Vorlesung: Statistik II für Wirtschaftswissenschaft

Prof. Dr. Helmut Küchenhoff

Institut für Statistik, LMU München

Sommersemester 2017

Konstruktion von statistischen Tests

1. Forschungshypothese
2. Operationalisierung über die zu beobachtende Zufallsvariable X und deren Parameter
3. Formulierung von H_0 typischerweise als Gegenteil der Forschungshypothese und H_1
4. Konstruktion bzw. Wahl einer geeigneten Testgröße $T(X) = T(X_1, ..., X_n)$ als Funktion der erhobenen Daten. Die Testgröße beinhaltet die Information der Daten bezüglich H_0.
5. Aus der Verteilung von $T(X)$ unter der Nullhypothese erhält man Ablehnbereich bzw. p-Wert
6. Entscheidungsregel: H_0 ablehnen, falls Testgröße im Ablehnbereich bzw. $p - Wert < \alpha$

Typen von Tests

- Ein–Stichproben–Fall vs. Zwei– oder Mehr–Stichproben–Fall
- Parametrisch vs. Non-Parametrisch
- Lageparameter, Verteilungen, Andere Parameter

Test auf den Erwartungswert

- Wir interessieren uns für den Erwartungswert μ einer metrischen Zufallsgröße.
  Beispiele: Alter, Einkommen, Körpergröße, Scorewert ...
- Wir können einseitige oder zweiseitige Hypothesen formulieren.
- Beispiele
  - Der Mittelwert der Länge eine Teils in der Produktion liegt bei 12.50 cm
  - Der Blutdruck einer Person wird durch eine Intervention niedriger

Zweiseitiger Gauss-Test auf den Erwartungswert μ

Voraussetzung: Stichprobenumfang n genügend groß (Faustregel $n > 30$)

2. X Zufallsgröße mit Erwartungwert μ.
3. Hypothese über μ:

$$\begin{aligned} H_0 &: \quad \mu = \mu_0 \\ H_1 &: \quad \mu \neq \mu_0 \end{aligned}$$

4. Testgröße: Normierter Mittelwert in der Stichprobe $X_1, \ldots, X_n$.

$$\begin{aligned} T &:= \frac{\bar{X} - \mu_0}{S} \sqrt{n} \\ S^2 &:= \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2 \end{aligned}$$

Bezeichnung für T: t-Wert oder z-Wert

Zweiseitiger Gauss-Test auf den Erwartungswert μ

5. (Approximative) Verteilung von T unter H_0

$$T \sim N(0,1)$$

6. Testentscheidung :

$$p - Wert = 2 \cdot [1 - \Phi(|T|)] = 2 \cdot \left[1 - \Phi\left(|\overline{X} - \mu_0| / \sqrt{\frac{S^2}{n}}\right)\right]$$

Φ ist die Verteilungsfunktion der Standardnormalverteilung
Ablehnung für

$$|T| > z_{1-\alpha/2}$$

$z_{1-\alpha/2}$ ist das (1-$\alpha/2$) - Quantil der Standardnormalverteilung

Einseitiger Gauss-Test auf den Erwartungswert μ

Voraussetzung: Stichprobenumfang n genügend groß (Faustregel $n > 30$)

2. X Zufallsgröße mit Erwartungwert μ.
3. Hypothese über μ:

$$\begin{aligned} H_0 &: \quad \mu \leq \mu_0 \\ H_1 &: \quad \mu > \mu_0 \end{aligned}$$

4. Testgröße: Normierter Mittelwert in der Stichprobe $X_1, \ldots, X_n$.

$$\begin{aligned} T &:= \frac{\bar{X} - \mu_0}{S}\sqrt{n} \\ S^2 &:= \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \end{aligned}$$

Bezeichnung für T: t-Wert oder z-Wert

Einseitiger Gauss-Test auf den Erwartungswert μ

5. (Approximative) Verteilung von T unter H_0

$$T \sim N(0,1)$$

6. Testentscheidung :

$$p - Wert = [1 - \Phi(T)] = \left[1 - \Phi\left((\overline{X} - \mu_0)/\sqrt{\frac{S^2}{n}}\right)\right]$$

