Mathematik II für Biologen

Tests für Erwartungswert & Median

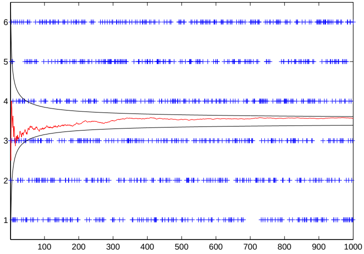
Stefan Keppeler

20. Juni 2014

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

- Zufallsvariable $X_1, \dots, X_n$, iid, d.h. insbesondere
  $$E[X_i] = \mu \quad \text{und} \quad \mathrm{Var}(X_i) = \sigma^2 \quad \text{für alle } i \text{ gleich}$$
- $\overline{X} := \frac{1}{n}(X_1 + \ldots + X_n) \quad \underset{\text{ZGS}}{\Longrightarrow} \quad \overline{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$
- **Beispiel** n-faches Würfeln:
  $$\mu = E[X_i] = 3{,}5 \quad \text{und} \quad \sigma^2 = \mathrm{Var}(X_i) \approx 3$$
  $$\Rightarrow \quad P\left[3{,}5 - 2\frac{\sqrt{3}}{\sqrt{n}} \leq \overline{X} \leq 3{,}5 + 2\frac{\sqrt{3}}{\sqrt{n}}\right] \approx 95\%$$

```
N=1000;
n=1:N;
X=unidrnd(6,1,N);
Xquer=cumsum(X)./n;
plot(n,X,'+')
hold on
plot(n,Xquer,'r')
plot(n,3.5+2*sqrt(3./n),'k')
plot(n,3.5-2*sqrt(3./n),'k')
hold off
```



Beispiel: (vgl. Vorlesung 6: Binomialverteilung & Binomialtest)

- Die Eichung einer Waage soll überprüft werden.
- Sollwert $20\,\mathrm{kg}$
- $n = 10$ Messungen ergeben (in kg):
  $20{,}1 \quad 20{,}3 \quad 20{,}9 \quad 19{,}2 \quad 20{,}8 \quad 20{,}1 \quad 20{,}2 \quad 20{,}4 \quad 20{,}2 \quad 20{,}3$
- Damals: Vorzeichentest (verschenkt viel Information)

Fragen:

- Wo liegen die Messwerte – ist der Erwartungswert oder Median ihrer Verteilung vielleicht 20?
- Mit welcher Genauigkeit kann man diese Aussage machen?

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Betrachte Messungen als Werte von Zufallsvariablen $X_1, \ldots, X_n$.

Modellannahmen:

- $X_1, \ldots, X_n$ sind iid mit
- $E[X_i] = \mu$ (unbekannt)
- $\mathrm{Var}(X_i) = \sigma^2$
  (bekannt beim z-Test – unrealistisch; später: t-Test)
- $X_1, \ldots, X_n$ sind normalverteilt *oder*
  n ist groß und $X_1, \ldots, X_n$ sind nicht zu "unnormal"

z-Test (am Beispiel Waage)

0. Annahme: $\sigma = 0{,}5$ bekannt
1. $H_0 : \mu = 20$ ("Waage geeicht")
2. $H_A : \mu \neq 20$ ("Waage nicht geeicht")
3. Teststatistik: $Z = \dfrac{\overline{X} - 20}{0{,}5/\sqrt{10}}$, wobei $\overline{X} = \frac{1}{n}(X_1 + \ldots + X_n)$

   $\left(\text{allg.: } Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \text{ für } H_0 : \mu = \mu_0\right)$
4. Verteilung von Z, falls H_0 stimmt:
   $Z \sim \mathcal{N}(0,1)$ (zumindest ungefähr) **Begründung:**
5. Signifikanzniveau: z.B. $\alpha = 5\%$
6. Verwerfe falls $|Z| \geq 1{,}96$
   (d.h. Verwerfungsbereich $K = (-\infty, -1{,}96] \cup [1{,}96, \infty)$)
7. $\overline{X} = 20,25$ d.h. $Z \approx 1{,}58$ beobachtet
8. Da $1{,}58 < 1{,}96$ wird H_0 nicht verworfen

t-Test: Wie z-Test, aber σ wird durch

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

geschätzt.
(vgl. empirische Varianz & Standardabweichung aus Vorlesung 1)

