Vorlesung 14a

Statistische Tests

Wiederholung

A. Approximativ normalverteilte Teststatistiken

1.a Liegt ein bestimmter Anteil vor?

1.b. Sind zwei Anteile gleich?

2.a. Liegt ein bestimmter Erwartungswert
(“Lageparameter”) vor?

2.b. Sind zwei Erwartungswerte gleich?

1.a. Test der Hypothese $p = p_0$

(Test auf Vorliegen einer bestimmten Erfolgswahrscheinlichkeit oder eines “Typenanteils in der Populationsanteil” p_0)

$$\text{Teststatistik } D_0 := \frac{H - p_0}{\sqrt{\frac{1}{n} p_0 (1 - p_0)}}$$

mit $H = \widehat{p} \ldots$ der Typenanteil in der Stichprobe

(die relative Anzahl der Erfolge in den n Versuchen).

Für “großes” np und $n(1-p)$

ist H approximativ $N\Big(p, \frac{1}{n} p(1-p)\Big)$ verteilt,

und falls $p = p_0$, ist D_0 approximativ $N(0, 1)$-verteilt.

Sei $d_0 := \dfrac{h - p_0}{\sqrt{\frac{1}{n} p_0(1-p_0)}}$, und sei Z standardnormalverteilt.

$\mathbf{P}[|Z| \geq |d_0|]$ ist dann der p-Wert,

zu dem die Hypothese $p = p_0$ abgelehnt werden kann.

(Dies ist die Wahrscheinlichkeit, bei Zutreffen der Hypothese ein mindestens so weit aus dem Rahmen fallendes Ergebnis zu bekommen wie das beobachtete.)

Beispiel (nach A 45a, Stofl 15/16) A sei eine Teilmenge des Einheitsquadrats Q, der Flächeninhalt von A betrage $p = 0.195$. Die zufälligen Punkte $X_1, X_2, \ldots$ seien unabhängig und uniform verteilt auf Q. Bestimmen Sie mit der Normalapproximation die Wahrscheinlichkeit, dass die Trefferquote der Menge A

i) bei 100 Versuchen (d.h. 100 auf das Quadrat verteilten Punkten)

ii) bei 10000 Versuchen

mit einem Abstand zu p ausfällt, der mindestens 0.02 beträgt.

Lösung: $\mathbf{P}_p(|H-p| \geq 0.02) = \mathbf{P}\Big(|Z| \geq \dfrac{0.02}{\sqrt{\frac{1}{n}p(1-p)}}\Big)$,

$\sqrt{p(1-p)} = 0.396, \;\; 0.02/0.396 = 0.051.$

(i) $\mathbf{P}(|Z| \geq 0.051 \cdot 10) = 2\,\Phi(-0.51) = 0.61.$

(ii) $\mathbf{P}(|Z| \geq 0.051 \cdot 100) = 2\,\Phi(-5.1) = 3.4 \cdot 10^{-7}.$

Eine Variante des obigen Beispiels:

A sei eine Teilmenge des Einheitsquadrats Q; der Flächeninhalt von A sei unbekannt. Die zufälligen Punkte $X_1, X_2, \ldots$ seien unabhängig und uniform verteilt auf Q. Die beobachtete Trefferquote von A unter den ersten n Punkten sei 0.215. Was ist der p-Wert für die Ablehnung der Hypothese “ *der Flächeninhalt von A ist 0.195* ”

i) bei n=100 Punkten

ii) bei n=10000 Punkten ?

