
Fallzahlkalkulation

Dr. Dirk Hasenclever
IMISE, Leipzig

Hasenclever@IMISE.uni-Leipzig.de

Fallzahlkalkulation bei klinischen Studien

- Prospektive klinische Studie:
 - ▶ Vorab-Planung erforderlich, wie viele Patienten man zur Beantwortung der Studienfrage benötigt
 - ▶ Erhebliche Konsequenzen bzgl.
 - Machbarkeit
 - Aufwand
 - Logistik
 - Aussagekraft

Warum Fallzahlplanung?

zu kleine Studien:

- decken selbst große Unterschiede kaum auf
- bessere Behandlung wird evtl. nicht etabliert

Beispiel: $n_{\text{konventionell}} = n_{\text{innovativ}} = 5$

konventionell: 20% Überlebenswahrscheinlichkeit

innovativ: 80% Überlebenswahrscheinlichkeit

Die Studie hat nur eine Chance von 38%, dass der große Unterschied signifikant (d.h. als nicht zufällig eingestuft) wird.

Warum Fallzahlplanung?

zu große Studien:

- „verbrauchen“ mehr Patienten als nötig
- machen auch irrelevant kleine Effekte sichtbar

Beispiel: $n_{\text{konventionell}} = n_{\text{innovativ}} = 900$

konventionell: 80% Überleben, 50% Übelkeit

innovativ: 98% Überleben, 60% Übelkeit

Der Unterschied bei der Nebenwirkung wird mit $\approx 99\%$ Wahrscheinlichkeit sichtbar.

Aber schon 106 Patienten je Arm hätten zum Aufzeigen des primär relevanten Wirkungsunterschieds mit dieser Sicherheit gereicht.

Zutaten für die praktische Fallzahlplanung

Für die Fallzahlkalkulation ist festzulegen:

- Mit welchem **primären Endpunkt** soll der Erfolg einer Therapie gemessen werden?
- Mit welchem **statistischen Werkzeug** (Test) wird er ausgewertet?
- Welcher **Unterschied** soll aufgedeckt werden?
- Mit welcher **Varianz**, die diesen Unterschied „verwischt“, muss gerechnet werden?
- Wie groß dürfen die **Fehlerraten** der statistischen Diagnostik sein?
- Welcher Anteil von **Drop-outs** wird erwartet?

Planung: Endpunkt und Test

Der primäre Endpunkt soll ...

- den zu beurteilenden medizinischen Sachverhalt **adäquat** und mit hohem Informationswert abbilden
- gut **reproduzierbar** sein
- möglichst **einfach** und bei allen Patienten zu erheben sein

Die statistische Methode soll ...

- auf die **Fragestellung** abgestimmt sein
- die **Information** aus dem Endpunkt gut ausnutzen

Planung: Therapieunterschied

Klären:

- Was ist der kleinste klinisch **relevante** Unterschied?
- Welcher Unterschied wird **realistisch** erwartet?

→ Falls „erwartet“ < „relevant“:

Ist die Studie überhaupt sinnvoll?

Quellen zur Schätzung des Unterschieds:

- Niveau des Standards aus Erfahrung
- Niveau des experimentellen Arms z.B. aus Daten einer Pilotstudie, aus Theorie (Modellrechnungen), oder Effektgrößen vergleichbarer Studien

Frühe Studien der EORTC

- European Organisation for Research on Treatment of Cancer
- Studien bis 1973: 71
 - ▶ Nicht signifikant 66
 - ▶ Signifikant 5
 - Wenig Fortschritt?
- Mediane Fallzahl: 245
 - ▶ Mit 80% Power konnte nur ein Unterschied von >25% im Überleben aufgedeckt werden...

Freiman et al. 1973 NEJM 299,13:690

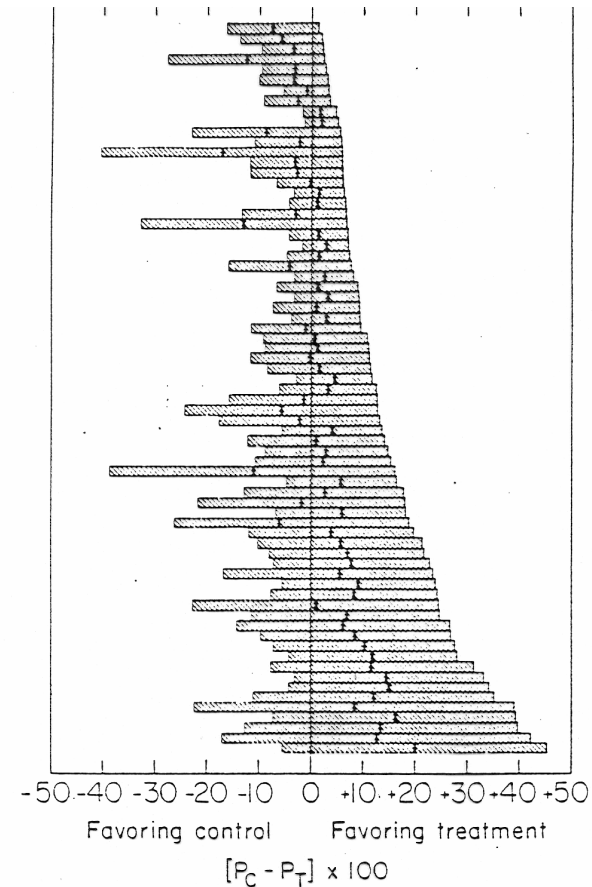


Figure 2. Ninety per Cent Confidence Limits for the True Percentage Difference for the 71 Trials.
The vertical bar at the center of each interval indicates the observed value, $\hat{P}_C - \hat{P}_T$, for each trial.

