Pflege Deine Vorurteile!

A statistician is a person
who draws a mathematically precise line
from an unwarranted assumption
to a foregone conclusion.

Statistik als wissenschaftliche Qualitätskontrolle

Quo vadis ?

Statistik als wissenschaftliche Qualitätskontrolle

Quo vadis ?

Science

PNAS
Proceedings of the National Academy of Sciences of the United States of America

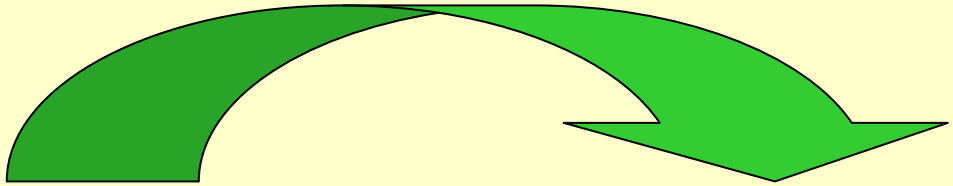
nature

1. Ein p-Wert ist ein p-Wert ist ein p-Wert

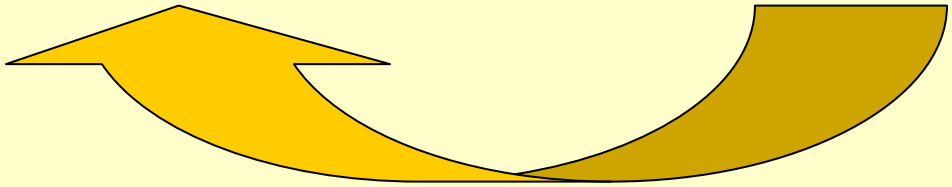
(und kein Test)

Zwei Arten logischen Schlußfolgerns

deduktiv

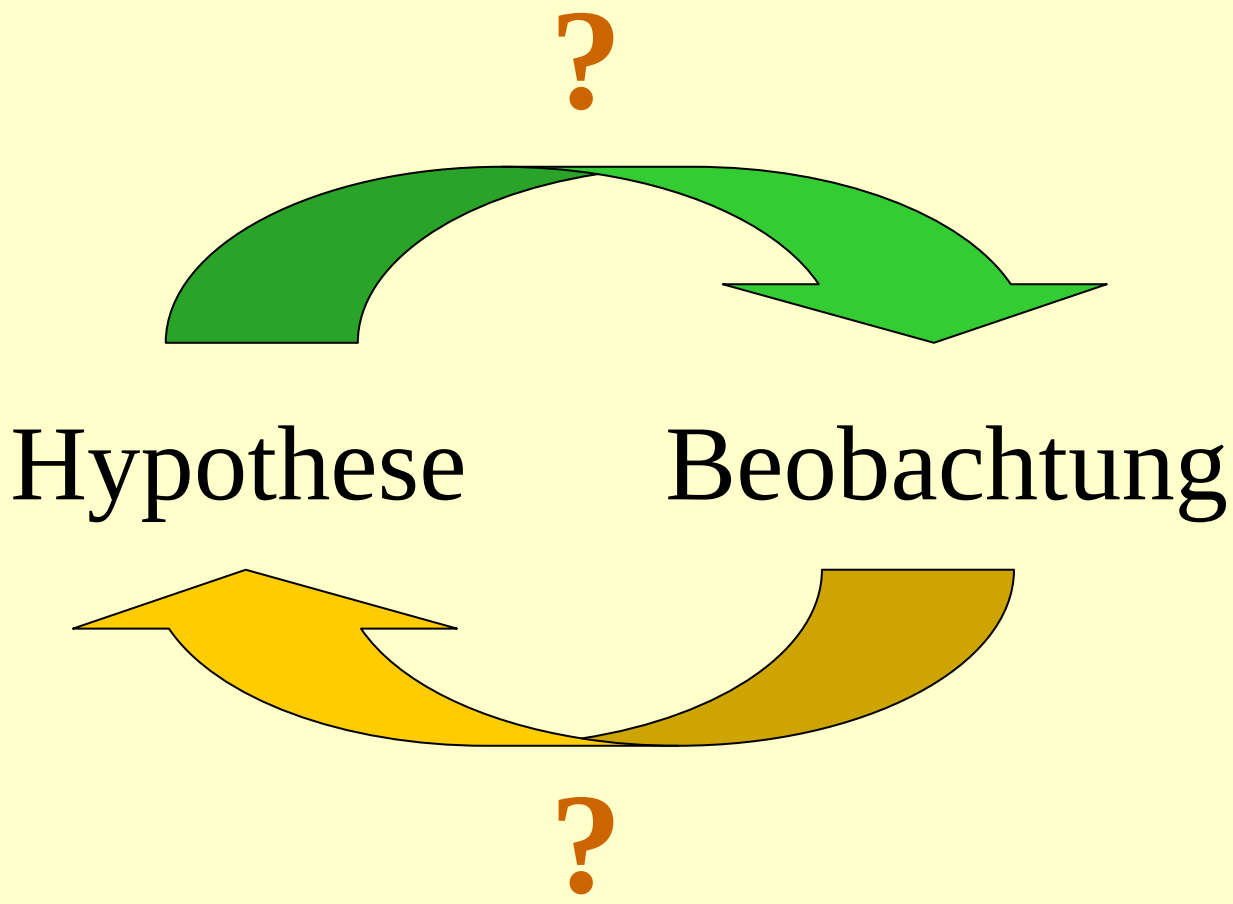


Ursache Wirkung



induktiv

Schlußfolgern in der Statistik



Schlußfolgern in der Statistik (induktiv)

Sir Ronald A. Fisher
(1890-1962)

Beobachtung

p-Wert

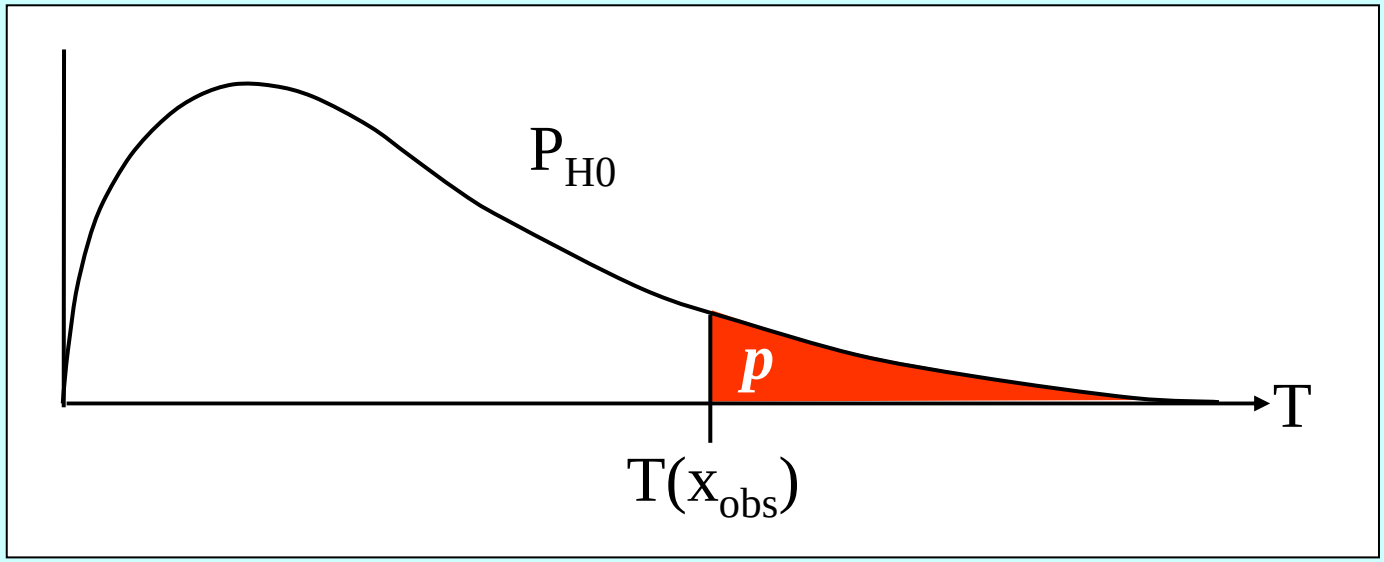
Hypothese

Der *p*-Wert

Hypothese	H0
Beobachtung	x_{obs}
Teststatistik	$T:x \rightarrow T(x)$

$$p = P_{H0}(x:T(x) \geq T(x_{obs}))$$

Der *p*-Wert



"... an informal index to be used as a measure of discrepancy between the data and the null hypothesis."