Lösung: Die obige Rechnung (mit $p = p_0 := 0.195$) ergibt die p-Werte

$$\mathbf{P}_{p_0}(|H - p_0| \geq 0.02) = \begin{cases} 0.61 & \text{für } n = 100 \\ 3.4 \cdot 10^{-7} & \text{für } n = 10000 \end{cases}$$

Noch eine Variante des obigen Beispiels:

A sei eine Teilmenge des Einheitsquadrats Q. Die zufälligen Punkte $X_1, X_2, \ldots$ seien unabhängig und uniform verteilt auf Q. Die beobachtete Trefferquote von A unter den ersten n Punkten sei H.

a) Geben Sie ein approximatives 95% - Konfidenzintervall für den Flächeninhalt p von A auf der Basis von H an.

b) Welche Intervallgrenzen $h \pm c$ ergeben sich für $h = 0.215$

i) bei n=100 Punkten

ii) bei n=10000 Punkten ?

Lösung: a) Für große n ist $\frac{H-p}{\sqrt{\frac{1}{n}H(1-H)}}$ approximativ $N(0,1)$ verteilt. Also ist ein approximatives 95%-Konfidenzintervall gegeben durch

$$I := \left[H - 2\sqrt{\frac{1}{n}H(1-H)}, H + 2\sqrt{\frac{1}{n}H(1-H)}\right]$$

b) Für $h = 0.215$ ist $\sqrt{h(1-h)} = 0.41$, also ergeben sich die Intervallgrenzen $h \pm c$ mit $c = \frac{2}{\sqrt{n}}0.41$

i) $c = 0.2 \cdot 0.41 = 0.082$ für $n = 100$

ii) $c = 0.02 \cdot 0.41 = 0.0082$ für $n = 10000$

1.b. Test der Hypothese $p_1 = p_2$

(Test auf Gleichheit von zwei Anteilen)

Teststatistik $D := \dfrac{H_1 - H_2}{\sqrt{\frac{1}{n_1} H_1(1-H_1) + \frac{1}{n_2} H_2(1-H_2)}}$

mit $H_1,\ H_2$ die Typenanteile in den beiden Stichproben

Für großes n ist $H_1 - H_2$ approximativ

$N\big(p_1 - p_2, \frac{1}{n_1} p_1(1-p_1) + \frac{1}{n_2} p_2(1-p_2)\big)$-verteilt,

und falls $p_1 = p_2$, ist D approximativ $N(0,1)$-verteilt.

Sei $d := \dfrac{h_1 - h_2}{\sqrt{\frac{1}{n_1}h_1(1-h_1) + \frac{1}{n_2}h_2(1-h_2)}}$,

und sei Z standardnormalverteilt.

$\mathbf{P}[|Z| \geq |d|]$ ist dann der p-Wert,

zu dem die Hypothese $p_1 = p_2$ abgelehnt werden kann.

(Dies ist die Wahrscheinlichkeit, bei Zutreffen der Hypothese ein mindestens so weit aus dem Rahmen fallendes Ergebnis zu bekommen wie das beobachtete.)

2.a. Test der Hypothese $\mu = \mu_0$

(Test auf Vorliegen eines bestimmtes Erwartungswerts
oder “Populationsmittelwertes” μ_0)

$$\text{Teststatistik } T := \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

$\bar{X}$ ist ein Schätzer für μ, S ist ein Schätzer für σ.

Für großes n ist

$\bar{X}$ approximativ $N(\mu, \sigma^2/n)$-verteilt,

und falls $\mu = \mu_0$, ist T approximativ $N(0, 1)$-verteilt.

Sei $t := \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ der beobachtete Wert von T,
und sei Z standardnormalverteilt.

$\mathbf{P}[|Z| \geq |t|]$ ist dann der p-Wert,

zu dem die Hypothese $\mu = \mu_0$ abgelehnt werden kann.

(Dies ist die Wahrscheinlichkeit, bei Zutreffen der Hypothese ein mindestens so weit aus dem Rahmen fallendes Ergebnis zu bekommen wie das beobachtete.)

Bei bekanntem σ kann man anstelle von t den Wert
$z := \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ verwenden.

Beispiel (A52 Stofl 2015/16).

