

8.9 Macht

a Bisher: H_0 vorausgesetzt.

Wahrscheinlichkeiten nur unter dieser Voraussetzung berechnet.

$$P^{(0)} \langle \text{Verwerfen von } H_0 \rangle = \alpha$$

Jetzt: Alternative voraussetzen.

Fehler: H_0 nicht verwerfen. = Fehler 2. Art

Wahrscheinlichkeit berechnen!

Berechnung von Wahrscheinlichkeiten setzt genau spezifizierte Hypothese voraus!

$$P^{(A)} \langle T \notin K \rangle = 1 - P^{(A)} \langle T \in K \rangle = \beta^{(A)}.$$

$1 - \beta^{(A)}$ = W. der „richtigen“ Entscheidung = **Macht**
soll möglichst hoch sein.

(Aufgepasst: In manchen Büchern wird
die Macht mit β bezeichnet.)

- b K ist der Verwerfungsbereich, hergeleitet aus der
Verteilung der Test-Statistik **unter der Nullhypothese**
Der Test (die Entscheidungsregel) bleibt.

- c **Beispiel Mastochsen:** 2 Fütterungsarten vergleichen.
„Wahrer“ Unterschied $\delta = 1.0 \text{ kg/Woche}$ (Vermutung).
Wie gross ist die Wahrscheinlichkeit, dass dieser Effekt, falls er richtig ist, sich durch den Versuch statistisch nachweisen lässt?
... d.h., dass die Testregel zur Verwerfung von H_0 führt?

Entscheidungsregel: H_0 verwerfen, falls $Z > 1.64$.

→ $U = Z \cdot \sqrt{1.8^2 \cdot 2/11} = Z \cdot 0.77$, also

$Z > 1.64$ → $U > 1.64 \cdot 0.77 = 1.26$.

Vt. von T unter H_A mit $\delta = 1.0\text{kg}$ ist

$$U = \bar{Y}_2 - \bar{Y}_1 \sim \mathcal{N} \left(\delta, \sigma_0^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \right) = \mathcal{N} \langle 1.0, 0.77^2 \rangle$$

Macht $1 - \beta^{(A)} = P^{(A)} \langle U > 1.26 \rangle$

$$= 1 - \Phi \left\langle \frac{1.26 - 1}{0.77} \right\rangle = 1 - \Phi \langle 0.34 \rangle = 0.37$$