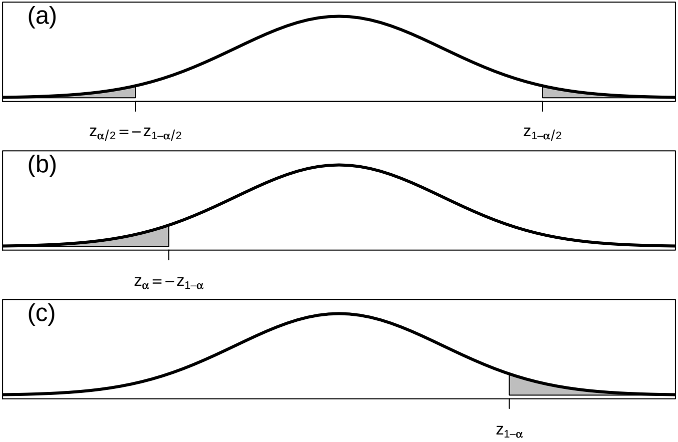
Φ ist die Verteilungsfunktion der Standard Normalverteilung
Ablehnung für

$$T > z_{1-\alpha}$$

$z_{1-\alpha}$ ist das (1-α) - Quantil der Standardnormalverteilung

Ablehnbereich einfacher Gauss-Test

Graphisch dargestellt liegt der kritische Bereich für die unterschiedlichen Fälle an den markierten Enden:



t–Test

Wird bei kleineren Stichproben verwendet.
Voraussetzung: X annähernd normalverteilt

1. X Zufallsgröße mit Erwartungwert μ.
2. Hypothese über μ:

$$\begin{aligned} H_0 &: \quad \mu = \mu_0 \\ H_1 &: \quad \mu \neq \mu_0 \end{aligned}$$

3. Testgröße: Normierter Mittelwert in der Stichprobe $X_1, \ldots, X_n$.

$$\begin{aligned} T &:= \frac{\bar{X} - \mu_0}{S}\sqrt{n} \\ S^2 &:= \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \end{aligned}$$

Bezeichnung für T: t-Wert oder z-Wert

t–Test

4. Verteilung von T unter H_0

$$T \sim t^{n-1}$$

t-Verteilung mit n-1 Freiheitsgeraden

5. Testentscheidung :

$$p - Wert = 2 \cdot [1 - F_{t;n-1}(|T|)]$$

$F_{t;n-1}$ ist die Verteilungsfunktion der t-Verteilung mit n-1 Freiheitsgeraden Ablehnung für

$$|T| > t_{1-\alpha/2}^{n-1}$$

$t_{1-\alpha/2}^{n-1}$ ist das (1-$\alpha/2$) - Quantil der t^{n-1} - Verteilung

Veränderung des Blutdruck nach einer Intervention

- Nullhypothese: Die Blutdruckdifferenz ist 0.

$$H_0 \quad : \quad \mu = 0$$
$$H_1 \quad : \quad \mu \neq 0$$

- Testgröße: Durchschnittliche Bluddruckdifferenz
- n= 22 $\longrightarrow$ zweiseitiger t -Test

Ergebnisse mit R

data: bdd
t = -1.8237, df = 21, p-value = 0.08246
Alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
-5.9034075 0.3870439
sample estimates:
mean of x
-2.758182

Vorzeichentest

Non-Parametrischer Test zur Lage einer Verteilung

2. Betrachtet wird der Median einer Verteilung von beliebiger Struktur
3. $H_0 : x_{med} = \delta_0$
   $H_1 : x_{med} \neq \delta$
4. T $=$ Anzahl der Werte $< \delta_0$
5. $T \sim B(n; 0.5)$
6. Testentscheidung

$$p - Wert = min(2 \cdot (1 - F_{B(n;0.5)}(T-1); 2 \cdot (F_{B(n;0.5)}(T))$$

$F_{B(n;0.5)}$: Verteilungsfunktion der Binomialverteilung

χ^2-Anpassungstest

Motivation

Wir wollen prüfen ob eine Zufallsgröße einer bestimmten Verteilung genügt.
Beispiel: Der Würfel ist fair (alle Zahlen habe die Wahrscheinlichkeit $1/6$)
Die Testgröße wird so konstruiert, dass sie die Abweichungen der unter H_0 erwarteten von den tatsächlich beobachteten absoluten Häufigkeiten misst.

