
Elementare Datentypen, Skalen und Messfehler

Dr. Dirk Hasenclever
University of Leipzig

[imise.](#)

Wie ist ein Datensatz formal “gegeben”?

■ Datentabelle

- Im einfachen Fall: rechteckige Datentabelle
allgemein: relationale Datenbank
- Zeilen sind Beobachtungseinheiten
- Spalten enthalten Variablen mit entsprechenden Messwerten

■ Essentielle Metadaten:

- Code-Listen,
- Skalentyp,
- Einheiten [cm oder inch?]

Rechteckige Datentabelle

Messgrößen = Variable

Beobachtungseinheiten (Patienten)

	male	rhb	rhct	albumin	resr	rcreat	defresp	iv	fftf	hdfail	cr
1	,00	11,90	35,00	4,00	.	,60	1	,00	98,49	,00	1,00
2	,00	14,70	44,00	4,00	.	1,00	1	,00	83,43	,00	1,00
3	,00	12,60	39,00	4,60	.	,90	1	,00	60,61	,00	1,00
4	,00	13,00	39,00	4,00	.	,90	1	,00	72,97	,00	1,00
5	,00	12,60	41,00	4,00	.	,80	1	,00	8,06	1,00	1,00
6	,00	12,90	38,00	4,30	.	,90	1	,00	47,62	,00	1,00
7	,00	11,50	34,00	4,20	.	,70	1	,00	27,13	,00	1,00
8	,00	11,40	33,00	4,30	47,00	,80	1	,00	41,63	,00	1,00
9	,00	12,80	38,00	3,90	.	,80	1	,00	80,99	,00	1,00
10	,00	11,30	34,00	3,90	.	1,10	1	,00	46,07	,00	1,00
11	,00	11,20	35,00	3,90	28,00	,70	1	,00	55,01	,00	1,00
12	,00	10,90	33,00	3,40	.	1,00	1	,00	133,25	,00	1,00
13	,00	11,20	35,00	4,20	.	,80	1	1,00	93,00	,00	1,00
14	,00	14,70	42,00	3,50	.	,80	1	,00	46,43	,00	1,00
15	,00	12,50	37,00	4,30	.	1,10	4	,00	59,22	,00	,00
16	,00	15,20	45,00	4,30	.	,80	1	,00	83,69	,00	1,00
17	,00	14,30	43,00	4,10	.	1,10	1	,00	41,63	,00	1,00
18	,00	12,90	40,00	3,60	.	1,00	4	,00	35,88	1,00	,00
19	,00	12,70	39,00	3,90	.	1,00	1	,00	101,48	,00	1,00
20	1,00	13,40	39,00	4,20	.	1,10	1	,00	64,88	,00	1,00
21	1,00	12,20	39,00	4,40	.	,80	1	,00	98,62	,00	1,00
22	1,00	12,10	37,00	4,20	.	,90	1	,00	93,72	,00	1,00

Essentielle Metadaten zur Datentabelle

Z.B.: Variable **male** 0=weiblich; 1=männlich Skala=nominal

	Name	TSD	Variablenlabel	Wertelabels	Fehlende Werte	SA	Meßniveau
1	male	N82	Geschlecht	{,00, weiblich}..	Kein	8R	Nominal
2	rhb	N82	Hämoglobin	Kein	Kein	8R	Metrisch
3	rhct	N82	Hämatokrit	Kein	Kein	8R	Metrisch
4	albumin	N12	Serum albumin	Kein	-10,00 - 1,00	8R	Metrisch
5	resr	N82	Blutsenkung 1h	Kein	Kein	8R	Metrisch
6	rcreat	N82	serum creatinine	Kein	Kein	8R	Metrisch
7	defresp	N20	Final outcome	{1, CR/CRu}...	Kein	8R	Ordinal
8	iv	N82	Stadium IV	{,00, Stadium II	Kein	8R	Ordinal
9	fftf	N82	Freedom from treatment failure	Kein	-19999,00 - ,00	8R	Metrisch
10	hdfail	N82	HD Progression/Rezidiv	{,00, zensiert}...	-1,00	8R	Nominal
11	cr	N82	complete remission	{,00, nein}...	Kein	8R	Nominal

Formale Daten allein sind nicht interpretierbar! I

- Eine **Datentabelle ohne Metadaten ist Zahlenmüll.**
- Aber auch eine **annotierte Datentabelle** ist **noch nicht interpretierbar**
- Für eine **seriöse Auswertung** benötigt man **zusätzlich**:
 - **(1) Angaben zur Messmethode für alle Variablen**
 - Beschreibung des Messverfahrens
 - Validierung des Messverfahrens
 - Verfahren der Qualitätskontrolle
 - Abschätzung des Messfehlers
 - Störgrößen, welche die Messgröße bzw. den Messfehler beeinflussen
 - Angaben zu Fehlmessungen (untere und obere Messschranke)
 - Einschätzung zur Selektivität von Fehlwerten
 - ...

Formale Daten allein sind nicht interpretierbar! II

- Für eine **seriöse Auswertung** benötigt man **des Weiteren**:
 - **(2) Angaben zur Auswahl der Stichprobe**
(d.h. der Beobachtungseinheiten)
 - Sind die Beobachtungseinheiten repräsentativ für eine Grundgesamtheit?
 - **Zufallsstichprobe** oder „**convenience sample**“
 - Sind interessierende Vergleichsgruppen **gleichartig ausgewählt** oder **potentiell verzerrt**?
 - ...
- Diese zusätzlichen Angaben findet man bei gut durchgeführten **prospektiven** Studien oder Experimenten in einem aussagefähigen **Studien- oder Projektprotokoll**.

Datentypen

- Je nach dem Datentyp der Variablen gibt es unterschiedliche statistische Beschreibungsmittel, Schätz- und Testverfahren
- NB: Messgrößen nehmen streng genommen nur diskrete Werte an.
 - (Selbst scheinbar kontinuierliche Messgrößen, werden als endliche Dezimalzahlen dokumentiert.)