Ergebnisse von Metaanalysen randomisierter Studien (nach Buyse 1991)

Tumor	Therapie	Zugewinn 5-Jahres Überlebensrate
Mamma Ca. > 50 Jahre	Hormontherapie Tamoxifen	+4,7%
Mamma Ca. < 50 Jahre	Chemotherapie CMF	+4,8%
Kolorektal Ca.	5-FU	+5,0%
Ovarial Ca.	Adriamycin (CAP versus CP)	+6%

Relevante kleine Unterschiede bei häufigen Krankheiten!

Planung: Varianzschätzung

Informationsquellen zur Varianzschätzung:

- vorhandener Datensatz
- Literatur (explizite Angabe, oder detektivisch berechnet z.B. aus berichtetem P-Wert, Mittelwertsdifferenz und Fallzahl)
- Vorversuch
- Berechnung aus einer mutmaßlichen Werteverteilung
- Simulation auf Grundlage eines Modells über das Zustandekommen der betrachteten Größe
-

Planung: Statistische Fehlergrößen

- α = Risiko für einen **Fehler 1. Art**
(Zufallsunterschied wird als essentiell interpretiert)
- β = Risiko für einen **Fehler 2. Art**
(relevanter Unterschied wird nicht erkannt)
Power = $1 - \beta$

ICH guideline E9 (Statistical Principles for Clinical Trials)

- type I error ... conventionally set at 5% or less
- type II error ... conventionally set at 10% or 20%
- alternative values ... may be acceptable or even preferable in some cases www.ich.org/ich5e.html

Planung: Drop-outs

- Drop-out-Rate bei routinemäßigen Nachkontrollen der Standardtherapie kann als **Planungsgrundlage** dienen
- Überlegen: Was sind mögliche **Ursachen** für Drop-out? Wie kann ihnen entgegengewirkt werden?
- Die Zahl zu rekrutierender Patienten wird so festgelegt, dass nach Abzug der angenommenen Drop-out-Rate die angestrebte **Zahl auswertbarer Fälle** erreicht wird.

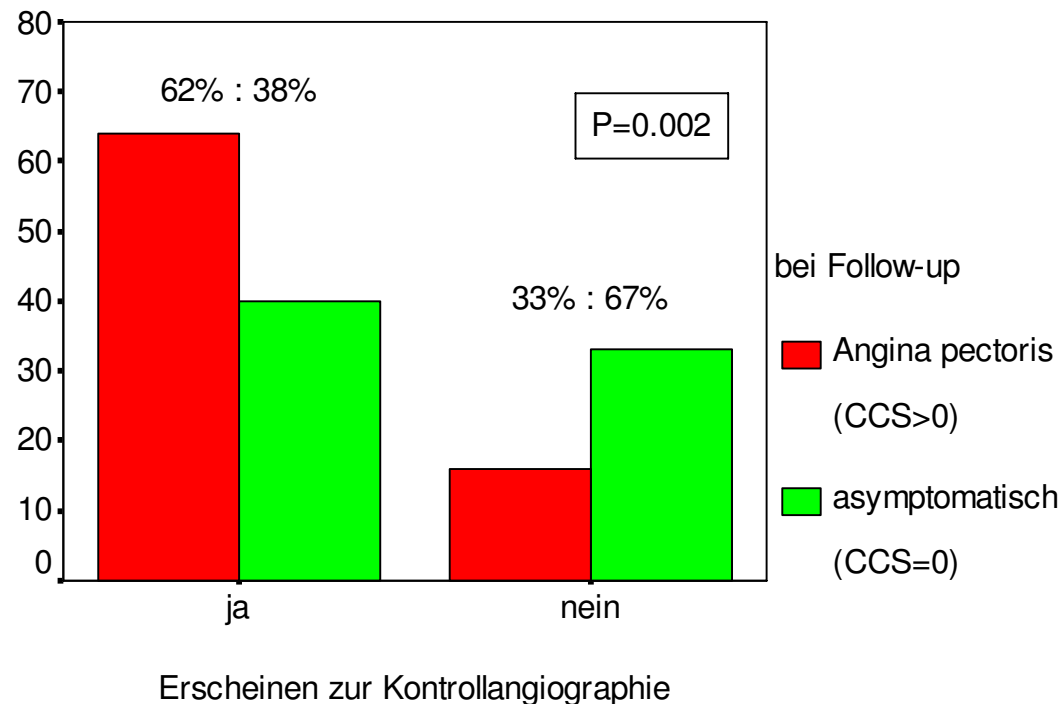
Beispiel: bei 20% Drop-out → $\text{Fallzahl} \times 1.25$

Bsp. informative Drop-outs

Achtung: Informative Drop-outs verzerren die Effekte.
Das lässt sich nicht durch höhere Fallzahlen korrigieren.