Goodman SN (1999) Ann Intern Med 130: 995-1004

Das Theorem von Bayes

Thomas Bayes (1702-1761)

$$P(H0|x_{obs}) \propto P(H0) \cdot p$$

"No test based upon the theory of probability can by itself provide any valuable evidence of the truth or falsehood of a hypothesis"

Neyman J, Pearson E (1933) Phil Trans R Soc A, 231:289-337

p-Wert als “Entscheidungskriterium“

$p=0.10$

unerwartet → “shows that the two groups are equivalent“

erwartet → “trend of borderline significance“

“not statistically significant, most probably because of small sample size“

p-Wert als “Entscheidungskriterium“

p=0.10

unerwartet → “shows that the two groups are equivalent“

erwartet → “trend of borderline significance“

“not statistically significant, most probably because of small sample size“

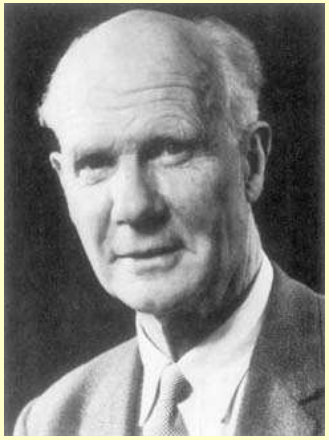
p=0.01

unerwartet → “in all likelihood represents a false positive“

“reflects unknown bias“

erwartet → “clearly demonstrates a treatment effect“

Schlußfolgern in der Statistik (deduktiv)



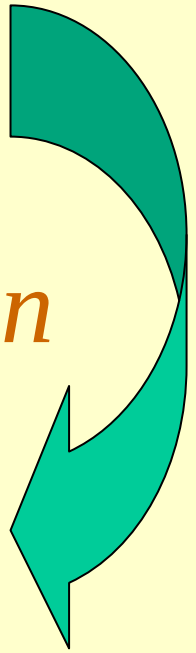
Egon Pearson (1895-1980)

Jerzy Neyman (1894-1981)

Hypothese

statistisches Testen

Beobachtung



Statistisches Testen

"Without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behavior with regard to them, in following which we insure that, in the long run of experience, we shall not often be wrong."

Neyman J, Pearson E (1933) Phil Trans R Soc A, 231:289-337

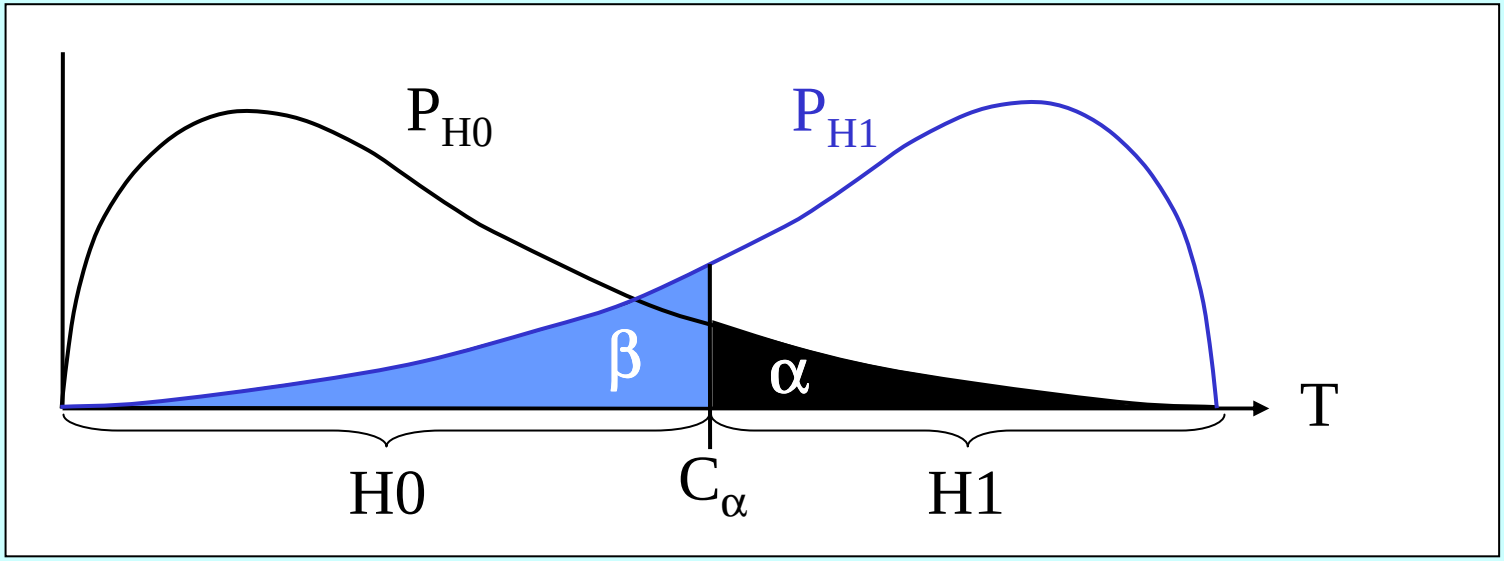
Statistisches Testen

Hypothesen	H0, H1
Beobachtungen	x
Teststatistik	$T:x \rightarrow T(x)$
Fehler 1. Art	α
Fehler 2. Art	β

Wähle C_α so, daß $P_{H0}(x:T(x)>C_\alpha) \leq \alpha$

$P_{H1}(x:T(x) \leq C_\alpha) \leq \beta$

Statistisches Testen



$$T(x_{obs}) \leq C_\alpha \longrightarrow H0$$

$$T(x_{obs}) > C_\alpha \longrightarrow H1$$

Das sogenannte „multiple Testproblem“

	p
Placebo *vs.* Behandlung A	0.125
Placebo *vs.* Behandlung B	0.015
Placebo *vs.* Behandlung C	0.045

Das sogenannte „multiple Testproblem“

	p	p_{crit}	
Placebo *vs.* Behandlung A	0.125	$\frac{0.05}{3}$	
Placebo *vs.* Behandlung B	0.015		
Placebo *vs.* Behandlung C	0.045		

Das sogenannte „multiple Testproblem“

	p	p_{crit}
Placebo *vs.* Behandlung A	0.125	$\frac{0.05}{3}$
Placebo *vs.* Behandlung B	0.015	
Placebo *vs.* Behandlung C	0.045	
Placebo *vs.* Behandlung D	0.020	$\frac{0.05}{2}$
Placebo *vs.* Behandlung E	0.005	

Das sogenannte „multiple Testproblem“

	p	p_{crit}	p_{crit}
Placebo *vs.* Behandlung A	0.125	$\frac{0.05}{3}$	$\frac{0.05}{5}$
Placebo *vs.* Behandlung B	0.015		
Placebo *vs.* Behandlung C	0.045		
Placebo *vs.* Behandlung D	0.020	$\frac{0.05}{2}$	
Placebo *vs.* Behandlung E	0.005		

Das sogenannte „multiple Testproblem“

	p	p_{crit}	p_{crit}
Placebo *vs.* Behandlung A	0.125	$\frac{0.05}{3}$	$\frac{0.05}{5}$
Placebo *vs.* Behandlung B	0.015		
Placebo *vs.* Behandlung C	0.045		
Placebo *vs.* Behandlung D	0.020	$\frac{0.05}{2}$	
Placebo *vs.* Behandlung E	0.005		

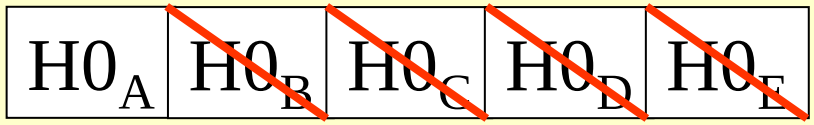
entweder

$H0_A$	~~$H0_B$~~	~~$H0_C$~~	~~$H0_D$~~	~~$H0_E$~~

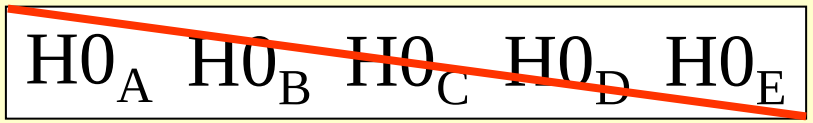
Das sogenannte „multiple Testproblem“

	p	p_{crit}	p_{crit}
Placebo *vs.* Behandlung A	0.125	$\frac{0.05}{3}$	$\frac{0.05}{5}$
Placebo *vs.* Behandlung B	0.015		
Placebo *vs.* Behandlung C	0.045		
Placebo *vs.* Behandlung D	0.020	$\frac{0.05}{2}$	
Placebo *vs.* Behandlung E	0.005		

entweder



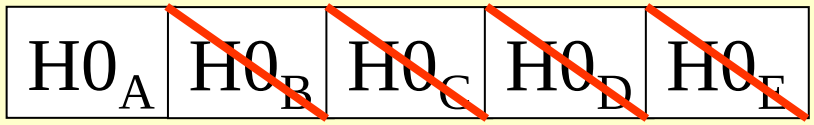
oder



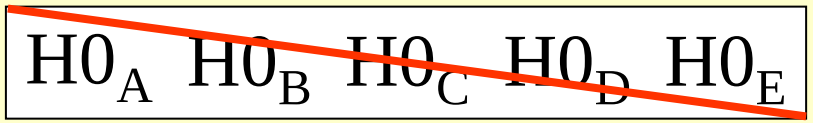
Das sogenannte „multiple Testproblem“

	p	p_{crit}	p_{crit}
Placebo *vs.* Behandlung A	0.125	$\frac{0.05}{3}$	$\frac{0.05}{5}$
Placebo *vs.* Behandlung B	0.015		
Placebo *vs.* Behandlung C	0.045		
Placebo *vs.* Behandlung D	0.020	$\frac{0.05}{2}$	
Placebo *vs.* Behandlung E	0.005		