Übersicht: Univariate Datentypen I

- **Kategoriale Daten** (Beobachtungseinheiten werden in Klassen eingeteilt.
 - **Nominal**
 - Binär (Mann/Frau)
 - Typen (Histologische Diagnosen, Automarken)
 - **Ordinal** (=angeordnet)
 - Binär (Erfolg j/n)
 - Score (Krankheitsstadien, Gewichtsklassen, Immunhistochemie)

Übersicht: Univariate Datentypen II

■ Numerische Daten

- Anzahlen (# Kinder, Minigolfscores)
- (Quasi)-Kontinuierlich
 - Verhältnisskala
 - Messwerte mit Einheit und Nullpunkt (kg, Zellkonzentrationen, °Kelvin)
 - Intervallskaliert
 - Skala ohne Nullpunkt. Nur Differenzen sind vergleichbar (VAS, Genexpression, °Celsius)
 - Prozentzahlen (% gefärbte Zellen pro Bildausschnitt, % Tumorzellen im Knochenmarkausstrich)
 - Zensierte Daten
 - Time to Event Daten (Überlebenszeiten)
 - Messwerte mit unterer und/oder oberer **Detektionsschwelle**
 - Z.B. Leukozyten unter Chemotherapie <500/ml als unterste Angabe
 -

Nominale Daten=ungeordnete Kategorien

- Beobachtungseinheiten werden in **ungeordnete Kategorien** eingeteilt.
- Bsp.: **Fixes vorgegebenes Klassifikationsschema**:
 - Farben (rot, grün, blau, ...) [Prototypen zur ostensiven Definition nötig]
 - Geschlecht (männlich, weiblich)
 - Histologischer Subtype einer Krebserkrankung
- Bsp.: **Offene Kategorienliste**
 - Behandelnde Klinik (Liste von möglichen Institutionen)
 - Geburtsort
 - Automarken vorbeifahrender Wagen

Nominale Daten=ungeordnete Kategorien II

- **Mathematische Struktur:** Kategorien bilden eine **ungeordnete Menge**
- **Relation zwischen Beobachtungen:** Gleichheit “=”
- **Deskriptive Auswertung:**
 - Tabelle,
 - Häufigkeiten,
 - Kuchendiagramm
 - **Lagemaß:** Modalwert, (**Mittelwert, Median nicht definiert**)
 - **Streuungsmaß:**
 - Häufigkeit, mit der zwei Beobachtungseinheiten gleich kategorisiert werden
 - Abweichung von Gleichverteilung (Shannon Information)
 - ...
- **Klassifikationsfehler:**
 - Reproduzierbarkeit
 - Übereinstimmung von Beurteilungen

Nominale Skala

(Abgeordnete nach Parteizugehörigkeit klassifiziert)

■ Sitzverteilung im 16. Deutschen Bundestag 2005

Endgültiges amtliches Ergebnis der Bundestagswahl 2005



CDU/CSU	226	36.8%
SPD	222	36.2%
FDP	61	9.9%
Linke	54	8.8%
Grüne	51	8.3%
Total	614	

Quelle: Bundeswahlleiter
Stand: 07.10.2005

bpb © 2005 Bundeszentrale
für politische Bildung

Histological Subtypes in Hodgkin's Lymphoma

“Original and expert diagnosis agreed for 245 of the 362 reviewed cases (68%).”

TABLE 1

Case Counts of Original and Review Histologic Diagnoses, and Positive Predictive Values of Original Histologic Diagnoses, Hodgkin Disease Diagnosed in Women in 1988–94, Greater Bay Area Cancer Registry^a

Registry classification (ICD-0-2)				Review classification (ICD-O-2 codes)																							
				9650		9652		9653		9657		9658		9659		9663		9664		9665		9666		9667			
				NOS		MC		LD		LP		LPd ^b		LPn ^c		NS		NScp ^d		NSlp		NSmc		NSsy ^e		Not HD	
				<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Total				14	4	35	10			1	—	1	—	13	4	210	58	12	3					65	18	11	3
9650	NOS	20	6	8	40	2	10							2	10	4	20	1	5					1	5	2	10
9652	MC	43	12	3	7	25	58							1	2	10	23						2	5	2	5	
9653	LD	3	1																						3	100	
9657	LP	6	2			3	50							3	50												
9658	LPd ^b	2	1							1	50														1	50	
9659	LPn ^c	5	1							1	20			4	80												
9663	NS	259	72	3	1	5	2							2	1	191	74	9	4					48	19	1	—
9664	NScp ^d	2	1															2	100								
9665	NSlp	1	—									1	100														
9666	NSmc	3	1													3	100										
9667	NSsy ^e	18	5													2	11							14	78	2	11

NOS: not otherwise specified; MC/mc: mixed cellularity; LD: lymphocyte depletion; LP/lp: lymphocyte predominance; NS: nodular sclerosis; HD: Hodgkin disease.

^a Covering Alameda, Contra Costa, Marin, Monterey, San Benito, San Francisco, San Mateo, Santa Clara, and Santa Cruz counties.

^b d: diffuse.

^c n: nodular.

^d cp: cellular phase.

^e sy: syncytial variant.

Messfehler bei nominalen Daten Glaser-S-2001-218

Gradeinteilung (Ordinalskala)

- Beobachtungseinheiten werden in vorgegebene endliche Liste **geordneter Kategorien** eingeteilt.
- Bsp.:
 - **Schulnoten (1-6)**
 - **Schwierigkeitsgrad (leicht, mittel, schwer)**
 - **Krankheitsstadium**
 - **CTC Toxizitätsscore**
 - **Glasgow Coma Score**

Grad-Skala

Glasgow Coma Scale für Erwachsene [Bearbeiten]

Punkte	Augen öffnen	Verbale Kommunikation	Motorische Reaktion
6 Punkte	-	-	befolgt Aufforderungen
5 Punkte	-	konversationsfähig, orientiert	gezielte Schmerzabwehr
4 Punkte	spontan	konversationsfähig, desorientiert	ungezielte Schmerzabwehr
3 Punkte	auf Aufforderung	unzusammenhängende Worte	auf Schmerzreiz Beugeabwehr (abnormale Beugung)
2 Punkte	auf Schmerzreiz	unverständliche Laute	auf Schmerzreiz Strecksynergismen
1 Punkt	keine Reaktion	keine verbale Reaktion	keine Reaktion auf Schmerzreiz

Schweregrad: Leicht: 14-15 Punkte ; Mittel: 9-13 Punkte ; Schwer: 3-8 Punkte



Vergrößerung auf **relevante** und **reproduzierbare** Unterschiede.
Vermeiden von Pseudogenauigkeit