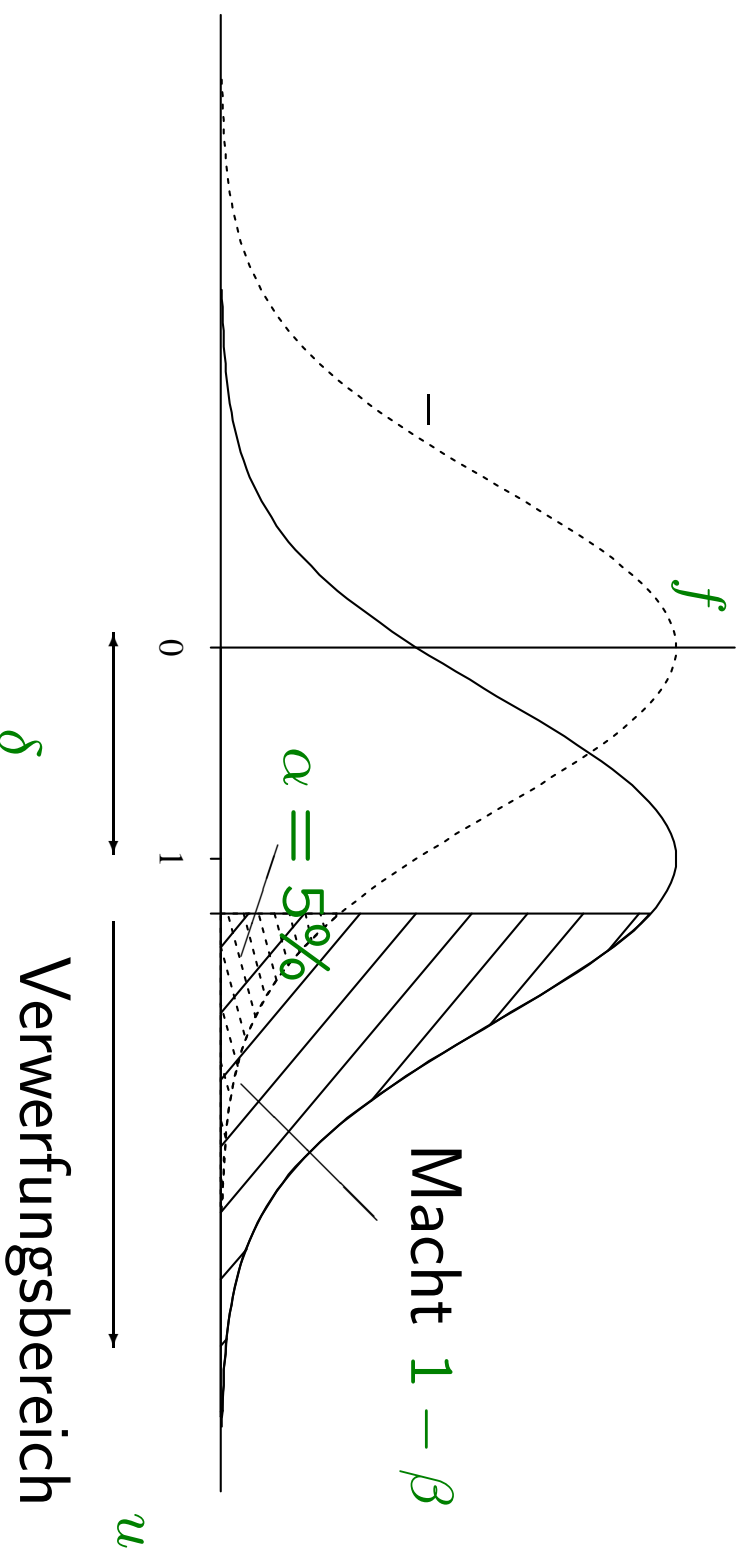
→ Wahrsch. eines „erfolgreichen Ausgangs“ des Versuchs

(statistischer „Beweis“ eines Effekts gelungen)

$$= 37\%. (!)$$

Verteilung der Test-Statistik U ...

... unter H_0 unter H_A



d Oft keine Vermutung betreffend Effekt δ .

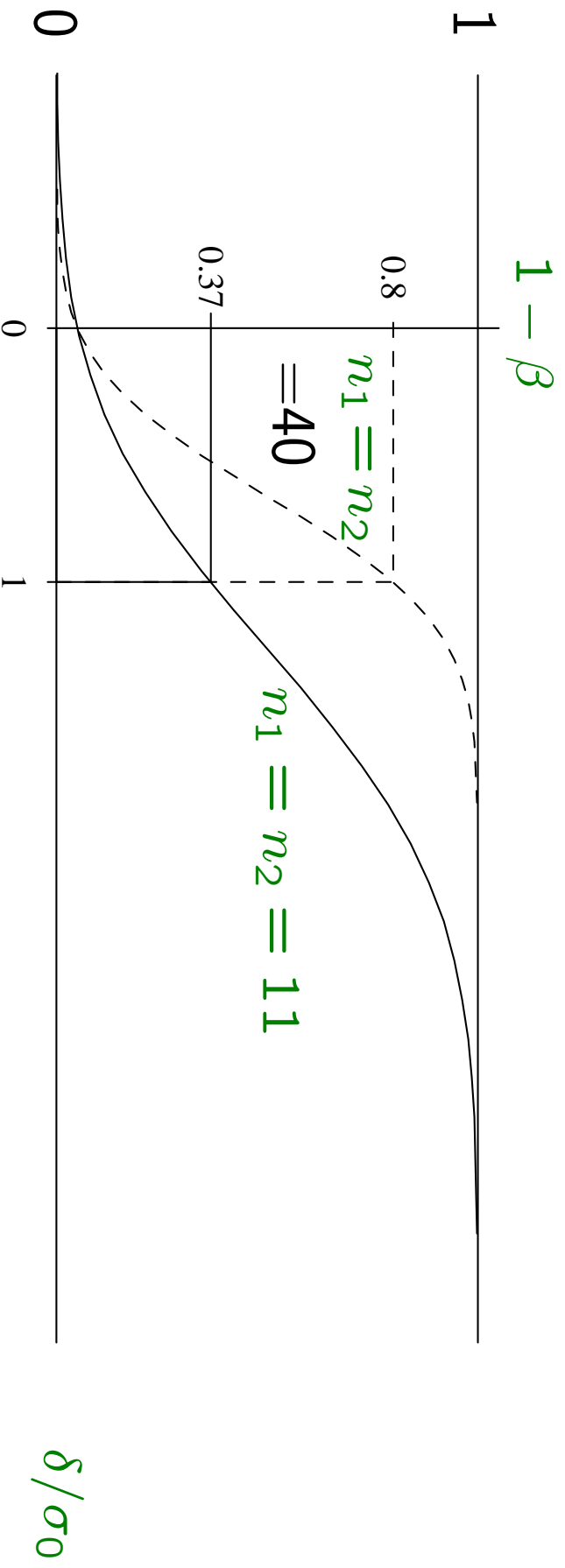
→ Macht als Funktion von δ .