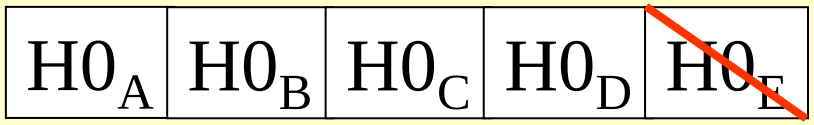
entweder



oder



aber nicht



2. Was ist schon/noch „normal“?

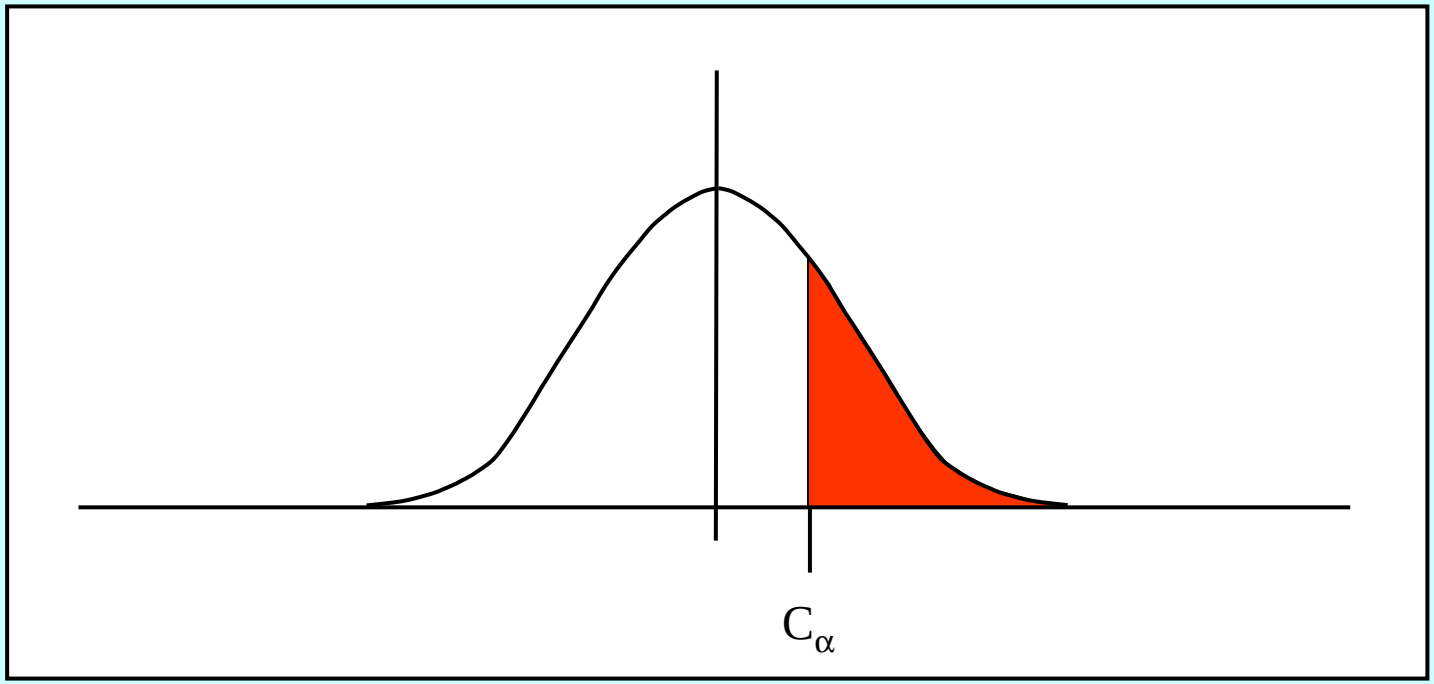
Verteilung: Fakt und Fiktion

Test

$P_{H0}(x:T(x)>C_\alpha)$

p-Wert

$P_{H0}(x:T(x)\geq T(x_{obs}))$



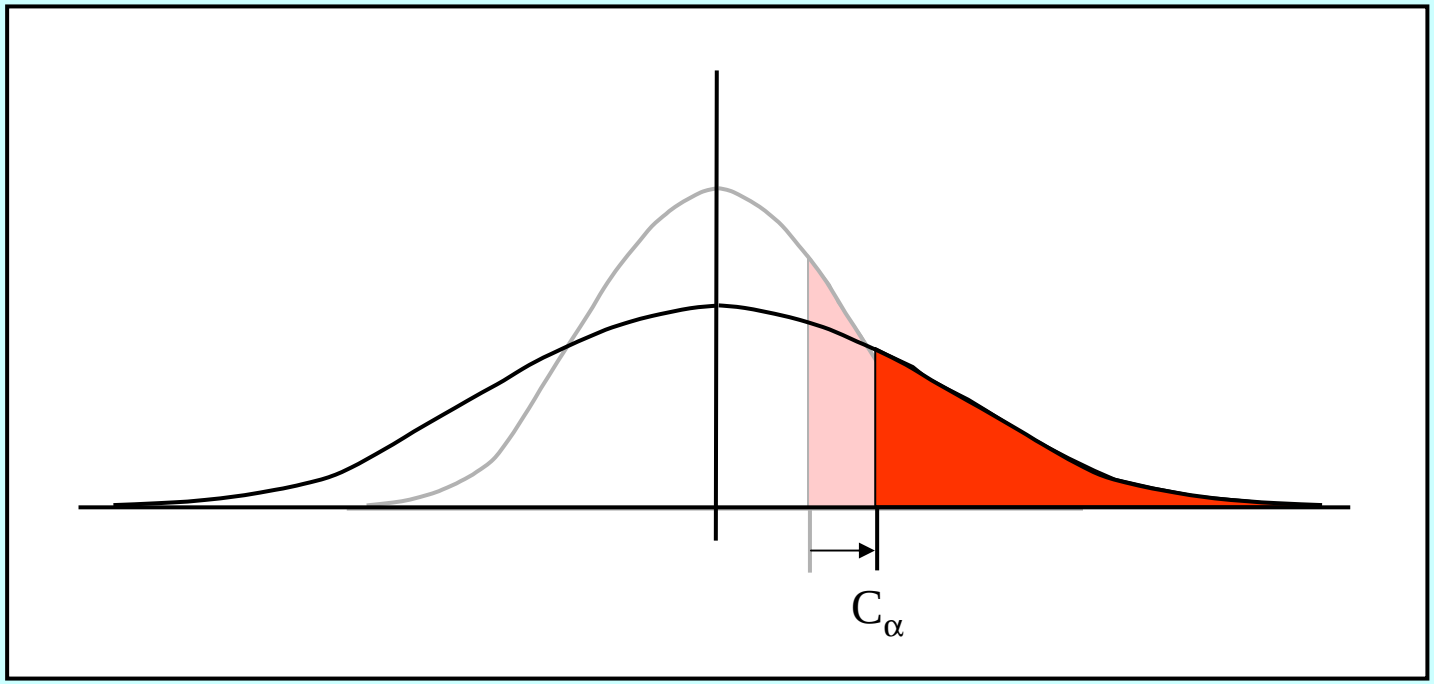
Verteilung: Fakt und Fiktion

Test

$P_{H0}(x:T(x)>C_\alpha)$

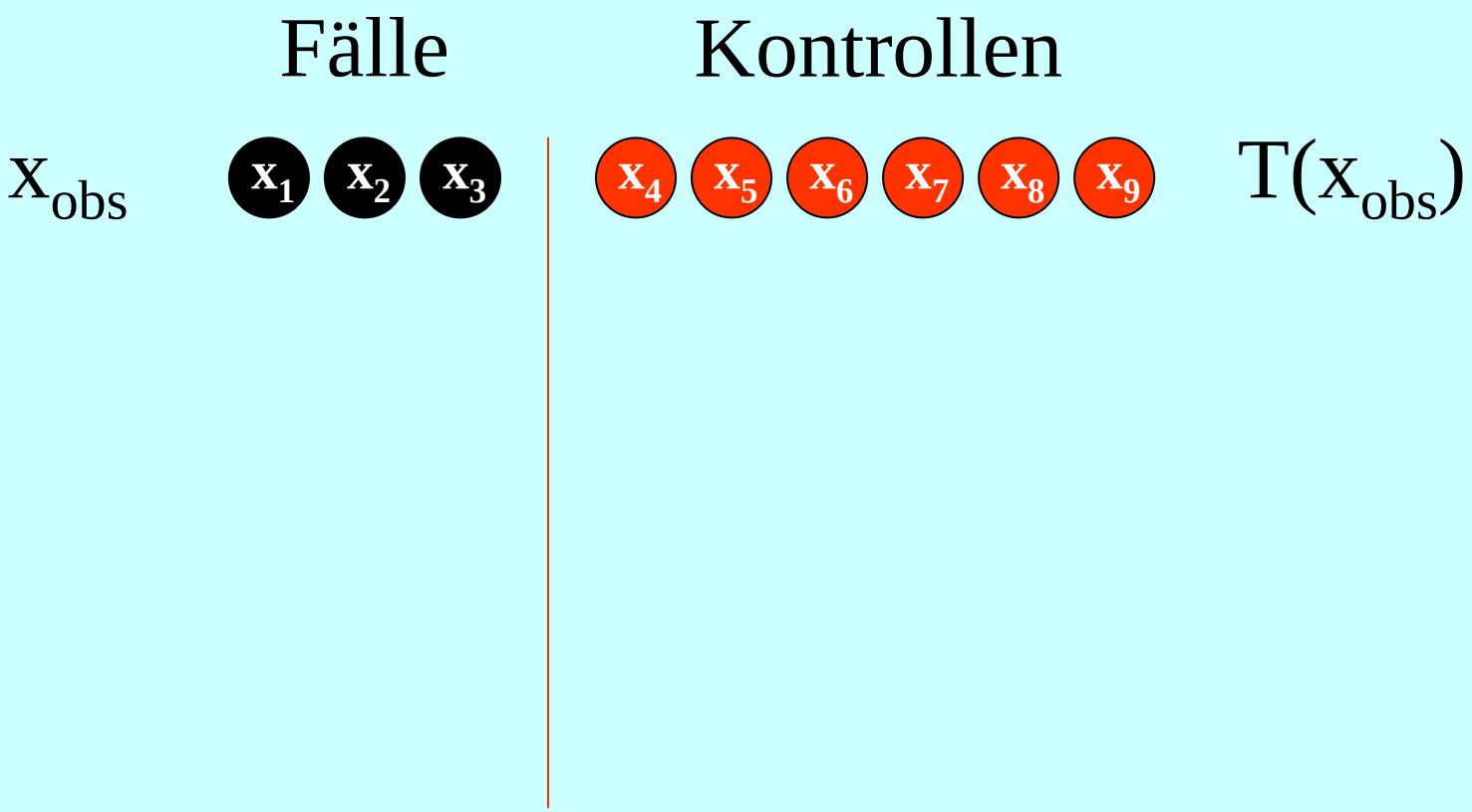
p-Wert

$P_{H0}(x:T(x)\geq T(x_{obs}))$