Gradeinteilung (Ordinalskala)

- **Mathematische Struktur:**

Kategorien bilden eine **total geordnete Menge**

- **Relation zwischen Beobachtungen:** Ordnung “<“, “>“, “=”

- **Deskriptive Auswertung:**

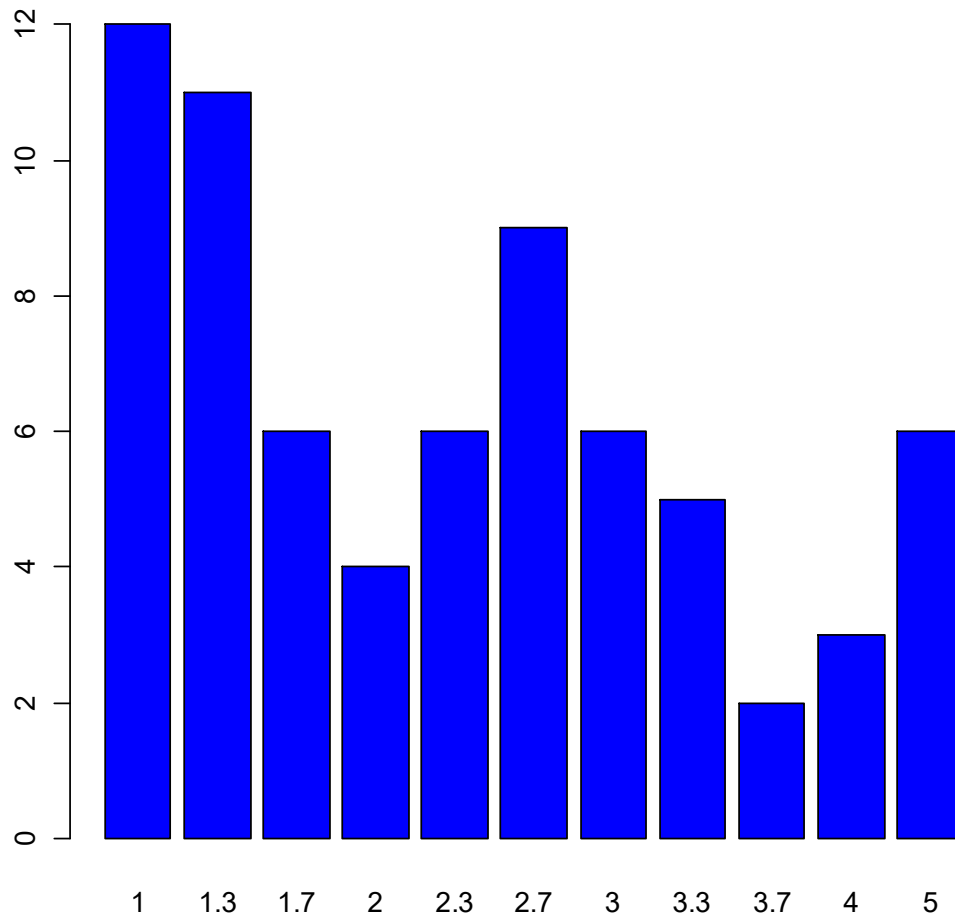
- Tabelle,
- Häufigkeiten, Kumulierte Häufigkeiten
(e.g.: % Studenten Note > 2.0)
- Quantile (q ist α -Prozent Quantil:= α -Prozent aller Werte $\leq q$)
- Lagemaß: Median (=50% Quantil)
- Streuungsmaß: Intervall [1.Quartil, 3.Quartil]
(Interquartildifferenz nicht definiert)

- **Klassifikationsfehler:**

- Reproduzierbarkeit
- Übereinstimmung von Beurteilungen
- Häufigkeit “mehr als eine Note Differenz”

Diplomnoten (Gradskala)

Diplom Nebenfachprüfungen Stand 2008-10-20



```
> median(noten)
[1] 2.3
```

```
> table(noten)
Noten (N=70)
```

1	1.3	1.7	2	2.3	2.7	3	3.3	3.7	4	5
12	11	6	4	6	9	6	5	2	3	6

Sinnvoll?

Setzt eigentlich Intervallskala voraus!

```
> mean(noten)
[1] 2.378571
```

“Messfehler”?

Intervallskalierte Daten I

Nur Differenzen sind interpretierbar

(Pseudo-)kontinuierliche Messgrößen ohne definierten Nullpunkt

■ **Bsp.:**

- **Jahresangaben (1914, 2001)**
- **Temperatur in °Celsius (Aber nicht in °Kelvin!)**
 - *"If it's twice as cold today as it was yesterday and it was zero degrees yesterday, how cold is it today?"*
- **Genexpression (=Konzentration von RNA in Gewebeprobe)**
 - Nur Verhältnisse von Signalen in unbekannter Einheit (d.h. Differenzen der Logarithmen) sind interpretierbar

Intervallskalierte Daten II

Nur Differenzen sind interpretierbar

(Pseudo-)kontinuierliche Messgrößen ohne definierten Nullpunkt

■ **Sonderfälle: Mögliche Werte nur innerhalb eines reellen Intervalls**

- **Prozentzahlen:**

- % Leukämischer Zellen in Knochenmarkausstrich
- % Gefärbter Zellen in Zählfenster in einem Gewebeschnitt
 - Problem: Messgenauigkeit hängt von Nenner ab und dieser kann von Messung zur Messung variieren...

- **Visuelle Analogskala (VAS)**

- Schmerz/Stimmung



(Studien zeigen: 1-2cm Unterschied auf einer 10cm Skala sind für Patienten bedeutsam: „Etwas besser geworden“.)

