

8. Tests

8.3 Weitere Beispiele und Begriffe

a **Beispiel Schlafverlängerung**

Nullhypothese: kein Unterschied

$$X_i \sim \mathcal{N}\langle 0, 1.5 \rangle; \quad X_i \text{ unabhängig}$$

b Alternativen: $(H_A)_\mu : X_i \sim \mathcal{N}\langle \mu, 1.5 \rangle$ mit $\mu > 0$ (einseitig).

Beobachtungen zusammenfassen! \bar{X} „Test-Statistik“ U .

c Verwerfungsbereich?

$$U = \bar{X} \sim \mathcal{N}\langle 0, 1.5/10 \rangle$$

falls H_0 gilt.

$$T = \bar{X} / \sqrt{0.15} \sim \mathcal{N}\langle 0, 1 \rangle \longrightarrow \text{Extreme Werte: } T \geq 1.64.$$

d $\bar{x} = 1.58$, $t = 1.58/1.5 = 1.05 < 1.64$.

Verwerfung. Effekt statistisch nachgewiesen.

f **Modern:** Statt t mit c zu vergleichen, können wir

$$P\langle T > 1.05 \rangle = P\langle \bar{X} > 1.58 \rangle = 0.15$$

ausrechnen.

W. von mindestens so extremen Werten wie der beobachtete.

P-Wert p

$$t > 1.64 \iff p < 0.05$$

8.4 Vorgehen bei einem stat. Test (Rezept)

- a **Problem** formulieren, in Worten.
- b **Nullhypothese** H_0 . Modell für die Beobachtungen.
Oft: Parametrische Familie, H_0 legt Parameter fest.
In der Regel möchte man H_0 gerne **widerlegen**.
- c **Alternativen** H_A werden in Betracht ziehen
(einseitig oder zweiseitig)
- d **Test-Statistik**.
Verteilung von U unter Alternativen soll möglichst stark
verschieden sein von der Vt. unter der Nullhypothese (Macht!).
Oft: Parameter testen: $H_0 : \theta = \theta_0$.
→ Test-Statistik $U = \hat{\theta}$.

Art des Experimentes, Beobachtungssituation [a]



Modelle für eine Beobachtung:



Nullhypothese [b]



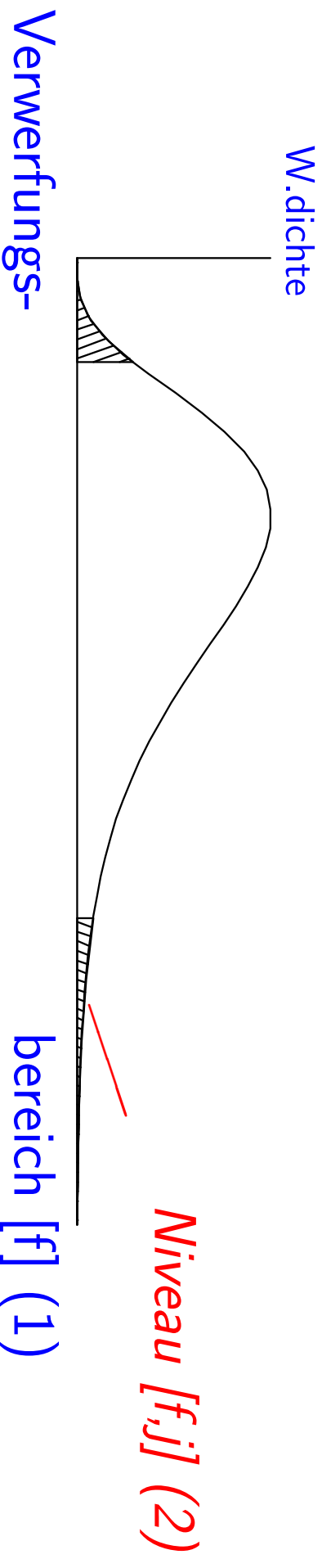
Alternativen [c]



Test-Statistik [d]

(1)

Verteilung der Test-Statistik unter der Nullhypothese [e]



- e **Verteilung** $\mathcal{F}_0\langle U \rangle$ von U unter H_0
Standardisierte Test-Statistik T .
- f Klassische Variante: **Verwerfungsbereich** bestimmen.
Willkürlich: Niveau α wählen.
„Kritischer Wert“, oft aus Tabelle oder Diagramm.
- g Bis hierher: **Test-Vorschrift** „fertig verpackt“
in den Lehr- und Nachschlagewerken nachzulesen.

