

Jahr zwei des neuen
Krebsregistergesetzes:

Wo stehen wir heute?

21. Informationstagung
Tumordokumentation der klinischen
und epidemiologischen Krebsregister

22.-24. April 2015
in Stuttgart

Workshop:

**Wir sind die Guten! -
Statistische
Methoden zum Vergleich von
Parametern und
Kenngrößen zur Ergebnisqualität**

Mittwoch, 22.04.2015
9:00h-12.30h

Die besten
GESCHICHTEN
schreibt wohl immer noch

das echte
LEBEN



Prof. Chi-Quadrat

Statistiker



CA Dr. A

A-Klinik



CA Dr. B

B-Klinik



Ass-Arzt Dr. Klein

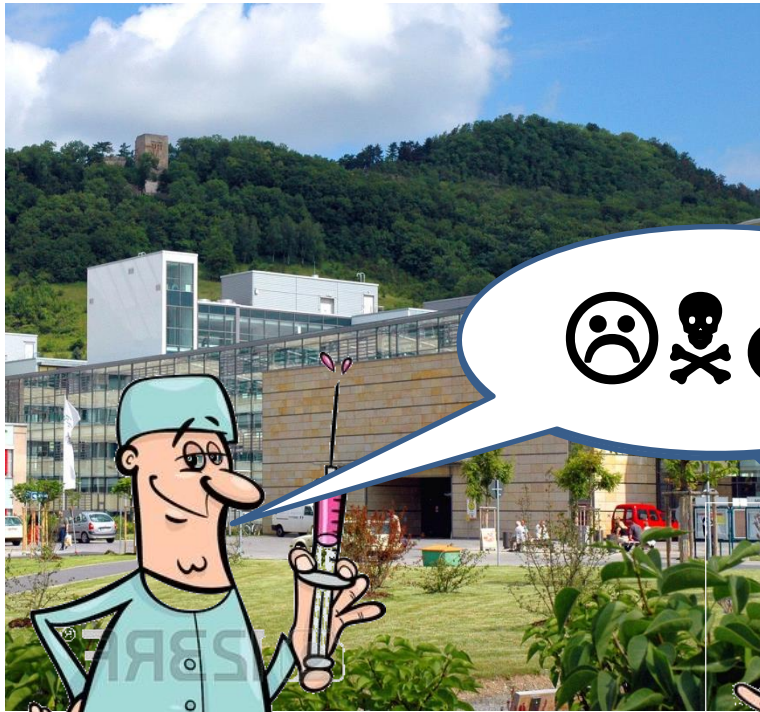


CA Dr. C

C-Klinik

Irgendwo in Deutschland ...







Beweise, dass wir die Guten sind!



☹️ ☠️
💣 ☠️
☹️ ☠️





**Versuchen
wir es...**

**Können Sie
helfen, den
Streit zu
schlichten?**




Das wollen wir betrachten...





Kreuztabellen: Statistik

Chi-Quadrat


Nichtparametrische Tests


 Eine Stichprobe...


 Unabhängige Stichproben...

 Verbundene Stichproben...

Mittelwertvergleiche

 T-Test bei unabhängigen Stichproben...

 T-Test bei verbundenen Stichproben...

 Einfaktorielle ANOVA...

Und das ebenfalls



Kaplan-Meier

Zeit: Faktor vergleichen...

Status: Speichern...

Ereignis definieren...

Faktor: Optionen...

Schichten:

Fallbeschriftung:

OK Einfügen Zurücksetzen Abbrechen Hilfe

Kaplan-Meier: Faktorstufen vergleichen

Teststatistiken

Log-Rang Breslow Tarone-Ware

Linearer Trend für Faktorstufen

Gemeinsam über Schichten Paarweise über Schichten

Für jede Schicht Paarweise für jede Schicht

Weiter Abbrechen Hilfe

**Vorher aber
etwas ganz
Wichtiges**



1. Machen Sie sich mit den Daten vertraut.
2. Stellen Sie diese grafisch dar.
3. Prüfen sie die untersuchten Merkmale auf Vollständigkeit und Plausibilität
4. Formulieren Sie Ihre Frage möglichst präzise:

Nicht:

- Welche Chemotherapie ist besser?

Sondern:

Welche der beiden Therapien hat

- Eine höhere Response-Rate
- Weniger Nebenwirkungen mit Grad >3
- Weniger Therapieabbrecher
- Längeres Überleben
- Längere Zeit bis zur Tumorprogression

Beispiel



Welche der beiden Therapien hat eine höhere Response-Rate ?

Gesamtgruppe	Response	Keine Response	Gesamt	Rate
Chemo A	20	20	40	50%
Chemo B	16	24	40	40%
	36	44	80	



A besser

Nun Aufteilung nach Geschlecht



Frauen	Response	Keine Response	Gesamt	Rate
Chemo A	2	8	10	20%
Chemo B	9	21	30	30%
	11	29	40	



B besser

Männer	Response	Keine Response	Gesamt	Rate
Chemo A	18	12	30	60%
Chemo B	7	3	10	70%
	25	15	40	



B besser

Historisches Beispiel

Sterblichkeit aufgrund von Tuberkulose in New York und Richmond aus dem Jahre 1910

	Bevölkerung		Todesfälle		Sterberate / 100 000	
	New York	Richmond	New York	Richmond	New York	Richmond
Gesamt	4 766 883	127 628	8 878	286	186	224

Nun Aufteilung nach Hautfarbe



	Bevölkerung		Todesfälle		Sterberate / 100 000	
	New York	Richmond	New York	Richmond	New York	Richmond
Weiß	4 675 174	80 895	8 365	131	179	162



	Bevölkerung		Todesfälle		Sterberate / 100 000	
	New York	Richmond	New York	Richmond	New York	Richmond
Farbig	91 709	46 733	513	155	560	332



Simpsons Paradoxon

	Therapie A (Therapieerfolge/Patienten)		Therapie B (Therapieerfolge/Patienten)	
kleine Nierensteine	93 % (81/87)	😊	87 % (234/270)	
große Nierensteine	73 % (192/263)	😊	69 % (55/80)	
zusammen	78 % (273/350)	😞	83 % (289/350)	😊

Scheinkorrelationen können entstehen, wenn Daten gruppiert ausgewertet werden, es aber innerhalb der Gruppen eine Ungleichverteilung einer wichtigen Größe gibt (die kein Confounder sein muss). Dieses Phänomen wurde unter dem Begriff Simpson-Paradoxon bekannt.



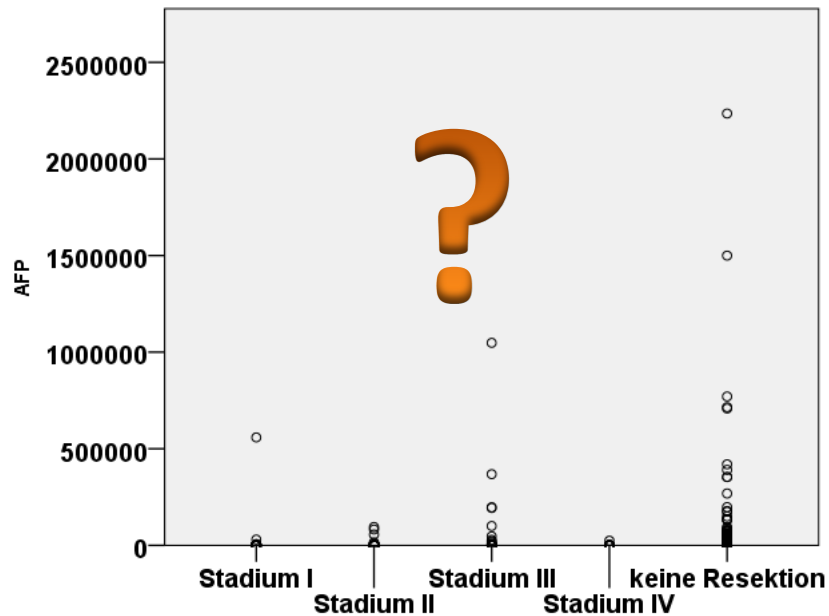
Vorsicht bei der Analyse von Subgruppen



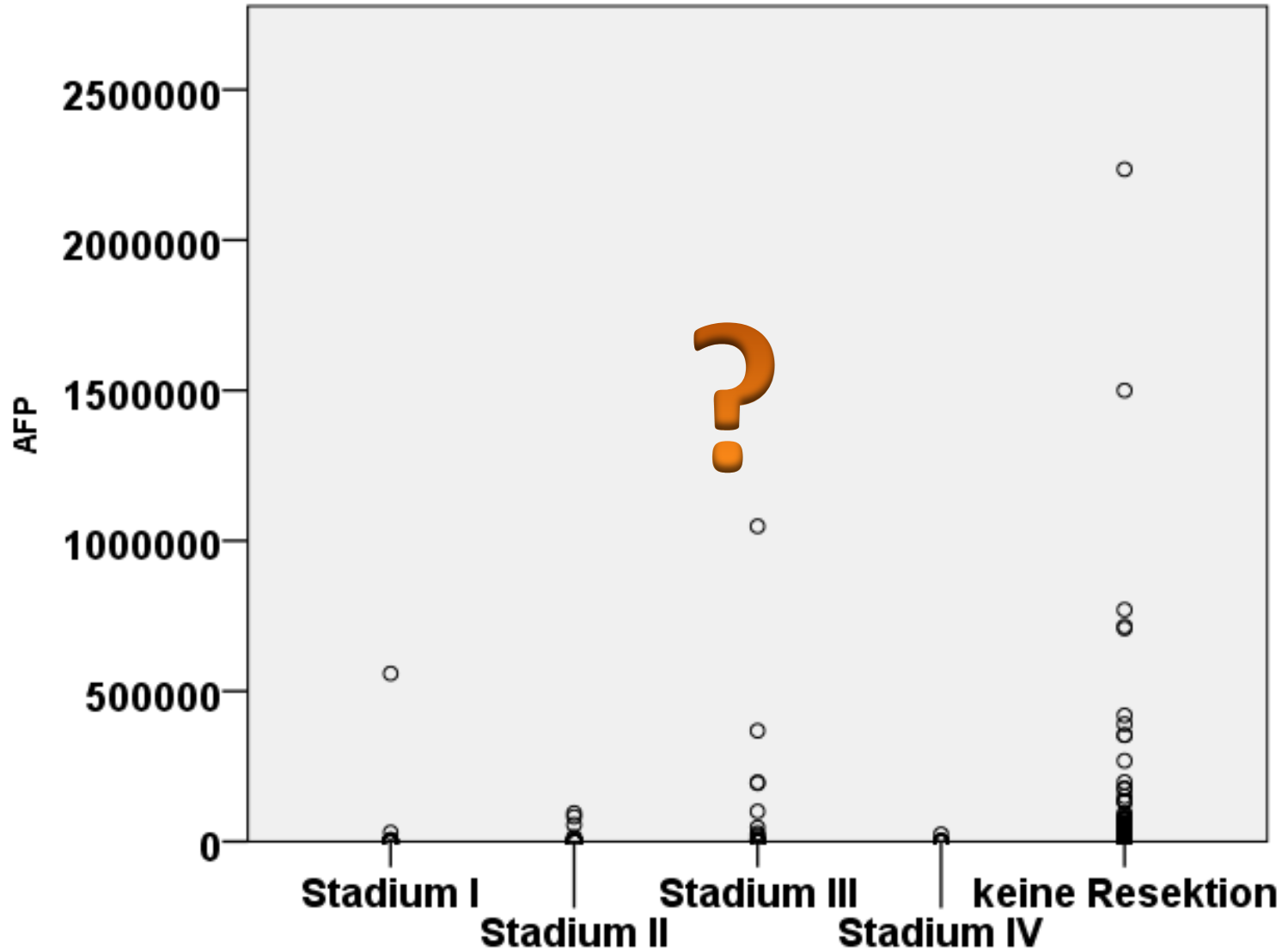
Manchmal muss man ein wenig nachdenken, bevor man sich ein gutes Bild von den eigenen Daten machen kann

Wir wollen uns anschauen, ob es einen Zusammenhang zwischen dem AFP-Wert und dem Tumorstadium in unseren Daten gibt.