Beispiel: Erscheinen zur Kontrollangiographie nach PTCA

(= perkutane transluminale coronare Angioplastie (PTCA) = ein Verfahren zur Erweiterung verengter Herzkranzgefäße, der Koronar-Arterien. Balondilatation.)



bei Follow-up

Angina pectoris
(CCS>0)

asymptomatisch
(CCS=0)

Telefonische
Nachfrage
Bei Nicht-
Erschienenen.

Determinanten der Fallzahl

Die Fallzahl steigt mit ...

- kleinerem gewünschten Signifikanzniveau α
- größerer gewünschter Power
- kleinerem aufzudeckendem Unterschied (annähernd quadratisch: halber Unterschied, vierfache Fallzahl)
- größerer Varianz des Endpunkts
- höherer Drop-out-Rate

Cave: Fallzahl sehr sensitiv gegenüber einigen Annahmen

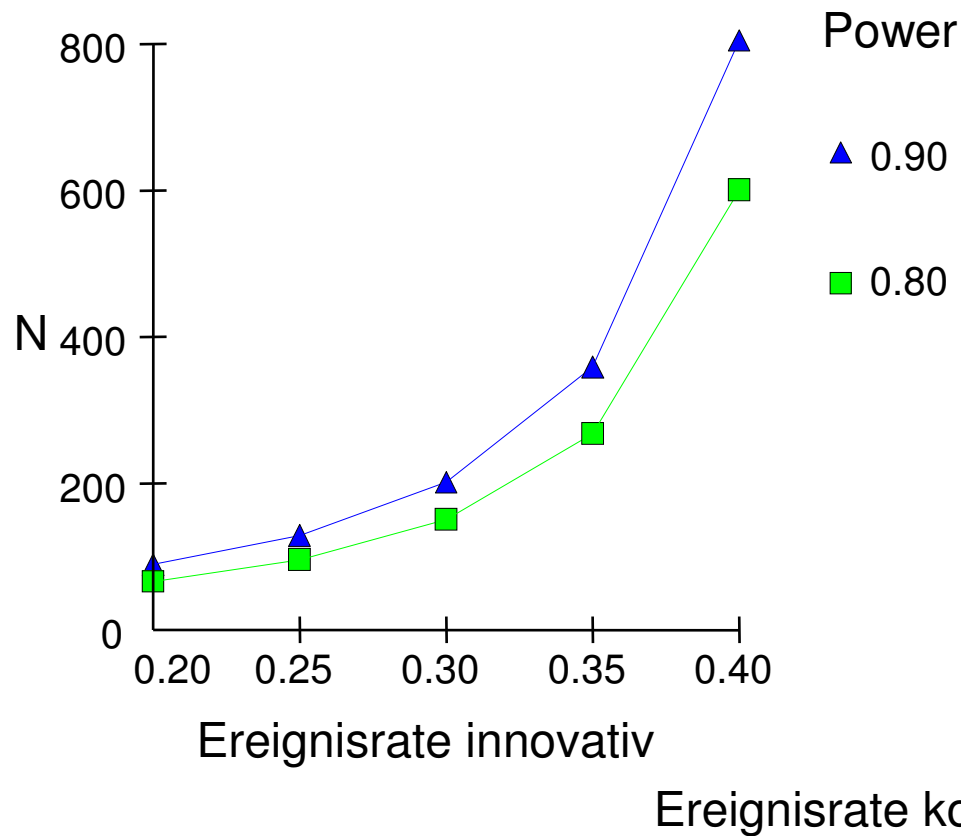
Deswegen Poweranalyse: Wie ändert sich die Power bei geplanter Fallzahl und variierenden Planungsparametern?

Beispiel: Fallzahlkalkulation

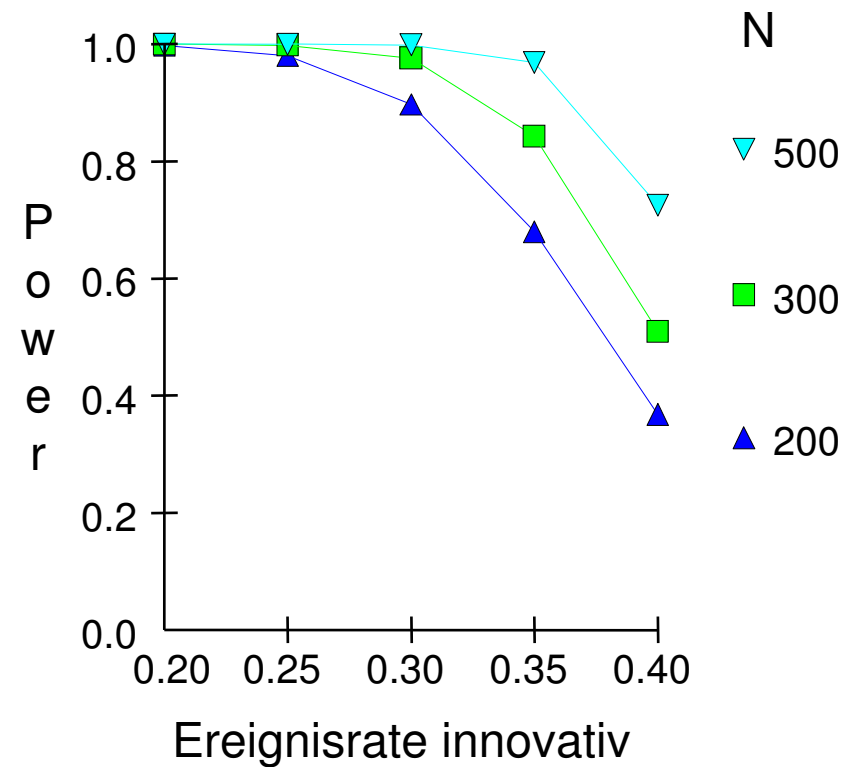
- **Fragestellung:** Kann die Zahl der Hypoglycämien bei Typ-1-Diabetikern durch eine innovative Schulung gesenkt werden?
 - **Endpunkt:** Wartezeit auf schwere Hypoglycämie
 - **Methode:** Kaplan-Meier Überlebenszeit, 2-seitig
 - **Beobachtungszeit** je Patient: 2 Jahre
 - Ereignisrate bei **Standardschulung:** 0.5/Jahr
 - angestrebt bei **innovativer** Schulung: 0.3/Jahr
 - **Fehlergrößen:** $\alpha=5\%$, $\beta=10\%$
 - erwartete **Drop-out-Rate:** 0.2/Jahr
- **benötigte Fallzahl insgesamt: N=202**