Art des Experimentes, Beobachtungssituation [a]



Modelle für eine Beobachtung:



Nullhypothese [b]



Alternativen [c]

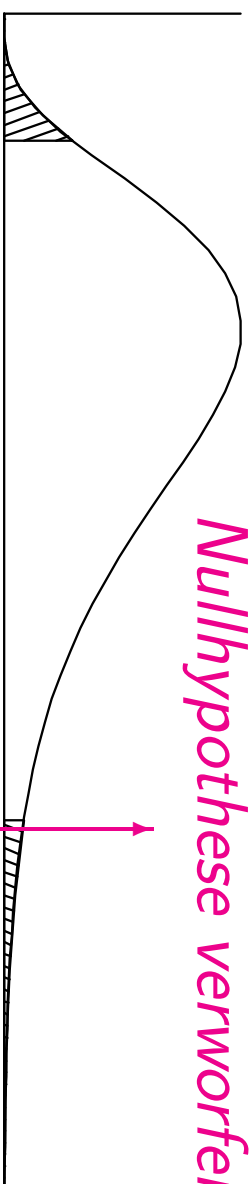


(1)

Test-Statistik [d]



Verteilung der Test-Statistik unter der Nullhypothese [e]



Nullhypothese verworfen! [i] (3)

Ausrechnen des Wertes der Test-Statistik [h] (3)



Messwerte [h] (3)

- h Beobachtungen \longrightarrow Wert t der (standard.) Test-Statistik T
- i Klassisch: **Entscheidung**:
Falls $t \in K$, wird die **Nullhypothese verworfen**.
- j Modern: P-Wert bestimmen. **Entscheidung**:
Falls $p < 5\%$, wird die **Nullhypothese verworfen**.

8.5 Tests für 1 Stichprobe oder 2 gepaarte

a **Problem „eine Stichprobe“**: $X_i, i = 1, 2, \dots, n$, unabhängig.

Kann ein bestimmter Lageparameter (Erwartungswert, Median) gleich einem vorgegebenen Wert μ_0 sein?

$$H_0 : \mu = \mu_0.$$

μ_0 : physikalische Konstante, chemische Konzentration, **Grenzwert**.

b **Problem „zwei gepaarte oder verbundene Stichproben“:**

- Blutdruck vor und nach Medikamenten-Einnahme,
- Schlaflänge bei zwei Schlafmitteln,
- Grösse von Vater und Sohn (Beobachtungseinheit = Familie)
- Anzahl beobachteter Verhaltenselemente von Mäusen bei Licht und bei Dunkelheit. „Verbundene Stichproben“

X_i = Differenz der Zielgrösse in den beiden Zuständen.
(evtl. nach Transformation).

Frage: Differenz „im wesentlichen gleich null“?

$$H_0 : \mu = 0.$$

d **z-Test**: besprochen.

H_0 : $X_i \sim \mathcal{N}\langle \mu_0, \sigma_0^2 \rangle$, $i = 1, \dots, n$, unabhängig;
die Varianz σ_0^2 sei bekannt

H_A : $X_i \sim \mathcal{N}\langle \mu, \sigma_0^2 \rangle$, $i = 1, \dots, n$, unabhängig, mit

(a) $\mu > \mu_0$; (b) $\mu < \mu_0$; (c) $\mu \neq \mu_0$

$$U : \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\mathcal{F}_0\langle U \rangle : \quad \bar{X} \sim \mathcal{N}\langle \mu_0, \sigma_0^2/n \rangle$$

$$K : \quad \text{Testgrösse standardisieren: } Z = \frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} \sim \mathcal{N}\langle 0, 1 \rangle$$

Kritische Werte für Z :

- (a) c so, dass $1 - \Phi\langle c \rangle = 0.05 \Rightarrow c = 1.64$; $K = \{Z \geq 1.64\}$
 (b) c so, dass $\Phi\langle c \rangle = 0.05 \Rightarrow c = -1.64$; $K = \{Z \leq -1.64\}$
 (c) c_1 so, dass $1 - \Phi\langle c_1 \rangle = 0.025 \Rightarrow c_1 = 1.96$;
 c_0 so, dass $\Phi\langle c_0 \rangle = 0.025 \Rightarrow c_0 = -c_1$; $K = \{|Z| \geq 1.96\}$

Kritische Werte für **standardisierte** Testgrösse unabhängig von n, μ_0, σ_0 .

- e **Beispiel Reifen** (nach Lehn+Wegmann, 1992).

Vergleich von 2 Profilen auf Winterreifen bezügl. Bremswirkung

10 Testfahrzeuge. $\sigma_0 = 3$ aus früheren Versuchen.

i	Profil A	Profil B	Differenz	Vorzeichen
1	44.5	44.9	0.4	+
2	55.0	54.8	-0.2	-
3	52.5	55.6	3.1	+
4	50.2	55.2	5.0	+
5	45.3	55.6	10.3	+
6	46.1	47.7	1.6	+
7	52.1	53.0	0.9	+
8	50.5	49.1	-1.4	-
9	50.6	52.3	1.7	+
10	49.2	50.7	1.5	+
mittlere Differenz			2.29	

$$z = 2.29 / (3 / \sqrt{10}) = 2.41 \in K:$$

- Die Reifenarten unterscheiden sich statistisch signifikant
in der Länge des Bremsweges.

- f Meistens: Nullhypothese $X_i \sim \mathcal{N}(\mu_0, \sigma^2)$, unabhängig, σ unbekannt.

