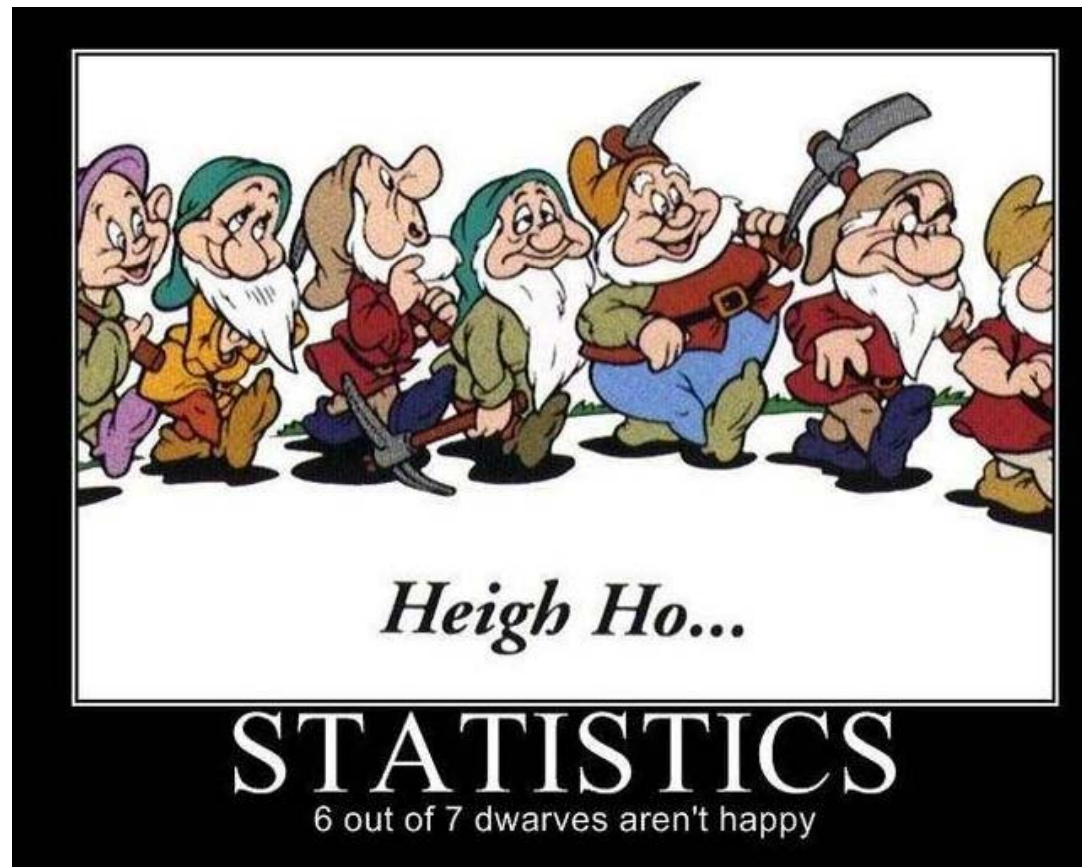


Statistik für Unbegabte (Mediziner und Apotheker)
Studie verstehen, Ergebnisse hinterfragen

T. Decker



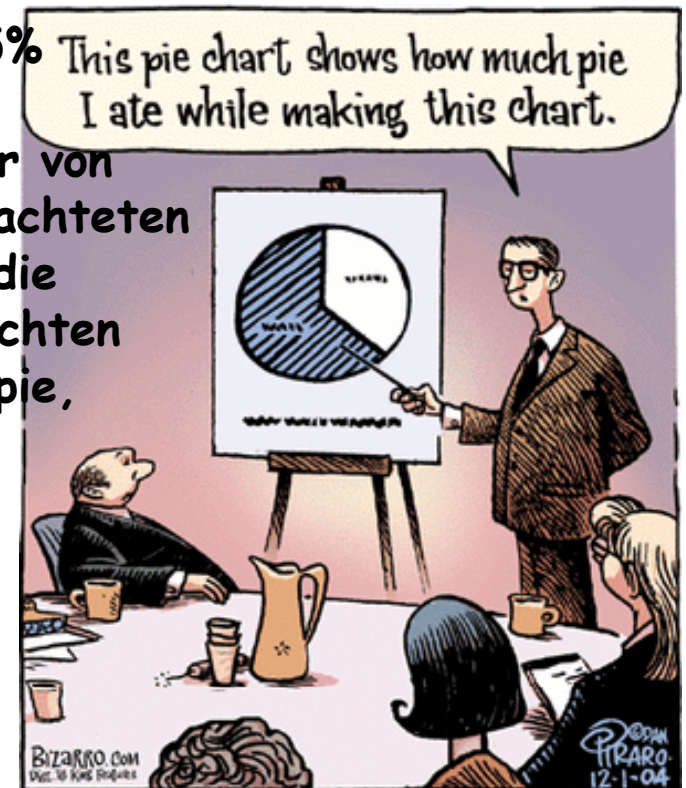


Folgende statistische Hypothese wird getestet:

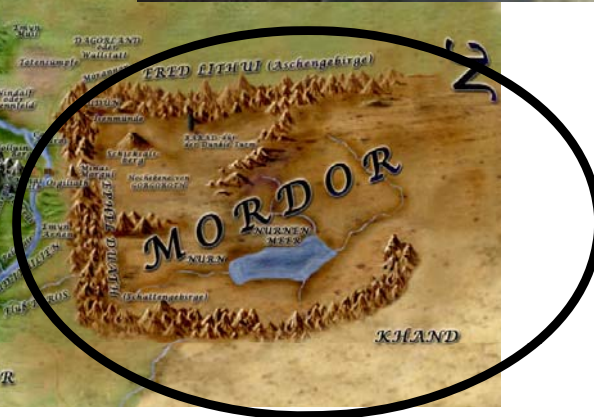
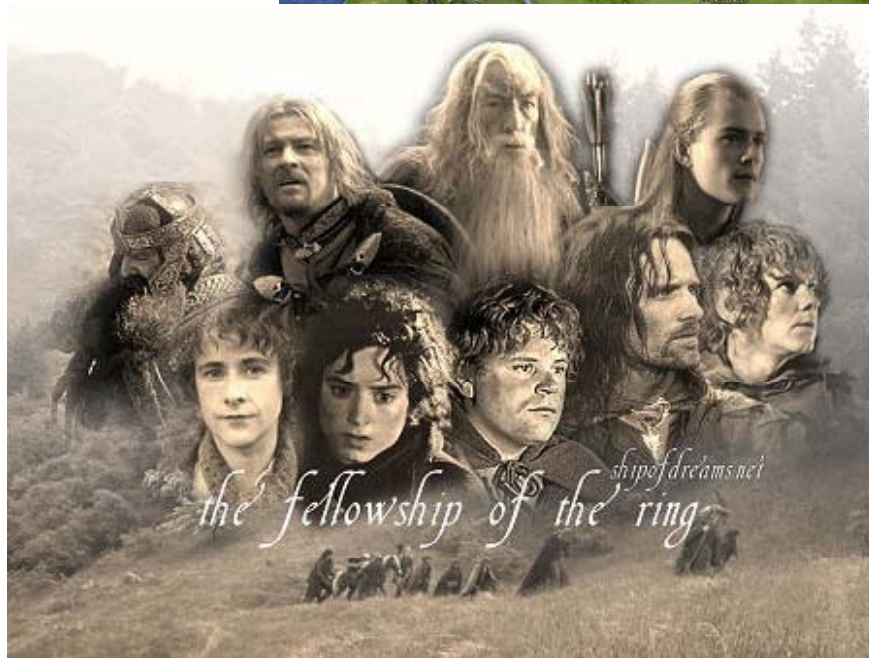
H_0 : Die wahre 6-Monats-PFS-Rate einer Vinorelbin-Everolimus-Kombinationstherapie ist kleiner gleich 35%

H_1 : Die wahre 6-Monats-PFS-Rate einer Vinorelbin-Everolimus-Kombinationstherapie ist größer gleich 55%

Für ein „einseitiges Signifikanzniveau von 0.05“, einer Power von 90% und 53 auswertbaren Patientinnen kann bei einer beobachteten 6-Monats-PFS-Rate von $\geq 47,2\%$ die Nullhypothese, dass die Vinorelbin-Everolimus-Kombinationstherapie bei der untersuchten Population nicht wirksamer ist als die Vinorelbin-Monotherapie, zurückgewiesen werden.



Statistik - die Spezies?



Deskriptive Statistik: Darstellung der Daten:

Achtung Fehlerbalken!!



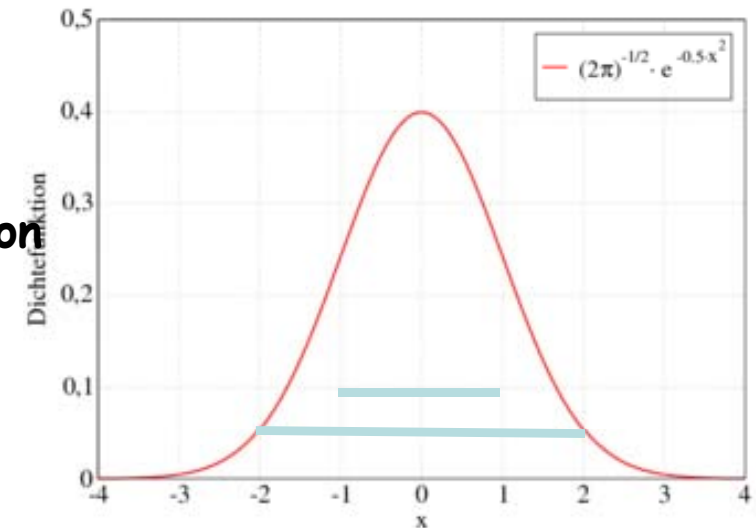
-**SD**: (Standartabweichung, Gauss/Normalverteilung)
-Faustregel: 2/3 der Beobachtungen in einer Population liegen im Bereich $\text{mean} \pm \text{SD}$ (95% $\text{mean} \pm 2\text{SD}$)
„Maß für die Streuung der Stichprobe“

-**SEM**:

-Standartabweichung des Mittelwerts: SD/\sqrt{n}
gibt an, wie genau der Mittelwert ist (hängt von SD und Probengröße n ab)

-**CI** (Konfidenzintervall)