Beispiel: Fallzahl- und Poweranalyse

Fallzahl vs. Unterschied



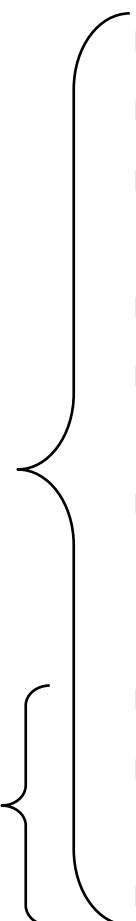
Power vs. Unterschied



Biometrie + Statistik

Biometrie = Anlegen
eines Meßwerkzeugs
an einen biologischen
Sachverhalt

Statistik = Rechnen mit
Wahrscheinlichkeiten

- 
- präzise Formulierung der klinischen Fragestellung
 - Prüfung möglicher Designs, Designoptimierung
 - Festlegung des Regimes der Erhebung endpunkt-relevanter Informationen und der Werkzeuge dafür
 - Definition des Endpunkts
 - Festlegung geeigneter Prozeduren, welche die valide Erhebung des Endpunkts sichern
 - Beschaffung von Informationen, welche die voraussichtliche Datenlage beim Endpunkt erahnen lassen
 - Festlegung der Auswertungsmethode
 - Zusammenstellung der quantitativen Parameter zur Fallzahlberechnung
 - Durchführung der Berechnung

Frequentistische Fallzahlplanung I

Einfachster Fall : $X \sim N(\theta, \sigma^2)$; σ^2 bekannt.

Einseitiges Testproblem : $H_0 : \theta \leq 0$ gegen $H_1 : \theta > 0$

Stichprobe des Umfangs n

Testgröße : $Z = \frac{\bar{X}_{(n)}}{\sigma} \sqrt{n} \sim N(0,1)$ unter der Nullhypothese. ("Gauss - Test")

Sei $\Phi(z)$ die Verteilungsfunktion der Standardnormalverteilung $N(0,1)$:

$\Phi(z) = \int_{-\infty}^z \phi(\varsigma) d\varsigma$; mit der Dichtefunktion $\phi(\varsigma) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\varsigma^2}{2}\right)$

Sei $Z_{1-\alpha}$ das $1 - \alpha$ Quantil der Standardnormalverteilung : $\Phi(Z_{1-\alpha}) = 1 - \alpha$

$\Rightarrow$ Lehne H_0 ab, falls $Z > Z_{1-\alpha}$

(N.B. : Falls Test zweiseitig sein soll : $|Z| > Z_{1-\alpha/2}$)

Frequentistische Fallzahlplanung II

Planungsvorgaben : Δ = relevanter Unterschied, für den der Fehler zweiter Art kontrolliert, bzw. die Power garantiert werden soll.

$1 - \beta$ = Powervorgabe

Verteilung der Teststatistik Z unter der konkreten Alternative H_Δ : $X \sim N(\Delta, \sigma^2)$

$$Z \sim N\left(\frac{\Delta}{\sigma}\sqrt{n}, 1\right)$$

Powerbedingung :

$$1 - \beta = \text{pr}(\text{Ablehnen } H_0 \mid H_\Delta) = \text{pr}(Z > Z_{1-\alpha} \mid H_\Delta) = 1 - \Phi\left(Z_{1-\alpha} - \frac{\Delta}{\sigma}\sqrt{n}\right) = \Phi\left(\frac{\Delta}{\sigma}\sqrt{n} - Z_{1-\alpha}\right)$$