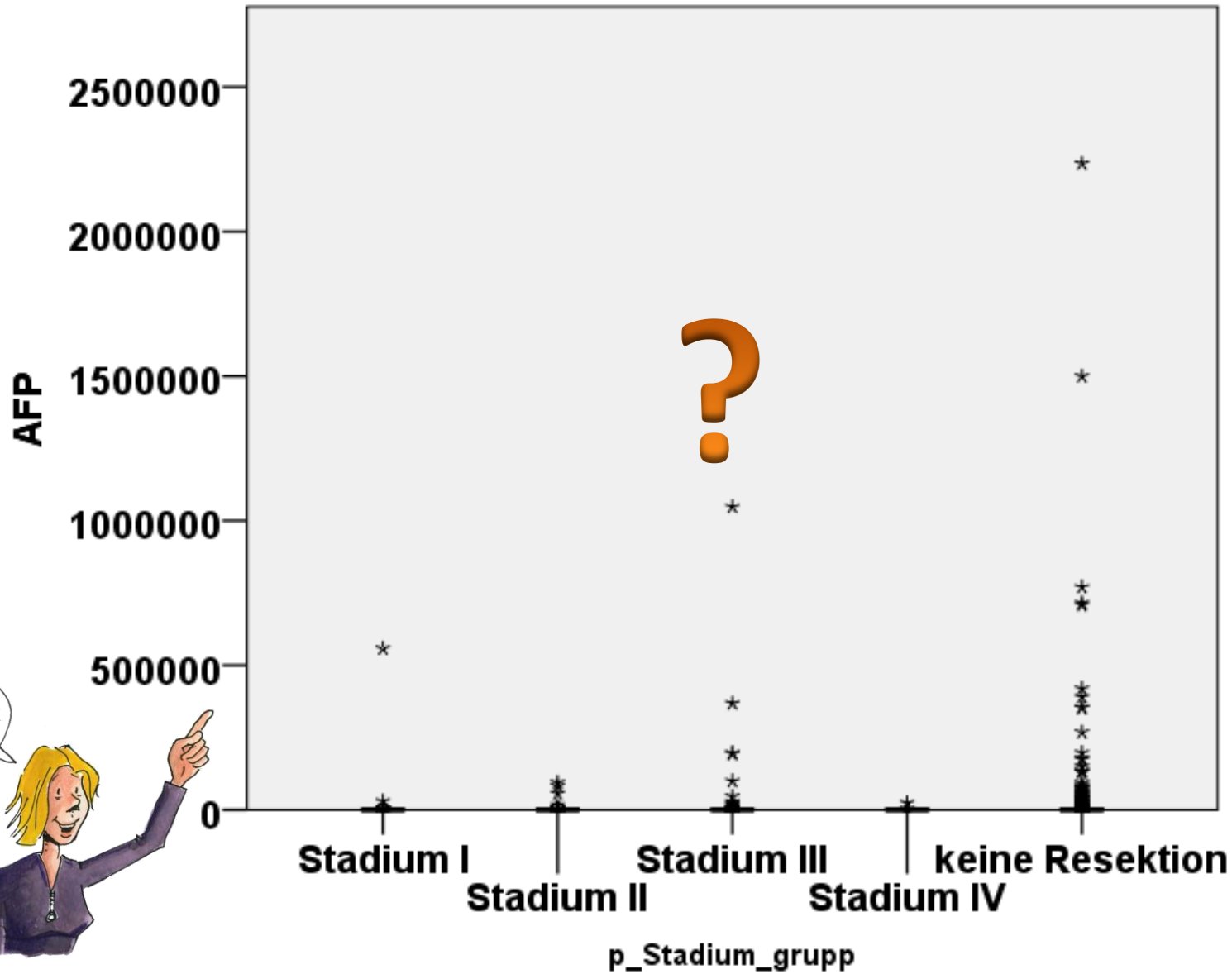
1. Versuch: Streudiagramm



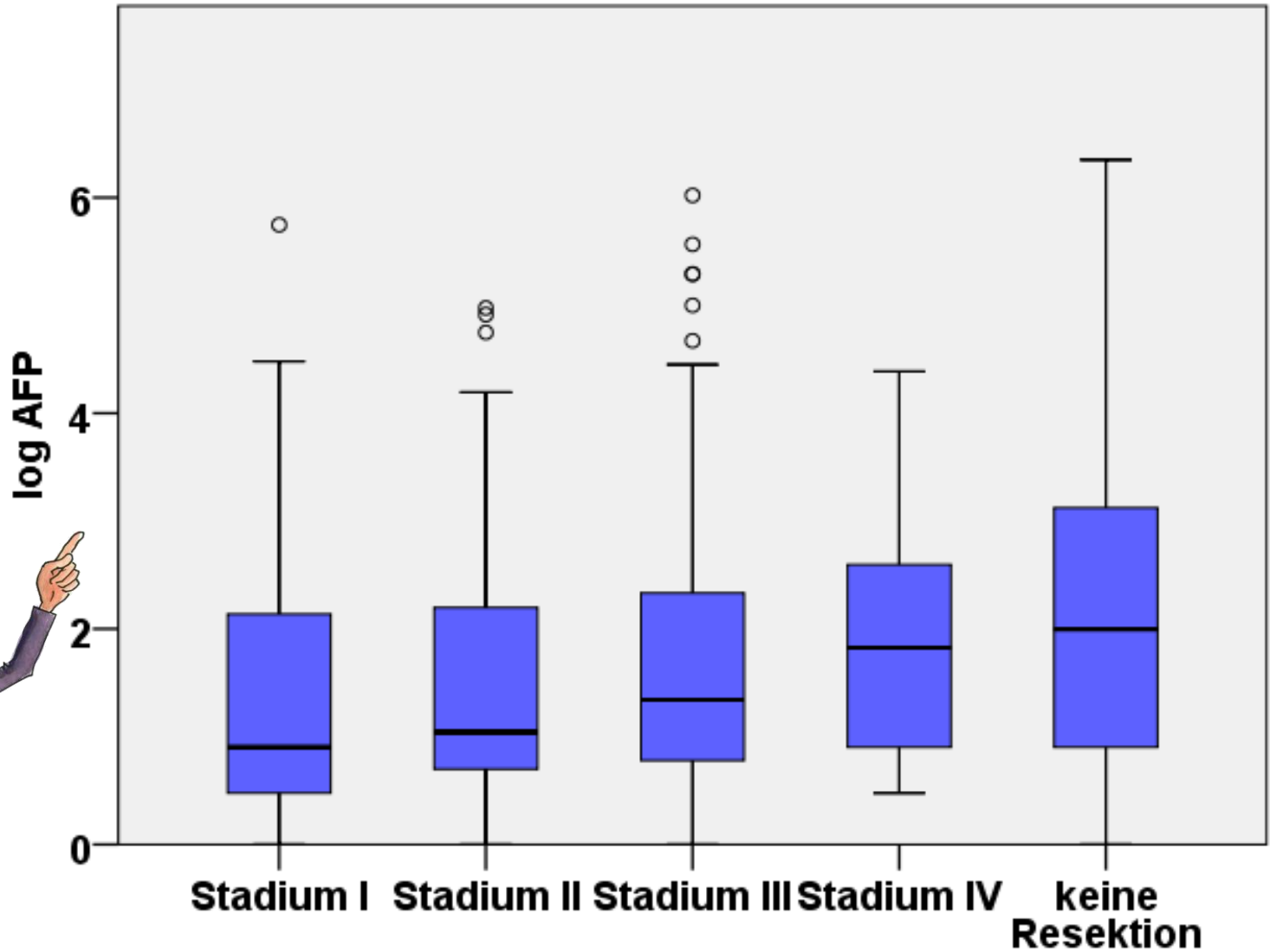
1. Versuch: Streudiagramm



2. Versuch: Boxplot



3. Versuch: Boxplot



**Lassen Sie uns nun
nachschaauen, welche
Qualitätsmerkmale in
Organzentren abgefragt
werden.**



Qualitätsparameter für Organzentren

Organ	Anzahl Prozeßqualität	Anzahl Ergebnisqualität
Darm	23	7
Pankreas	14	4
Leber	14	1
Magen	8	4
Haut	15	3
Brust	24	1
Lunge	17	3
Prostata	23	1
Kopf-Hals	14	1
Neuroonkologie	6	2

Qualitätsparameter für die Mortalität bei Darmkrebs	
Komplikationen bei...	
Revisionseingriffen (Kolon/Rektum)	
Postoperative Mortalität	
Anastomosen (getrennt nach Kolon/Rektum)	
Mortalität operativ	

Langzeitergebnisse uninteressant?



**Beginnen wir mit der
Untersuchung der
Anzahl von
Revisionseingriffen beim
Rektum.
Dies ist ein Merkmal der
Ergebnisqualität beim
Rektumkarzinom.**

KN	EB	Kennzahlen-Definition	Kennzahlenziel	Zähler	Grund- gesamt- heit (=Nenner)	Plausi- unklar	Soll- vorgabe
16	5.2. 7	Revisions- OP's Rektum	Möglichst niedrige Rate an Revisionsoperatione n nach elektiven Eingriffen	Revisionsoperationen infolge von perioperativen Komplikationen innerhalb von 30d nach elektiver OP	Elektive Rektum- Eingriffe	< 0,01%	≤ 10 %



Nun, der Statistiker bildet meist eine Nullhypothese, diese lautet hier: In allen drei Kliniken ist die Anzahl der Revisionseingriffe gleich. Diese versuchen wir nun mit statistischen Methoden zu widerlegen

Tun Sie's, sie werden sehen, ich bin der Beste!



Elektive Eingriffe beim Rektumkarzinom	Klinik A	Klinik B	Klinik C
Revisionseingriffe	8	4	0
Klinik ohne Revisionseingriff verlassen	42	46	8
Gesamt	50	50	8



Wir berechnen die relativen Häufigkeiten



Revisionseingriff	Klinik A		Klinik B		Klinik C	
	Anzahl	%	Anzahl	%	Anzahl	%
nein	42	88%	46	92%	8	100%
ja	8	16%	4	8%	0	0%
Total	50	100%	50	100%	8	100%



Schauen wir uns
an, wie genau
diese Schätzungen
sind

http://www.jaik.de/js/bin_konf.htm

Konfidenzintervall (oben) - Signifikanztest (unten)

Eingabe von X und n -> Konfidenzintervall

Einzugeben ist der **Stichprobenumfang n** und die **Zahl der Erfolge X**.
(z.B. 450 (=X) von 800 (=n) Salmanianern emfatieren)

X =	<input type="text" value="0"/>
n =	<input type="text" value="8"/>
Sicherheitswahrscheinlichkeit :	<input type="radio"/> 90% <input checked="" type="radio"/> 95% <input type="radio"/> 95,5% (2 σ)
... Konfidenzintervall ...	<input type="text" value="[0 ; 0.3244]"/> Klinik C \Rightarrow [0% bis 32 %]
	[0.0834 ; 0.2851] Klinik A \Rightarrow [8% bis 29 %]
	[0.0315 ; 0.1884] Klinik B \Rightarrow [3% bis 19 %]



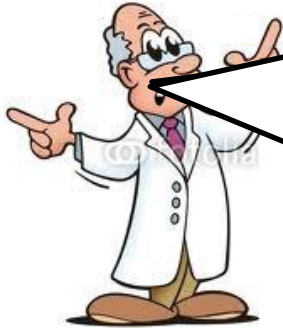
Wir sehen, da gibt es
große
Überschneidungen

Klinik B

Klinik A

Klinik C





**Die Nullhypothese lautet: Alle 3 Raten
sind gleich.**

**Wir wollen sehen, wie wahrscheinlich
diese Annahme ist. Wenn die
Wahrscheinlichkeit weniger als 5%
beträgt, dann lehnen wir die
Nullhypothese ab.**

Dazu machen wir einen χ^2 -Test

<http://www.quantpsy.org/chisq/chisq.htm>

CALCULATION FOR THE CHI-SQUARE TEST

An interactive calculation tool for chi-square tests of goodness of fit and independence

will let you know if there is a problem. In the 2 x 2 case of the chi-square test of independence, expected frequencies less than 5 are usually considered acceptable if Yates' correction is employed.

[Curriculum vitae](#)

[Selected publications](#)

[Supplemental material
for publications](#)

[Online utilities](#)

[Mediation & moderation
material](#)

[PSY-GS 321: Multilevel
Modeling](#)

[Vanderbilt Psychological
Sciences](#)

[Vanderbilt Quantitative
Methods](#)

[Organizations](#)

[Friends and colleagues](#)

[Contact me](#)

	Gp 1	Gp 2	Gp 3	Gp 4	Gp 5	Gp 6	Gp 7	Gp 8	Gp 9	Gp 10	
Cond. 1:	8	4	0								12
Cond. 2:	42	46	8								96
Cond. 3:											0
Cond. 4:											0
Cond. 5:											0
Cond. 6:											0
Cond. 7:											0
Cond. 8:											0
Cond. 9:											0
Cond. 10:											0
	50	50	8	0	0	0	0	0	0	0	108

Output:

Chi-square: 2.7

degrees of freedom: 2

p-value: 0.25924026

© 2010-2015,
Kristopher J. Preacher

> 0,05

Folglich: Nullhypothese (alle drei sind gleich) beibehalten

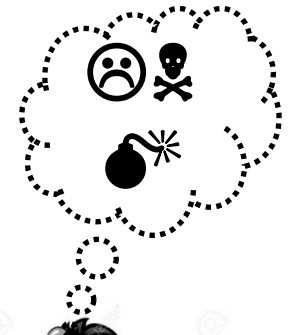


Das mache ich sehr gerne.