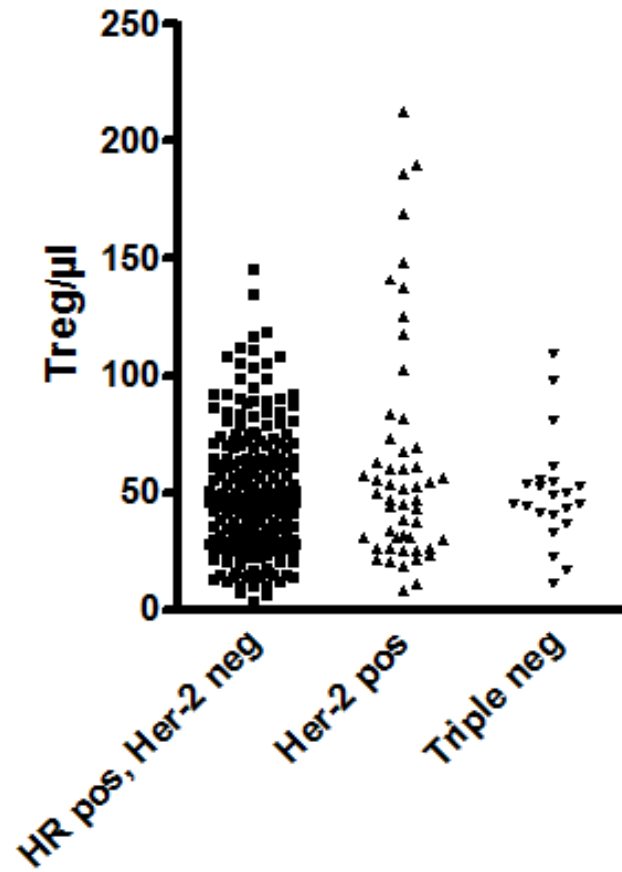
-Das von der Probe kalkulierte Intervall enthält den Populationswert mit 95% Sicherheit (besonders hilfreich zur Abschätzung der „statistischen Signifikanz“)



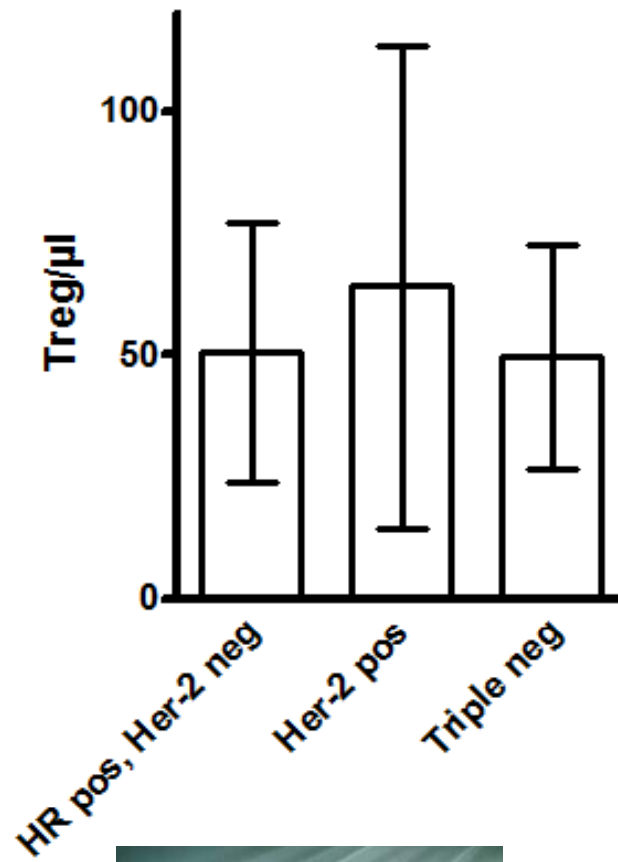
Daten - Präsentation...