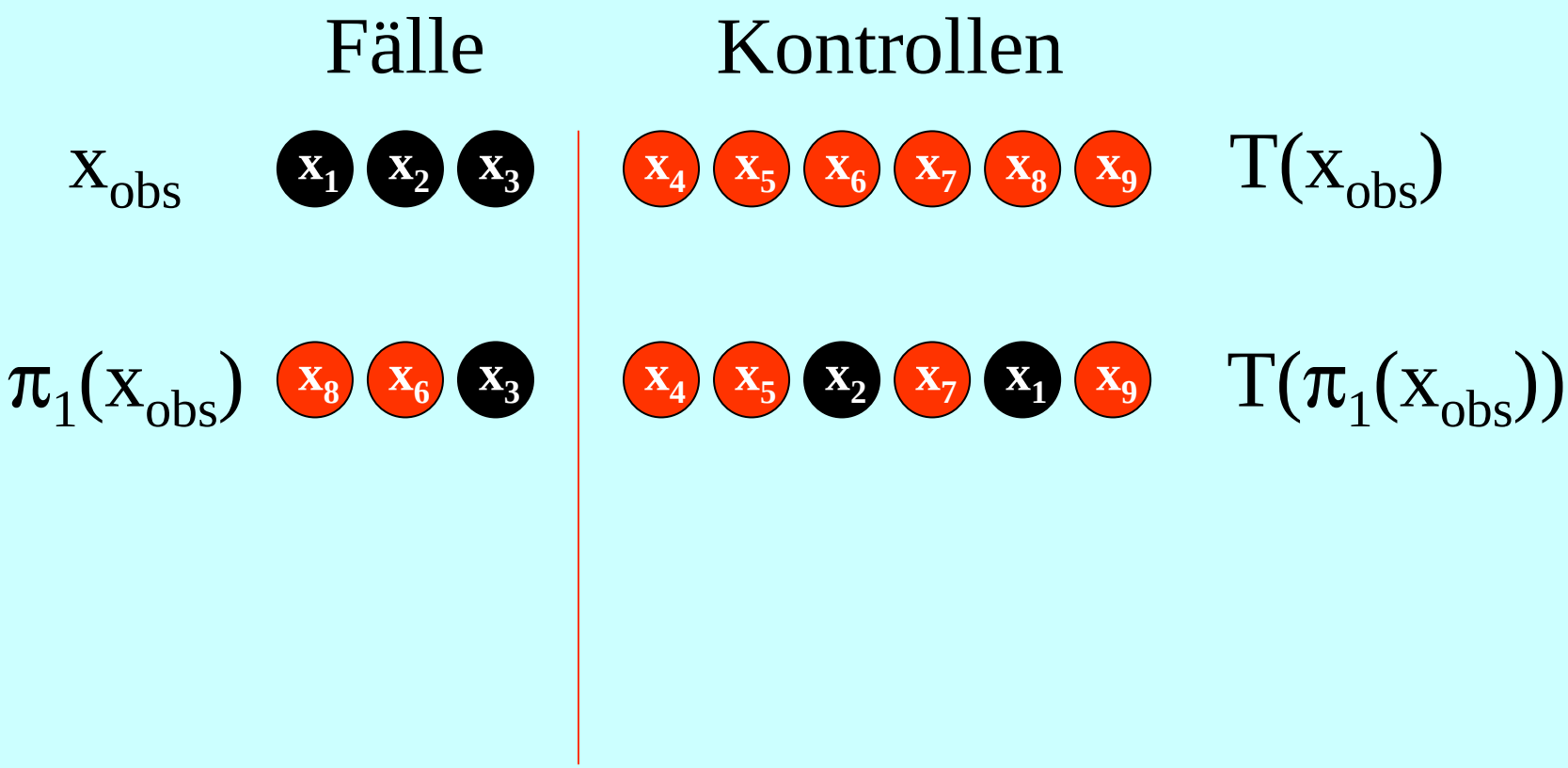
Randomisierungs-Tests

„statistics without tears“



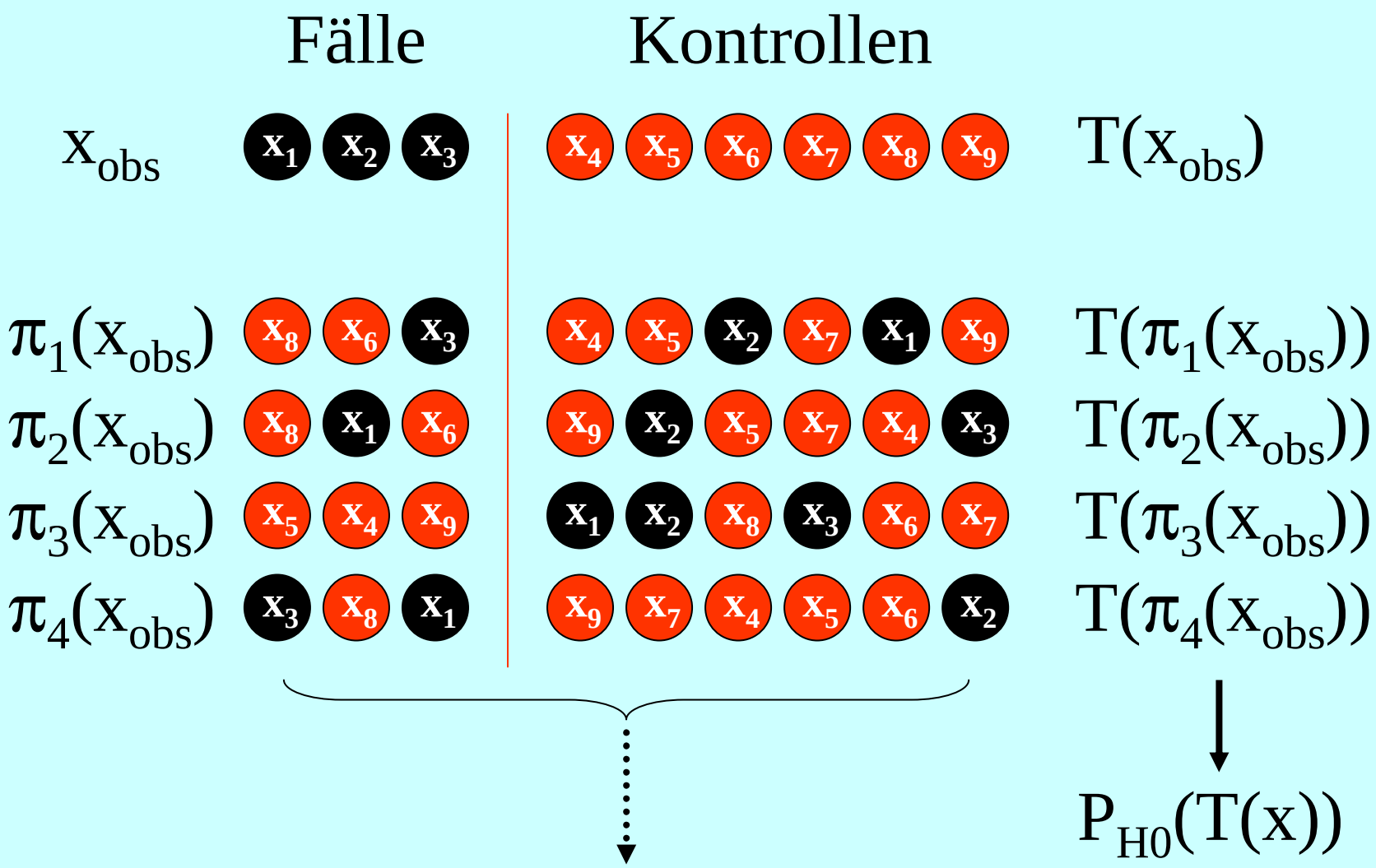
Randomisierungs-Tests

„statistics without tears“



Randomisierungs-Tests

„statistics without tears“



3. Ein p-Wert mißt keine Effektgröße

(signifikant ist nicht gleich „signifikant“)

Kleine Studie, großer Effekt ...

	Erfolg	∅Erfolg	Σ
Verum	40	10	50
Placebo	25	25	50
Σ	65	35	100

χ^2=8.62, 1 df, p=0.004 OR=4.000 CI: 1.517-10.749