Doch zunächst noch einige Bemerkungen zu dem gerade durchgeführten Test.



Lieber Herr Professor,
Sie wollen doch nicht abstreiten, dass
0% viel, viel besser ist als 16%!
Wiederholen Sie doch Ihren Test mit
meinen Daten und denen des
Kollegen aus der A-Klinik.



1. Bemerkung

	Gp 1	Gp 2	Gp 3	Gp 4	Gp 5	Gp 6	Gp 7	Gp 8	Gp 9	Gp 10
Cond. 1:	8	4	0							12
Cond. 2:	42	46	8							96
Cond. 3:										0
Cond. 4:										0
Cond. 5:										0
Cond. 6:										0
Cond. 7:										0
Cond. 8:										0
Cond. 9:										0
Cond. 10:										0
	50	50	8	0	0	0	0	0	0	108

Output:

Chi-square: 2.7
degrees of freedom: 2
p-value: 0.25924026
Yates' chi-square: 1.183
Yates' p-value: 0.55349642

Status: At least one expected frequency is less than 1 !!

2. Bemerkung

	Gp 1	Gp 2	Gp 3	Gp 4	Gp 5	Gp 6	Gp 7	Gp 8	Gp 9	Gp 10	
Cond. 1:	8	4	0								12
Cond. 2:	42	46	8								96
Cond. 3:											0
Cond. 4:											0
Cond. 5:											0
Cond. 6:											0
Cond. 7:											0
Cond. 8:											0
Cond. 9:											0
Cond. 10:											0
	50	50	8	0	0	0	0	0	0	0	108

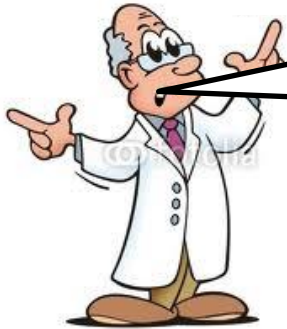
Output:

Chi-square: 2.7
degrees of freedom: 2
p-value: 0.25924026
Yates' chi-square: 1.183
Yates' *p*-value: 0.55349642

Status: **At least one expected frequency is less than 1 !!**



???



Der Erwartungswert ist diejenige Anzahl von Fällen, die dann aufgetreten wäre, wenn unsere Nullhypothese genau zuträfe.

also wenn tatsächlich die Re-Operationen pro Klinik genau gleich häufig aufgetreten wären.

Revisionseingriff	Klinik A		Klinik B		Klinik C	
	Anzahl	EW	Anzahl	EW	Anzahl	EW
nein	42	44,4	46	44,4	8	7,1
ja	8	5,6	4	5,6	0	0,9
Gesamt	50	50	50	50	8	8

Falls wir einen statistisch signifikanten Wert $p < 0,05$ erhalten hätten, wäre dieser nicht verwertbar gewesen, da (Faustregel) 95 % aller Erwartungswerte > 5 sein sollten. Hier ist $1/6 (=17\%) < 5$.



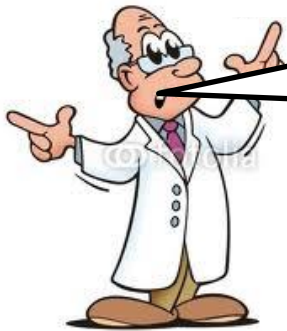
Wir merken
uns:

Revisionseingriff	Klinik A		Klinik B		Klinik C	
	Anzahl	EW	Anzahl	EW	Anzahl	EW
nein	42	44,4	46	44,4	8	7,1
ja	8	5,6	4	5,6	0	0,9
Gesamt	50	50	50	50	8	8

Bei kleinen Fallzahlen sollte man sich mit der Beschreibung der Daten begnügen.

In unserem Beispiel:

In der C-Klinik wurden 8 Patienten elektiv wegen eines Rektumkarzinoms operiert. Bei diesen wurde kein Revisionseingriff durchgeführt.



Wenn wir nur 2 Kliniken betrachten, erhalten wir eine Vierfeldertafel. Für diese – und nur für diese – gibt es den exakten Fisher-Test. Dieser funktioniert auch bei kleinen Zahlen.

<http://www.langsrud.com/fisher.htm>

Fisher's Exact Test

COMPUTE

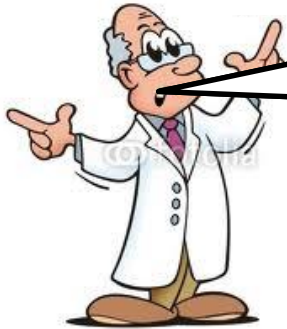
42	8
8	0

Fisher's Exact Test
<http://www.langsrud.com/fisher.htm>

TABLE = [42 , 8 , 8 , 0]
Left : p-value = 0.280091508329541
Right : p-value = 1
2-Tail : p-value = 0.5831196155095191

3 p-Werte, darf ich einen aussuchen ?





Auch dieser Test zeigt keinen Unterschied zwischen den Kliniken A und C.

Bisher haben wir also noch keinen Sieger im Wettbewerb um die beste Ergebnisqualität.

Fisher's Exact Test

COMPUTE

42	8
8	0

```
Fisher's Exact Test
http://www.langsrud.com/fisher.htm
-----
TABLE = [ 42 , 8 , 8 , 0 ]
Left   : p-value = 0.280091508329541
Right  : p-value = 1
2-Tail : p-value = 0.5831196155095191
-----
```

Nullhypothese beibehalten
 $p > 0,05$

<http://www.langsrud.com/fisher.htm>



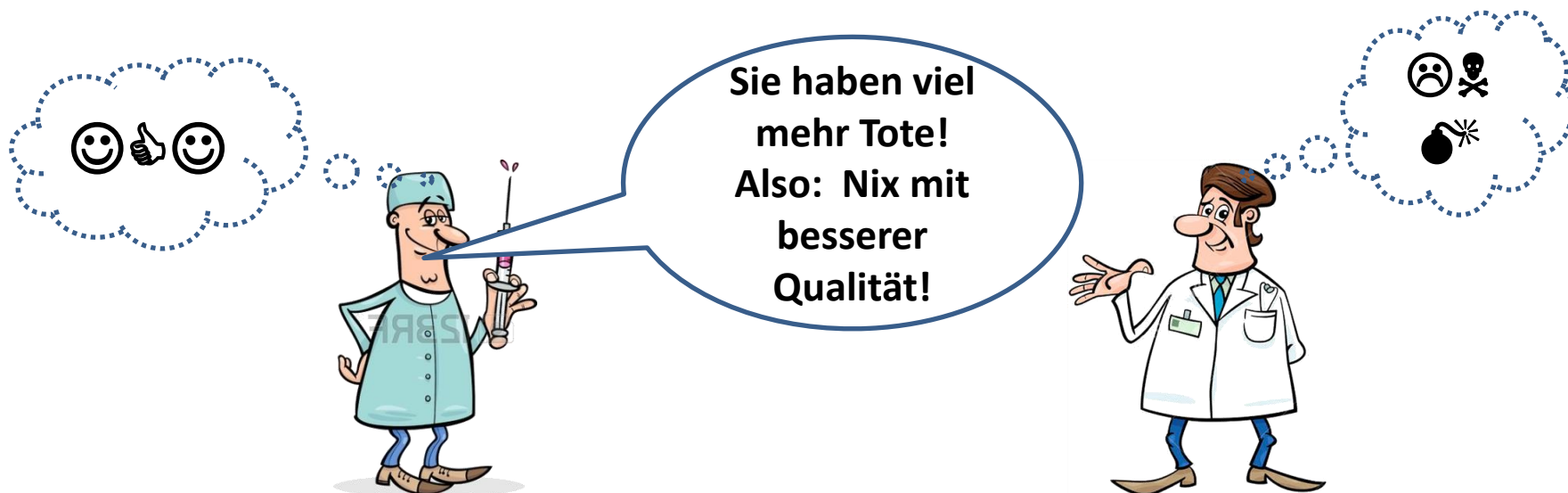
Was ist mit der Letalität? Da sind die Zahlen größer



Betrachten wir die Letalität



In der Klinik postoperativ verstorben	Klinik A		Klinik B		Klinik C	
	Anzahl	%	Anzahl	%	Anzahl	%
nein	116	97%	93	93%	18	90%
ja	4	3%	7	7%	2	10%
Total	120	100%	100	100%	20	100%





Hier geht es um die 30-Tage-Letalität,
n i c h t um die Klinikletalität.
Den Unterschied zeige ich Ihnen.

KN	EB	Kennzahl- n- Definition	Kennzahlenziel	Zähler	Grund- gesamt- heit (=Nenner)	Plausi- unklar	Soll- vorgabe
16	5.2.7	Mortalität post- operativ	Möglichst niedrige Rate an postoperativ verstorbenen Patienten nach elektiven Eingriffen	Postoperativ verstorbenen nach elektiven Eingriffen innerhalb von 30d	Elektiv operierte Patienten	< 0,01%	≤ 5 %

Tag postoperativ	Klinik A verstorben	Klinik B verstorben	Klinik C verstorben
1			
2		1	
...			
8	1	1	
9			1
10			
11	1		
12	1		1
13		1	
...			
20	1		
21	1*	1	
..			
28	1**	1	
29			
30			
31			
32		1***	
33			
34			
35		1***	

- Nach Entlassung in Geriatrie verstorben
- ** Nach Entlassung in Reha verstorben

- *** gehört nicht zur 30-Tage-Letalität



Nach Herstellen der Beobachtungsgleichheit haben wir folgende Situation:

30-Tage-Letalität	Klinik A		Klinik B		Klinik C	
	Anzahl	%	Anzahl	%	Anzahl	%
nein	114	95%	95	95%	18	90%
ja	6	5%	5	5%	2	10%
Total	120	100%	100	100%	20	100%



**Wir merken
uns:**

Bei postoperativer Sterblichkeit ist es wichtig

- **Klinikletalität**
- **30-Tage-Letalität**

zu unterscheiden.

Es können erhebliche Unterschiede auftreten.





**Aber ! Meine
Patienten sind
alle viiiiiel älter**



**Alter ist
doch kein
Qualitäts-
merkmal**



**Betrachten wir die
Altersverteilungen,
Aber vorher
Wiederholung
einiger Grundlagen**



Merkmal

```
graph TD; A[Merkmal] --> B[diskret]; A --> C[stetig];
```

diskret

endlich (oder abzählbar
unendlich) viele
Ausprägungen/Werte

Beispiele aus der Tumordokumentation:

- Geschlecht,
- Histologie-Codes,
- Tumorstadium
- Prozeduren,
- Blutgruppen,
- Anzahl Blutkonserven,
- Postoperative Sterblichkeit
- ...

stetig

alle (reellen) Zahlen eines
Intervalls können mögliche
Ausprägungen/Werte sein

Beispiele aus der Tumordokumentation:

- Alter,
- Tumormarker,
- Operationsdauer,
- Temperatur
- Überlebenszeit
- ...

Merkmal

```
graph TD; A[Merkmal] --> B[nominalskaliert]; A --> C[ordinalskaliert]; A --> D[intervallskaliert];
```

nominalskaliert

Ausprägungen sind
Namen, Ordnung nicht
sinnvoll

Beispiele:

- Geschlecht,
- Histologie-Codes,
- Prozeduren,
- Blutgruppen,
- ...

ordinalskaliert

Ausprägungen/
Werte können
geordnet, aber
Abstände nicht
interpretiert werden

Beispiele:

- Tumorstadium
- Anastomosen-
Insuffizienzen
- Lokalisation im
Kolon
- Komplikationen
- ...

intervallskaliert

Ausprägungen/Werte
sind Zahlen, deren
Abstände
interpretiert werden
können

Beispiele:

- Alter,
- Tumormarker,
- Operationsdauer,
- Temperatur
- ...