Die Größe X eines Stücks in einer seriellen Fertigung sei die Summe aus einem Normwert μ_0 und einem “zufälligen Fehler” mit Standardabweichung σ. Von Stück zu Stück seien die Fehler unabhängig und identisch verteilt. Denken wir uns μ_0 als 5, und stellen wir uns vor, es sei bekannt, dass $\sigma = 0.1$ ist. Ein Qualitätskontrolleur entnimmt eine Stichprobe vom Umfang 30 und misst den Mittelwert 5.1. Wie wahrscheinlich ist eine mindestens so große Abweichung unter der beschriebenen Hypothese?

Lösung: $z = \frac{5.1 - 5.0}{0.1/\sqrt{30}} = 5.48$

$\mathbf{P}(|Z| \geq 5.48) = 2\,\Phi(-5.48) = 4.3 \cdot 10^{-8}$

Mit einer derart extremen Abweichung wäre

unter der Hypothese $\mu = \mu_0$ einmal in 20 Mio. zu rechnen!

2.b. Test der Hypothese $\mu_X = \mu_Y$

(Test auf Gleichheit zweier Erwartungswerte
oder “Populationsmittelwerte” μ_1 und μ_2)

Teststatistik $T := \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n_x} + \frac{S_Y^2}{n_y}}}$

$\bar{X} - \bar{Y}$ ist ein Schätzer für $\mu_X - \mu_Y$. Für großes n ist $\bar{X} - \bar{Y}$ approximativ $N(\mu_X - \mu_Y, \sigma_X^2/n_x + \sigma_Y^2/n_y)$-verteilt, und falls $\mu_1 = \mu_2$, ist T approximativ $N(0, 1)$-verteilt.

Sei $t := \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}}$

der beobachtete Wert von T, und sei Z

standardnormalverteilt.

$\mathbf{P}[|Z| \geq |t|]$ ist dann der p-Wert,

zu dem die Hypothese $\mu_X = \mu_Y$ abgelehnt werden kann.

(Dies ist die Wahrscheinlichkeit, bei Zutreffen der Hypothese ein mindestens so weit aus dem Rahmen fallendes Ergebnis zu bekommen wie das beobachtete.)

Beispiel (A46 ii) Stofl 2016/17). Aus zwei Populationen wurden Stichproben des Umfangs 30 bzw. 50 entnommen und die Werte x_i bzw. y_j eines reellen Merkmals gemessen. Die Stichprobenmittelwerte und -standardabweichungen waren

$m_x = 10, m_y = 14, s_x = 5, s_y = 8.$

Zu welchem p-Wert lässt sich die Hypothese der Gleichheit der beiden Populationsmittelwerte ablehnen?

Lösung: $t = \dfrac{4}{\sqrt{\frac{25}{30} + \frac{64}{50}}} = 2.75$

$\mathbf{P}(|Z| \geq 2.75) = 2\,\Phi(-2.75) = 0.006$

B. Permutationstests

3. Entsprechen die Anteile in der Stichprobe einem rein zufälligen Ziehen?

4. Kommen zwei Stichproben aus derselben Verteilung, oder aus gegenseitig verschobenen Verteilungen?

3. Fishers exakter Test.

$g \ldots$ Größe der Gesamtpopulation

$w \ldots$ Anzahl von Typ 1-Individuen in der Gesamtpopulation

$K \ldots$ Anzahl von Typ 1-Individuen in der n-Stichprobe

(gezogen ohne Zurücklegen)

Teststatistik $D := K - n\frac{w}{g}$

Unter der Hypothese des rein zufälligen Ziehens ist K

Hyp(n, g, w)-verteilt

Sei $d := k - n\dfrac{w}{g}$, und sei Y Hyp(n, g, w)-verteilt.

$\mathbf{P}[|Y - n\frac{w}{g}| \geq |d|]$ ist dann der p-Wert,

zu dem die Hypothese abgelehnt werden kann.

(Dies ist die Wahrscheinlichkeit, bei Zutreffen der Hypothese ein mindestens so weit aus dem Rahmen fallendes Ergebnis zu bekommen wie das beobachtete.)