Kleine Studie, großer Effekt ...

	Erfolg	∅Erfolg	Σ
Verum	40	10	50
Placebo	25	25	50
Σ	65	35	100

χ^2=8.62, 1 df, p=0.004 OR=4.000 CI: 1.517-10.749

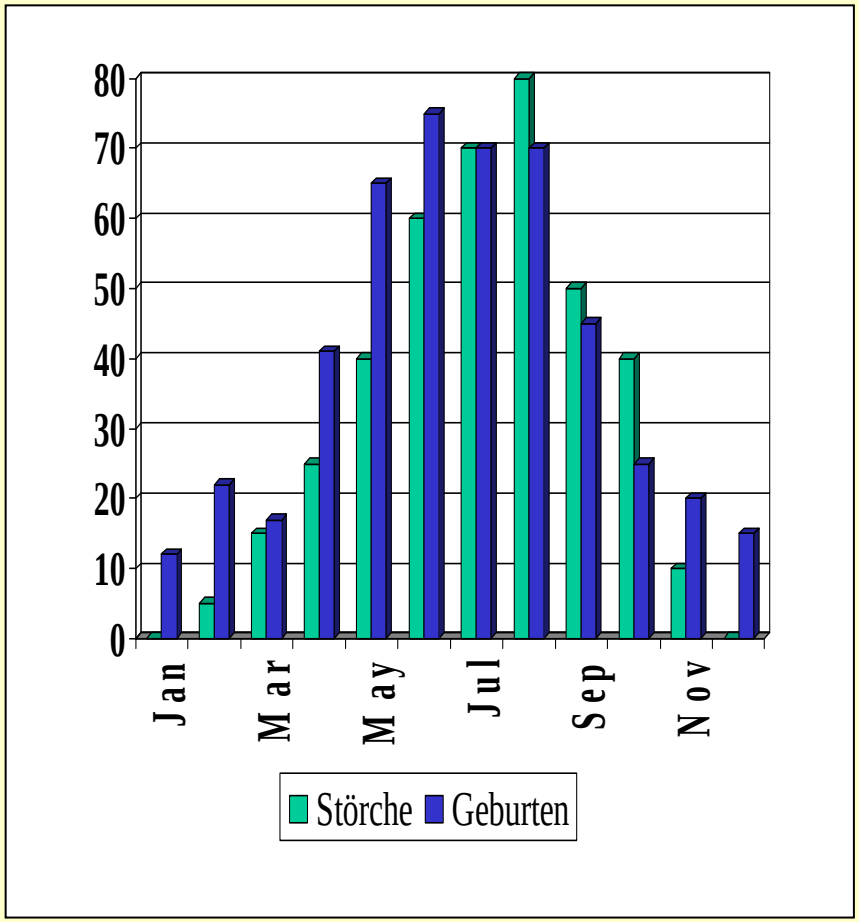
	Erfolg	∅Erfolg	Σ
Verum	2648	2352	5000
Placebo	2500	2500	5000
Σ	5148	4852	10000

χ^2=8.62, 1 df, p=0.004 OR=1.126 CI: 1.040-1.219

4. *post hoc ergo propter hoc*

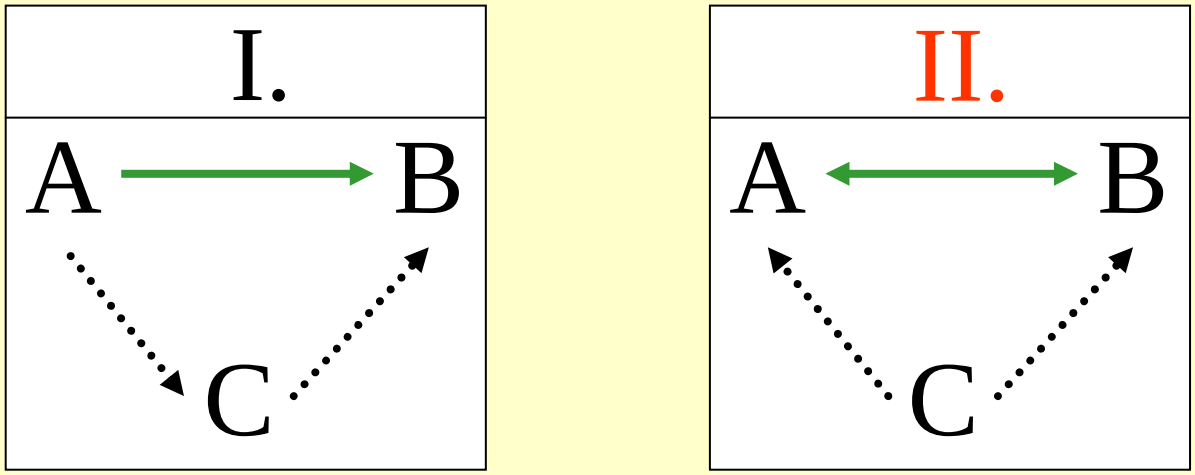
(die Sache mit dem Klapperstorch)