Typen von Merkmalen/Kenngrößen/Variablen

Diskret*	endlich oder abzählbar unendlich viele Ausprägungen/Werte
stetig	alle (reellen) Zahlen eines Intervalls können mögliche Ausprägungen/Werte sein
nominalskaliert	Ausprägungen sind Namen, Ordnung nicht sinnvoll
ordinalskaliert	Ausprägungen/Werte können geordnet, aber Abstände nicht interpretiert werden
intervallskaliert	Ausprägungen/Werte sind Zahlen, deren Abstände interpretiert werden können

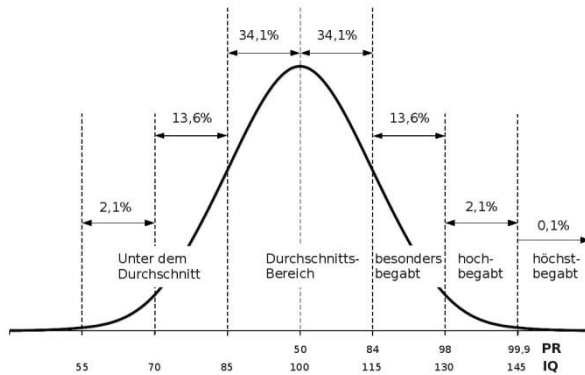
* Sonderfall: Diskret und 0-1 (ja/nein) codiert



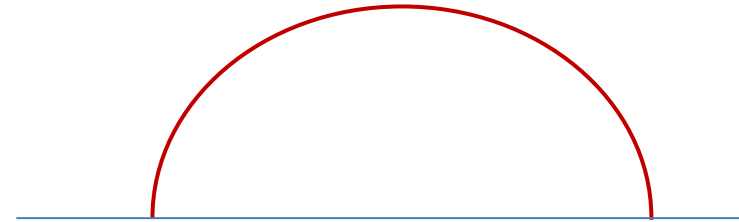
Im SPSS sieht das so aus:

AFP	 Skala
Alter	 Skala
pT_Kat	 Ordinal
pN_K	 Ordinal
M_Kat	 Ordinal
p_Stadium_...	 Ordinal
Therapie	 Nominal
Kompl	 Nominal
Name	 Nominal

Typen von Verteilungen



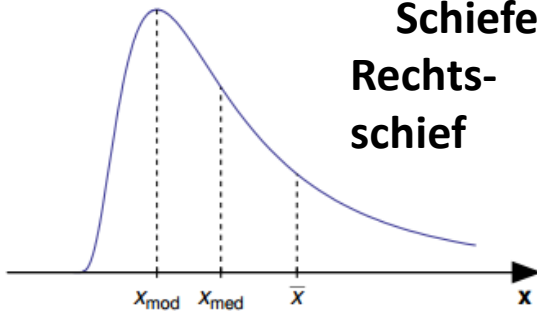
Normalverteilung



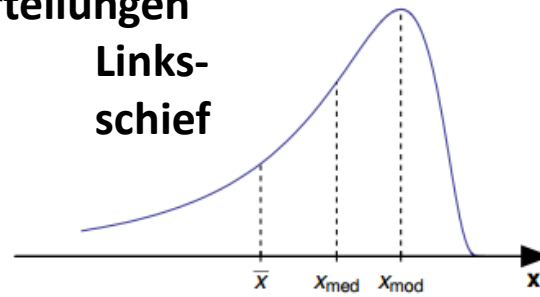
Symmetrische Verteilung

Schiefe Verteilungen

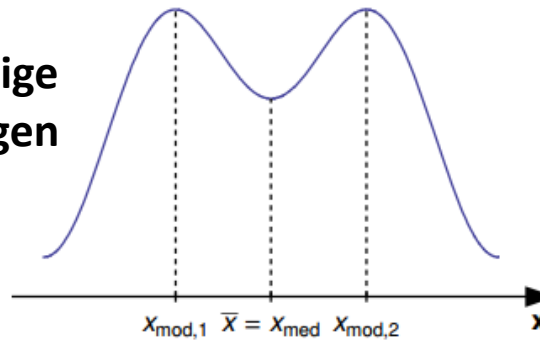
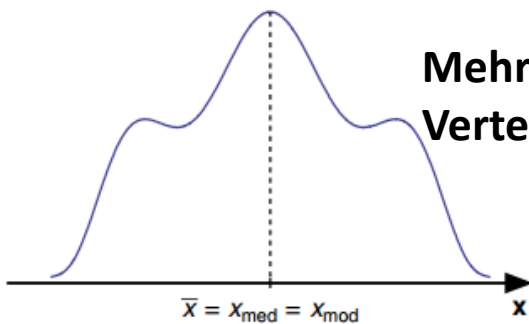
Rechts-schief



Links-schief

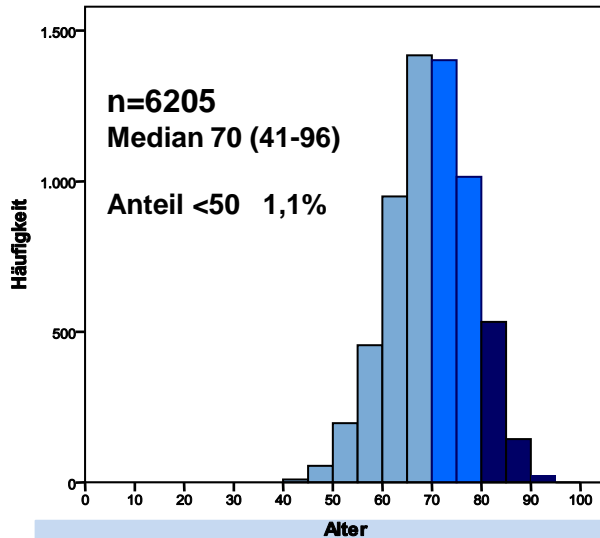


Mehrgipflige Verteilungen



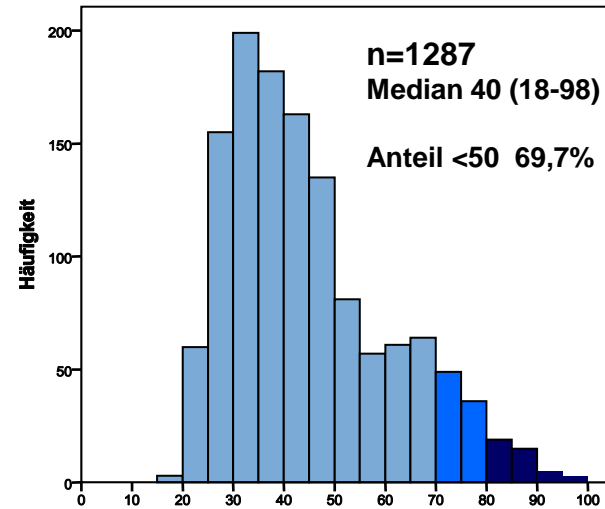
Alters-Verteilungen aus dem Tumorregister Bayreuth

Prostata



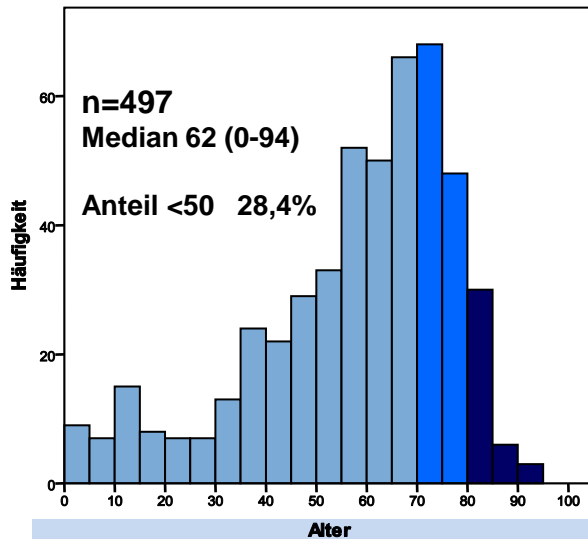
Annähernd symmetrisch

Cervix
uteri



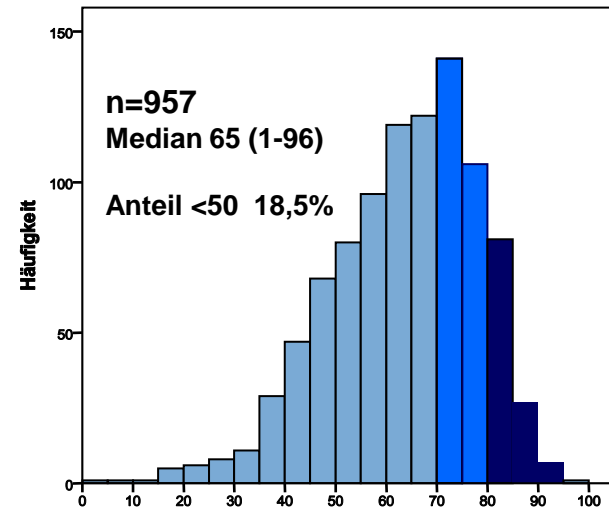
linksschief

Gehirn



mehrgipflig

Ovar



rechtsschief



Da gibt es eine
Reihe von
Möglichkeiten.
Lassen Sie mich
einige beschreiben.
Die erste Methode
kennen Sie schon,
wir machen einen
 χ^2 -Test.
Dazu bilden wir
Altersklassen.

Klingt ziemlich
kompliziert. Wie
vergleicht man
nun die
Altersverteilungen
?



Alter	Klinik A		Klinik B		Klinik C	
	Anzahl	%	Anzahl	%	Anzahl	%
<80 Jahre	80	67%	73	73%	15	75%
80 Jahre und älter	40	33%	27	27%	5	25%
Total	120	100%	100	100%	20	100%

χ^2 -Test:
Kein
signifikanter
Unterschied

Alter	Klinik A		Klinik B	
	Anzahl	%	Anzahl	%
<65 Jahre	56	47%	57	57%
75-84 Jahre	38	32%	30	30%
85 Jahre und älter	26	22%	13	13%
Total	120	100%	100	100%

χ^2 -Test:
Kein
signifikanter
Unterschied

Alter	Klinik A		Klinik B	
	Anzahl	%	Anzahl	%
<65 Jahre	27	23%	30	30%
65-74 Jahre	29	24%	27	27%
75-84 Jahre	38	32%	30	30%
85 Jahre und älter	26	22%	13	13%
Total	120	100%	100	100%

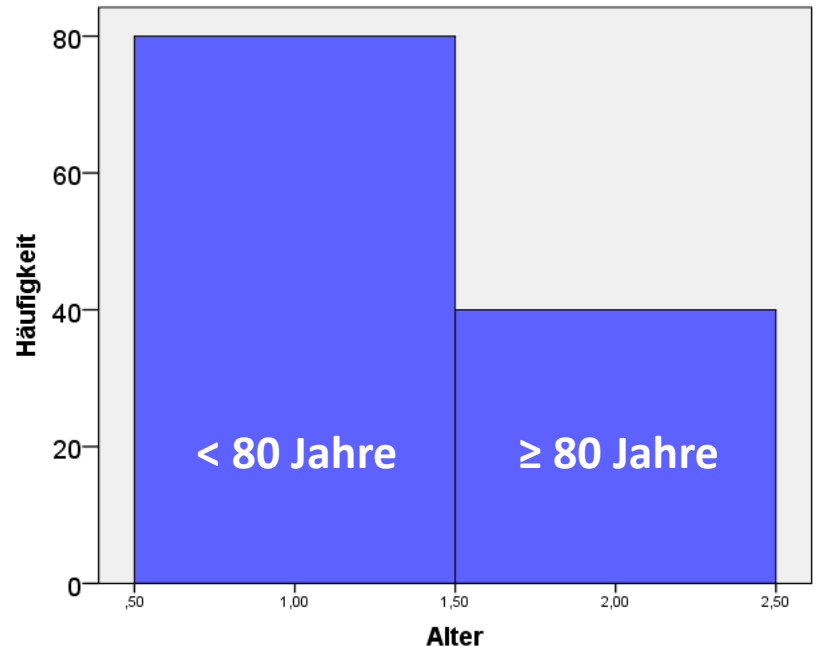
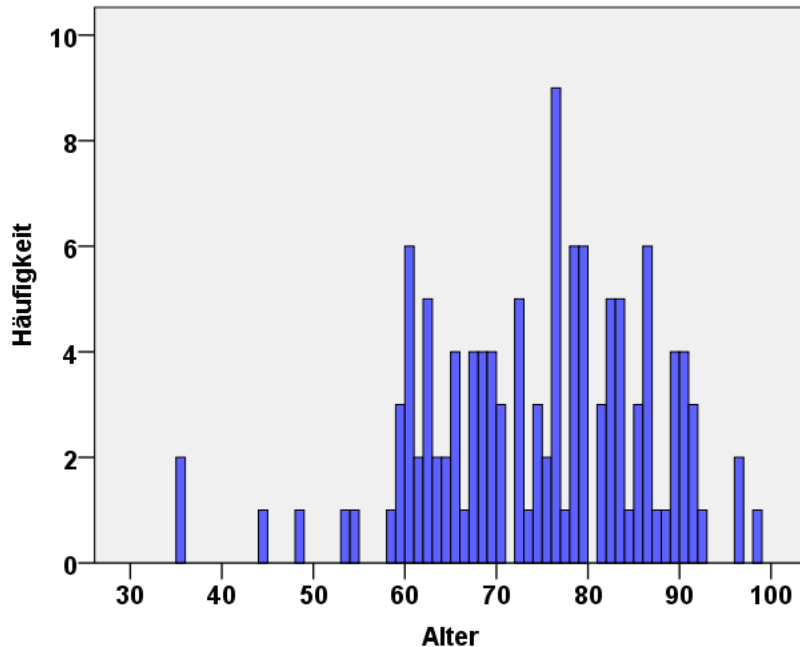
χ^2 -Test:
Kein
signifikanter
Unterschied



Durch die Klasseneinteilung geht natürlich eine Menge Information verloren.