$n_1 = n_2 = n$. H_0 verworfen, wenn $U > 1.64 \cdot \sigma_0 \sqrt{2/n}$

Unter H_A gilt $U \sim \mathcal{N}(\delta, \sigma_0^2 2/n)$

$$\begin{aligned}
 1 - \beta^{(A)} &= P^{(A)} \left\langle U > 1.64 \cdot \sigma_0 \sqrt{2/n} \right\rangle \\
 &= 1 - \Phi \left\langle \frac{1.64 \cdot \sigma_0 \sqrt{2/n} - \delta}{\sigma_0 \sqrt{2/n}} \right\rangle \\
 &= 1 - \Phi \left\langle 1.64 - \frac{\delta}{\sigma_0} \sqrt{n/2} \right\rangle.
 \end{aligned}$$

Die Macht nimmt zu mit „standardisiertem Effekt“ δ/σ_0 und Stichprobenumfang n .



Macht $1 - \beta^{(A)}$ als Funktion von δ heisst auch **Gütefunktion**.
 $\beta^{(A)}$, als Funktion von δ heisst auch Operations-Charakteristik.

- e Wie gross muss n sein, damit bei $\delta = 1.0$ und $\sigma_0 = 1.8$
 $1 - \beta^{(A)} = 0.80$ wird?

$$\begin{aligned} 1 - \Phi \left(1.64 - \frac{1.0}{1.8} \sqrt{n/2} \right) &= 0.8 \quad \implies \\ 1.64 - 0.556 \sqrt{n/2} &= -0.84 \\ n &= 2 \cdot \left(\frac{2.48}{0.556} \right)^2 = 39.2 \end{aligned}$$

Es braucht (etwa) 40 Tiere pro Gruppe.

→ Versuchsplanung, Berechnung des Stichprobenumfangs.

- *f* Man muss σ_0 ungefähr kennen und δ festlegen.
Vorversuch?
Nichtparametrischer Test?

- *g* Macht (-funktion) als Qualitätskriterium zur Wahl eines Tests.
Optimieren! \longrightarrow Mathematische Statistik.
Optimale Schätzungen \longrightarrow Testgrößen für optimale Tests.

8.11 Sinn und Unsinn statistischer Tests

- a Fachzeitschriften: Man darf nur über Effekte schreiben, die statistisch auf dem 5%-Niveau signifikant sind.
 - **Test dient als Filter gegen wilde Spekulationen.**

Perversion: Untersuchte Fragen, für die ein signifikantes Ergebnis erwartet werden kann – auch uninteressante ...

Statistiker unterstützen diesen Unsinn nicht.

Relevanz wichtiger als Signifikanz!

b Unterschied zwischen (statistischer) „Signifikanz“ und (praktischer) „Relevanz“

Zu kleine und zu grosse Stichproben!

Mit grossen Stichproben kann man Bagatelleffekte, die eigentlich niemanden interessieren, stolz als „statistisch signifikant“ nachweisen und publizieren.

c* Konsequenz? Statistische Tests sind unsinnig!

→ Man müsste $H_0 : \theta \leq$ Schwellenwert γ prüfen!

Regel für Entscheidung zwischen $H_0 : \theta \leq \gamma$ und $H_1 : \theta > \gamma$.

d Wieso Tests ausführlich behandeln?