Von Störchen und Babys



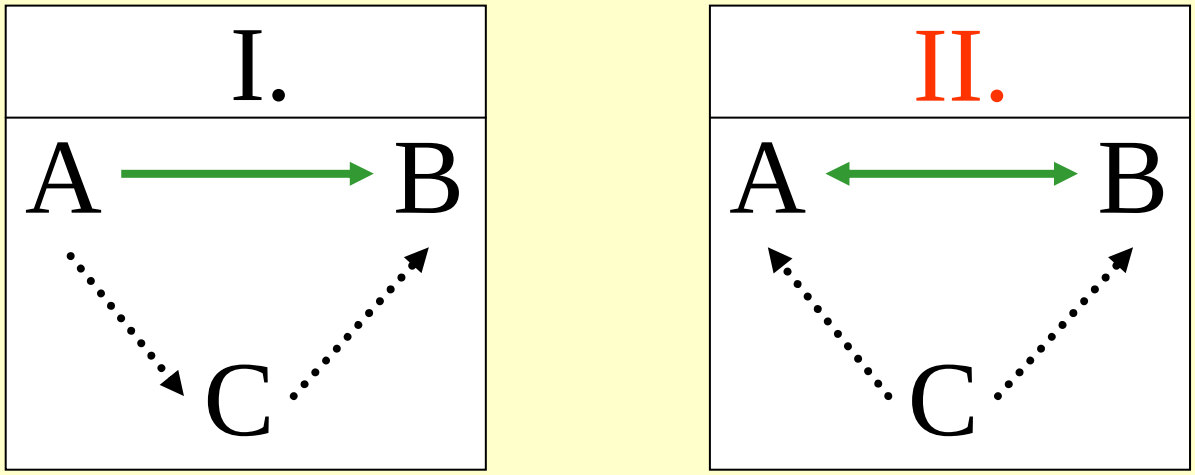
r = 0.898

Scheinkorrelation/assoziation



A: Geschlecht, B: Verhalten, C: Erziehung

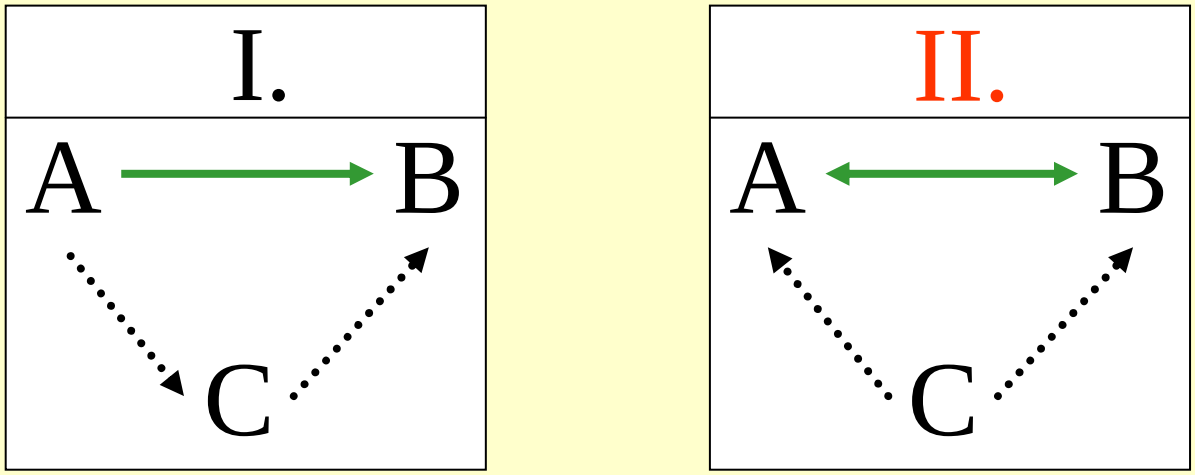
Scheinkorrelation/assoziation



A: Geschlecht, B: Verhalten, C: Erziehung

A: Therapieform, B: Morbidität, C: Mobilität

Scheinkorrelation/assoziation

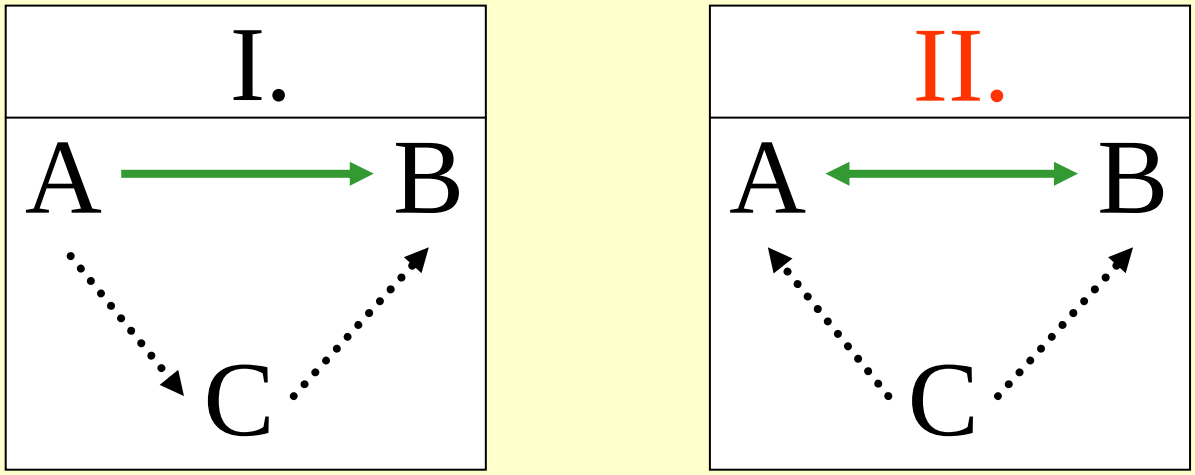


A: Geschlecht, B: Verhalten, C: Erziehung

A: Therapieform, B: Morbidität, C: Mobilität

A: Ernährung, B: Lebensdauer, C: Sozialisation

Scheinkorrelation/assoziation



A: Geschlecht, B: Verhalten, C: Erziehung

A: Therapieform, B: Morbidität, C: Mobilität

A: Ernährung, B: Lebensdauer, C: Sozialisation

A: Mobiltelefonieren, B: Schlafstörungen, C: Lebensweise

Alle Confounder bedacht?

Alle Confounder bedacht?

Alle Confounder bedacht?

... Augen links

Alle Confounder bedacht?

... Augen rechts

„The DDT ban myth“

Malaria-Prävalenz in Sri Lanka

1948	2,800,000
1958	Beginn des DDT Einsatzes
1962	*Silent Spring* (Rachel Carson)
1963	17
1964	Verbot von DDT
1968	1,000,000
1969	2,500,000

„The DDT ban myth“

Malaria-Prävalenz in Sri Lanka

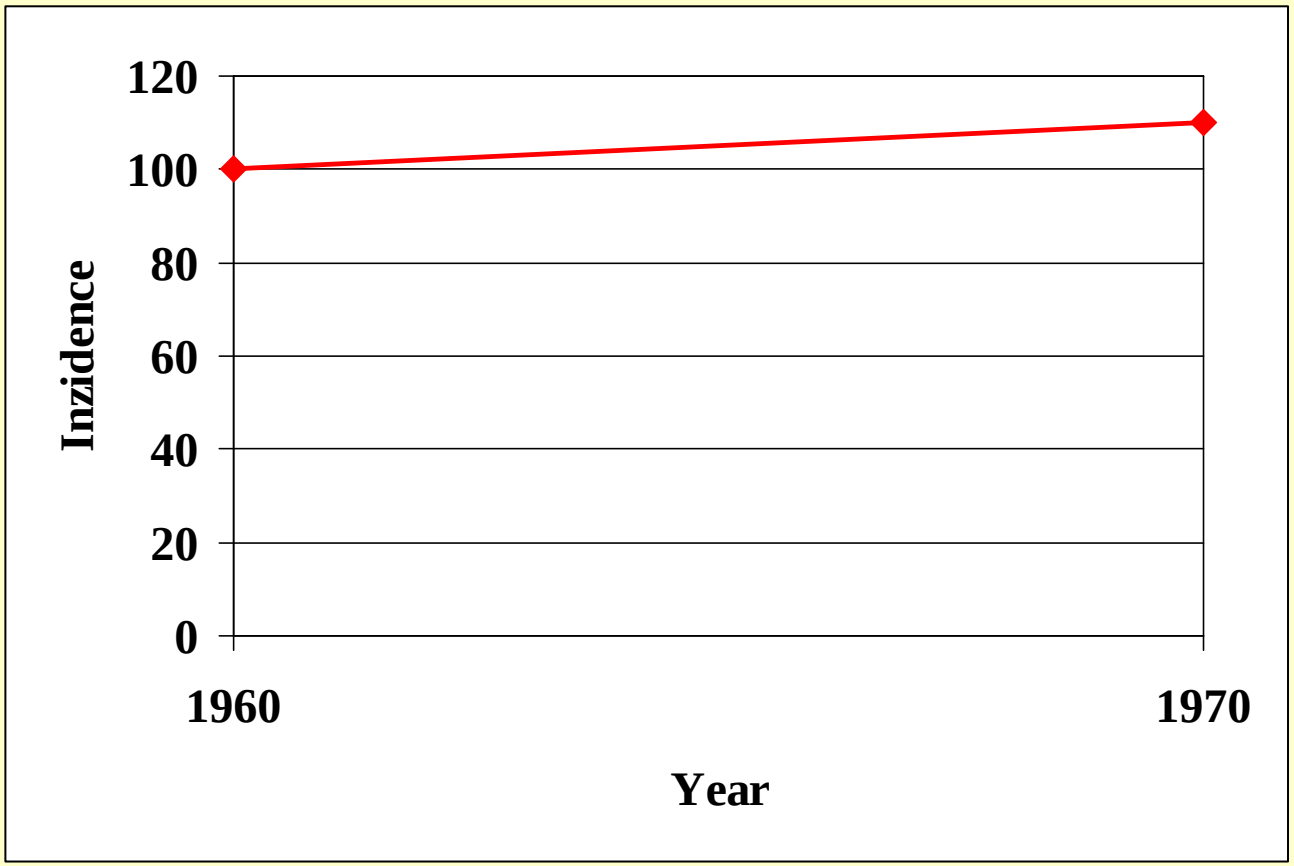
1948	
1958	Be… DDT Einsatzes
1962	*…pring* (Rachel Carson)
1963	17
1964	Verbot von DDT
1968	1,000,000
1969	2,500,000