Auch sollte man die gewählte Klasseneinteilung begründen, z.B. damit, dass Patienten ≥ 80 Jahre eine besondere (Risiko-) Gruppe darstellen.

Es gibt Tests die keine Klasseneinteilung erfordern.





Wie vergleichen wir nun unsere beiden Verteilungen. Prinzipiell gibt es mehrere Möglichkeiten, z.B.:

1. Vergleich der Mediane
2. Vergleich der Rangfolge der Messwerte
3. Vergleich der Mittelwerte

Was bitte ist ein Median?



Median nennt man einen Wert, wenn genauso viele kleinere wie größere Werte existierten.

Beispiel

98 87 84 61 79 81 60 69 78 91 72

sortiert

60 61 69 72 78 79 81 84 87 91 98

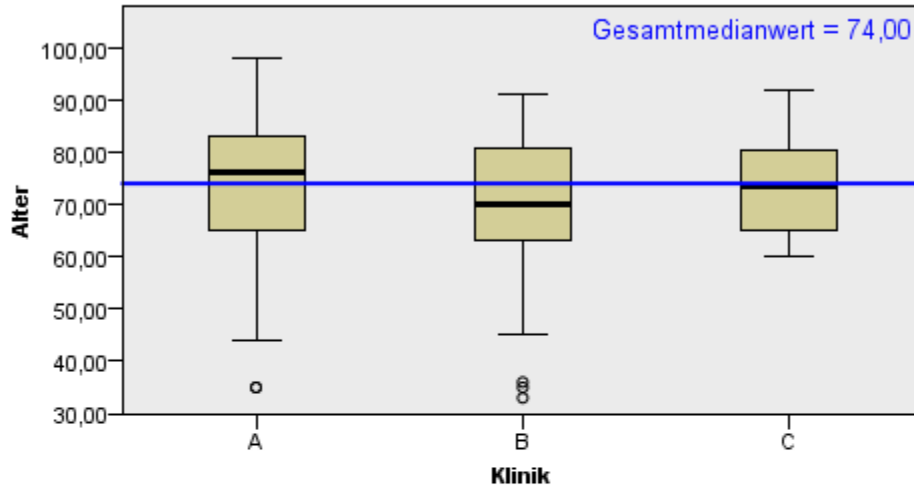


Median

Der **Median-Test** untersucht, ob zwei oder mehr unabhängige Stichproben aus Grundgesamtheiten mit gleichem Median stammen.

1. Vergleich der Mediane : Mediantest

Mediantest unabhängiger Stichproben



Da der Test **keine Annahmen hinsichtlich der Häufigkeitsverteilung der Daten** voraussetzt, zählt er zu den **nicht-parametrischen Verfahren**.

Gesamtanzahl	240
Median	74,000
Teststatistik	2,429
Freiheitsgrade	2
Asymptotische Sig. (zweiseitiger Test)	,297

Nullhypothese beibehalten

2. Vergleich der Rangfolge der Messwerte

Mann-Whitney-U-Test

für 2 Stichproben

Kruskal-Wallis-Test

für mehr als 2 Stichproben

Stichprobe Rot

98 87 84 61 79

Stichprobe Blau

90 88 75 83 54 60 81

sortiert

	54	60	61	75	79	81	83	84	87	88	90	98
Rang	1	2	3	4	5	6	7	8	9	10	11	12
Rang- summe	Rot : 3+5+8+9+12=37							Blau : 1+2+4+6+7+10+11=41				

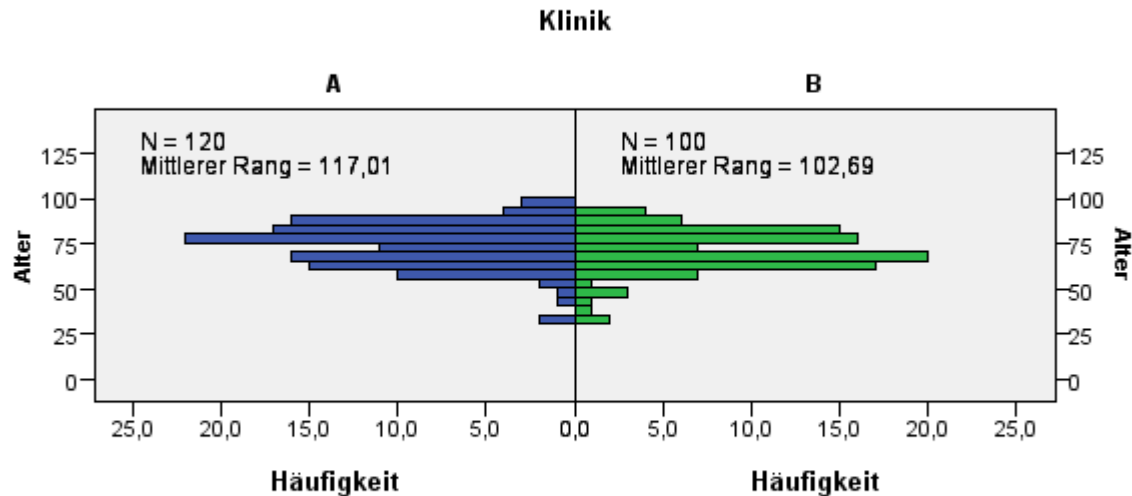


Beide Tests sind nichtparametrische Testverfahren,
da sie keine Annahmen über die Häufigkeitsverteilung machen.



2. Vergleich der Rangfolge der Messwerte

Mann-Whitney-U-Test



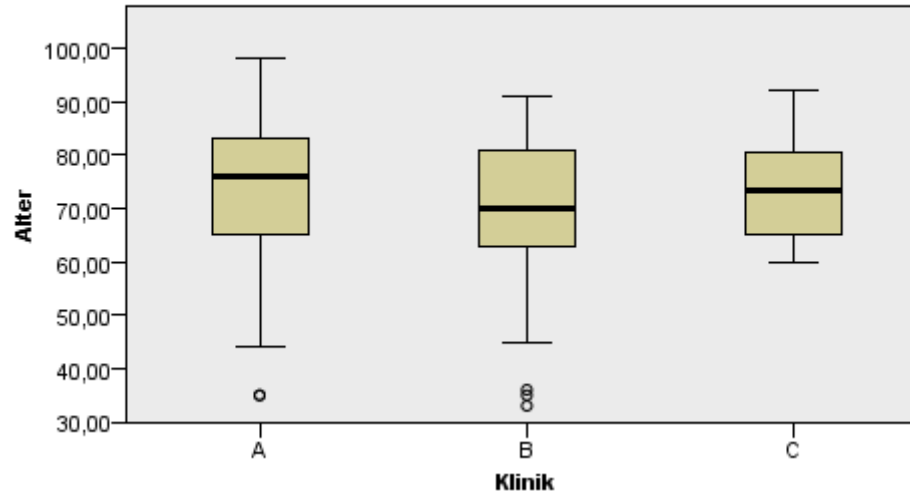
Gesamtanzahl	220
Asymptotische Sig. (zweiseitiger Test)	,096

Nullhypothese
beibehalten

- Voraussetzung : unabhängige Stichproben (Gruppen oder Messreihen)
- ordinalskalierten Variable

2. Vergleich der Rangfolge der Messwerte

Kruskal-Wallis-Test



Gesamtanzahl	240
Teststatistik	2,914
Freiheitsgrade	2
Asymptotische Sig. (zweiseitiger Test)	,233

**Nullhypothese
beibehalten**

- **Voraussetzung : unabhängige Stichproben** (Gruppen oder Messreihen)
- **ordinalskalierten Variable**

2. Vergleich der Mittelwerte: T-Test

Der T-Test prüft ob sich die Mittelwerte zweier unabhängiger Stichproben unterscheiden.

Voraussetzung ist, dass die Daten der Stichproben einer

- normalverteilten Grundgesamtheit entstammen.

Der klassische t-Test setzt voraus, dass beide Stichproben aus Grundgesamtheiten mit

- gleicher Varianz entstammen.
- Der Welch-Test oder t-Test nach Satterthwaite ist eine Variante, die die Gleichheit der Varianzen nicht voraussetzt.



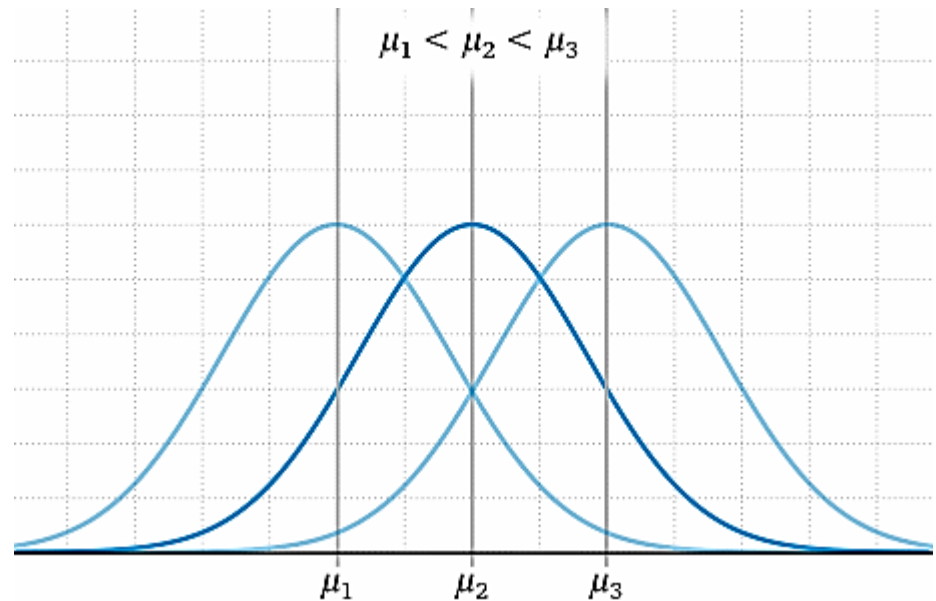
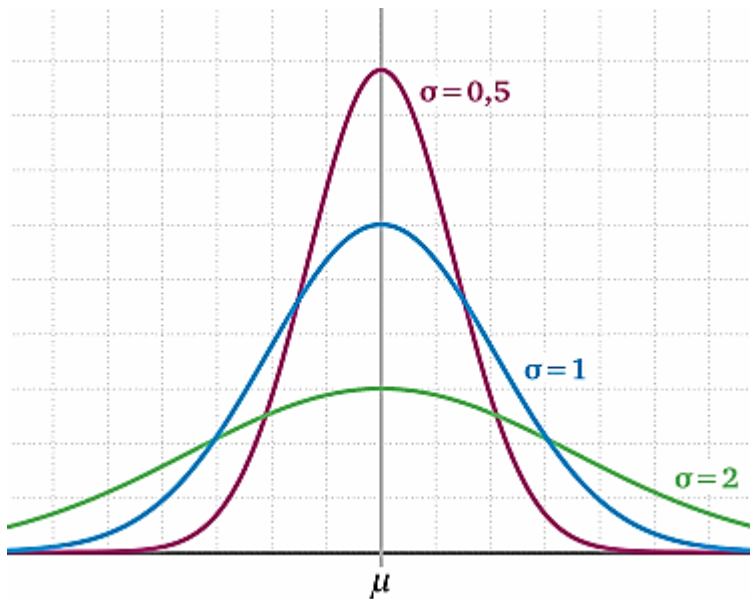
Der T-Test macht Annahmen über die Form (Normalverteilung) und Varianz der Verteilung, deshalb nennen wir ihn einen **Parametrischen Test**

Die Normalverteilung

Manche Verteilungen lassen sich ganz einfach beschreiben.
Für uns ist die Normalverteilung wichtig. Sie lässt sich durch die Parameter

- Mittelwert μ
- Standardabweichung σ eindeutig festlegen.

Beispiele:





Prüfen wir ob das Alter in den beiden Kliniken normal verteilt ist. Dazu benutzen wir den **Kolmogorov-Smirnov-Test**.
Alter ist ein stetiges Merkmal, dies ist die Voraussetzung für diesen Test.