Nicht eine einzige Verteilung, sondern

alle Normalverteilungen mit Erwartungswert μ_0 .

σ^2 : **Stör-Parameter, nuisance parameter.**

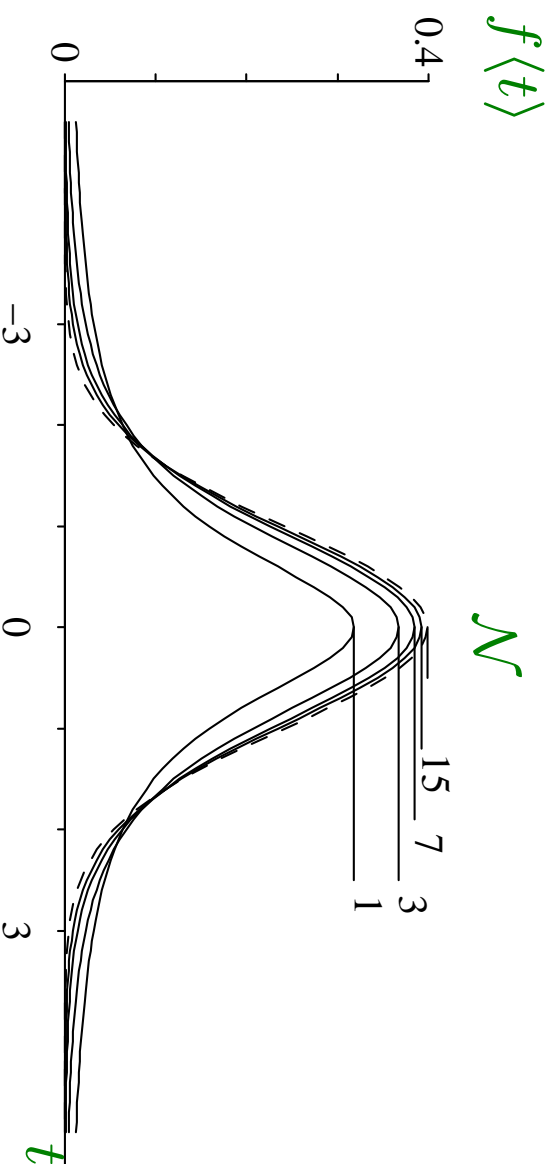
σ durch den Schätzwert $\hat{\sigma} = S$ ersetzen!

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{\bar{X} - \mu_0}{\sqrt{\frac{1}{n(n-1)} \sum (X_i - \bar{X})^2}}$$

g S schwankt zufällig. T nicht normalverteilt.

Verteilung von T hängt nur vom Stichprobenumfang n ab.
 $m = n - 1$ „Freiheitsgrade“.

Verteilung von T : grössere Varianz und langschwänziger.



t-Test von Student.

- i S/\sqrt{n} : Schätzung der Standard-Abweichung σ_0/\sqrt{n} des arithmetischen Mittels \bar{X} , „Standardfehler“, standard error se.

- j Beispiel Reifen: $S^2 = \frac{1}{9} \left((0.4 - 2.29)^2 + (0.2 - 2.29)^2 + \dots + (1.5 - 2.29)^2 \right) = 3.31^2$, Standardfehler

$$se = 3.31/\sqrt{10} = 1.048, t = 2.29/1.048 = 2.18.$$

$$m = 10 - 1. \text{ Krit. Wert } 2.26 > |t|. H_0 \text{ nicht verworfen!}$$

Einseitiger Test: Kritischer Wert 1.83. Signifikant!

Interpretation?

k **Vorzeichen-Test** (oder Zeichen-Test).

Testgrösse = Anzahl Beobachtungen grösser als μ_0
 = Anzahl positive **Vorzeichen** von $X_i - \mu_0$.

l H_0 : $X_i \sim \mathcal{F}_0$ mit Median μ_0 , unabhängig.

H_A : $X_i \sim \mathcal{F}$ mit Median μ ;

(a) $\mu > \mu_0$, (b) $\mu < \mu_0$, (c) $\mu \neq \mu_0$.

U : $U = \text{Anzahl } \{i \mid X_i > \mu_0\}$

$\mathcal{F}_0\langle U \rangle$: $P_0\langle X_i > \mu_0 \rangle = 1/2$, $U \sim \mathcal{B}\langle n, 1/2 \rangle$.

K : Tabelle!

m **Beispiel Reifen**: 8 positive Vorzeichen von $n = 10$.

Tabelle: $c_0 = 1$, $c_1 = n - c_0 = 9$.

Nullhypothese durch den Vorzeichentest nicht abgelehnt.

- o Keine Normalverteilung vorausgesetzt!
Aber: Information „verschenkt“.

- p **Rangsummen-Test von Wilcoxon für gepaarte Stichproben**
oder Vorzeichen-Rangsummen-Test
(signed rank test, one-sample Wilcoxon test).

H_0 : $X_i \sim \mathcal{F}_0$, unabh., \mathcal{F}_0 stetig und symmetrisch bez. μ_0

H_A : $X_i \sim \mathcal{F}$, unabhängig, \mathcal{F} symmetrisch bez.