Resistenzbildung !

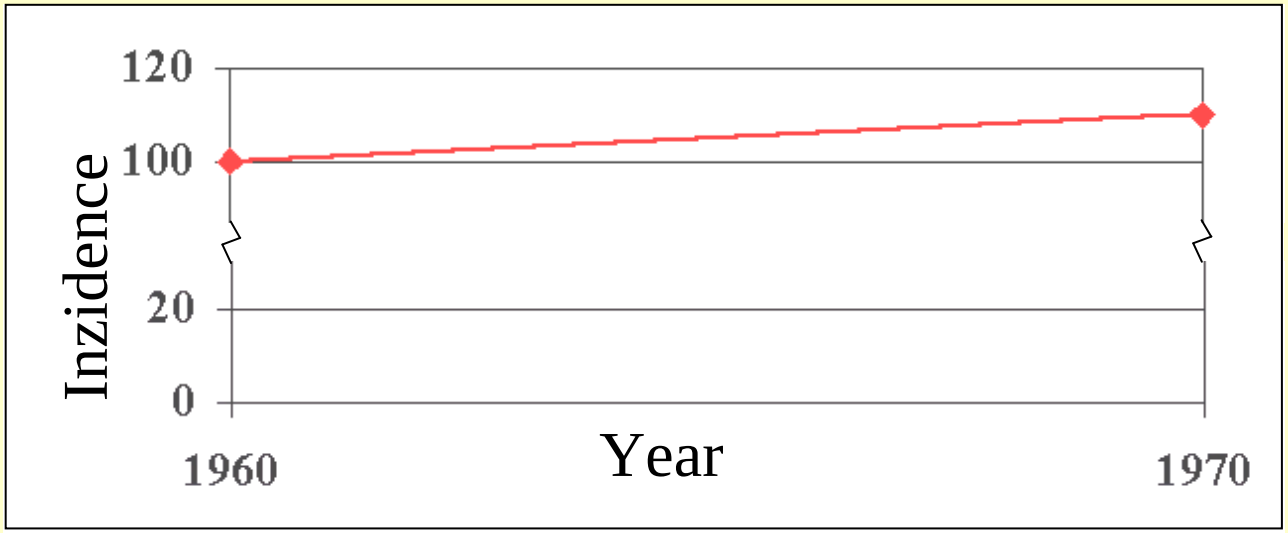
5. Skalieren, Normieren, Standardisieren

(Bilder sagen mehr als tausend Worte)