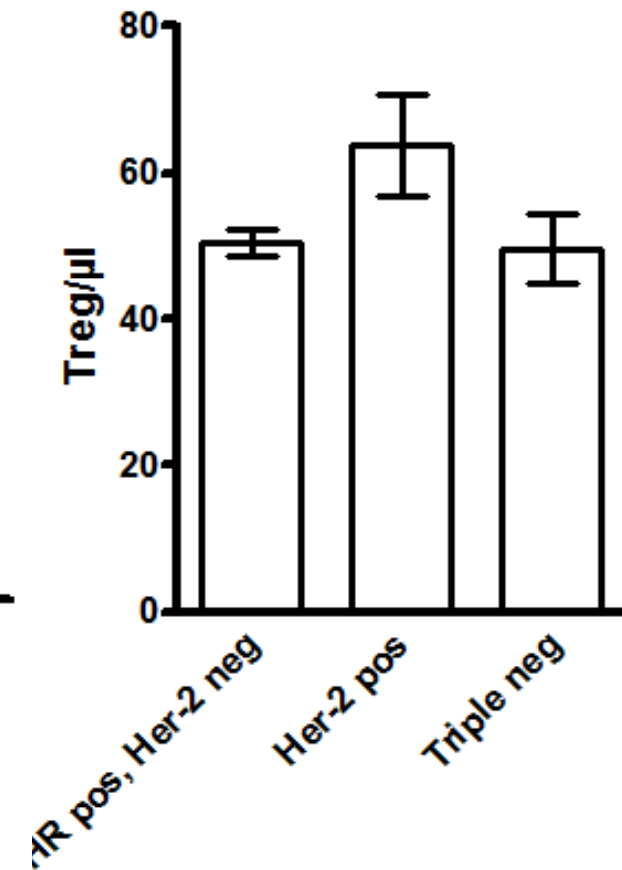
Scattergramm



+/- SD



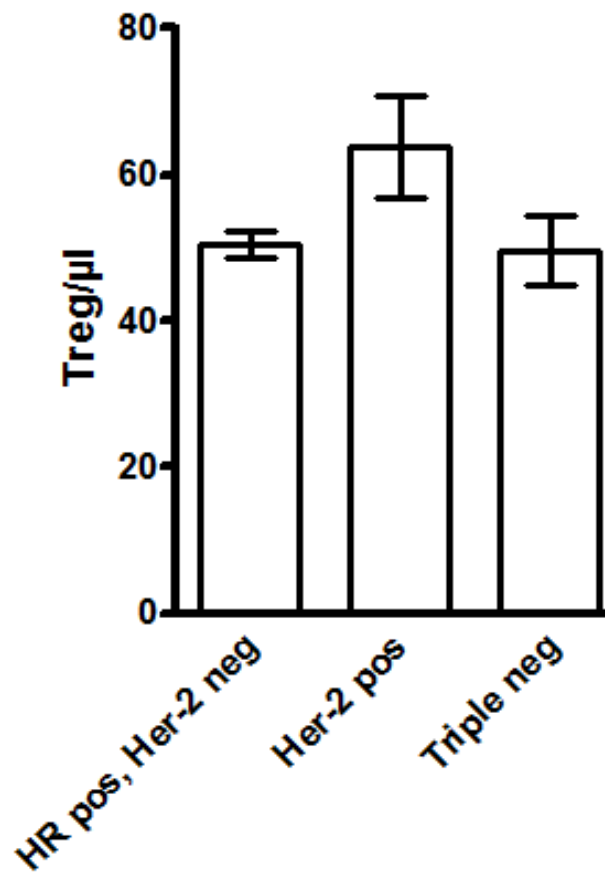
+/- SEM



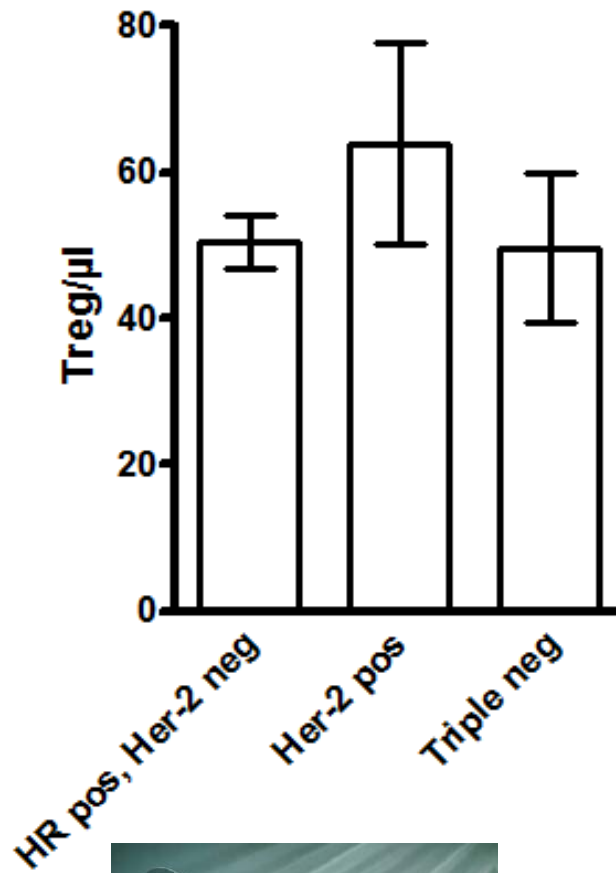
Daten - Präsentation...