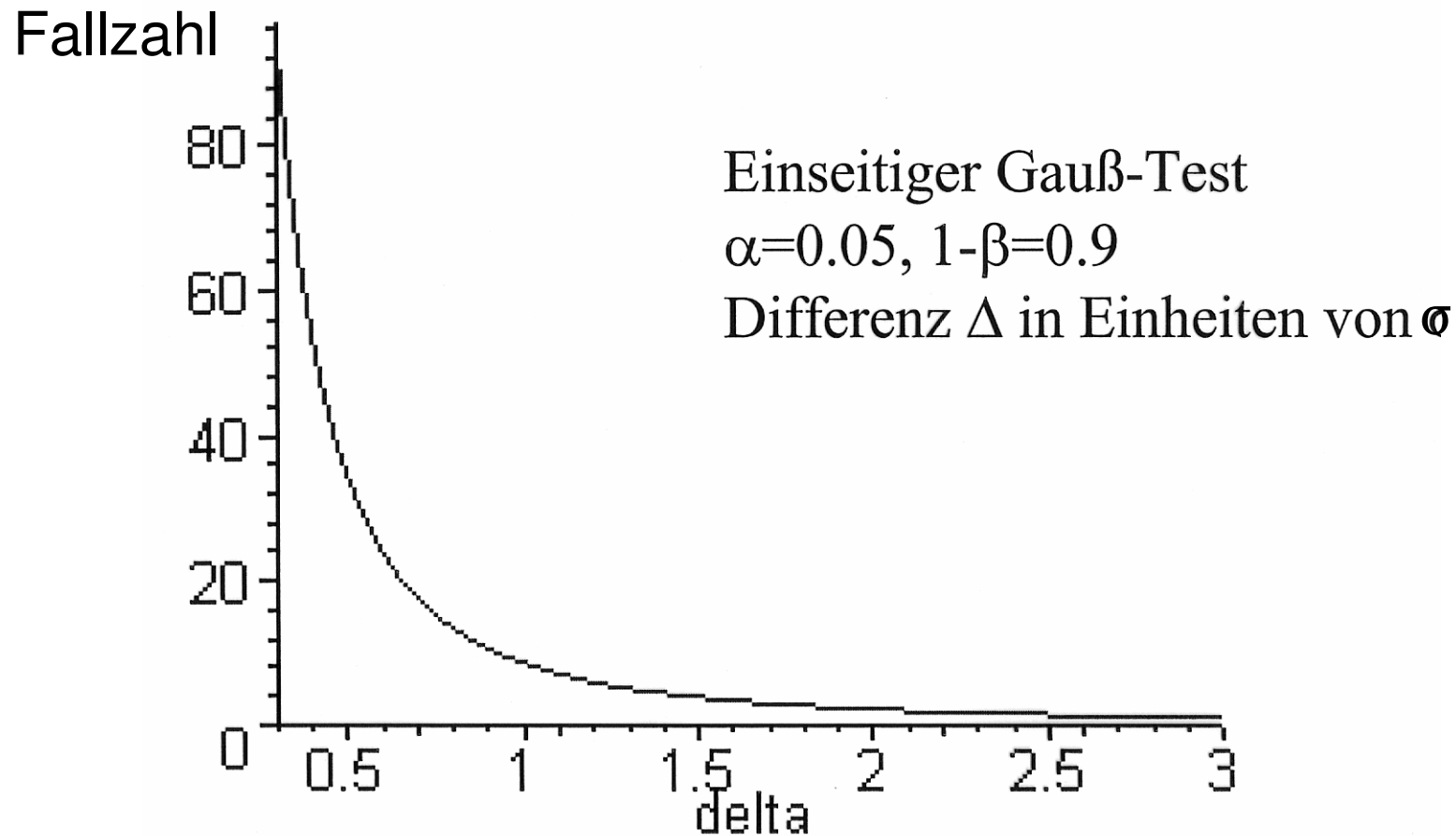
Also (Φ^{-1} auf beiden Seiten) : $Z_{1-\beta} := \Phi^{-1}(1 - \beta) = \frac{\Delta}{\sigma}\sqrt{n} - Z_{1-\alpha}$

Auflösen nach n ergibt die erforderliche Fallzahl :

$$n = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2}{\left(\frac{\Delta}{\sigma}\right)^2}$$

$\left(\frac{\Delta}{\sigma}\right)$ heißt auch der effektive Unterschied (Δ in Einheiten der Standardabweichung).

Frequentistische Fallzahlplanung III



Frequentistische Fallzahlplanung IV

- Falls man σ nicht kennt, muss man
 - Für Planungszwecke σ vorab-abschätzen („guestimate“) und
 - in der Formel statt der Normalverteilung die t-Verteilung verwenden.
- Ähnliche (meist kompliziertere Formeln) für andere Variablen-Typen (z.B. binär) und Design- Typen (z.B. Zwei Gruppen-Vergleich)
- **Paradox der Studienplanung:**
 - Um eine Studie optimal planen zu können, sollte man besser ihr Ergebnis kennen...!