Beachte:

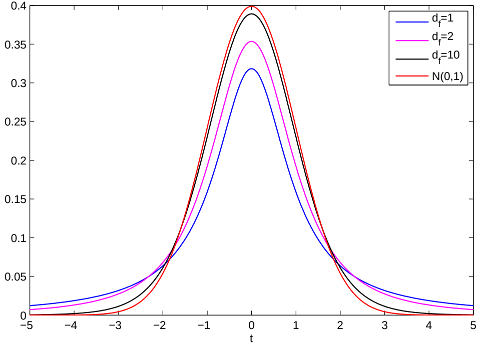
- s ist Zufallsvariable, und
- damit hängt Wert von den Daten ab
- es gilt $E[s^2] = \sigma^2$ (Erwartungstreue, vgl. Vorlesung 7)

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

t-Test (am Beispiel Waage)

1. $H_0 : \mu = 20$ ("Waage geeicht")
2. $H_A : \mu \neq 20$ ("Waage nicht geeicht")
3. Teststatistik: $T = \dfrac{\overline{X} - 20}{s/\sqrt{10}}$, wobei
   $\overline{X} = \frac{1}{n}(X_1 + \ldots + X_n)$ und $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}$
   $\left(\text{allg.: } T = \dfrac{\overline{X} - \mu_0}{s/\sqrt{n}} \text{ für } H_0 : \mu = \mu_0\right)$
4. Verteilung von T, falls H_0 stimmt, heißt
   (Studentsche) t-Verteilung mit $d_f = n - 1$ Freiheitsgraden.
   Tabelliert: Werte $t_{d_f,q}$ mit $P[T \leq t_{d_f,q}] = q$ (Quantile)

Dichte der Studentschen t-Verteilung



MATLAB-Code

```
t=-5:0.02:5;
plot(t,tpdf(t,1),'b')
hold on
plot(t,tpdf(t,2),'m')
plot(t,tpdf(t,10),'k')
plot(t,normpdf(t,0,1),'r')
hold off
legend('d_f=1','d_f=2','d_f=10','N(0,1)');
xlabel('t')
```

t-Test (am Beispiel Waage, Forts.)

5. Signifikanzniveau: z.B. $\alpha = 5\%$
6. Verwerfe falls $|T| \geq t_{9;\,97{,}5\%} \approx 2{,}26$
7. $\overline{X} = 20,25$, $s \approx 0{,}46$
   d.h. $T \approx 1{,}72$ beobachtet
8. Da $1{,}72 < 2{,}26$ wird H_0 nicht verworfen

außerdem: (zweiseitiges) $(1-\alpha)$-Vertrauensintervall für μ

$$\left[\overline{X} - \frac{t_{d_f,1-\alpha/2} \cdot s}{\sqrt{n}}\,,\ \overline{X} + \frac{t_{d_f,1-\alpha/2} \cdot s}{\sqrt{n}}\right] \approx [19{,}92\,,\ 20{,}58]$$

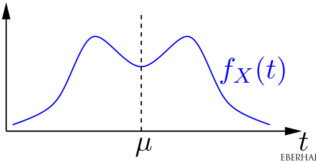
Begründung:

Analog für z-Test:
Ersetze dazu $s \to \sigma$ und, z.B. für $\alpha = 5\%$, $t_{d_f,1-\alpha/2} \to 1{,}96$.