(a) $\mu > \mu_0$; (b) $\mu < \mu_0$; (c) $\mu \neq \mu_0$.

U : Nach folgendem Rezept zu bilden:

1. Bilde $X'_i = X_i - \mu_0$, streiche die negativen Vorzeichen, bilde die Ränge R_i .

$$R_i = \text{rank} \left(|X'_i| \mid |X'_1|, |X'_2|, \dots, |X'_n| \right).$$

Beispiel Reifen: $\mu_0 = 0$, $X_i - \mu_0$ Differenz der Bremswege.

Ränge R_i : 2, 1, 8, 9, 10, 6, 3, 4, 7, 5.

2. Summiere die R_i , für die $X_i - \mu_0 > 0$

Beispiel: $2+8+9+10+6+3+7+5=50$.

Sei $V_i = 1$, falls $X_i - \mu_0 > 0$ ist, $V_i = 0$ sonst.

$$U^+ = \sum_{i=1}^n V_i R_i.$$

Beispiel: $u^+ = 1 \cdot 2 + 0 \cdot 1 + 1 \cdot 8 + 1 \cdot 9 + 1 \cdot 10 + 1 \cdot 6 + 1 \cdot$

$3 + 0 \cdot 4 + 1 \cdot 7 + 1 \cdot 5 = 50$.

$U^- = \sum_{i=1}^n (1 - V_i) R_i$. $U^+ + U^- = \sum_{i=1}^n R_i = n(n+1)/2$.

$\mathcal{F}_0\langle U \rangle$: Verteilung von U^+ oder U^- hängt **nicht** von der Verteilung \mathcal{F} ab (solange sie stetig ist).

K : Bis $n = 30$ aus Tabelle:

- (a) c aus der Zeile „ c , einseitig“;
 $K = \{U^+ \geq n(n+1)/2 - c\} = \{U^- \leq c\}$;
- (b) c aus der Zeile „ c , einseitig“; $K = \{U^+ \leq c\}$;
- (c) c_0 aus der Zeile „ c_0 , zweiseitig“;
 $K = \{U^+ \leq c_0\} \cup \{U^+ \geq n(n+1)/2 - c_0\}$
 $= \{\min\langle U^+, U^- \rangle \leq c_0\}$.

Für grössere n : Approximation durch die Normalvert.

n	1	2	3	4	5	6	7	8	9	10
c , einseitig	-	-	-	-	0	2	3	5	8	10
c_0 , zweiseitig	-	-	-	-	-	0	2	3	5	8
n	11	12	13	14	15	16	17	18	19	20
c , einseitig	13	17	21	25	30	35	41	47	53	60
c_0 , zweiseitig	10	13	17	21	25	29	34	40	46	53
n	21	22	23	24	25	26	27	28	29	30
c , einseitig	58	65	73	81	89	98	107	116	126	137
c_0 , zweiseitig	67	75	83	91	100	110	119	130	140	151

Beispiel Reifen: $u^+ = 50$, $u^- = 5$.

Tabellenwert $c_0 = 8$, $n(n+1)/2 - c_0 = 47 < u^+$ ($u^- \leq 8$).

Nullhypothese **verworfen**.

U ist asymptotisch normalverteilt.

$$\begin{aligned} \mathcal{E}\langle U^+ \rangle &= \frac{n(n+1)}{4}, & \text{var}\langle U^+ \rangle &= \frac{n(n+1)(2n+1)}{24} \\ Z^+ &= \frac{U^+ - \mathcal{E}\langle U^+ \rangle}{\sqrt{\text{var}\langle U^+ \rangle}} = \frac{U^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \approx \mathcal{N}\langle 0, 1 \rangle \end{aligned}$$

Verwerfungsbereich (zweiseitig): $|Z^+| \geq 1.96$.

- q Keine („**Bindungen**“), **ties**) oder Nullen!
 Nullen ($X_i = \mu_0$) weglassen.
 (Vert. von $T =$ bedingte Vert., gegeben die Anzahl Nullen).
 Bindungen: Ränge aufteilen. Asymptot. Näherung mit korrig. Varianz.

Herleitung der Verteilung von U^+ :

Wie verteilen sich die Vorzeichen auf die Ränge, wenn H_0 gilt?

$ x'_i $	0.2	0.4	0.9	1.4	1.5	1.6	1.7	3.1	5.0	10.3
Rang $\langle x'_i \rangle$	1	2	3	4	5	6	7	8	9	10
Vorz $\langle x'_i \rangle$	-	+	+	-	+	+	+	+	+	+
v_i	0	1	1	0	1	1	1	1	1	1

3. Zeile zufällig, Bernoulli-vert.

Jede Folge von 10 Vorzeichen hat gleiche Wahrscheinlichkeit $1/2^{10}$.

→ Verteilung von U^+ .