Minimales n für den Vierfeldertest

Tabelle 157. Minimale Stichprobenumfänge ($n_1 = n_2$) für den Zweistichproben-Fisher-Test bei einseitiger Fragestellung; nach Haseman, J. K. (1978): Exact sample sizes for use with the Fisher-Irwin-Test for 2×2 tables. Biometrics 34:106–109 [der auch Stichprobenumfänge für die Power = 0,5 gibt, die im allgemeinen im Bereich 1/2 bis 1/3 der oberen Zahlen (Power = 0,9) liegen]

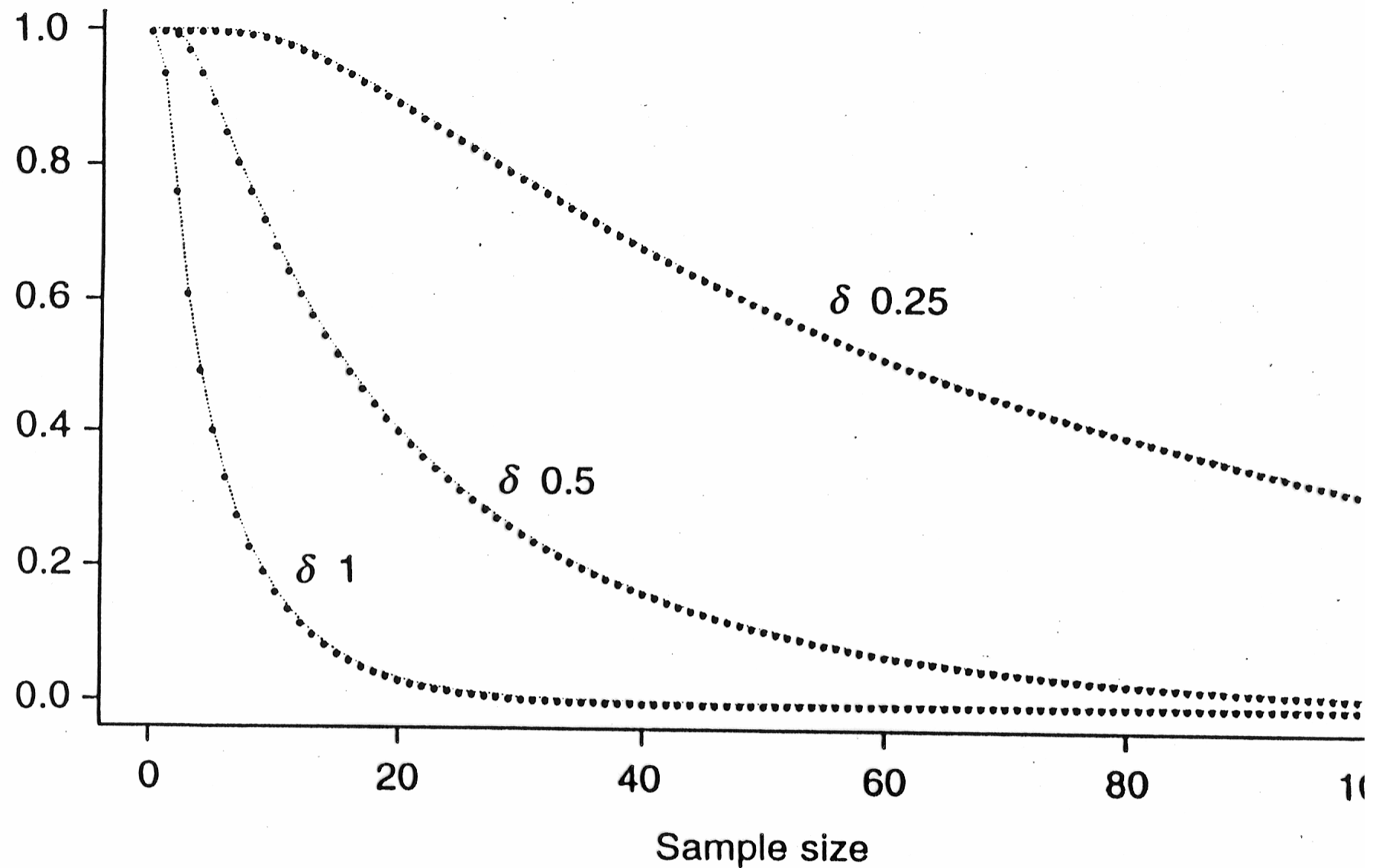
oberes Dreieck: $\alpha = 0,01$; obere Zahl: Power = 0,9
unteres Dreieck: $\alpha = 0,05$; untere Zahl: Power = 0,8

$\pi_1 \backslash \pi_2$	0,95	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05
0,95		745	130	60	37	25	18	14	12	9	8
0,9	503		344	108	56	35	24	18	13	9	9
0,8	371	232		269	86	46	29	20	15	12	9
0,7	89	173	503		143	69	42	28	18	13	12
0,6	67	74	338	393		113	55	34	22	16	9
0,5	42	56	249	408	609		77	43	28	18	14
0,4	34	39	97	302	475	129		60	35	22	15
0,3	25	30	73	111	163	229	171		77	42	24
0,2	20	25	47	84	116	163	129	609		69	35
0,1	18	19	36	41	53	84	111	143	503		25
0,05	14	13	23	23	31	41	53	86	108	334	
	10	12	18	23	30	47	73	113	173	269	130
	9	10	15	23	30	47	73	113	173	269	101
	8	10	12	18	25	39	56	89	130	173	232
	6	8	10	12	17	25	39	56	89	130	173
	5	6	8	10	13	18	25	34	42	50	583
	5	5	6	9	11	14	20	34	67	371	

Sachs

Fallzahlplanung Likelihoodparadigma I

- Statt die Fallzahl zu bestimmen, mit der man vorgegebene Fehler erster und zweiter Art garantieren kann, werden im Likelihoodparadigma die beiden folgende Größen kontrolliert:
- (Angenommen eine der beiden Punkthypothesen H_0 und H_A sei wahr)
 - Wahrscheinlichkeit M **starke Evidenz für die falsche Hypothese** zu beobachten.
 - Wahrscheinlichkeit W dafür, dass am Ende der Studie **keine starke Evidenz für die richtige Hypothese** vorliegt.