Der Test wird zunächst für kategoriale Größen definiert. Bei stetigem Größen kann der Test angewendet werden, wenn die Stichprobe $\mathbf{X}$ in k (oft willkürlich gewählten) Klassen eingeteilt wird..

χ^2-Anpassungstest

1. Die diskrete Zufallsgröße X mit möglichen Werten 1,...,k hat eine bestimmte Verteilung $F_0(x)$
2. $H_0 : P(X = i) = \pi_i$
   $H_1 : P(X = i) \neq \pi_i$ für mindestens ein i
3. Konstruktion der Testgröße

$$T(\mathbf{X}) = \sum_{i=1}^{k} \frac{(N_i - n\pi_i)^2}{n\pi_i}$$

wobei

- N_i die absolute Häufigkeit der Stichprobe $\mathbf{X}$ für die i-te Klasse angibt
- π_i die Wahrscheinlichkeit, dass X in die Klasse i fällt
- n die Größe der Stichprobe beinhaltet.

χ^2-Anpassungstest

4. Verteilung der Testgröße

$$T_{H_0} \sim \chi^2_{k-1}$$

Die χ^2-Verteilung gilt nur asymptotisch und ist zumeist hinreichend genau, wenn höchstens 1/5 der erwarteten Klassenbesetzungen $n\pi_i$ kleiner als 5 und alle $n\pi_i$ größer als 1 sind.

5. Testentscheidung
Kritischer Bereich: $K = (c_{k-1;1-\alpha}; \infty)$

Approximativer Test auf Erwartungswert–Differenz bei unabhängigen Stichproben

1. X und Y sind zwei Größen mit Erwartungswerten μ_X und μ_Y
2. $X_1, \ldots, X_{n_X}$ und $Y_1, \ldots, Y_{n_Y}$ **unabhängige** Stichproben
3. $H_0 : \mu_X = \mu_Y$
   $H_1 : \mu_X \neq \mu_Y$
4. Testgröße: standardisierte Differenz der Mittelwerte

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}}$$

5. $T \sim N(0,1)$ bei großen Stichprobenumfängen
   (Faustregel: Stichprobenumfänge $n_X, n_Y > 30$)

Approximativer Test auf Erwartungswert–Differenz bei unabhängigen Stichproben

6. Testentscheidung :

$$p - Wert = 2 \cdot [1 - \Phi(|T|)]$$

   Φ ist die Verteilungsfunktion der Standardnormalverteilung
   Ablehnung für

$$|T| > z_{1-\alpha/2}$$

   $z_{1-\alpha/2}$ ist das (1-$\alpha/2$) - Quantil der Standardnormalverteilung

Die entsprechenden einseitigen Tests sind analog zum approximativen Gauss-Test (verwende $1 - \alpha$ Quantile)

Beispiel: Radio-Hördauer Ost-West

- Hören Personen in den alten Bundesländern im Schnitt mehr Radio?
  X : Hördauer im den alten Bundesländern,
  Y: Hördauer in den neuen Bundesländern

$$\begin{aligned} H_0 &: \quad \mu_X - \mu_Y \leq 0 \\ H_1 &: \quad \mu_X - \mu_Y > 0 \end{aligned}$$

- Befragung unter 253 Personen aus den alten Bundesländern und 932 Personen aus den neuen Bundesländern
  - unverbundene Stichproben $X_1, \ldots, X_{253}$ und $Y_1, \ldots, Y_{932}$
  - Stichprobengrößen $n_X = 253, n_Y = 932 > 30$
- Durchschnittliche Hördauer:
  11.4 h (Standardabweichung 8.4 h) in den alten Bundesländern
  9.5 h (Standardabweichung 8.4 h) in den neuen Bundesländern

Beispiel: Radio-Hördauer Ost-West

- Signifikanzniveau: $\alpha = 0.1$
- Differenz der Radio-Hördauer

$$\overline{X} - \overline{Y} = 11.4 - 9.5 = 1.9$$

- Testgröße

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}} = 1.9/0.65 = 2.9$$

- p–Wert : 0.001865813
- H_0 wird abgelehnt, Personen aus den alten Bundesländern hören signifikant länger Radio.