Mit ihm kann man testen, ob eine bekannte Verteilung vorliegt:

Kolmogorov-Smirnov-Test: Optionen [X]

-Hypothetische Verteilungen-

Normal

Verteilungsparameter

Stichprobendaten verwenden
 Benutzerdefiniert

Mittelwert: 0 Standardabw.: 1

Gleichverteilung

Verteilungsparameter

Stichprobendaten verwenden
 Benutzerdefiniert

Min: 0 Max: 1

Exponentialverteilung **P**oisson-Verteilung

Mittelwert

Stichprobenmittelwert
 Benutzerdefiniert

Mittelwert: 0

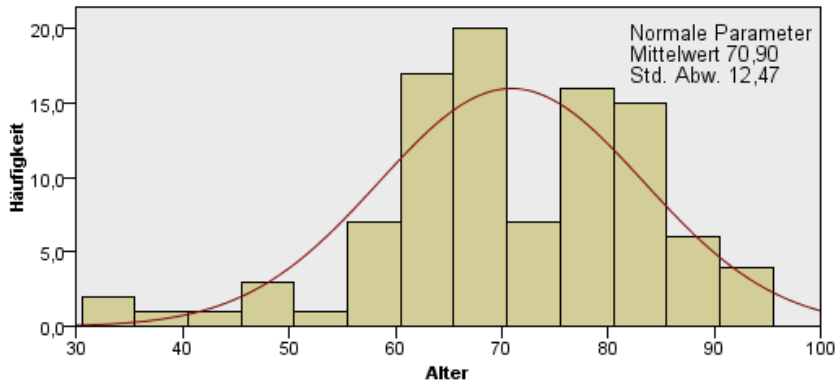
Mittelwert

Stichprobenmittelwert
 Benutzerdefiniert

Mittelwert: 0

OK Abbrechen Hilfe

Kolmogorov-Smirnov-Test einer Stichprobe

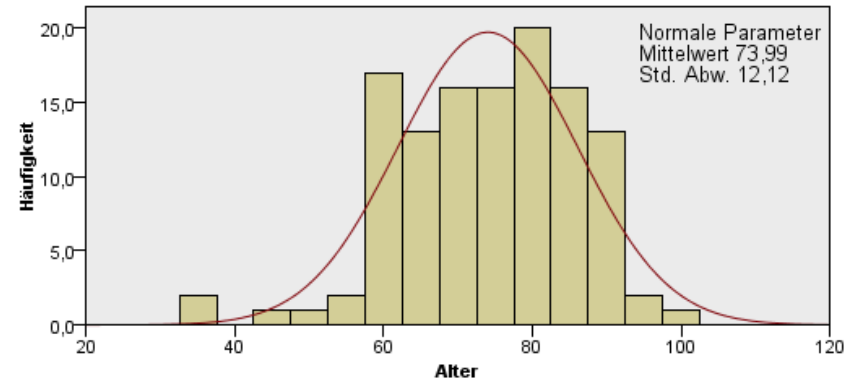


Asymptotische Sig. (zweiseitiger Test)

,426

Nullhypothese beibehalten
Alter in Klinik A ist normalverteilt

Kolmogorov-Smirnov-Test einer Stichprobe



Asymptotische Sig. (zweiseitiger Test)

,388

Nullhypothese beibehalten
Alter in Klinik B ist normalverteilt



**Wir dürfen
parametrische
Tests anwenden,
machen wir den
T-Test**

T-Test bei unabhängigen Stichproben

Gruppenstatistiken

	Klinik	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Alter	A	120	73,992	12,1195	1,1064
	B	100	70,900	12,4677	1,2468

		Levene-Test der Varianzgleichheit	
		F	Signifikanz
Alter	Varianzen sind gleich	,015	,903
	Varianzen sind nicht gleich		

Varianzen sind gleich

**Nullhypothese beibehalten
Mittelwerte des Alters sind gleich**

T-Test für die Unabhängigkeit						
T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
					Untere	Obere
1,860	218	,064	3,0917	1,6626	-,1851	6,3684
1,855	208,661	,065	3,0917	1,6669	-,1944	6,3777



Dazu brauchen wir
die
Einfaktorielle
Varianzanalyse
ANOVA

(ANalysis Of
Variance)

Können wir auch
mehr als 2
Mittelwerte
vergleichen?



Einfaktorielle Varianzanalyse ANOVA (ANalysis Of Variance)

Alter

	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
A	120	73,992	12,1195	1,1064	71,801	76,182	35,0	98,0
B	100	70,900	12,4677	1,2468	68,426	73,374	33,0	91,0
C	20	73,900	8,7714	1,9613	69,795	78,005	60,0	92,0
Gesamt	240	72,696	12,0811	,7798	71,160	74,232	33,0	98,0

Test der Homogenität der Varianzen

Alter

Levene-Statistik	df1	df2	Signifikanz
1,448	2	237	,237


**Nullhypothese beibehalten
Varianzen sind gleich**

**Nullhypothese beibehalten
Mittelwerte des Alters sind gleich**

Einfaktorielle ANOVA

Alter

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	553,004	2	276,502	1,909	,151
Innerhalb der Gruppen	34329,792	237	144,851		
Gesamt	34882,796	239			





Diese ganze
Testerei ist
Unsinn!

Jawohl, jedes Kind
sieht, dass es
Unterschiede
gibt zwischen 0%
und 10%.



Sie
schulden
uns eine
Erklärung

Einverstanden

Sie haben alle Recht.

- 
1. Es ist tatsächlich unsinnig, etwas – in unserem Fall die Altersverteilungen – mit 8 verschiedenen Tests zu vergleichen. Dies geschah hier nur aus didaktischen Gründen.
- 

Richtig ist, sich vorher den geeigneten Test auszusuchen und das Ergebnis zu akzeptieren.

- 
2. Vermeiden Sie möglichst Prozentangaben bei weniger als 100 Fällen. Und – wenn unumgänglich – geben Sie wenigstens keine Nachkommastellen an.
- 

Prozent kommt aus dem italienischen *per cento*, „von Hundert“, wenn man bei insgesamt < 100 Fällen diese Bezeichnung anwendet, begeht man also bereits einen statistisch nicht gerechtfertigten Fehler auf eine größere Grundgesamtheit zu schließen.



3. Statistische Tests dienen dazu, zu prüfen, ob Einzel-Beobachtungen allgemeingültig sind.



Solange Sie sich darauf beschränken, Ihre Beobachtung zu beschreiben, machen sie nichts falsch.

Sie dürfen also sagen:

„In der C-Klinik musst im vergangenen Jahr von 8 Patienten mit Rektumkarzinom keiner revidiert werden, in der A-Klinik waren es 8 von 50.“

N i c h t sagen dürfen Sie hingegen:

„In der C-Klinik werden weniger Rektumkarzinome revidiert als in der A-Klinik“, denn die Tests haben gezeigt, dass diese Verallgemeinerung nicht zulässig ist. Es wäre durchaus möglich, dass die Beobachtungen im nächsten Jahr ganz anders sind.

4. Wir hatten sehr kleine Zahlen

Ihre Stichprobenumfänge (8 bzw. 50 Patienten) waren klein, bei z.B. 40 und 250 Patienten (also dem 5-fachen Stichprobenumfang) hätte das Ergebnis anders ausgesehen.

**Fisher's
Exact Test**

COMPUTE

0	40
40	210

Fisher's Exact Test
<http://www.langsrud.com/fisher.htm>

TABLE = [0 , 40 , 40 , 210]
Left : p-value = 0.001637407619077632
Right : p-value = 1
2-Tail : p-value = 0.0024400274843181697



Nullhypothese ablehnen, $p=0,002$. Die Revisionsrate ist unterschiedlich und zwar deutlich geringer in der C-Klinik

O.K.,
Sie kriegen
noch eine
Chance.

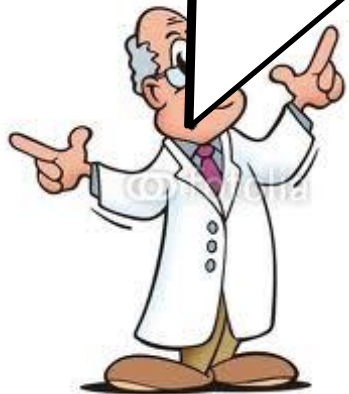
Aber diesmal keine
Tricks, von wegen
kleine Fallzahl oder
nicht normalverteilt
und solcher
Schnickschnack.

Finden Sie
raus, wer
die besten
Überlebens-
kurven hat

Einverstanden



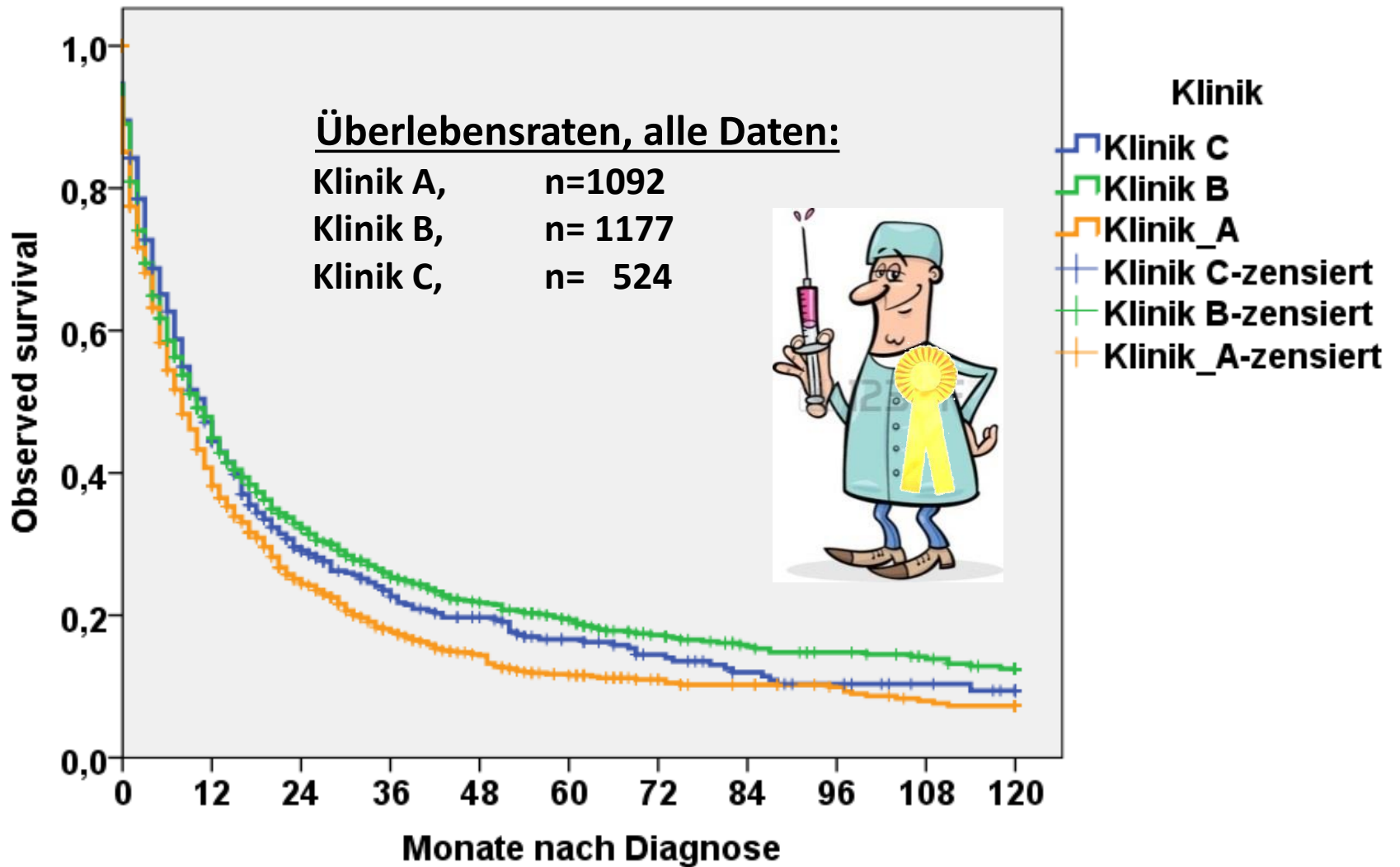
Heute wartet kaum einer so lange, deshalb wird das Überleben meist nach Kaplan-Meier berechnet und da benutzt man zum Berechnen der Unterschiede besser einen Computer.



! Wenn keine zensierten Daten vorliegen, ist die Überlebensrate einfach die relative Häufigkeit der Lebenden. !

Kann man den Unterschied von Überlebensraten auch mit dem χ^2 -Test prüfen?





Ich empfehle den Log-Rang-Test (auch Mantel-Cox genannt). Er wird in der Medizin am häufigsten verwendet. Außerdem sind die Unterschiede im Ergebnis minimal. Ich zeige Ihnen dies (nur aus didaktischen Gründen !!!)

Kaplan-Meier: Faktorstufen vergleichen

Teststatistiken

Log-Rang Breslow Tarone-Ware

Linearer Trend für Faktorstufen

Gemeinsam über Schichten Paarweise über Schichten

Für jede Schicht Paarweise für jede Schicht

Weiter Abbrechen Hilfe

Und welchen der drei Tests sollen wir nun nehmen? Denn wir haben gelernt ...



! Richtig ist, sich vorher den geeigneten Test auszusuchen und das Ergebnis zu akzeptieren. !