Intervallskalierte Daten

- **Mathematische Struktur:** Potentielle Messwerte liegen auf einer **affinen Gerade**
- **Relation zwischen Beobachtungen:** Ordnung “<“, “>“, “=” + **Differenzbildung sinnvoll**
 - Differenzen zwischen beliebigen Beobachtungen sind vergleichbar.
(Vorsicht bei Intervalldaten am Rande.)
- **Deskriptive Auswertung:**
 - Histogramm, empirische Verteilungsfunktion
 - **Lagemaß:** Mittelwert, Median
 - **Streuungsmaß:**
 - Varianz, Standardabweichung, Median Absolute Deviation (MAD)
- **Messfehler:**
 - Reproduzierbarkeit bei Mehrfachmessung

Verhältnisskala

- **Mathematische Struktur:** Potentielle Messwerte bilden Teilmenge eines **Körper** (meist $\mathbb{R}^+$ oder Intervall $[a,b]$ in $\mathbb{R}$)
- **Relation zwischen Beobachtungen:** Ordnung “<“, “>“, “=“ +
Addition, Subtraktion,
Multiplikation und Division
 - Differenzen und Verhältnisse (“Doppelt so groß”) zwischen beliebigen Beobachtungen sind interpretierbar.
- **Deskriptive Auswertung:**
 - Histogramm, empirische Verteilungsfunktion
 - **Lagemaß:** Mittelwert, Median
 - **Streuungsmaß:**
 - Varianz, Standardabweichung, Median Absolute Deviation (MAD)
- **Messfehler:**
 - Reproduzierbarkeit bei Mehrfachmessung

Symmetrische Verteilung

■ Beschreibung

- Lagemass

- Mittelwert: $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$

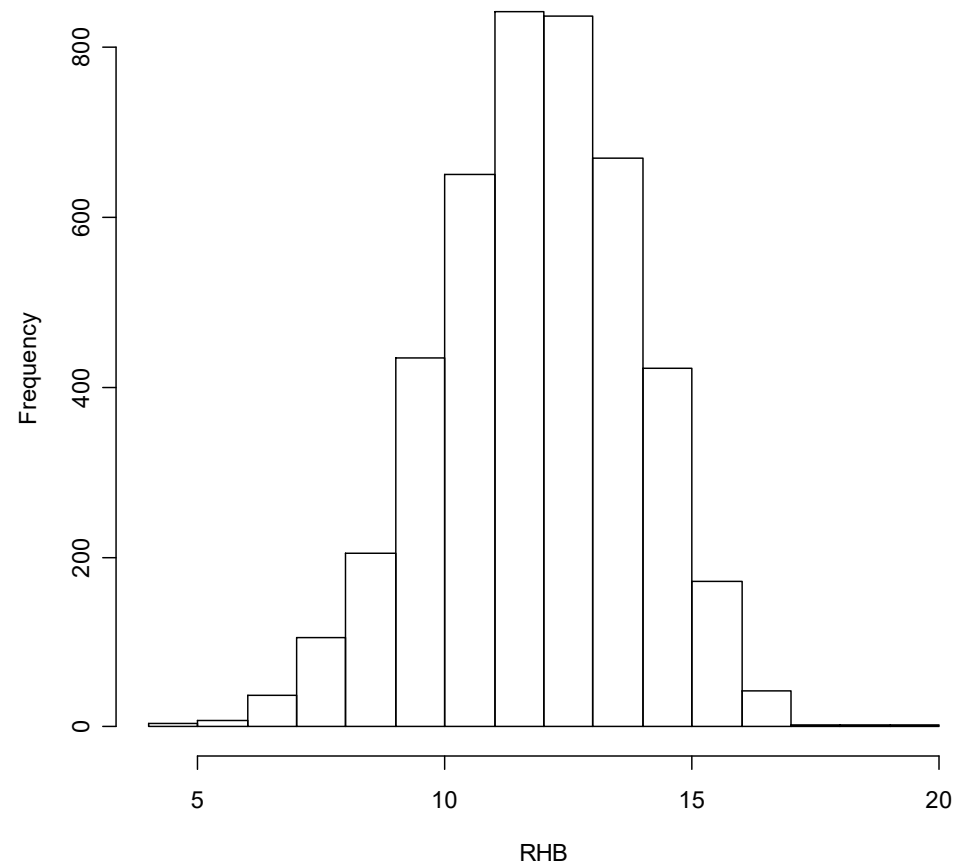
- (=Median=Mode!)

- Streuung

- Standardabweichung:

$$sd = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

Histogramm von Hämoglobinkonzentrationen bei Diagnose eines Hodgkin Lymphoms



Symmetrische Verteilung mit Ausreißern

■ Beschreibung

• Robustes Lagemass

– Trimmed MEAN

– `> mean(x)`

– `[1] 12.24481`

– `> mean(x,trim=0.1)`

– `[1] 12.13530`

• Streuung

– `mad`=Median absolute deviation

$$mad = 1.4826 \cdot median(|x_i - median(x_i)|)$$

`> sd(x)`

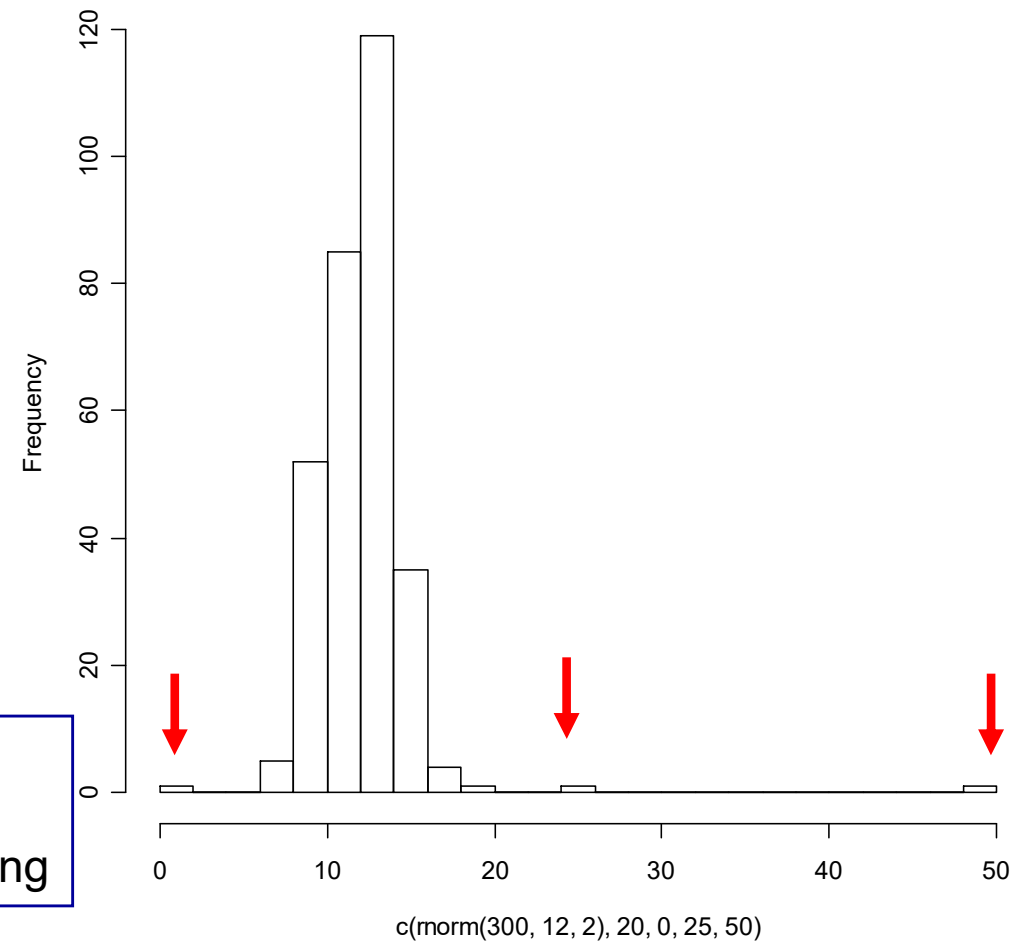
`[1] 3.149656`

`> mad(x)`

`[1] 1.979288`

Faktor, damit
`mad=sd` für
Normalverteilung

Histogram of `c(rnorm(300, 12, 2), 20, 0, 25, 50)`



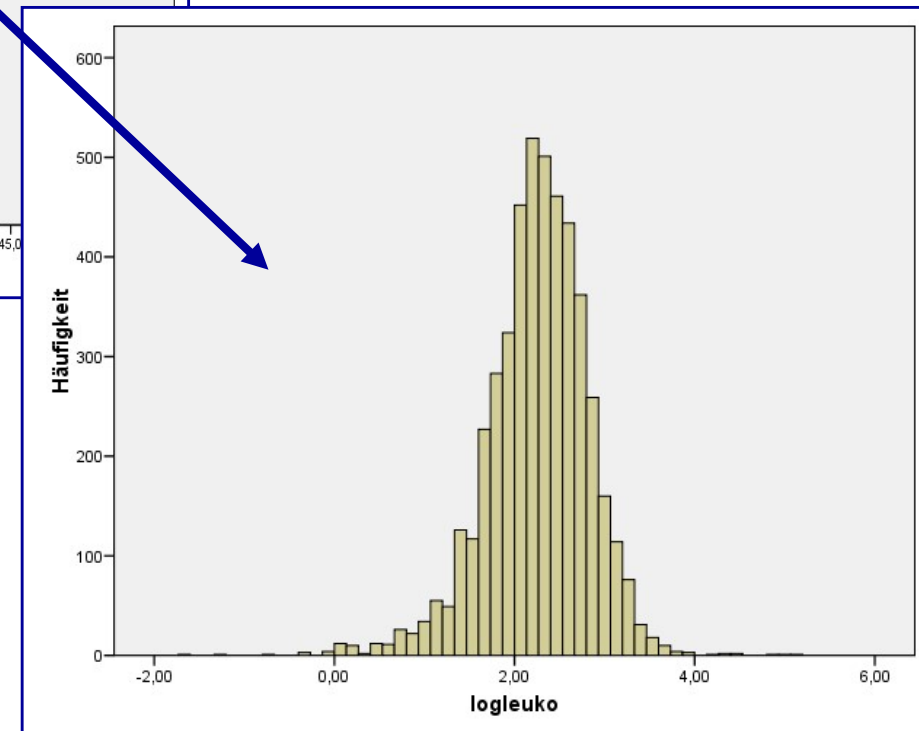
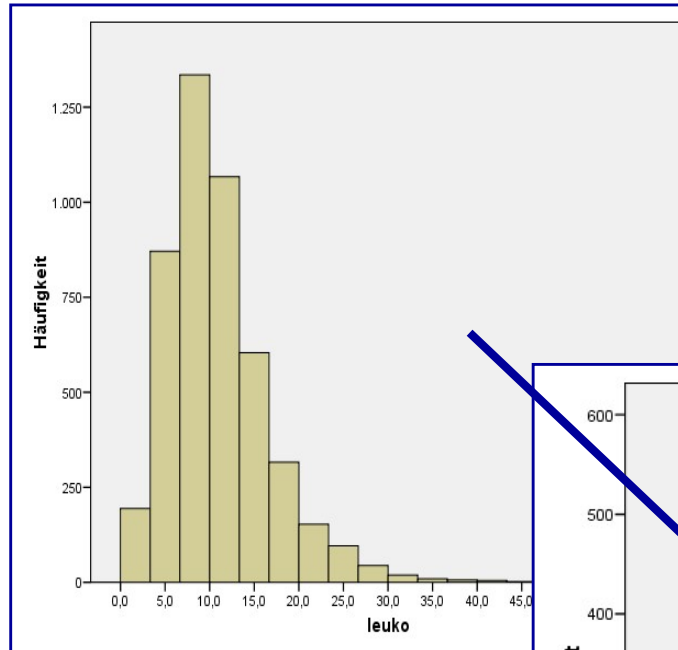
Transformierbare Asymmetrische Verteilung

■ Transformiere auf eine symmetrische Verteilung

- Z.B. Logarithmieren
 - Zellzahlen (z.B. Leukozyten)
 - Zellsekrete
- Andere Transformationen
 - Logit für Raten
 - Wurzel ..

■ Logarithmieren oft indiziert, falls

- Verhältnisskala und
- Standardabweichung in der Größenordnung des Mittelwerts
- nur positive Werte sinnvoll.