Doppelter t-Test auf die Erwartungswertdifferenz bei unabhängigen Stichproben

1. Vergleich zweier Mittelwerte
2. X und Y sind zwei Größen mit Erwartungswerten μ_X und μ_Y
   X und Y sind **normalverteilt**.
3. $H_0 : \mu_X = \mu_Y$
   $H_1 : \mu_X \neq \mu_Y$
4. Testgröße: Normierte Differenz der Mittelwerte

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}}$$

Doppelter t-Test auf die Erwartungswertdifferenz bei unabhängigen Stichproben

5

$$p - Wert = 2 \cdot [1 - F_{t;k}(|T|)]$$

$F_{t;k}$ ist die Verteilungsfunktion der t-Verteilung mit k Freiheitsgeraden

$$k = \frac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right)^2}{\frac{1}{n_X - 1} \cdot (\frac{s_X^2}{n_X})^2 + \frac{1}{n_Y - 1} \cdot (\frac{s_Y^2}{n_Y})^2}$$

Ablehnung für

$$|T| > t^k_{1-\alpha/2}$$

$t^k_{1-\alpha/2}$ ist das (1-α/2) - Quantil der t^k - Verteilung

Tests auf Erwartungswertdifferenz bei abhängigen Stichproben

1. Gegeben ist eine verbundene Stichprobe
   $X_1, \dots, X_n$ und $Y_1, \dots, Y_n$
2. Bilde die Differenz

$$D_i \quad = \quad X_i - Y_i \quad i = 1, \dots, n$$

3. Berechne Standardabweichung der Differenz

$$s_d \quad = \quad \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (d_i - \bar{d})^2}$$

4. Führe einen Test auf den Erwartungswert von D durch
   - $n > 30 \longrightarrow$ Gauß-Test
   - D normalverteilt $\longrightarrow$ t-Test

Der Wilcoxon Test für unabhängige Stichproben

Test ist identisch mit dem Mann-Whitney-U-Test

1. Unterschied in der Lage zweier Verteilungen
2. X und Y sind zwei Größen mit Medianen med_X und med_Y
3. $H_0 : \text{med}_X = \ med_Y$ vs. $H_1 : \text{med}_X \neq \text{med}_Y$
4. Testgröße Gegeben zwei unabhängige Stichproben X_i und Y_i
   Grundidee: Betrachte die Ränge aus allen Beobachtungen X_i und Y_j und bezeichne diese mit $\text{rg}(X_i)$ und $\text{rg}(Y_j)$, z.B.
   $X_1 = 3,\ X_2 = 5,\ Y_1 = 6,\ Y_2 = 1,\ Y_3 = 4\ \Rightarrow$
   $\text{rg}(X_1) = 2,\ \text{rg}(X_2) = 4,\ \text{rg}(Y_1) = 5,\ \text{rg}(Y_2) = 1\ ,\ \text{rg}(Y_3) = 3$

   $$T = \sum_{i=1}^{m} \text{rg}(X_i)$$

   Die exakte Verteilung von T kann berechnet werden. Für hinreichend große n und m kann sie durch eine NV approximiert werden. Ablehnung von H_0 für große und kleine T.