- r Test-Statistiken, deren Verteilung **nicht** von einem konkreten **parametrischen** Modell für die Beobachtungen abhängt, und die entsprechenden Tests nennt man **nicht-parametrisch** (oder verteilungsfrei). **Rangtest**.
- Genaueres, Allgemeineres unter den Stichworten **Randomisierungs-Tests** oder **Permutations-Tests**.
- s Wilcoxon-Test nützt absolute Grösse der positiven gegenüber den negativen $X_i - \mu_0$ aus. Nullhypothese ist „schärfer“ als beim Vorzeichentest: Es wird **Symmetrie vorausgesetzt**.
- Für Differenzen der Paare bei verbundenen Stichpr. naheliegend.
- t Welchen Test wählen?

8.6 Interpretation von Testergebnissen

- a Verwerfung der Nullhypothese. Interpretation?
- (1) Ein Effekt ist nachgewiesen, eine „Alternative“ ist richtig,
 - (2) Nullhypothese auf andere Weise verletzt:
 - Daten enthalten systematischen Fehler,
 - X_i sind nicht unabhängig,
 - nicht normalverteilt (beim t-Test),
 - nicht symmetrisch verteilt (beim Rangsummen-Test),
 - nicht alle gleich verteilt ($\text{med}\langle X_i \rangle$ versch. beim Vorz.t.),
 - (3) es ist zufällig das unwahrscheinl. Ereignis K eingetreten.

(3) ist unvermeidbar. W. durch Signifikanzniveau kontrolliert.
Man möchte (2) vermeiden, um (1) schliessen zu können.
Aber wie?

b (A) $X_i \sim \mathcal{N}(\mu, \sigma_0^2)$, σ_0^2 bekannt \rightarrow z-Test.

(B) $X_i \sim \mathcal{N}(\mu, \sigma_0^2)$, σ_0^2 unbekannt \rightarrow t-Test.

(C) $X_i \sim \mathcal{F}$ symmetrisch um μ , sonst bel. \rightarrow Rangsummen-Test.

(D) $X_i \sim \mathcal{F}_i$ mit Median μ , sonst beliebig \rightarrow Vorzeichen-Test.

Unabhängigkeit der X_i .

- c ● Annahme, dass die Nullhypothese oder eine Alternative „gilt“, ist **Voraussetzung** für die Anwendung des Tests.
- d ● Gesamte Nullhypothese führt zu „statistischem Widerspruch“. Abweichungen von den anderen Annahmen wegdiskutieren (soweit möglich überprüfen).

- e Tests verwenden, die **möglichst wenige Voraussetzungen** brauchen!

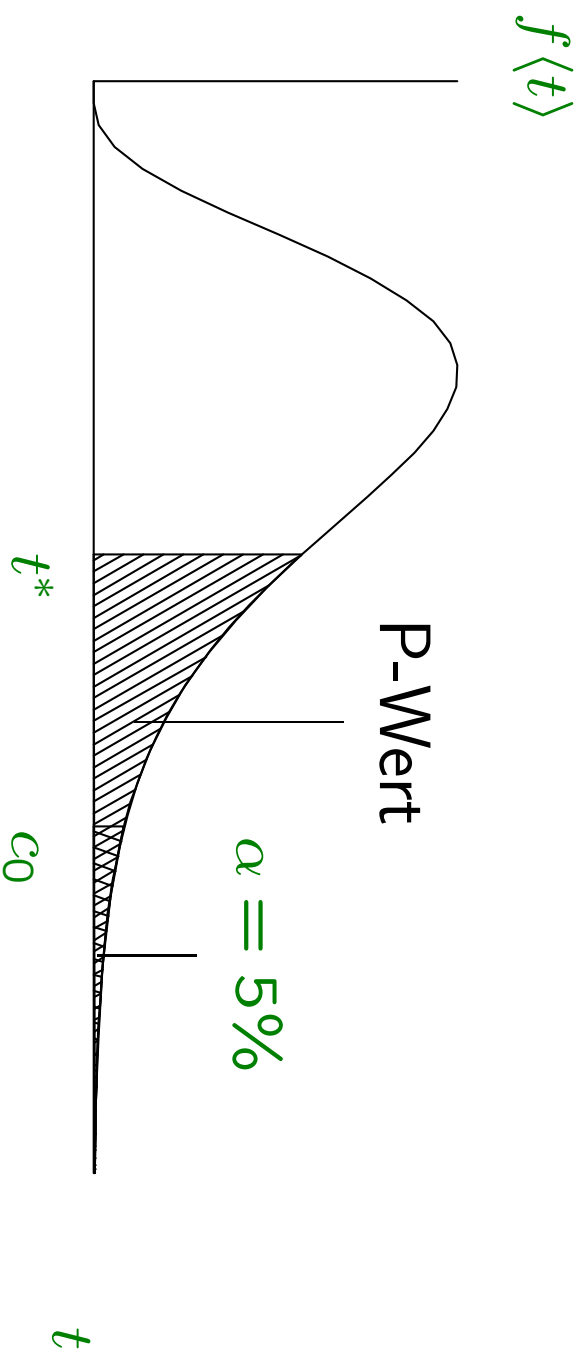
Also immer Rangsummen-Test statt t- oder z-Test, immer Vorzeichen-Test statt Rangsummen-Test.