Royall

Figure 4.1 *Probability of weak evidence for normal mean: $\frac{1}{8} < L_2/L_1 < 8$
Hypothesized means differ by δ standard deviations.*

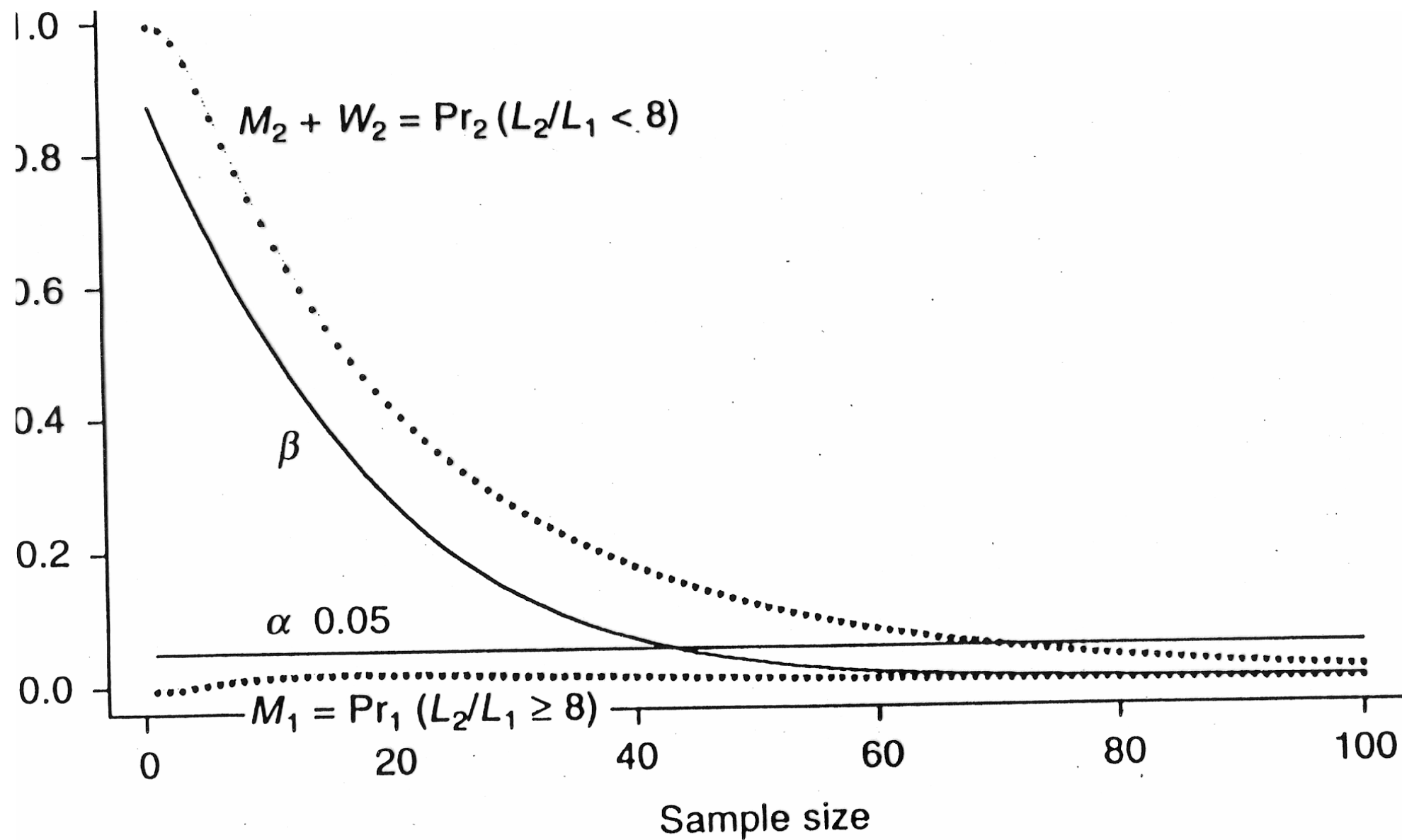


Figure 4.3 *Probabilities of undesirable results for normal mean.*

Royall

Fallzahlplanung Likelihoodparadigma II

Einfachster Fall: $X \sim N(\theta, \sigma^2)$; σ^2 bekannt.