Tests für den Median

- Vorzeichentest: siehe Vorlesung 6
  Annahmen: $X_1, \dots, X_n$
  - unabhängig
  - mit gleichem Median μ.
- Wilcoxon-(Rangsummen-)Test
  Annahmen: $X_1, \dots, X_n$
  - unabhängig
  - mit gleicher Dichte, $f_{X_i} = f_{X_j}$,
  - die symmetrisch um den Median μ ist.

  z.B. $X_j \sim \mathcal{N}(\mu, \sigma_j^2)$ oder auch



Wilcoxon-Test (am Beispiel Waage)

1. $H_0 : \mu = 20$ ("Waage geeicht")
2. $H_A : \mu \neq 20$ ("Waage nicht geeicht")

3. & 7. Teststatistik / Berechnung derselben
   - Bilde Differenzen $d_i = X_i - 20$
   - Ordne $|d_i|$ aufsteigend und vergebe Ränge $1, \ldots, n = 10$
   - $U^- =$ Summe der Ränge, für die $d_j < 0$
     (oder $U^+ =$ Summe der Ränge, für die $d_j > 0$ ...egal)

X_i	20,1	20,3	20,9	19,2	20,8	20,1	20,2	20,4	20,2	20,3
d_i	0,1	0,3	0,9	−0,8	0,8	0,1	0,2	0,4	0,2	0,3
Rang($\|d_i\|$)	1,5	5,5	10	8, 5	8,5	1,5	3,5	7	3,5	5,5

$U^- = 8,5 \quad (U^+ = 46,5)$

4. Keine explizite Formel für Verteilung von $U^{\pm}$,
   aber: Falls H_0 gilt, sollte $U^- \approx U^+ \approx \frac{n(n+1)}{4} = 27{,}5$ sein.
   Hälfte der Summe der Ränge, da Vorzeichen
   der d_i wie durch Münzwurf bestimmt.
5. Signifikanzniveau: z.B. $\alpha = 5\%$
6. Verwerfe H_0, falls $\min(U^-, U^+) \leq 8$
   (kritischer Wert tabelliert)
   bzw. $K = [0, 8] \cup [47, 55]$
8. $U^- = 8{,}5 \notin K$, also wird H_0 nicht verworfen (*gerade so...*)*

* Übrigens: Mit $19{,}2 \to 19{,}3$ hätte der Wilcoxon-Test bereits verworfen, der t-Test noch nicht.

Wahl des Tests für den Lageparameter (Erwartungswert/Median)

- $X_i \sim \mathcal{N}(\mu, \sigma^2)$ unabhängig, σ^2 bekannt $\longrightarrow$ z-Test
- $X_i \sim \mathcal{N}(\mu, \sigma^2)$ unabhängig, σ^2 unbekannt $\longrightarrow$ t-Test
- X_i unabhängig mit gleicher Dichte, symmetrisch um Median $\rightarrow$ Wilcoxon-Test
- X_i unabhängig mit gleichem Median $\rightarrow$ Vorzeichen-Test

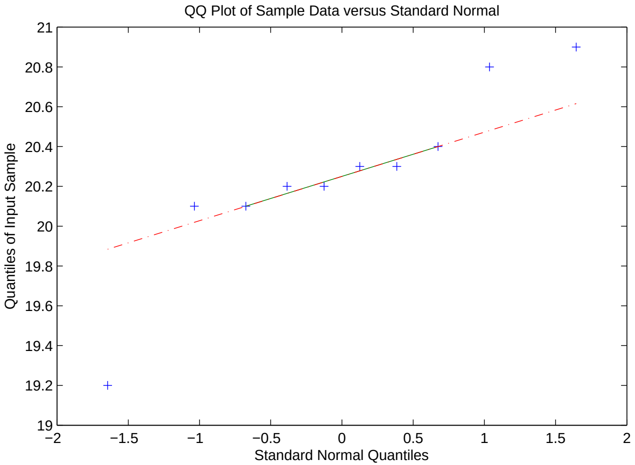
Stärke der Voraussetzungen

- Falls Daten tatsächlich normalverteilt, so hat t-Test größte Macht (leicht größer als Wilcoxon)
- Bei Abweichungen von Normalverteilung hat t-Test oft wesentlich geringere Macht als Wilcoxon (und nie größere)

Konsequenz: Bevorzuge meist Wilcoxon gegenüber t-Test.

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Übrigens: QQ-Plot für die Waage-Daten...



...vielleicht nicht so gut normalverteilt.