χ^2-Unabhängigkeitstest

1. Sind zwei kategoriale Zufallsgrößen unabhängig? Unterscheiden sich zwei Anteile?
2. Zwei Zufallsgrößen X und Y mit k bzw. l Ausprägungen

$$p_{ij} = P(X = i, Y = j)$$
$$p_{i\bullet} = P(X = i) \quad p_{\bullet j} = P(Y = j)$$

3. Hypothesen:

$$\begin{aligned} H_0 &: \quad \text{X und Y sind stochastisch unabhängig} \\ & \quad\; p_{ij} = p_{i\bullet} \cdot p_{\bullet j} \text{ für alle } i = 1, \ldots, k, \; j = 1, \ldots, l \\ H_1 &: \quad \text{X und Y sind stochastisch abhängig} \\ & \quad\; p_{ij} \neq p_{i\bullet} \cdot p_{\bullet j} \text{ für mindestens eine ij-Kombination} \end{aligned}$$

χ^2-Unabhängigkeitstest

4. Prüfgröße:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

5. Verteilung:

$$\chi^2 \sim \chi^2_{(k-1)(l-1)}$$

Annahmebereich

$$\chi^2 \leq c_{1-\alpha,\,(k-1)(l-1)}$$

Dabei ist $c_{1-\alpha,\,(k-1)(l-1)}$ das

- $(1-\alpha)$-Quantil der χ^2-Verteilung
- mit $(k-1)\cdot(l-1)$ Freiheitsgraden.

Beispiel: χ^2-Unabhängigkeitstest

$$e_{ij} = \frac{n_{i\bullet} n_{\bullet j}}{n}$$

Geschlecht \ Fußball	ja	nein	Summe
m	87	10	97
w	23	31	54
Summe	110	41	151

Erwartete Besetzungszahlen bei Unabhängigkeit

	ja (j=1)	nein (j=2)
m (i=1)	$\frac{97 \cdot 110}{151} \approx 71$	$\frac{97 \cdot 41}{151} \approx 26$
w (i=2)	$\frac{54 \cdot 110}{151} \approx 39$	$\frac{54 \cdot 41}{151} \approx 15$

$$\begin{aligned} \chi^2 &= \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \\ &\approx \frac{(87-71)^2}{71} + \frac{(10-26)^2}{26} + \frac{(23-39)^2}{39} + \frac{(31-15)^2}{15} \\ &\approx 37.09 \end{aligned}$$

Beispiel: χ^2-Unabhängigkeitstest

- Signifikanzniveau: $\alpha = 0.01$
- Überprüfung mit Faustregel:
  Erwartete Besetzungszahlen $e_{ij} \geq 5$ ✓
- Bestimmung der Freiheitsgrade: $k = l = 2$

$$\text{Freiheitsgrade} = (k-1)\cdot(l-1) = (2-1)\cdot(2-1) = 1$$

$$q_{1-0.01;\,(2-1)(2-1)} = q_{0.99;\,1} \approx 6,63$$

- H_0 wird abgelehnt

Unabhängigkeit und Differenz von Anteilen

Die beide Fragen:

- Gibt es Unterschiede in den Anteilen von $Y = 1$ zweier Gruppen ?
- Gibt es einen Zusammenhang zwischen Gruppen–Zugehörigkeit und einem binären Merkmal Y ?

sind äquivalent.

Differenz von Anteilen bei abhängigen Stichproben

Voraussetzungen:

- X und Y sind zwei Bernoulli-Größen mit

$$\begin{aligned} p_X &= P(X = 1) \\ p_Y &= P(Y = 1) \end{aligned}$$

- $(X_1, Y_1), \ldots, (X_n, Y_n)$ **abhängige, verbundene** Stichproben
- Absolute Häufigkeiten werden in einer Kontingenztafel festgehalten

	Y=0	Y=1
X=0	n_{11}	n_{12}
X=1	n_{21}	n_{22}

Hier kann der χ^2 -Unahängigkeitstest angewendet werden
Für kleine Stichproben: Fisher-Test

Zusammenfassung

- Konstruktion von statistischen Tests verläuft nach einfachen Prinzipien
- Hervorragende Übersicht und Darstellung in Fahrmeier et al. (2016)
- Viele weitere Tests vorhanden
- Immer Angabe von Schätzern und Konfidenzintervallen (nicht nur p–Werte!)