- Filter gegen wilde Spekulationen
- Grundlage für das Verständnis der Beziehung zwischen Wahrscheinlichkeits-Modellen und empirischen Daten.
- Grundlage für **Vertrauensintervall**.

e Überprüfung von Annahmen

Sind Daten normalverteilt? \longrightarrow Test!

Nullhypothese: $x_i \sim \mathcal{N}(\mu, \sigma^2)$ für irgendein μ, σ .

Man will Nullhypothese beweisen!

Das geht nicht.

Allenfalls wieder als **Konvention** verwenden,
gegen die Anwendung falscher statistischer Methoden.

Man müsste nachweisen, dass Abweichung von der Annahme
„noch im ungefährlichen Bereich“ ist.

Wir werden eher grafische Methoden
für die Überprüfung von Annahmen verwenden.

Regel der Art „Prüfe, ob normalverteilt,
falls akzeptiert, verwende t-Test, sonst Rangsummentest“
sollte vermieden werden!

Lieber immer Rangsummen-Test, da „meistens“ besser
und nie viel schlechter als t-Test

f Äquivalenz-Tests

Nachweis, dass 2 Medikamente (Messmethoden) **gleich** sind!
Beweis der Nullhypothese ! ?

Man kann nachweisen, dass der **Unterschied kleiner als Δ_0** ist.

Nullhypothese: $|\Delta| \geq \Delta_0$, Alternative: $< \Delta_0$.

Figur! \rightarrow Für ein vorgegebenes n
kann der Verwerfungsbereich $= \emptyset$ sein!

Man braucht für kleine Δ_0 sehr grosse Stichproben
(oder sehr kleines σ).

g

Entscheidungen

Bei Tests spielen Null- und Alternativ-Hypothese nicht die gleiche Rolle.

Einhaltung von Grenzwerten, Qualitätskontrolle

Zwei Parteien mit gegensätzlichen Interessen.

Welches ist die Nullhypothese? Verschiedene Wahl!

Unentschieden, falls \bar{X} im (zweiseitigen) Annahmebereich der Nullhypothese.

- h Neutrale Formulierung: **Entscheidung**, ob $\mu < \mu_0$ **oder** $\mu > \mu_0$
Entscheidungsregel $K(\bar{X}) = 0$ oder $= 1$.
- 2 mögliche Fehler:
 $(\mu < \mu_0, \text{ aber } K = 1)$ oder $(\mu > \mu_0, \text{ aber } K = 0)$
Für gegebenes μ kann man Wahrsch.en berechnen.
 - Fehler können verschieden viel „kosten“.
 - Verallgemeinerung auf > 2 Entscheidungsmöglichkeiten
→ **Entscheidungs-Theorie**,
eng verwandt mit **Diskriminanzanalyse** → Multivariate Stat.

- i Schöne Formulierung löst das Problem nicht!
Grenzwert eingehalten? \rightarrow unentschieden, falls \bar{X}
im (zweiseitigen) Annahmebereich der Nullhypothese.

Wann passiert das?

Wahres μ nahe beim Grenzwert und n klein.

μ ist nie exakt = Grenzwert.

Also muss man „einfach“ n gross genug wählen!

Überflüssig, falls μ weit vom Grenzwert weg ist.

j **Sequentielle Tests**

- Wähle n (klein), wende Test an.
- Falls Ergebnis nicht eindeutig, nochmals n Beob.
- Wiederholen!
(Irgendwann muss man in der Praxis aufhören.)

→ Stichprobenumfang hängt vom Ergebnis ab.

Grosser Aufwand nur, wenn nötig!

Achtung! „Irrtumswahrscheinlichkeit“

resp. Wahrscheinlichkeit für die beiden Arten von Fehlern
verändert sich! → Kritische Werte anpassen!

Merkpunkte **Bedeutung von statistischen Tests**

- Eigentlich prüfen statistische Tests keine Effekte, sondern, ob der **Stichprobenumfang** n **gross genug** war, um den (evtl. minimalen) Effekt nachzuweisen.
- Deshalb sind **Vertrauensintervalle sinnvoller als Tests!**
- Tests zur Überprüfung von Annahmen sind von fraglichem Wert.
- Genauere Fragestellungen: \longrightarrow Äquivalenztests, sequentielle Tests
 \longrightarrow Entscheidungs-Theorie