Hypothesenpaar: $H_1: \theta = \theta_1$ gegen $H_2: \theta = \theta_1 + \delta$

Stichprobe des Umfangs n , Beobachtung $\bar{x}$

$$\text{Maß für Evidenz: } LR = \frac{f(\bar{x} | \theta_1 + \delta, \sigma)}{f(\bar{x} | \theta_1, \sigma)} = \exp\left[\frac{n\delta}{\sigma^2}\left(\bar{x} - \left(\theta_1 + \frac{\delta}{2}\right)\right)\right]$$

$$\text{Unter } H_1 \text{ berechne: } \text{pr}_1(LR \geq k) = \text{pr}_1\left(\frac{n\delta}{\sigma^2}\left(\bar{x} - \left(\theta_1 + \frac{\delta}{2}\right)\right) \geq \ln(k)\right)$$

$$\ln() \quad | \quad = \text{pr}_1\left(\bar{x} \geq \frac{\sigma^2}{n\delta} \ln(k) + \left(\theta_1 + \frac{\delta}{2}\right)\right)$$

$$\begin{aligned} \bar{x} \sim N\left(\theta_1, \frac{\sigma^2}{n}\right) \quad | \quad &= \text{pr}_1\left(\frac{\bar{x} - \theta_1}{\sigma/\sqrt{n}} \geq \frac{\sigma \ln(k)}{n\delta} + \frac{\delta\sqrt{n}}{2\sigma}\right) \\ &= 1 - \Phi\left(\frac{\sigma \ln(k)}{n\delta} + \frac{\delta\sqrt{n}}{2\sigma}\right) = \Phi\left(-\frac{\sigma \ln(k)}{n\delta} - \frac{\delta\sqrt{n}}{2\sigma}\right) \end{aligned}$$

Dies ist die Wahrscheinlichkeit $M_1(k)$ für "misleading evidence für H_2 " falls H_1 wahr.

Fallzahlplanung Likelihoodparadigma II

Wegen $\Phi(z) = 1 - \Phi(-z)$ gilt :

$$\text{pr}_1(\text{LR} \leq k) = 1 - \Phi\left(-\frac{\sigma \ln(k)}{n\delta} - \frac{\delta\sqrt{n}}{2\sigma}\right) = \Phi\left(\frac{\sigma \ln(k)}{n\delta} + \frac{\delta\sqrt{n}}{2\sigma}\right)$$

$$\text{pr}_1(\text{LR} \leq \frac{1}{k}) = \Phi\left(\frac{\delta\sqrt{n}}{2\sigma} - \frac{\sigma \ln(k)}{n\delta}\right) \quad | \quad \ln(1/k) = -\ln(k)$$

ist die Wahrscheinlichkeit für starke Evidenz für richtige Hypothese.

$$W_1(k) = \text{pr}_1\left(\frac{1}{k} \leq \text{LR} \leq k\right) = \Phi\left(\frac{\sigma \ln(k)}{n\delta} + \frac{\delta\sqrt{n}}{2\sigma}\right) - \Phi\left(\frac{\delta\sqrt{n}}{2\sigma} - \frac{\sigma \ln(k)}{n\delta}\right)$$

Dies ist die Wahrscheinlichkeit für "weak evidence",
d.h. starke Evidenz für keine der beiden Alternativen.

Wähle n so, dass diese Wahrscheinlichkeiten akzeptabel.

Fallzahlplanung Bayesianisch I

Einfachster Fall : $X \sim N(\theta, \sigma^2)$; σ^2 bekannt. θ unbekannt.

Stichprobe des Umfangs n , Beobachtung $\bar{x}$

Der Prior für θ sei auch Normalverteilt $\theta \sim N(\theta_0, \varphi_0^2)$

Posterior : $\theta \sim N(\theta_1, \varphi_1^2)$

mit $\frac{1}{\varphi_1^2} = \frac{1}{\varphi_0^2} + \frac{1}{(\sigma^2/n)}$ ("Präzisionen addieren sich")

$\theta_1 = \frac{\frac{1}{\varphi_0^2}}{\frac{1}{\varphi_1^2}} \theta_0 + \frac{\frac{1}{(\sigma^2/n)}}{\frac{1}{\varphi_1^2}} \bar{x}$ ("Präzisionsgewichteter Mittelwert")

Die Normalverteilung ist mit sich selbst konjugiert :

Normaler Prior + Normale Daten = Normaler Posterior!

Fallzahlplanung Bayesianisch II

Prädiktive Verteilung :

Gegeben $X \sim$ Dichte $f(x | \theta)$ und Prior $\pi(\theta)$ für θ

Was glaubt man über eine zukünftige Realisierung x von X ? (Vorhersage)

$$\text{pr}(x | \pi(\theta)) = \int_{\theta \in \Theta} f(x | \theta) \pi(\theta) d\theta$$

CAVE : $\text{pr}(x | \pi(\theta))$ ist viel breiter als $\pi(\theta)$

Bsp. : $X \sim N(\theta, \sigma^2)$ Prior : $\theta \sim N(\theta_0, \varphi_0^2)$

$\text{pr}(x | \pi(\theta)) = N(\theta_0, \varphi_0^2 + \sigma^2)$ bei einer Beobachtung und

$\text{pr}(x | \pi(\theta)) = N(\theta_0, \varphi_0^2 + \frac{\sigma^2}{n})$ bei n Beobachtungen.

Fallzahlplanung Bayesianisch III

- Mittels der Prädiktiven-Verteilung möglicher Beobachtung kann man nun z.B.
 - bzgl. eines plausiblen Priors
 - die prädiktive Wahrscheinlichkeit in Abhängigkeit von n dafür ausrechnen,
 - dass nach Vorliegen der Daten die resultierende Posteriorverteilung
 - $\text{pr}(\theta \geq \theta_0) \geq 1 - \alpha$ erfüllt.
- und danach die Fallzahl auswählen.