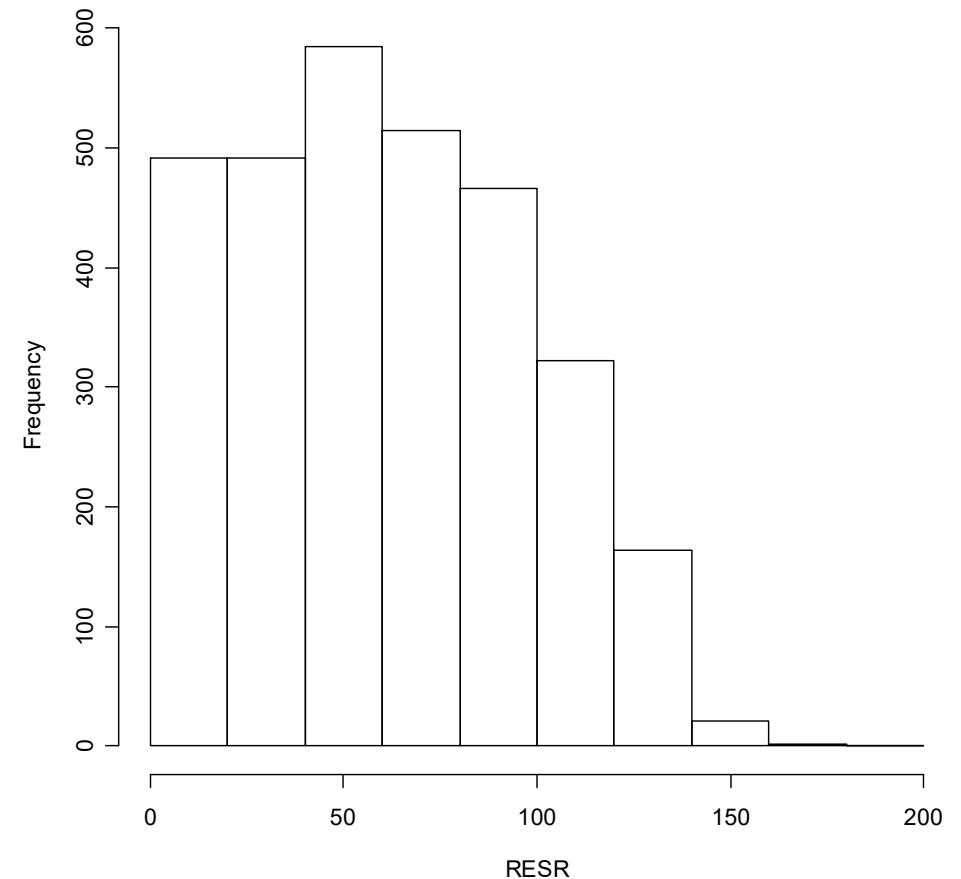


Erratische Aymmetrische Verteilung

■ Beschreibung

- Lagemass
 - Median
- Streuung (Skala)
 - Interquartil-Abstand
 - mad

Histogramm von Blutsenkung (1h)
bei Diagnose eines Hodgkin Lymphoms



Volumenmessung von Schilddrüsenknoten mittels Ultraschall:

~50% Messfehler in Studie mit drei unabhängigen Beobachtern

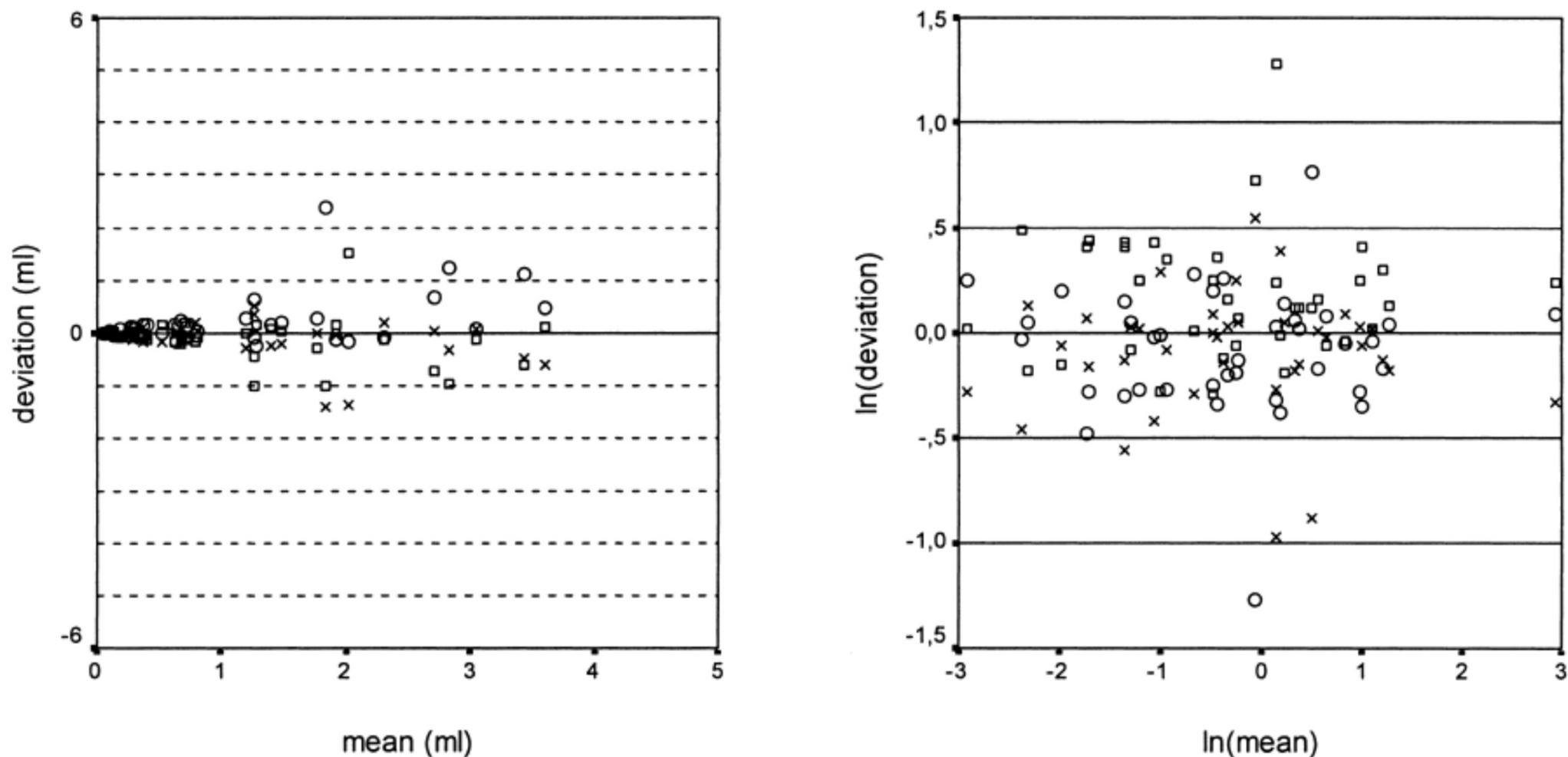


FIG. 1. Deviation of the nodule volumes determined by the three investigators from the mean nodule volume. The x-axis represents the mean nodule volumes in milliliters on the original scale (left) and on the logarithmic scale (right). The y-axis represents the deviation from the mean in milliliter on the original scale (left) and on the logarithmic scale (right) for the three observers. Volumes in Figure 1 are estimated by the ellipsoid method. $\square$, Investigator 1; $\times$, Investigator 2; $\circ$, Investigator 3.