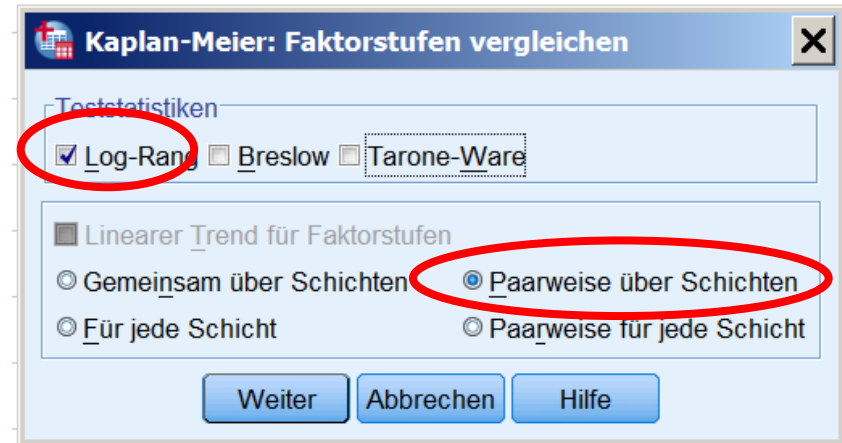
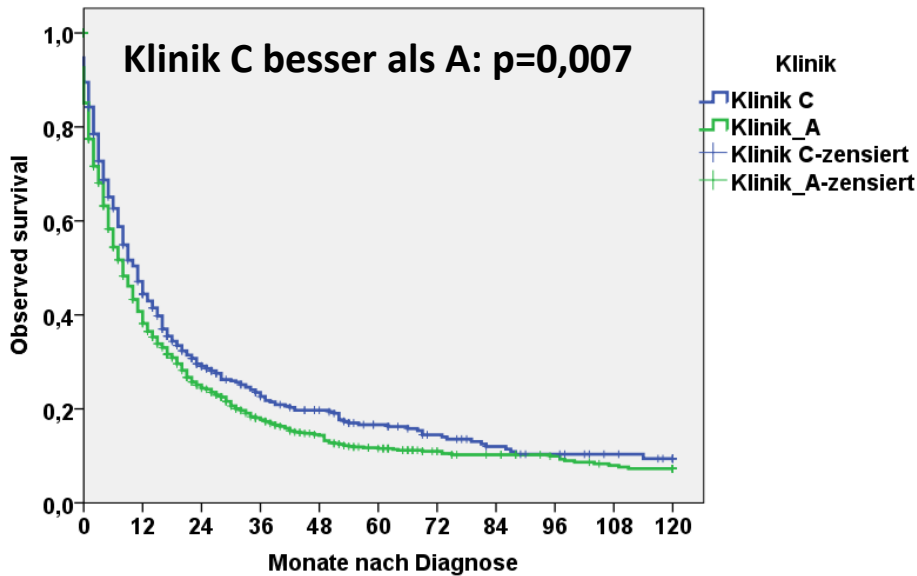
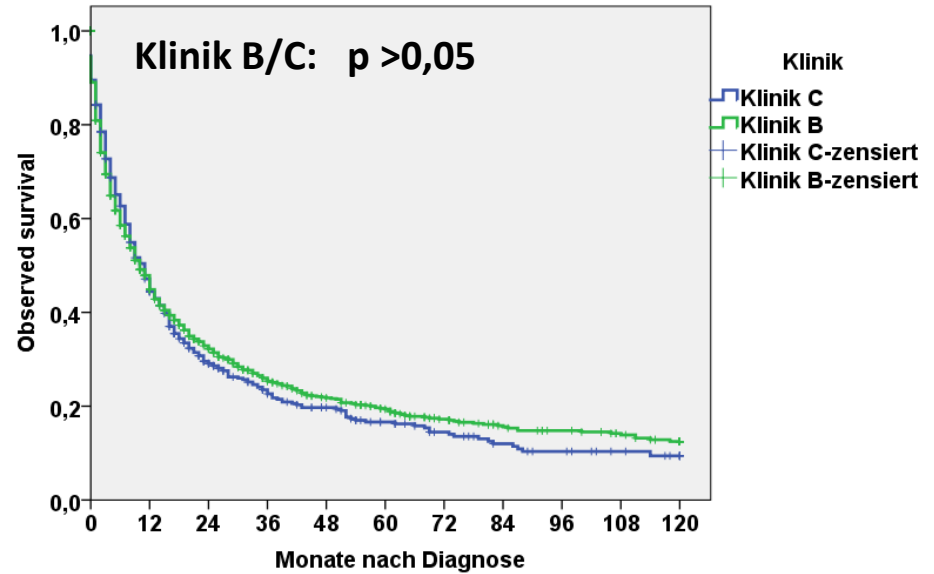
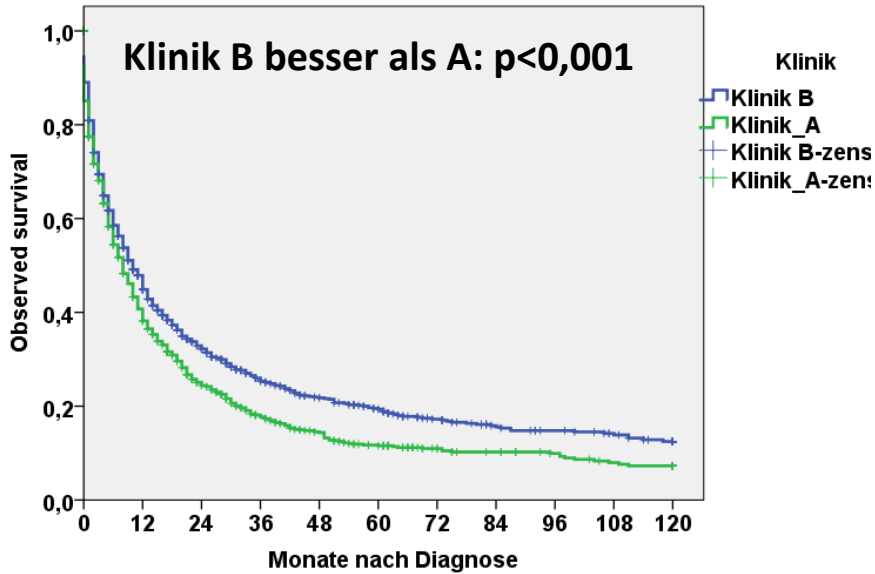
Klinik		Klinik_A	Klinik B
		Sig.	Sig.
Log Rank (Mantel-Cox)	Klinik A		
	Klinik B	0,000	
	Klinik C	0,007	0,500
Breslow (Generalized Wilcoxon)	Klinik A		
	Klinik B	0,001	
	Klinik C	0,002	0,682
Tarone-Ware	Klinik A		
	Klinik B	0,000	
	Klinik C	0,003	0,958

Ergebnis:

- Klinik A unterscheidet sich von B und C jeweils statistisch signifikant.
- Klinik B und C unterscheiden sich nicht statistisch signifikant.

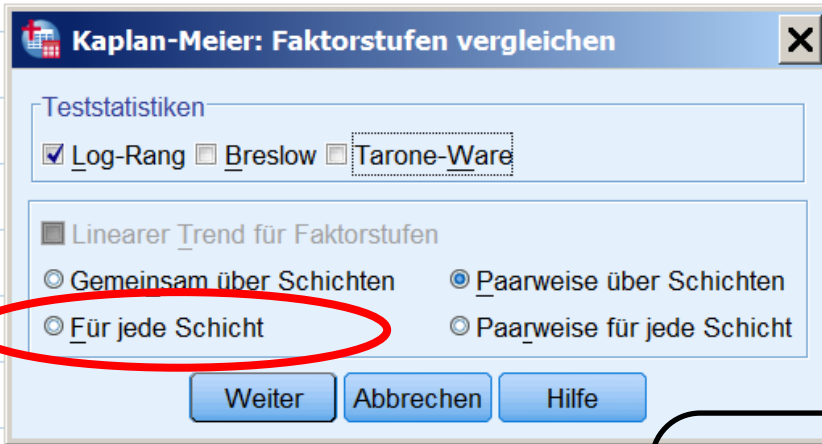
Die Art des Unterschiedes (besser ↔ schlechter) müssen wir aus den Kurven ablesen.

Ergebnis:



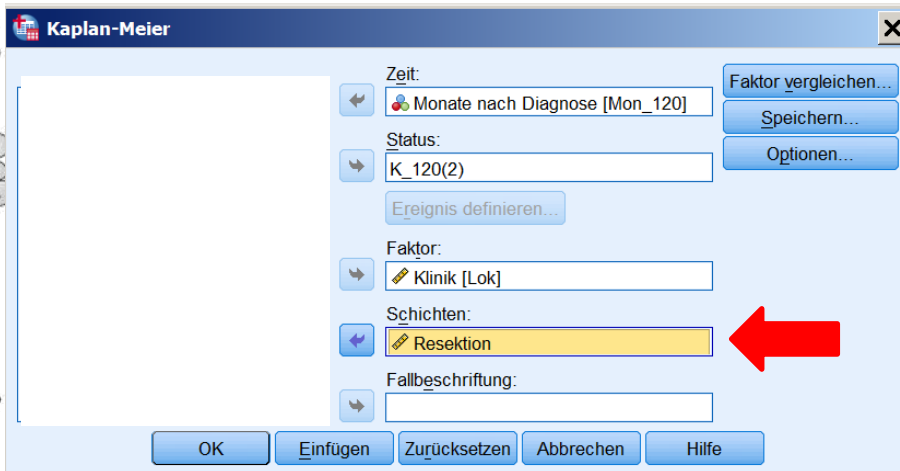


**Wieder kein
Gewinner**



Können Sie das auch rechnen?

Eine Schicht entsteht, wenn wir die Fälle aus jeder Klinik unterteilen, z.B. in Resezierte und Nicht-Resezierte.



Teststatistik:

Paarweise Vergleiche		Log Rank (Mantel-Cox)	
		Klinik A	Klinik B
		Sig.	Sig.
Resezierte	Klinik A		
	Klinik B	0,000	
	Klinik C	0,431	0,000
Nicht-Resezierte	Klinik A		
	Klinik B	0,000	
	Klinik C	0,000	0,663

Ergebnis:

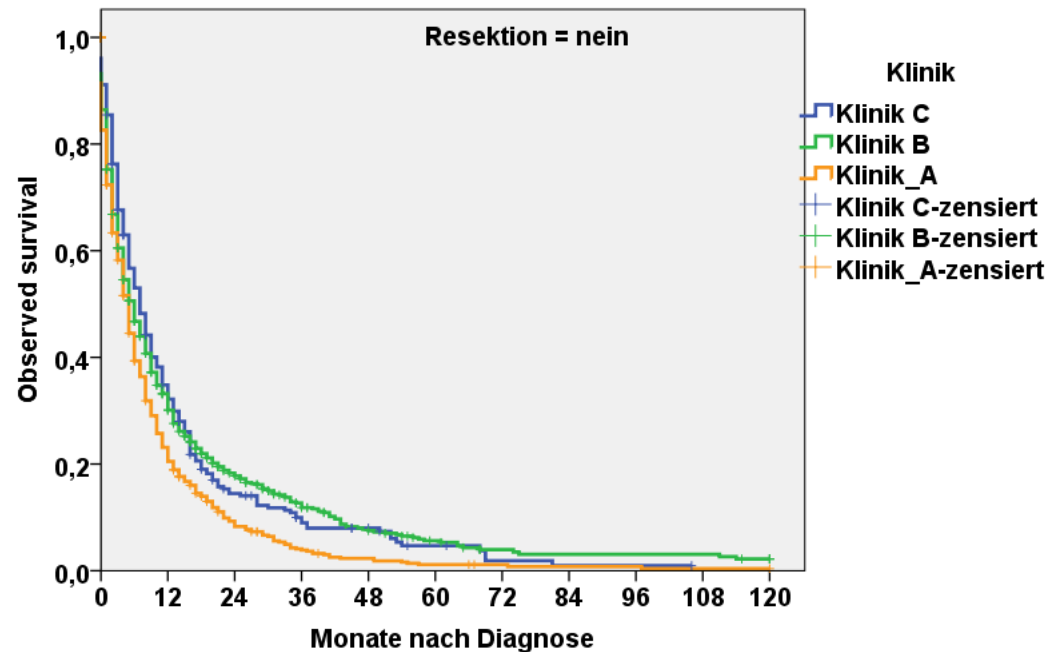
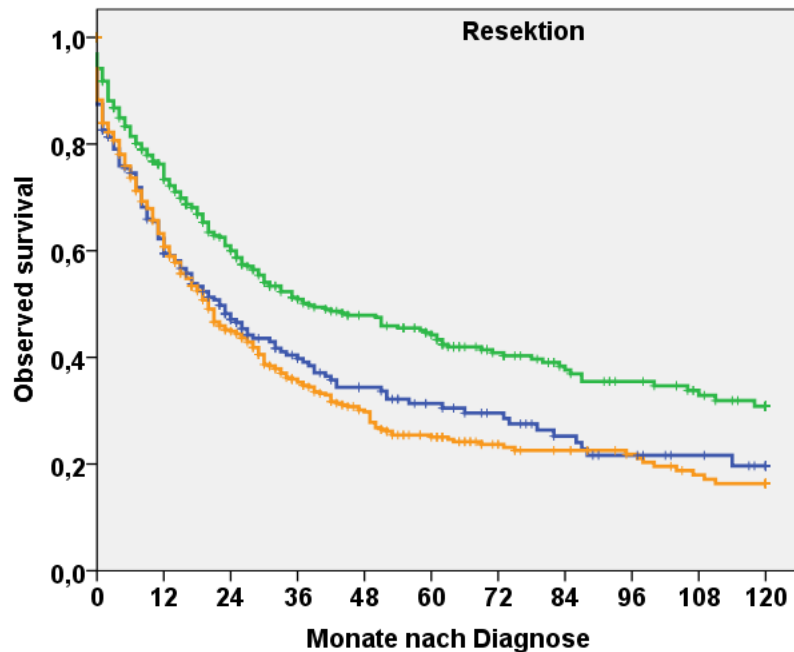
Bei Resezierten:

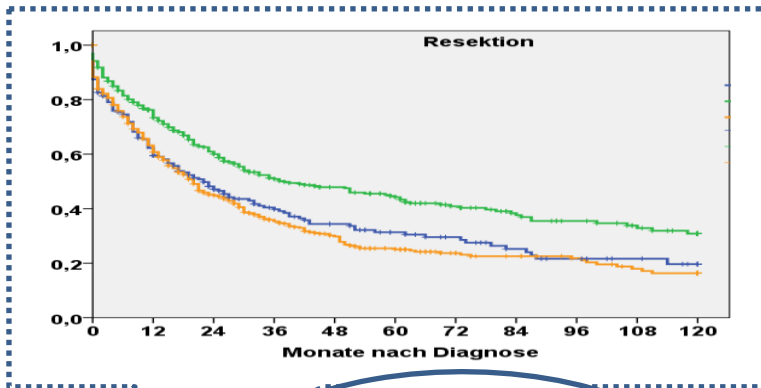
B besser A ($p < 0,001$) und B besser C ($p < 0,001$)

A/C: $p > 0,005$

Bei Nicht-Resezierten B besser A ($p < 0,001$) und C besser A ($p < 0,001$)

B/C: $p > 0,005$





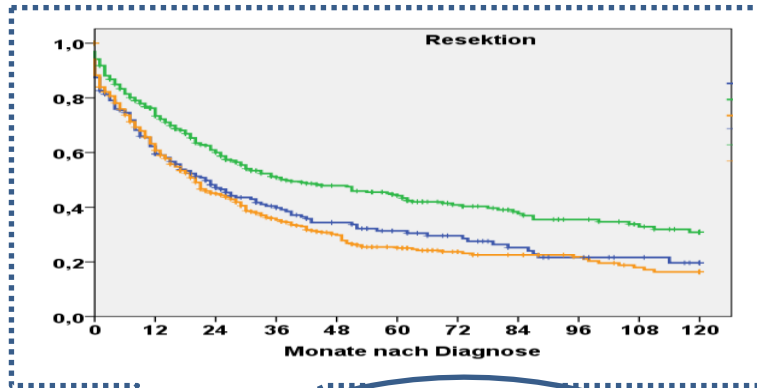
**Ich
reseziere
am besten**



???




???



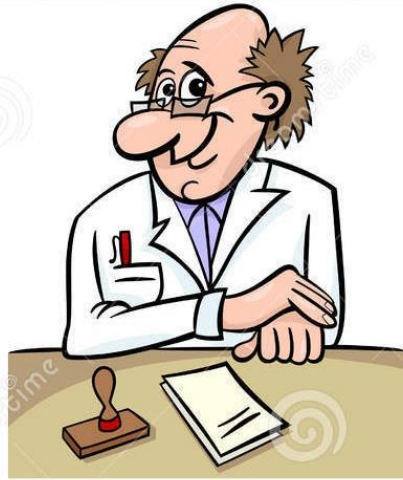
War doch klar, oder?



A photograph of a sunset over the ocean. The sun is low on the horizon, creating a bright, golden path of light across the water. The sky is filled with soft, colorful clouds in shades of orange, pink, and purple. The water is dark blue with gentle ripples.

*... und wenn sie nicht
gestorben sind*

Ende





**Nun ein wenig
Wiederholung**



Einfaktorielle Varianzanalyse ANOVA (ANalysis Of Variance)

ONEWAY deskriptive Statistiken

Alter

	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
LTX	182	58,92	8,816	,653	57,63	60,21	16	72
lok Therapie	226	64,21	8,905	,592	63,05	65,38	37	85
Gesamt	408	61,85	9,238	,457	60,95	62,75	16	85

Test der Homogenität der Varianzen

Alter

Levene-Statistik	df1	df2	Signifikanz
1,928	1	406	,166

Nullhypothese beibehalten
Varianzen sind gleich

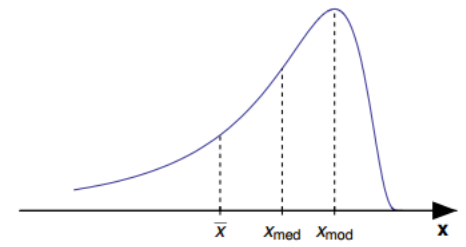
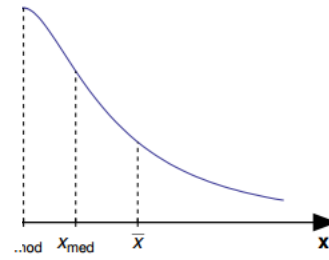
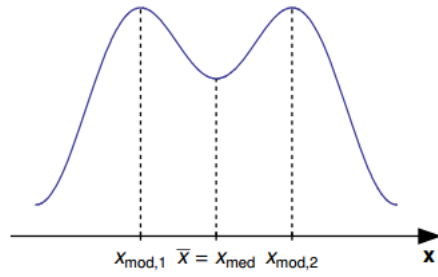
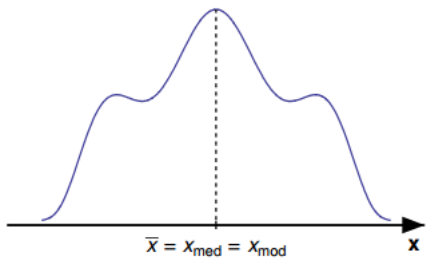
Nullhypothese beibehalten
Mittelwerte des Alters sind gleich

Einfaktorielle ANOVA

Alter

	Quadratsumme	df	Mittelwert	Signifikanz
Zwischen den Gruppen	2810,448	1	2810,448	,000
Innerhalb der Gruppen	31910,728	406	78,598	
Gesamt	34731,176	407		







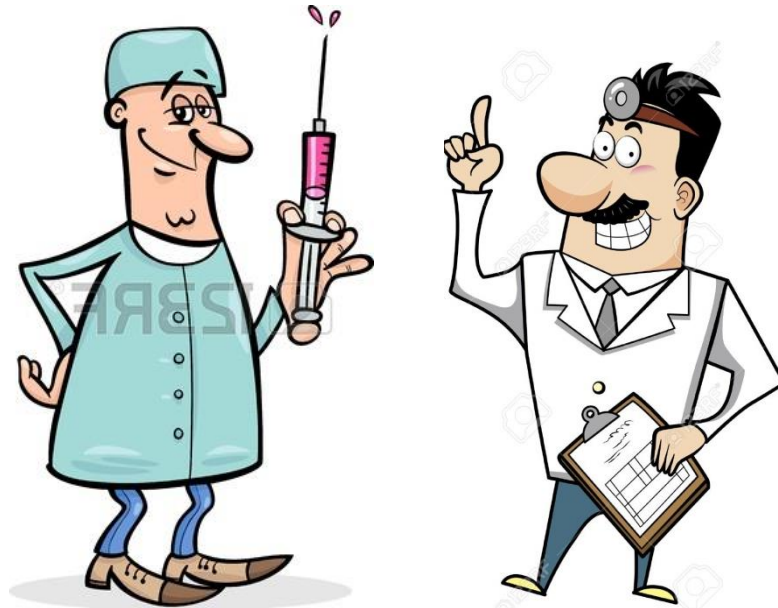
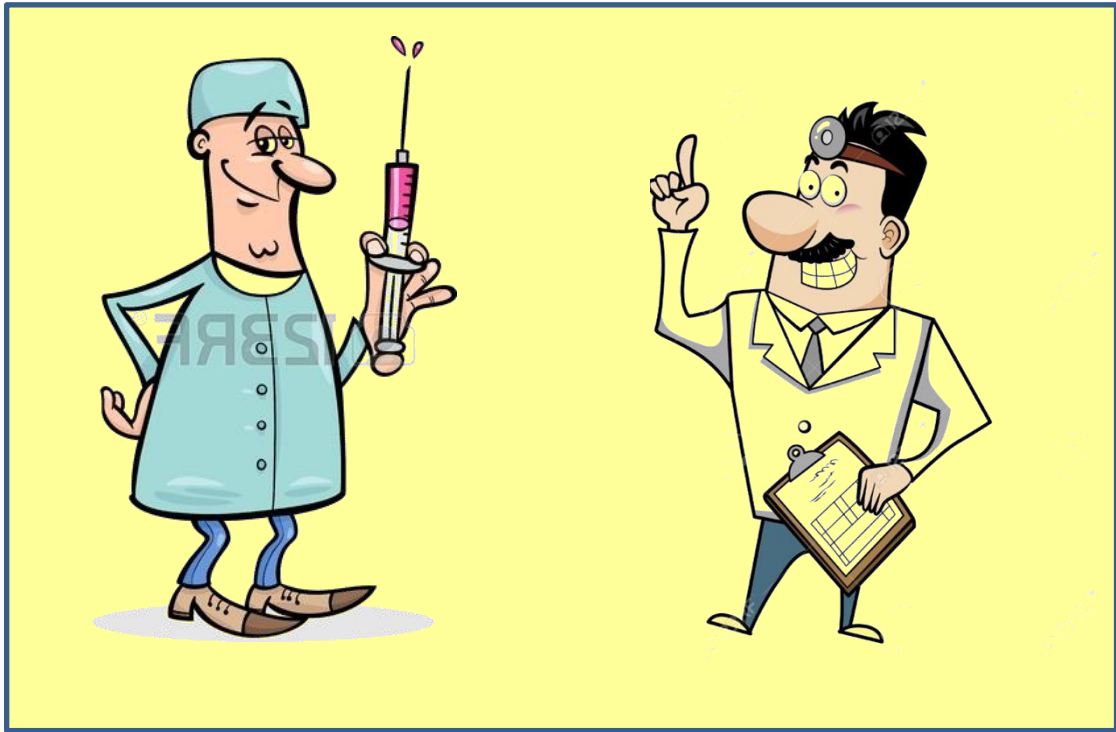
Typen von Merkmalen/Kenngrößen/Variablen

diskret	endlich oder abzählbar unendlich viele Ausprägungen/Werte
stetig	alle (reellen) Zahlen eines Intervalls können mögliche Ausprägungen/Werte sein
nominalskaliert	Ausprägungen sind Namen, Ordnung nicht sinnvoll
ordinalskaliert	Ausprägungen/Werte können geordnet, aber Abstände nicht interpretiert werden
intervallskaliert	Ausprägungen/Werte sind Zahlen, deren Abstände interpretiert werden können
qualitativ	endlich viele Ausprägungen/Werte, typischerweise nominalskaliert, gegebenenfalls auch ordinalskaliert
quantitativ	Ausprägungen/Werte stellen Ausmaß bzw. Intensität











Können Sie
helfen, den
Streit zu
schlichten?

Versuchen
wir es...



mo.

Grundlagen – Typen von Merkmalen
Vergleich

Tag postoperativ	Klinik A verstorben	Klinik B verstorben	Klinik C verstorben
1			
2		1	
...			
8	1	1	
9			1
10			
11	1		
12	1		1
13		1	
...			
20	1		
21	1*	1	
..			
28	1**	1	
29			
30			
31			
32		1***	
33			
34			
35		1***	

- Nach Entlassung in Geriatrie verstorben
- ** Nach Entlassung in Reha verstorben

- *** gehört nicht zur 30-Tage-Letalität

Klinik A verstorben	Klinik B verstorben	Klinik C verstorben
	1	
1	1	
		1
1		
1		1
	1	
1		
1*	1	
1**	1	
	1***	
	1***	

- Nach Entlassung in Geriatrie verstorben
- ** Nach Entlassung in Reha verstorben

- *** gehört nicht zur 30-Tage-Letalität