- f Nein! Information ausnützen. Macht = W . für Fehler 2. Art.
Die allzu grosszügige Lockerung der Voraussetzungen bezahlt man mit einem Verlust an Macht.

- g Konkrete Empfehlung für das Testen eines Lageparameters:
 - Rangsummen-Test von Wilcoxon anwenden, falls Verteilung symmetrisch ist wegen
 - theoretischen Überlegungen (z.B. Differenzen),
 - empirischen Resultaten in grossen Datensätzen,
 - Vorwissen (frühere Studien mit ähnlichen Daten);
 - sonst Vorzeichen-Test durchführen;

- t-Test (und z-Test) vermeiden.
Gewinn an Macht für normalverteilte Beob. minim.
Wenn nicht genau normalvert., hat der Rangsummen-Test
meistens grössere Macht als der t-Test
(Ausnahme: Sehr kleine Stichproben).

8.7 Der P-Wert



c Signifikanz: Test-Statistik rechts von c_0 ($\in K$)

\Leftrightarrow P-Wert kleiner als 5% – und umgekehrt.

P-Wert: verfeinertes „**Mass der Signifikanz**“.

$P = 6\%$: knapp nicht signifikant, Nachdenken „erlaubt“.

$P = 74\%$: kein Hinweis auf Abweichung von der Nullhypothese.

d Literatur, ältere Konvention:

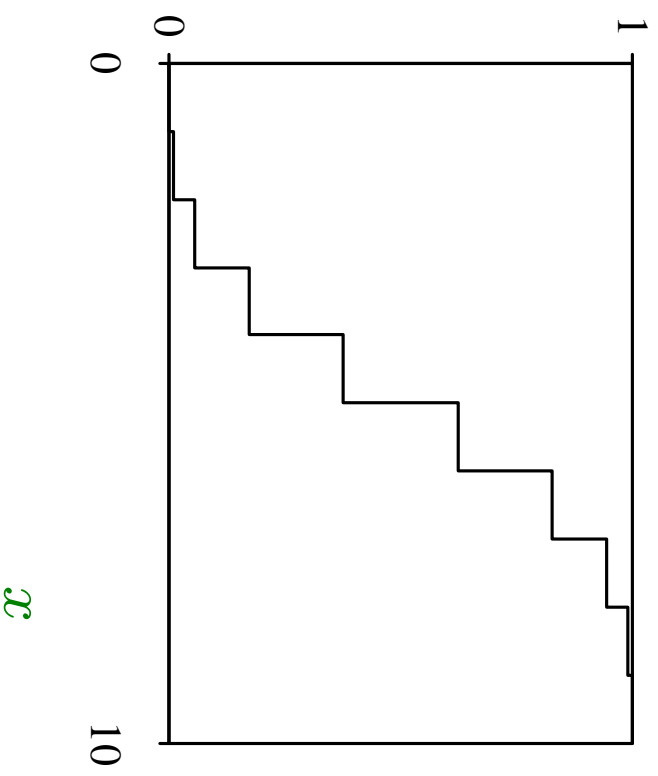
$P > 0.05$	nicht signifikant	(n.s.)
$0.05 \geq P > 0.01$	schwach signifikant	*
$0.01 \geq P > 0.001$	stark signifikant	**
$0.001 \geq P$	sehr stark signifikant	***
$z = 2.63^{**}$		
$z = 1.46$	(n.s.)	

e Hinweise:

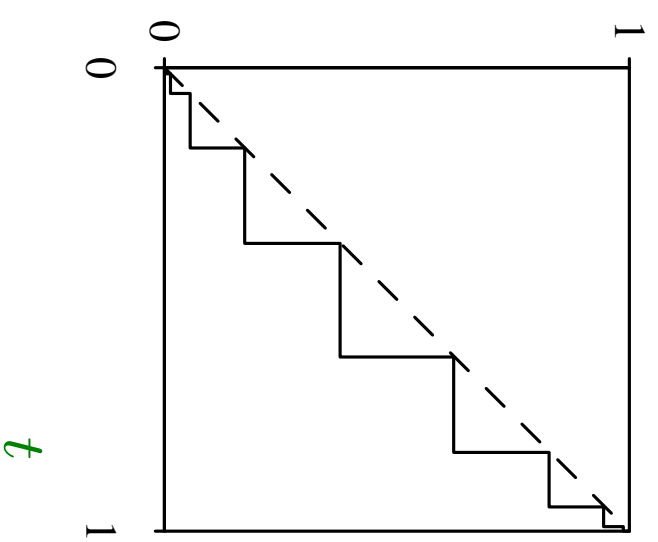
- P-Wert = transformierte Test-Statistik mit uniformer Vert.
- Mit P-Wert kann man einfacher entscheiden als mit t .
Computer muss mehr arbeiten.
- Achtung: Wegen Faulheit der Programmierer oft unexakte Tests!
(Asymptotische Näherung auch bei kleinen Stichproben!)
- P-Wert aus verschiedenen Tests ?!
- P-Wert und Wahrscheinlichkeit!
„Die Wahrscheinlichkeit für die Nullhypothese ist 10%.“
BITTE NICHT!
„Die Irrtums-Wahrscheinlichkeit ist 3%.“ (auch nicht.)