Beispiel (A 47 b, Stofl 2016/17) 6 Objekte werden (ohne Mehrfachbelegungen) auf die Plätze $\{1,\dots,18\}$ gesetzt.

(i) Was ist (bei rein zufälliger Wahl) der Erwartungswert der Anzahl der Objekte, die auf der Platzmenge $M := \{1,\dots,10\}$ landen?

(ii) Es ist nur eines der sechs Objekte auf M gelandet. Zu welchem p-Wert können Sie unter Verwendung von Fishers exaktem Test die Hypothese der reinen Zufälligkeit verwerfen?

Lösung: (i) $n\frac{w}{g} = 6 \cdot 10/18 = 10/3$

(ii) $k = 1$

$$\mathbf{P}\left(|Y - \tfrac{10}{3}| \geq \tfrac{10}{3} - 1\right) = \mathbf{P}(Y \in \{0, 1, 6\})$$

$$= \frac{1}{\binom{18}{6}}\left(\binom{10}{0}\binom{8}{6} + \binom{10}{1}\binom{8}{5} + \binom{10}{6}\binom{8}{0}\right)$$

$$= \frac{1}{18564}(28 + 560 + 210) = 0.043$$

4. Wilcoxons Rangsummentest

Teststatistik $W := \sum_{i=1}^{n_x} R(X_i)$

mit $R(X_i) :=$ Rang von X_i in $\{X_1, \ldots, Y_{n_y}\}$.

Unter der Hypothese “X_i und Y_j kommen aus derselben (kontinuierlichen) Verteilung”

ist W so verteilt wie die Summe von n_x rein zufällig ohne Zurücklegen aus $\{1, 2, \ldots, n_x + n_y\}$ gezogenen Zahlen.

Sei $w = \sum\limits_{i=1}^{n_x} R(x_i)$ der beobachtete Wert von W, und S sei so verteilt wie die Summe von n_x rein zufällig ohne Zurücklegen aus $\{1, 2, \ldots, n_x + n_y\}$ gezogenen Zahlen.

Sei a der minimal und b der maximal mögliche Wert von S.

$2 \min(\mathbf{P}(S \leq w), \mathbf{P}(S \geq w))$ ist dann der p-Wert, zu dem die Hypothese abgelehnt werden kann.

(Dies ist die Wahrscheinlichkeit, bei Zutreffen der Hypothese ein mindestens so weit aus dem Rahmen fallendes Ergebnis zu bekommen wie das beobachtete.)

Beispiel (A 47 c, Stofl 2016/17)

(i) Für wieviele dreielementige Mengen aus $\{1, 2, \dots, 30\}$ ist die Summe ihrer Elemente ≤ 10?

(ii) 3 Objekte wurden (ohne Mehrfachbelegungen) auf die Plätze $\{1, \dots, 30\}$ gesetzt, sie fielen auf die Plätze $7, 2$ und 1. Zu welchem p-Wert können Sie unter Verwendung des Wilcoxon-Rangsummentests die Hypothese der reinen Zufälligkeit (zugunsten einer "Tendenz an die Ränder", d.h. zu den Plätzen mit den kleinen bzw zu den Plätzen mit den großen Rängen) verwerfen?

(i) Dies sind die Mengen $\{1,2,3\}, \{1,2,4\}, \{1,2,5\}$, $\{1,2,6\}, \{1,2,7\}, \{1,3,4\}, \{1,3,5\}, \{1,3,6\}$, $\{1,4,5\}, \{2,3,4\}, \{2,3,5\}$, also 11 Stück.

(ii) 11+11=22 dreielementige Teilmengen von $\{1,\ldots,30\}$ haben mindestens so extreme (d.h. nahe beim Minimum oder Maximum liegende) Rangsummmen wie $\{1,2,7\}$. Es gibt insgesamt $\binom{30}{3} = 4060$ dreielementige Teilmengen. Damit ergibt sich als p-Wert: $\frac{22}{4060} = 0.0054$.