bei Routinedaten
- Auch Routinedaten brauchen eine hohe Datenqualität

9/30/2004

Interpretation  
Routinedatenbanken

31



**Das gesamte Modell  
kann nicht besser sein  
als die Datenqualität!**

**Danke für Ihre  
Aufmerksamkeit!**

9/30/2004

Interpretation  
Routinedatenbanken

32