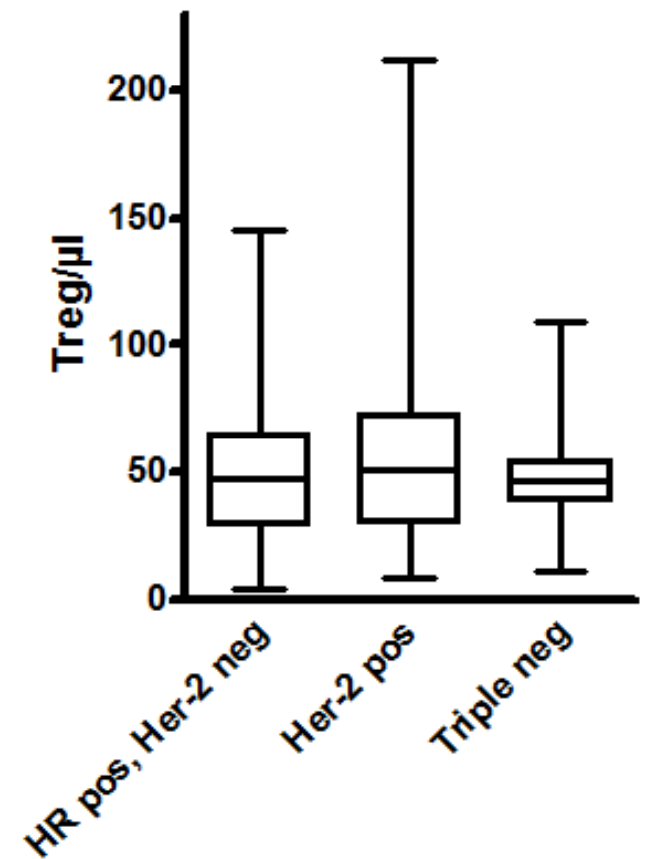
+/- SEM



+/- CI



Box-whisker

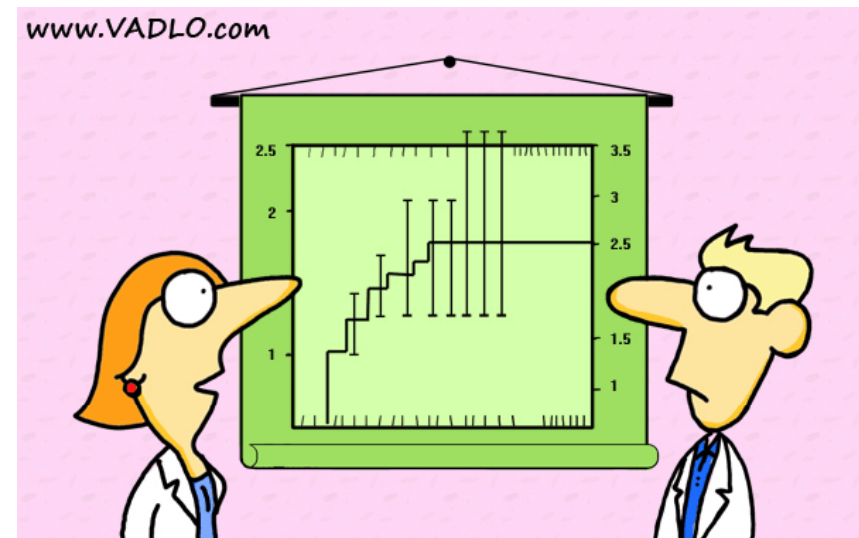


Error bars - „Regeln“ fürs **clevere** Publizieren



Wenn man kleine Unterschiede hervorheben möchte (muss):
SEM zeigen und hoffen, dass die Leser glauben, es handelt sich um SD

Wenn man große Unterschiede verschleiern möchte:
SD zeigen und hoffen, dass die Leser denken, es seien SEM



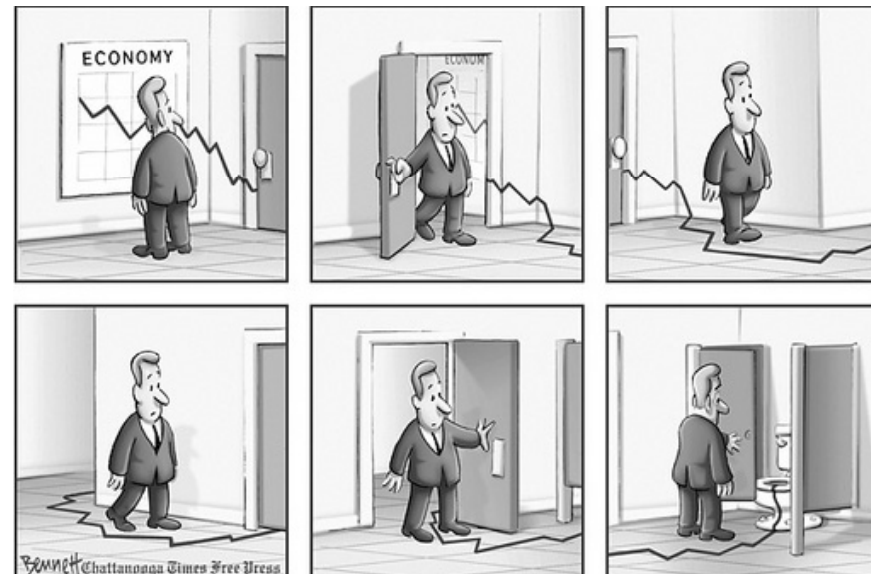
“Did you really have to show the error bars?”

Error bars - „Regeln“ fürs richtige Publizieren



Wenn man die Variationsbreite der Daten zeigen möchte:
Scatter blot, box-whisker, (mean +/- SD)

Wenn man zeigen möchte, wie genau man den Mittelwert gef. hat:
Mean +/- (95% CI),



P-Werte und Signifikanz

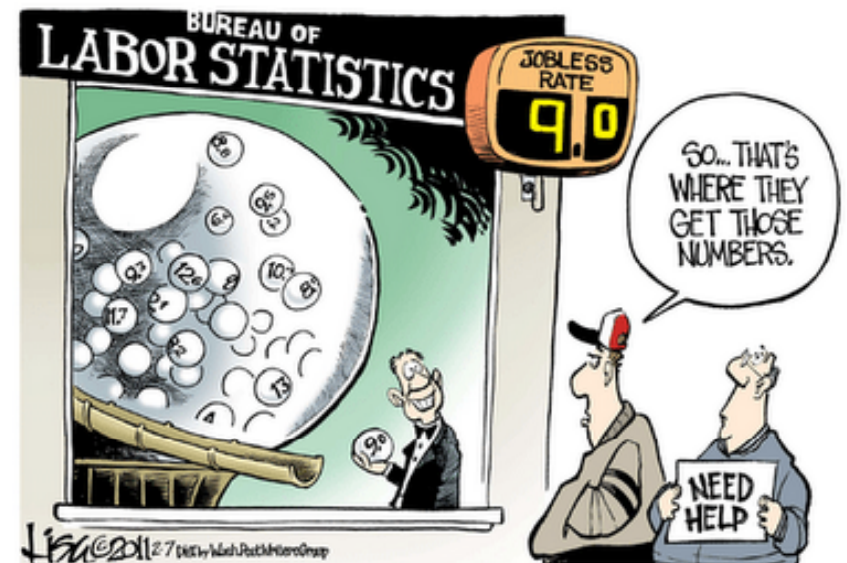


Beispiel: „Coin-Flipping“: 20 Würfe, 16 x Kopf;
Null-Hypothese: fairer Würfel;

Wie groß ist die Chance (bei einem fairen Würfel)
, dass bei 20 Würfeln entweder 16 oder mehr oder 4 oder weniger
Köpfe kommen?