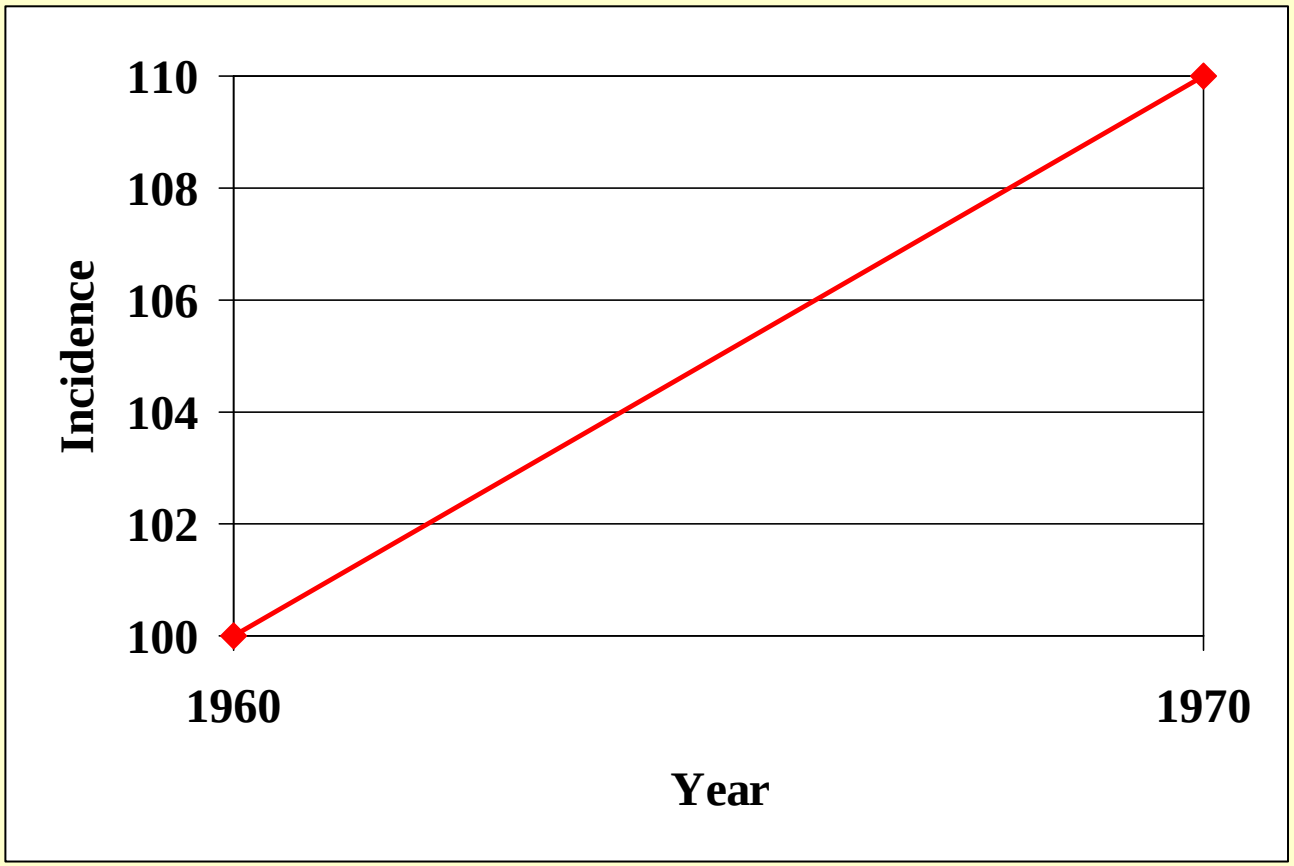
Schwache Trends, starke Trends



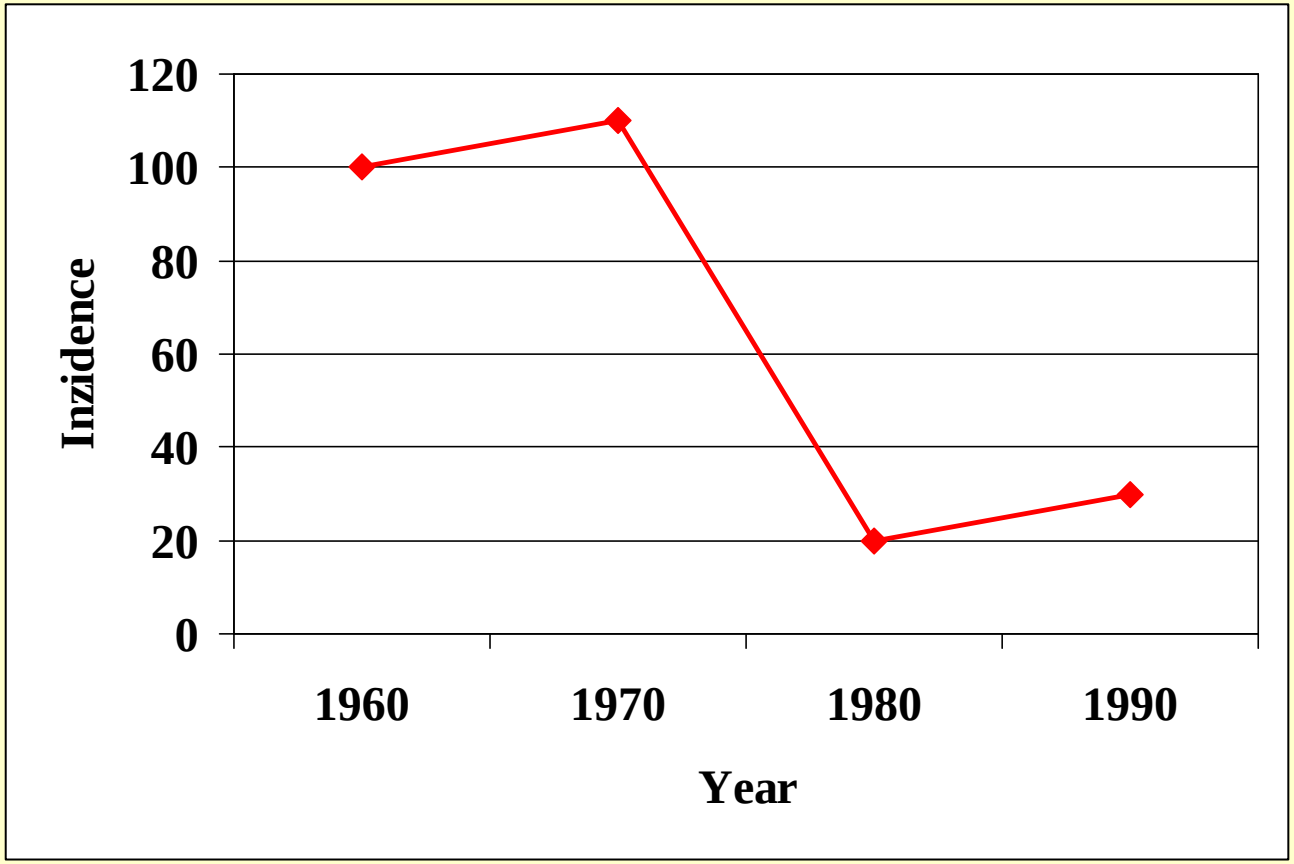
Schwache Trends, starke Trends



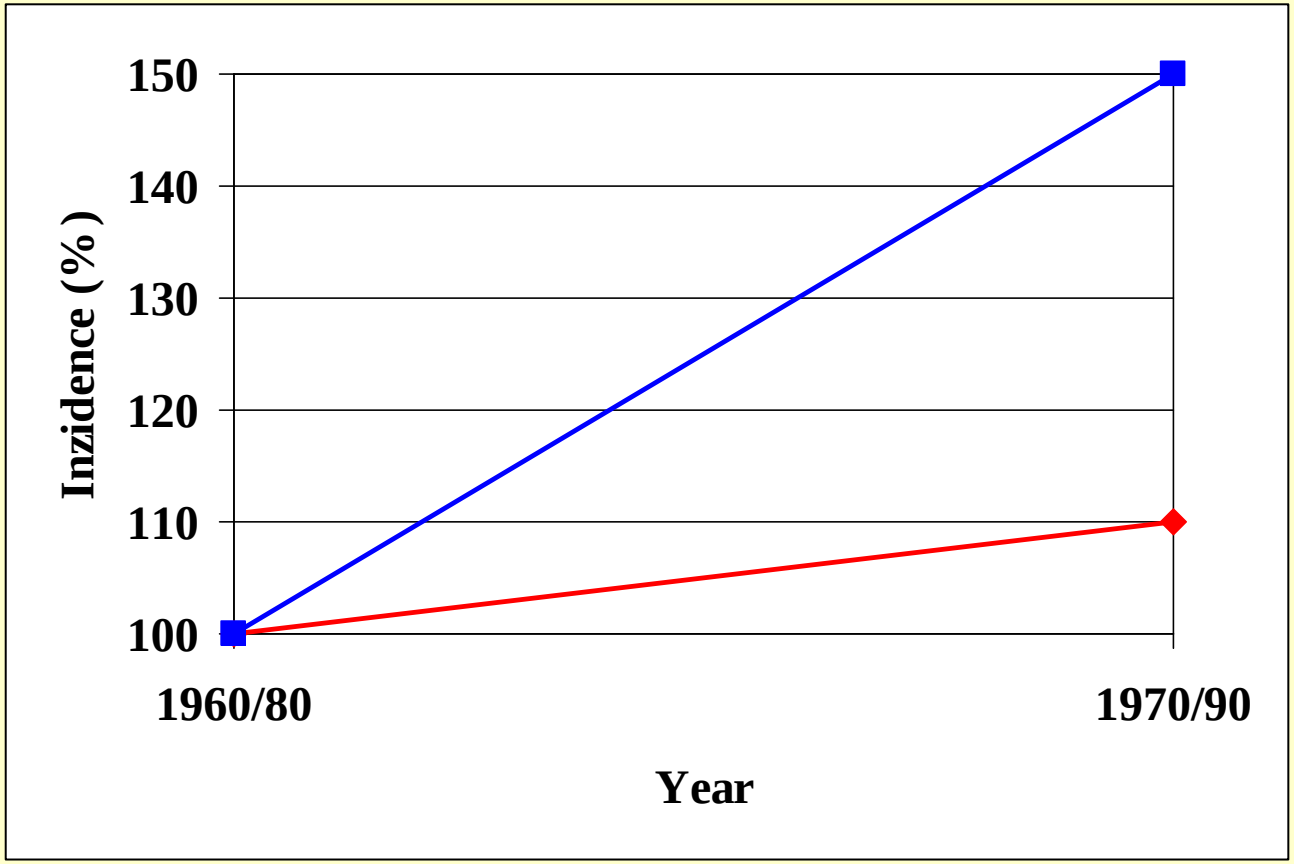
Schwache Trends, starke Trends



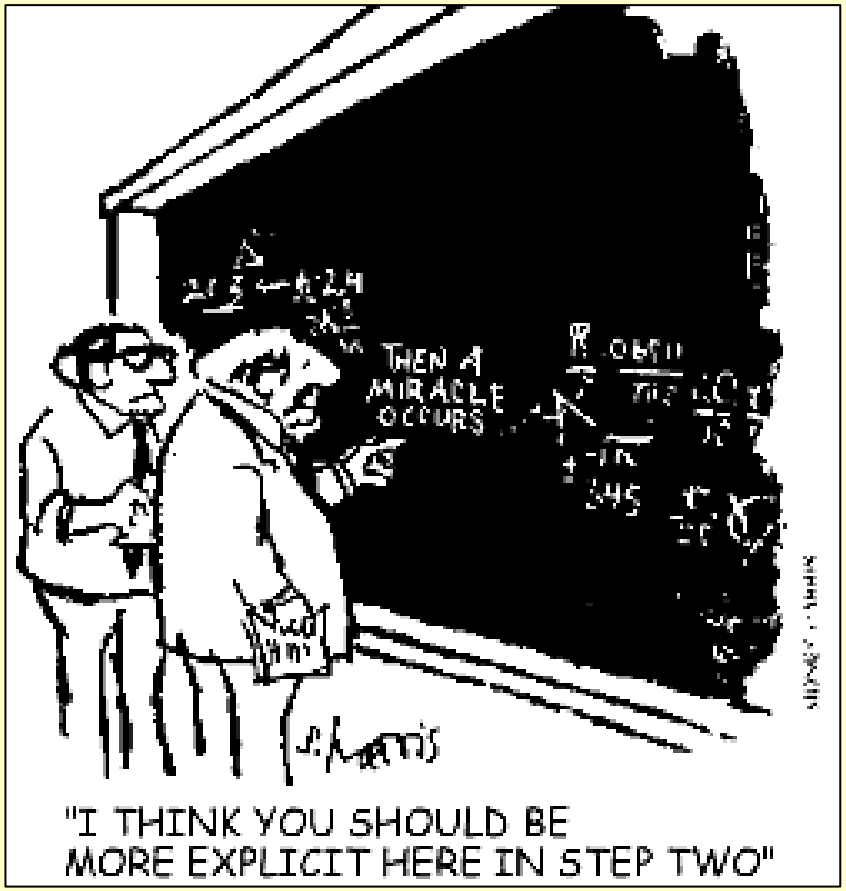
Schwache Trends, starke Trends



Schwache Trends, starke Trends



The End



"I THINK YOU SHOULD BE
MORE EXPLICIT HERE IN STEP TWO"