Normalbereiche von Laborwerten

- Zur Interpretation von Laborwerten benötigt man **Normalbereiche**.
- Diese **variieren von Labor zu Labor**
- Diese können von **Alter, Geschlecht, und evtl. weiteren Faktoren abhängen**.

■ Institut für Laboratoriumsmedizin
Universität Leipzig

■ Hämoglobin (konventionelle Einheit)

- Nabelschnur: 13,5-20 g/dl
- <14d: 13,4-19,8 g/dl
- 14d-1m: 10,7-17,1 g/dl
- 1-2m: 9,4-13,0 g/dl
- 2-4m: 10,3-14,1 g/dl
- 4-6m: 11,1-14,1 g/dl
- 6-9m: 11,4-14,0 g/dl
- 9-12m: 11,3-14,1 g/dl
- 0,5-5y: 11,0-14 g/dl
- 5-9y: 11,5-14,5 g/dl
- 9-12y: 12,0-15,0 g/dl
- 12-14y: **M:** 12,0-16,0 g/dl; **F:** 11,5-15,0 g/dl
- 15-17y: **M:** 12,3-16,6 g/dl; **F:** 11,7-15,3 g/dl
- 18-44y: **M:** 13,2-17,3 g/dl; **F:** 11,7-15,5 g/dl
- 45-64y: **M:** 13,1-17,2 g/dl; **F:** 11,7-16 g/dl
- >65y: **M:** 12,6-17,4 g/dl; **F:** 11,7-16,1 g/dl

Laborwerte aus Hepatitis Studie am KKSL

Normalbereich
Miterfassen!

Obere Messschwelle

Untere
Detektionsschwelle

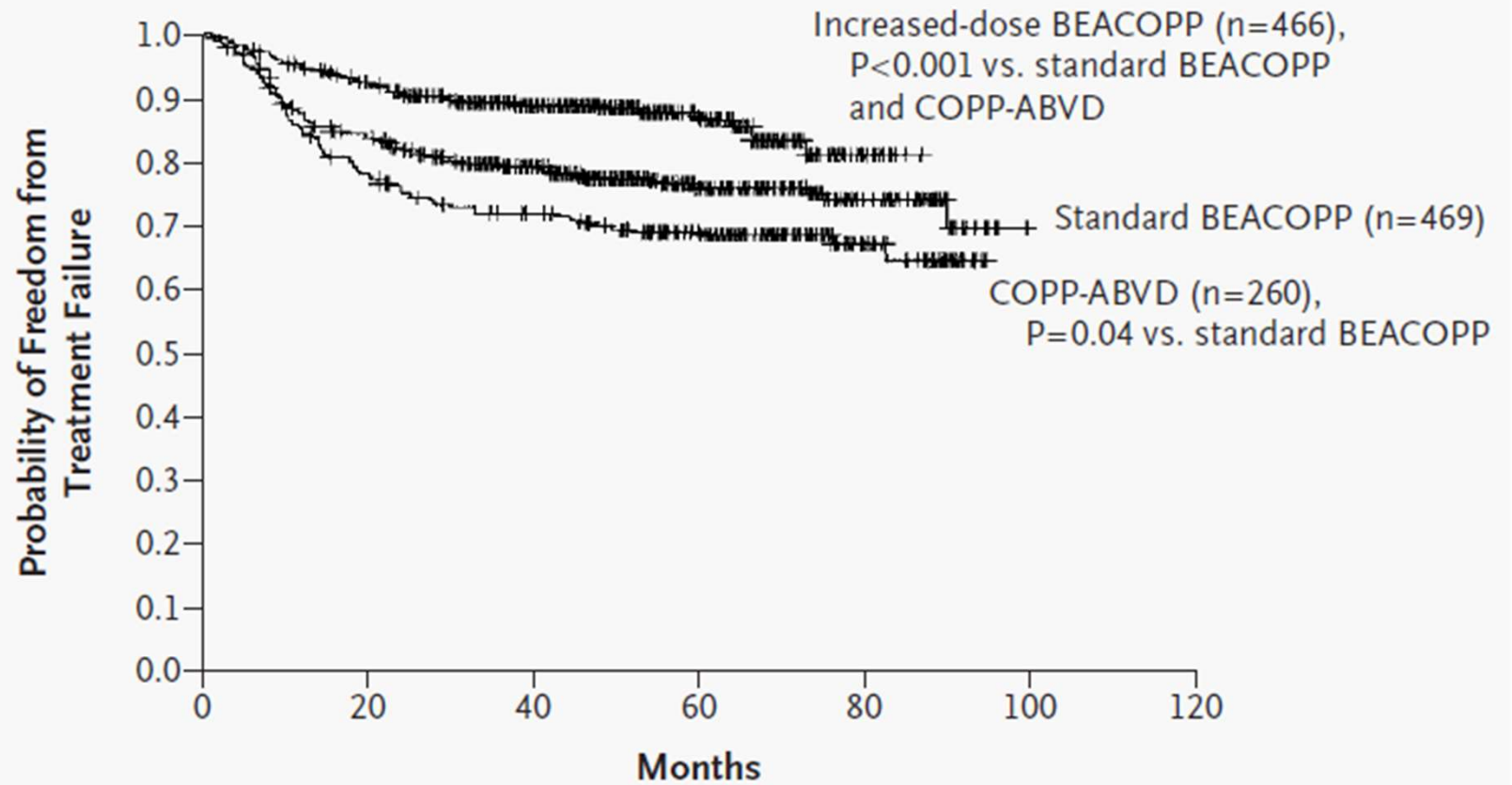
Normalbereich,
Detektionsschwelle
Obere Messschwelle
kann Laborspezifisch
sein...