(binomial Verteilung.... 1,19% oder 0,019)

Kontext!! (Deutsche Mark versus Münze in Zaubershow)





P-Werte Verständnisprobleme:

-Null Hypothese ist gewöhnlich das Gegenteil von dem, was die Studie zeigen möchte (z.B. neue Therapie und Standard sind gleich wirksam, ein ev. Unterschied zwischen den Gruppen besteht zufällig)

„Knoten im Hirn“

„Rückwärts-Logik: Null Hypothese wird festgelegt, Gruppe wird beobachtet, es wird die Wahrscheinlichkeit bestimmt, mit der beobachtete Ergebnis zufällig erreicht wird,

-Wenn diese Wahrscheinlichkeit kleiner ist, als ein vorher festgelegter Wert (Signifikanzniveau), spricht man von statistisch signifikant

Trick: one tailed / two tailed

Vorteil: Ergebnis schneller „signifikant!! Aber:

One tailed nur, wenn vorher schon klar ist, dass der Unterschied nur in einer Richtung liegen kann (z.B. Größe vor und nach Wachstumshormon)

Statistisch signifikant.....



-Häufig missverstanden! (statistisch signifikant nicht gleich bedeutsam, wichtig)!!!

-Entscheidung treffen!, Null Hypothese annehmen / verwerfen.."

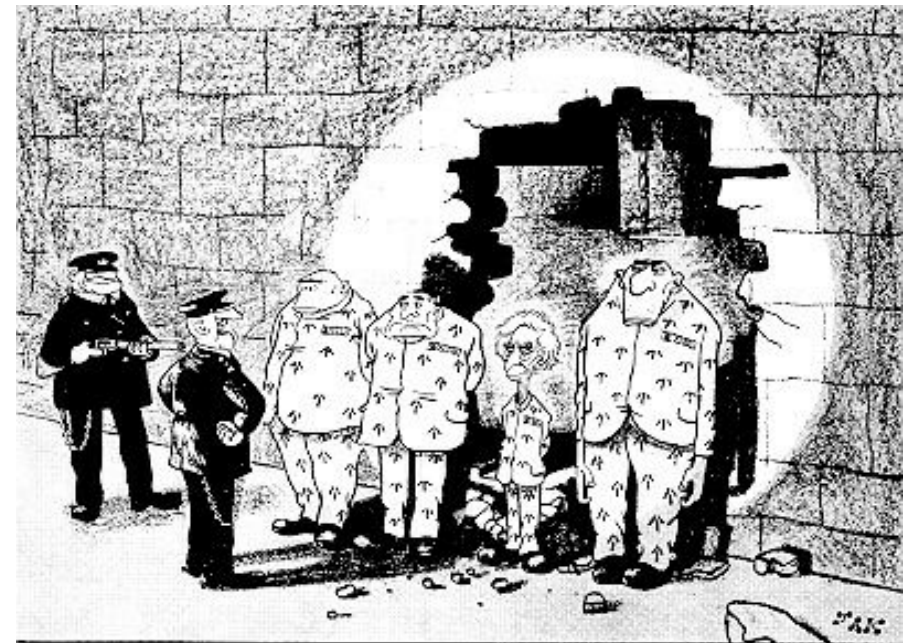
-p-Wert Schwelle = Signifikanz-level alpha (0.05)

-Wenn die Wahrscheinlichkeit p (wird von den aktuellen Daten berechnet) < 0.05 , dann:

-Signifikant, sehr signifikant,

-extrem signifikant?

-- fast signifikant? ($p = 0,051..$)



"ALL RIGHT! FOR THE LAST TIME. WHO'S THE BRAINS BEHIND THIS?"

Statistisch signifikant.....





-Type I Fehler:

-Passiert, wenn die Nullhypothese abgelehnt wird, obwohl sie zutrifft

-Type II Fehler

-Passiert, wenn die Nullhypothese nicht abgelehnt wird, obwohl sie falsch ist

-Niedrige Signifikanzlevel alpha: wenige Type I / mehr Typ 2 Fehler

	-Schuldig	Unschuldig
	Typ I -Korrekt	korrekt
	-Korrekt	Typ II



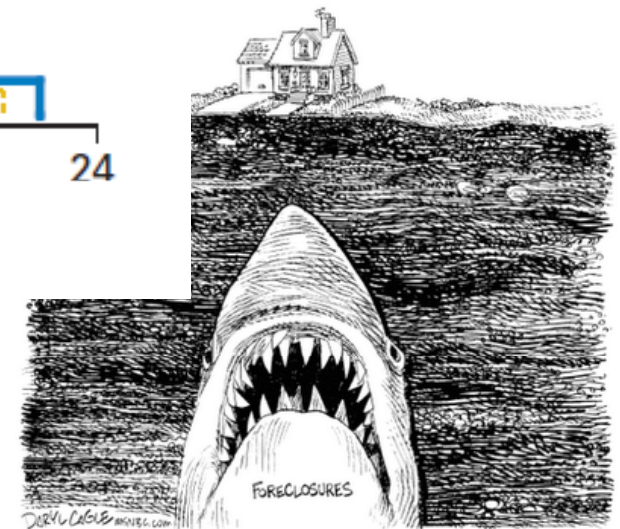
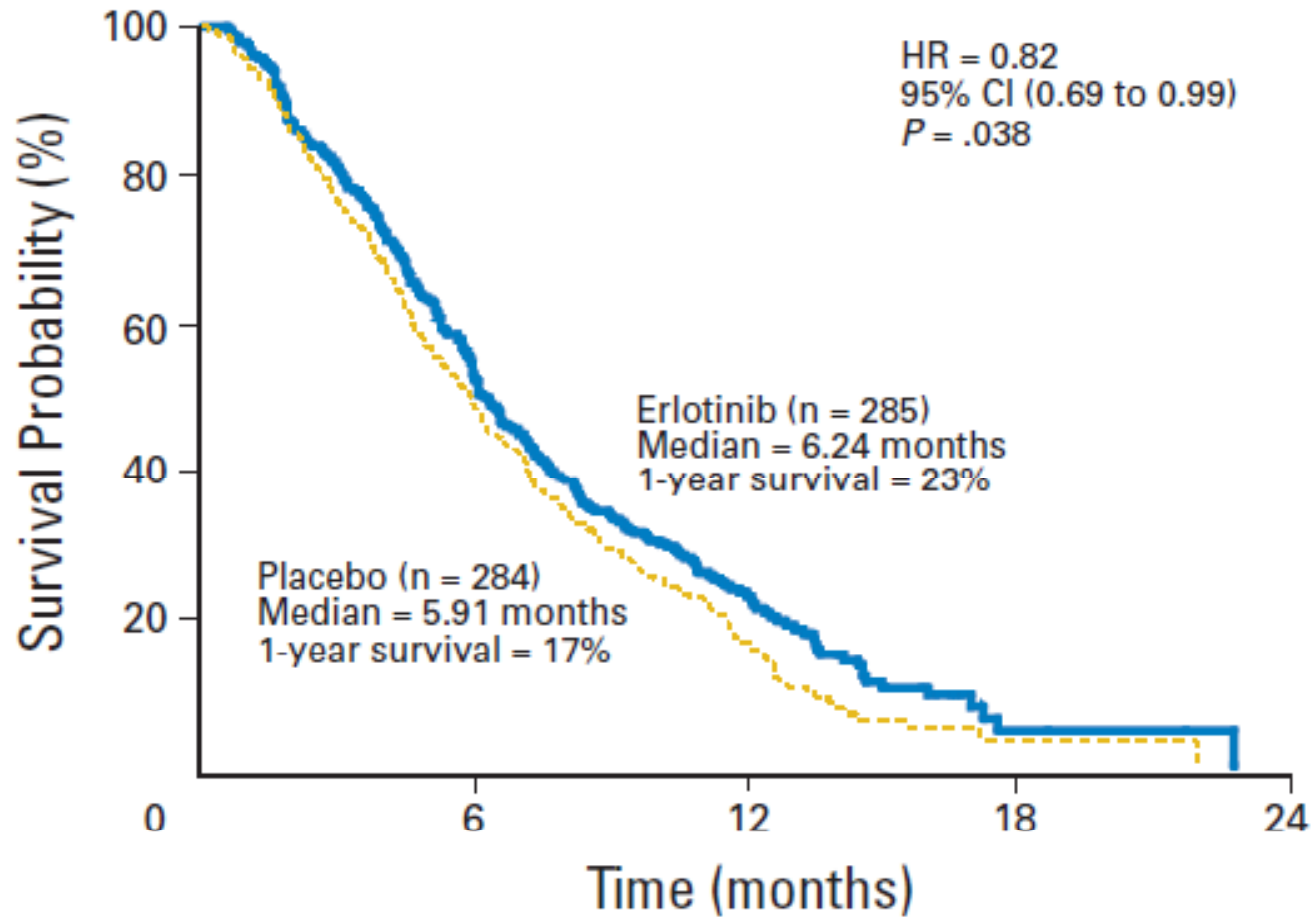
Statistisch signifikant.....



-Wenn ein Ergebnis „statistisch signifikant“ ist, nicht aufhören **A**

-Statistisch überraschend
-(aber)

-Bayesianische
-Der K...
die Int...
genetis...
-(Beisp...





-Wenn ein Ergebnis „statistisch nicht signifikant“ ist, nicht aufhören zu denken!

-Statistisch nicht sign. heisst, dass die erzielten Resultate durchaus auftreten können, wenn H_0 stimmt

-Insbesondere bei großen Konfidenzintervallen könnten die Daten einfach unklar sein (zu kleine Stichprobe, „zu wenig Power“)

-Faustregel: Stichprobe $\times 4$; CI wird um den Faktor 2 kleiner



„Power“ Wenn zwischen zwei Behandlungen ein def. Unterschied besteht (z.B. OS 10%): Wie groß ist die Chance ein statistisch signifikantes Resultat zu erhalten (Stichprobenumfang)