- Bei diskret vert. T ist der P-Wert auch diskret verteilt.

$$F(X)$$



$$F(PW)$$



8.8 Vergleich von zwei quantitativen Stichproben

a **Einige Fragestellungen:**

- Unterscheiden sich Puppengewichte von Fliegen (der gl. Art), die sich in gleich grossen Tieren verschiedener Wirtsarten entwickeln?
- Haben Raucher und Nichtraucher unterschiedl. Blutdruck?
- Unterscheiden sich Schnecken auf Mager- & Fettwiesen?
- Führt Düngersorte A zu höherem Ertrag als Dünger B?

Je eine Stichprobe aus zwei verschiedenen Grundgesamtheiten.
Zielgrösse Y messen.

- b Unterscheidung von geparteten Stichproben:
Stichprobenumfänge n_1 und n_2 können
für unabh. Stichproben verschieden sein;
bei geparteten Stichproben immer gleich gross.

d **Modell:** $Y_{1,1}, Y_{1,2}, \dots, Y_{1,n_1}$ $Y_{2,1}, Y_{2,2}, \dots, Y_{2,n_2}$

$$Y_{1,i} \sim \mathcal{F}_1, \quad i = 1, 2, \dots, n_1;$$

$$Y_{2,i} \sim \mathcal{F}_2, \quad i = 1, 2, \dots, n_2;$$

$$Y_{1,1}, Y_{1,2}, \dots, Y_{1,n_1}, Y_{2,1}, \dots, Y_{2,n_2} \quad \text{alle unabhängig.}$$

H_0 : $\mathcal{F}_1 = \mathcal{F}_2$. Alle Y haben gleiche Verteilung.

H_A : \mathcal{F}_1 und \mathcal{F}_2 unterscheiden sich

durch eine Verschiebung δ : $F_2(x) = F_1(x - \delta)$

Speziell: $Y_{1,i} \sim \mathcal{N}\langle \mu, \sigma^2 \rangle$ $Y_{2,i} \sim \mathcal{N}\langle \mu + \delta, \sigma^2 \rangle$.

δ interessierender Parameter, μ, σ Stör-Parameter.

e $\hat{\delta} = \bar{Y}_{2,\cdot} - \bar{Y}_{1,\cdot} = (Y_{2,1} + \dots + Y_{2,n_2})/n_2 - (Y_{1,1} + \dots + Y_{1,n_1})/n_1$

f **Zwei-Stichproben-z-Test.**

Beispiel Mastochsen. 2 Futterarten vergleichen.

Zielgrösse: mittlere wöchentliche Gewichtszunahme in 1 Monat.

Erfahrung: Solche Gewichtszunahmen $\approx \mathcal{N}(\dots, (\sigma_0 = 1.8\text{kg})^2)$.

Einseitige Alternative.

\mathfrak{g} $H_0 : Y_{k,i} \sim \mathcal{N}\langle \mu, \sigma_0^2 \rangle$ (*i.i.d.*);

$H_A : Y_{1,i} \sim \mathcal{N}\langle \mu_1, \sigma_0^2 \rangle; \quad Y_{2,i} \sim \mathcal{N}\langle \mu_1 + \delta, \sigma_0^2 \rangle$

einseitig: $\delta > 0$.

$U : U = \bar{Y}_{2,\cdot} - \bar{Y}_{1,\cdot}$,

$\mathcal{F}_0\langle U \rangle : \bar{Y}_{k,\cdot} \sim \mathcal{N}\langle \mu, \sigma_0^2/m_k \rangle$, also

$$U = \bar{Y}_{2,\cdot} - \bar{Y}_{1,\cdot} \sim \mathcal{N}\langle 0, \sigma_0^2(1/n_1 + 1/n_2) \rangle$$

$$Z = \frac{\bar{Y}_{2,\cdot} - \bar{Y}_{1,\cdot}}{\sqrt{\sigma_0^2(1/n_1 + 1/n_2)}} \sim \mathcal{N}\langle 0, 1 \rangle$$

K : einseitig: $K = \{Z \geq 1.64\}$.

h Mastochsen:

„extensiv“	2.7	2.7	1.1	3.0	1.9	3.0	3.8	3.8	3.8	0.3	1.9	1.9
„intensiv“	6.5	5.4	8.1	3.5	0.5	3.8	6.8	4.9	9.5	6.2	4.1	

$$\bar{y}_{2,\cdot} - \bar{y}_{1,\cdot} = 5.39\text{kg} - 2.37\text{kg} = 3.02\text{kg}$$

$$z = 3.02 / \sqrt{1.8^2 \cdot 2/11} = 3.93 > 1.64.$$

Nullhypothese (wichtig) verworfen.