Keine Mischung von
Zahlen und Strings

PATNO	LEUK	LEUK_R	BILI	BILI_R	HBSAG	HBSAG_R	HBS	HBS_R
S1003	6.82	4,0-9,0	275.80	2,0-20,5	10.21	<1	<1,00	<10
S4003	8.89	4,0-9,0	200.60	2,0-20,5	212.14	<1	1.04	<10
S2002	7.92	4,0-9,0	274.80	2,0-20,5	851.96	<1	<1,00	<10
S3001	7.77	4,0-9,0	100.50	2,0-20,5	>1000,0	<1	<1,00	<10
S2001	5.36	4,0-9,0	148.50	2,0-20,5	>1000,0	<1	<1,00	<10
S4002	6.96	4,0-9,0	549.60	2,0-20,5	>1000,0	<1	<1,00	<10
S1001	5.47	4,0-9,0	203.40	2,0-20,5	>1000,0	<1	<1,00	<10
S4001	8.94	4,0-9,0	419.20	2,0-20,5	>1000,0	<1	4.25	<10
S1002	5.19	4,0-9,0	141.00	2,0-20,5	>1000,0	<1	<1,00	<10
S2004	7.59	4,0-9,0	363.20	2,0-20,5	>1000,0	<1	7.55	<10
S4004	6.83	4,0-9,0	202.20	2,0-20,5	>1000,0	<1	12.23	<10
S1004	7.00	4,0-9,0	147.50	2,0-20,5	>1000,0	<1	3.66	<10
S4006	7.02	4,0-9,0	415.80	2,0-20,5	>1000,0	<1	4.36	<10
S2006	11.22	4,0-9,0	455.70	2,0-20,5	>1000,0	<1	1.58	<10
S2007	7.50	4,0-9,0	277.60	2,0-20,5	>1000,0	<1	<1,00	<10
S1012	8.49	4,0-9,0	48.40	2,0-20,5	>1000,0	<1	<1,00	<10
S2005	9.94	4,0-9,0	341.10	2,0-20,5	789.34	<1	4.84	<10
S4005	5.49	4,0-9,0	586.70	2,0-20,5	>1000	<1	1.85	<10
S1009	4.76	4,0-9,0	31.50	2,0-20,5	>1000,0	<1	1.16	<10
S1006	5.80	4,0-9,0	45.80	2,0-20,5	2.19	<1	15.57	<10
S1011	6.44	4,0-9,0	60.50	2,0-20,5	>1000,0	<1	<1,00	<10
S1010	7.62	4,0-9,0	142.70	2,0-20,5	>1000,0	<1	<1,00	<10
S1013	6.83	4,0-9,0	65.90	2,0-20,5	>1000,0	<1	<1,00	<10
S1005	4.83	4,0-9,0	181.80	2,0-20,5	>1000,0	<1	<1,00	<10
S1007	5.29	4,0-9,0	133.80	2,0-20,5	>1000,0	<1	<1,00	<10
S1008	6.21	4,0-9,0	176.60	2,0-20,5	4.42	<1	25.38	<10

Time to event – Daten

- Zeit von Start der Beobachtung bis Auftreten eines Ereignisses
- Bsp: Diagnose Krebs bis Tod
 - Wenn geheilt – Ereignis (lange nicht) beobachtbar
- Komplexe Datenstruktur
 - Zeit bis letzte Beobachtung
 - Indikator, ob Ereignis eingetreten ist



Number at Risk

COPP-ABVD	260	239	216	162	60	1
Standard BEACOPP	469	438	342	163	72	0
Increased-dose BEACOPP	466	442	352	153	21	0

Diehl-V et al: N Engl J Med 2003;348:2386-95.

Datenerfassung

Das CRF-Design hat einen direkten Einfluss auf die Qualität der erhobenen Daten!

4 Datenerfassungsarten zur Dokumentation:

- **Text** _____ (Name behandelnder Arzt, Kommentarfelder)

→ **sonstige Freitexte vermeiden!!!**

- **Zeit- und Datumsangaben** → Format vorgeben!

$\overline{T} \overline{T} / \overline{M} \overline{M} / \overline{J} \overline{J} \overline{J} \overline{J}$ (Deutschland/Europa)

$\overline{D} \overline{D} / \overline{M} \overline{O} \overline{N} / \overline{Y} \overline{Y} \overline{Y} \overline{Y}$ (international)

$\overline{M} \overline{M} / \overline{D} \overline{D} / \overline{Y} \overline{Y} \overline{Y} \overline{Y}$ (Amerika!)

Datenerfassung II

■ Werte / Zahlen → quantitative Merkmale

- Laborwerte / Größe / Gewicht / Blutdruck / Dosierung / ...

→ Einheiten vorgeben, z.B.:

	Meßwert	Einheit (Auswahl aus Vorgaben)		andere Einheit
Hämoglobin	_____	☐ mmol/l	☐ g/dl	☐ _____
Kreatinin (Serum)	_____	☐ µmol/l=nmol/ml	☐ mg/dl	☐ _____

- **Zählwerte** (Anzahl Tage, Tabletten,...) → Stellenanzahl vorgeben, z.B. Tage

■ Auswahllisten → qualitative Merkmale

- Kodierung automatisch in DB, z.B. **Entscheidungsfragen**: ☐ ja ☐ nein ☐ unbekannt

Mehrfachauswahl: ☐ Lokalisation I ☐ Lokalisation II ☐ Lokalisation III

Sonderform Matrix:

	nie	manchmal	oft
Frage 1	☐	☐	☐
Frage 2	☐	☐	☐
...	☐	☐	☐

- **Kodierung durch Arzt**, z.B. Diabetes mellitus (1 = ja, 2 = nein, 9 = unbekannt)
- Kodierung durch Dokumentar (bei komplizierten Zusammenhängen)

Lernziele

- Fähigkeit, Skalentyp von Variablen zu erkennen und zu benennen.
- Sensibilisierung für Bedeutung von Messfehlern.
- **Hausaufgabe:** Bitte zu jedem Anstrich im Datentypen Überblick (Folien 7+8) ein **eigenes Beispiel** aufschreiben.