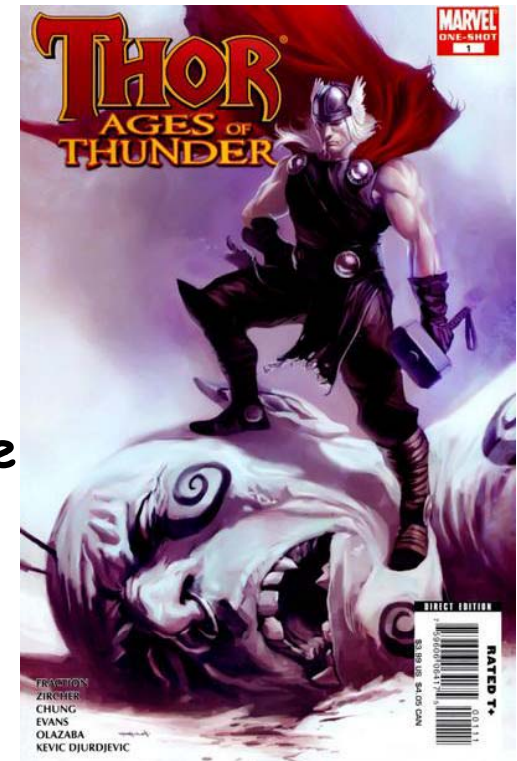
-Der Stichprobenumfang wird so gewählt, sodaß die Chance ein statistisch signifikantes Resultat zu erhalten z.B. 80% beträgt; je größer der Unterschied, desto geringer die Fallzahl

-Je größer die Power, desto geringer Typ-2 Error

-Power = $C/C+D$

Wird (vom PC) kalkuliert: sample size, SDs, Signifikanzlevel, Größe des erwarteten Effekts

	H0 ablehnen	H0 annehmen
-H0 true	A	B
-H0 false	C	D



Multiple Vergleiche



-Wenn Du Deine Daten lang genug quälst, zeigen sie was Du willst

-X unabhängige Vergleiche aus der selben Stichprobe:
Chance, dass mindestens 1 Ergebnis signifikant ist: $1 - 0,95^X$

-Bei 13 Vergleichen: 50% Chance auf ein statistisch sign. Ergebnis (aus Zufall)

-Bonferroni-Korrektur:

-Statistisch signifikant, wenn $p < 0,05/\text{Anzahl der Vergleiche}$

-Beispiel T-REX:

-G1 v G2/G3; N0 v N1 v N2, T1 v T2-4, her-2 pos

-v Hr pos v triple neg., ductal vs lobulär

-P=0,008, p=0,04; (7 Tests: Sig Niveau: 0,007)



Multiple Vergleiche



JELLY BEANS CAUSE ACNE!

SCIENTISTS! INVESTIGATE!

BUT WE'RE PLAYING MINECRAFT! ... FINE.

WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE ($P > 0.05$).

THAT SETTLES THAT. I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.

SCIENTISTS! BUT MINECRAFT!

WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE ($P < 0.05$).

WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE ($P > 0.05$).

WHOA!

WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE ($P > 0.05$).

WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE ($P > 0.05$).

News

GREEN JELLY BEANS LINKED TO ACNE!

95% CONFIDENCE

ONLY 5% CHANCE OF COINCIDENCE!

SCIENTISTS...



Welcher Test?



Kontinuierliche Variable von einer Gauss Verteilung
(z.B. Leukozytenzahl; mean +/-SD)

- 2 Gruppen: **T Test** (paired, unpaired) (ggf. Korrektur für multiples testen)
- >2 Gruppen: **ANOVA**

Kontinuierliche Daten von „nicht-Gauss Verteilungen“
(weniger Power! im T-Test Beispiel $p=0.008$ versus $p=0.29$)
(z.B. Gleason Score: median +/- 25er/75er Perzentilen)

- 2 Gruppen: **Mann-Whitney, Wilcoxon**
- >2 Gruppen: **Kruskal-Wallis, Friedman`s**



Überlebenszeit

Kaplan-Meier Kurve, Median survival, % Überleben nach x Jahren

Log-rank Test

Binomial

AML Heilung ja/nein nach 5 Jahren: (Kontingenz-Tafeln)

- 2 Gruppen: Fisher, Mc Nemars
- >2 Gruppen: Chi square



- „Don` t forget the big picture“

- **Sinn und Zweck**: Extrapolieren von einer limitierten Stichprobe, um generelle Aussagen zu machen

- „**Müll rein - Müll raus**“ : Datenqualität / Design / Manipulation



- Problem der sehr großen (Mini-Unterschiede statistisch signifikant) und sehr kleinen Stichproben (underpowered)

- $P < 0,05$ ist nicht heilig !

- Statistisch signifikant heißt nicht, wissenschaftlich relevant

- Absence of evidence is not evidence of absence



-
- Problem der multiplen Vergleiche
 - Planung (Stichprobenumfang, Tests etc. vorher!!)
 - Konfidenzintervalle sind einfacher zu handhaben als p-Werte
 - Skeptisch sein! Viele (die meisten publizierten?) (statistisch signifikanten) Ergebnisse sind falsch (TypI error), kleine Stichprobe, Bias, multiple Vergleiche, „flexible Datenanalyse“,



Das Monty Hall Problem

3 Türen, hinter einer Tür ist ein Auto, hinter 2 Türen eine Ziege

Du wählst eine Tür.....



Der Showmaster öffnet eine andere Tür und zeigt Dir, dass dahinter
Kein Auto ist - er bietet Dir an, Deine Wahl zu wechseln

Was nun? Annehmen? Falle?



Statistik - intuitiv daneben?



Auflösung:

Türe wechseln!!!

Erklärung

z.B. auf youtube: Monthy Hall