(Sinn eines solchen Tests? Schätzproblem!)

i **t-Test.**

$$H_0 : Y_{k,i} \sim \mathcal{N}\langle \mu, \sigma^2 \rangle \quad (i.i.d)$$

$$H_A : Y_{1,i} \sim \mathcal{N}\langle \mu, \sigma^2 \rangle, Y_{2,i} \sim \mathcal{N}\langle \mu + \delta, \sigma^2 \rangle,$$

$$\delta \neq 0 \text{ (oder } \delta > 0 \text{ oder } \delta < 0),$$

unabhängig.

Man ersetzt in der standardisierten Statistik für den z-Test die Varianz σ_0^2 durch eine Schätzung,

$$\hat{\sigma}^2 = \frac{1}{n_1 + n_2 - 2} \left(\sum_{i=1}^{n_1} (Y_{1,i} - \bar{Y}_{1,\cdot})^2 + \sum_{i=1}^{n_2} (Y_{2,i} - \bar{Y}_{2,\cdot})^2 \right) .$$

t-Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden.

j Test mit weniger Voraussetzungen!

Rangsummen-Test von Wilcoxon, Mann und Whitney, U-Test

H_0 : $Y_{k,i} \sim \mathcal{F}$ (i.i.d.); \mathcal{F} bel. Vert.

H_A : $Y_{1,i} \sim \mathcal{F}_1, Y_{2,i} \sim \mathcal{F}_2, F_2\langle x \rangle = F_1\langle x - \delta \rangle, \delta \neq 0.$

U : 1. Bestimme Rang $R_{k,i}$ bezügl. „vereinigten Stichproben“

2. $U^{(1)} = \sum_{i=1}^{n_1} R_{1,i}$ (oder $U^{(2)} = \sum_{i=1}^{n_2} R_{2,i}$)

$$(U^{(1)} + U^{(2)}) = (n_1 + n_2)(n_1 + n_2 + 1)/2.)$$

$$T^{(k)} = U^{(k)} - n_k(n_k + 1)/2$$

$$(T^{(1)} + T^{(2)}) = n_1 n_2.)$$

Zweiseitige Fragestellung: $T = \min\langle T^{(1)}, T^{(2)} \rangle.$

$\mathcal{F}_0\langle U \rangle$: Verteilung von $T^{(k)}$ hängt nur von n_1, n_2 ab.

Grosse n_1, n_2 : Asymptotische Näherung.

K : Tabelle.

<i>n</i> ₁	1	2	3	4	5	6	7	8	9	10	<i>n</i> ₂	11	12	13	14	15	16	17	18	19	20
1	--	--	--	--	--	--	--	0	0	0	--	--	--	--	--	--	--	2	2	2	--
2	--	--	--	--	0	--	1	2	3	3	0	1	1	1	1	1	1	2	2	7	8
3	--	--	--	0	1	1	2	4	4	5	3	4	4	5	5	6	6	7	7	13	14
4	--	--	0	1	2	2	3	4	4	5	6	7	8	9	10	11	11	11	12	13	14
5	--	0	1	2	3	3	5	6	7	8	9	11	12	13	14	15	17	17	18	19	20
6	--	1	2	3	5	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27	27
7	--	1	3	5	6	8	10	12	14	14	16	18	20	22	24	26	28	30	32	34	34
8	--	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41	41
9	--	0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48	48
10	--	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55	55
11	--	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62	62
12	--	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69	69
13	--	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76	76
14	--	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	69	74	78	83	83
15	--	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90	90
16	--	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98	98
17	--	2	6	11	17	22	28	34	39	45	51	57	63	69	75	81	87	93	99	105	105
18	--	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112	112
19	--	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119	119
20	--	2	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127	127

k **Beispiel Mastochsen.**

Daten·10	
ext.	3 11 19 19 19 27 27 30 30 38 38
int.	5 35 38 41 49 54 62 65 68 81 95
Ränge	
ext.	1 3 5 5 5 7.5 7.5 9.5 9.5 13 13
int.	2 11 13 15 16 17 18 19 20 21 22
	Σ 79 174

$u^{(1)} = 79$ und $u^{(2)} = 174$. (Kontrolle: $u^{(1)} + u^{(2)} = 22 \cdot 23/2$.)

$t^{(1)} = 79 - 11 \cdot 12/2 = 13$, $t^{(2)} = 174 - 11 \cdot 12/2 = 108$,

$t = \min\langle t^{(1)}, t^{(2)} \rangle = 13$.

$K = \{T \leq 30\}$. Nullhypothese wird verworfen.

m **Vergleich der Streuungen**

n Grafischer Vergleich mit **box-plots**.

Unterschiede zwischen den Medianen von 2 Gruppen signifikant?

Kerben. Es soll gelten:

Wenn sich die Kerben von zwei Kisten nicht überschneiden,
dann ist der Unterschied zwischen den Gruppen signifikant.

→ „Notched box plots“, gekerbte Kisten-Diagramme.

(Testregel nicht nach dem üblichen Schema!)

