Wahrscheinlichkeitstheorie und Statistik

14. Woche

Dávid Tóth (BME SZIT)

2., 5. Dezember 2024

Hypothesentests - Grundprinzipen

Hypothesentests

Hypothesentests gehören zu den grundlegenden Methoden der Statistik. Sie haben das Ziel, Annahmen über Messwerte mit wahrscheinlichkeitstheoretischen Methoden zu überprüfen.

Solche Annahmen oder *Nullhypothesen* sind typischerweise:

- Der Erwartungswert der Stichprobevariablen ist genau/mindestens 2.
- Bei zwei verschiedenen Stichproben sind die Erwartungswerte der zugrunde liegenden Vertleilungen gleich.
- Die Daten folgen einer Normalverteilung.
- Die Stichprobenvaribalen sind unabhängig. (Bei einer *einfachen* Stichprobe ist diese Annahme in der Definition enthalten, aber (im Allgemeinen) nicht alle Stichproben sind einfach!)

Hypothesentests - Grundprinzipen

Hier beschäftigen wir uns mit den ersten zwei Typen: unsere Hypothesen beziehen sich immer auf den/die Parameter der Verteilung/-en.

Wir testen folgenderweise: nach der Messung wird überprüft, ob die gemessenen Daten der Hypothese widersprechen, d.h., ob unter der Annahme, dass die Hypothese gilt, die gemessenen Daten sehr unwahrscheinlich sind. Sind die Messwerte sehr unwahrscheinlich, wird die Hypothese *verworfen*, sonst wird sie *angenommen*.

Hypothesentests - Grundprinzipen

Wenn die Nullhypothese formuliert wurde, dann treffen wir unsere Entscheidung folgenderweise:

- Wir betrachten die Menge $\mathfrak{X}$ aller möglichen Stichprobenwerte, also ist $\mathfrak{X} = \operatorname{ran} \mathbf{X}$, wo $\mathbf{X} = (X_1, \ldots, X_n)$ eine Stichprobe ist.
- Diese Menge wird in zwei disjunkte Teilmengen aufgeteilt:

$$\mathfrak{X} = \mathfrak{X}_{\mathrm{An}} \cup \mathfrak{X}_{\mathrm{Ab}}$$

  wo $\mathfrak{X}_{\mathrm{An}}$ der *Annahmebereich* (elfogadási tartomány) und $\mathfrak{X}_{\mathrm{Ab}}$ der *Ablehnungsbereich* (kritikus tartomány) sind.
- Wir führen das Experiment durch und erhalten eine Realisierung $x_1, \ldots, x_n$.
- Wenn die Daten im Annahmebereich liegen, dann *nimmt* man die Nullhypothese *an* (elfogadjuk a nullhipotézist), anderfalls *verwirft* man die Nullhypothese (elvetjük a nullhipotézist)

Hypothesentests - Grundprinzipen

Es gibt also zwei Fragen, die überlegt werden müssen:

- Nach welchen Prinzipen wird die Nullhypothese H_0 aufgestellt?
- Wie wird der entsprechende Annahmebereich konstruiert?

Hypothesentests - Grundprinzipien

Grunprinzipien bei der Aufstellung der Nullhypothese:

1. Die Nullhypothese H_0 wird oft (aber überhaupt nicht immer) so gewählt, dass die zugrunde liegende Verteilung eindeutig bestimmt wird, falls H_0 richtig ist.

Beispiel. Wir finden eine Münze auf der Strasse und möchten überprüfen, ob die Münze regulär ist. Die Aussage "die Münze ist regulär" ist eine gute Nullhypothese hier, weil falls sie richtig ist, dann sind die Stichprobenvariablen Indikatorvariablen eines Ereignisses von Wahrscheinlichkeit $1/2$. Andererseits ist "die Münze ist nicht regulär" beispielsweise keine gute Nullhypothese, weil falls sie richtig ist, so wissen wir nur, dass die Wahrscheinlichkeit eines Kopfs nicht $1/2$ ist, aber nichts anderes.

Hypothesentests - Grundprinzipien

2. Es gibt zwei Arten von möglichen Fehlern bei der Entscheidung:
   - *Fehler erster Art:* Die Nullypothese H_0 wird verworfen, obwohl sie eigentlich richtig wäre.
   - *Fehler zweiter Art:* Die Nullhypothese wird angenommen, obwohl sie eigentlich falsch ist.

Die Alternativhypothese H_1 ist das Gegenteil der Nullhypothese, d.h. einfach $\neg H_0$. Ein Fehler zweiter Art kann also auch so beschrieben werden, dass die Nullhypothese angenommen wird, obwohl die Alternativhypothese richtig ist.

Die Nullhypothese wird typischerweise so gewählt, dass ein Fehler erster Art schlechter ist, als ein Fehler zweiter Art. Unsere Methode erlauben nämlich, die Wahrscheinlichkeit eines Fehlers erster Art zu kontrollieren, während die Wahrscheinlichkeit eines Fehlres zweiter Art nur begrenzt beeinflußt werden kann.

"Die Schuldigen freizusprechen ist besser als die Verurteilung der Unschuldigen."

Hypothesentests - Grundprinzipien

Beispiel. Ein neues Medikament wird getestet. Hier ist die gute Nullhypothese

$$H_0 : \text{"das Medikament ist unwirksam oder schädlich"},$$

und die gute Alternativhypothese ist

$$H_1 : \text{"das Medikament ist wirksam"}.$$

Wäre nämlich H_0 richtig, so hätte ihre Ablehnung (also die Akzeptanz des Medikamentes) fatale Konsequenzen. Ein solches Ereignis muss also sehr unwahrscheinlich sein. (Wirtschaftlich ist es natürlich auch nicht günstig, wenn das Medikament wirksam ist aber nicht akzeptiert wird.)

Hypothesentests

Die Konstruktion des Annahmebereiches (bzw. des Ablehnungsbereiches) hängt von der konkreten Situation ab.

Beispiel. (u-Test) Wir kaufen jeden Tag Brot im Laden, und in den letzten Tagen scheint das Ein-Kilo-Brot kleiner als früher. Wir möchten also testen, ob das Gewicht des Brots im Durchschnitt (also im Erwartungswert) wirklich 1 kg ist.

Wir messen also das Gewicht der gekauften Brote in $n = 25$ nacheinanderfolgenden Tagen und finden, dass das durchschnittliche Gewicht 0,98 kg ist.

Was wäre die gute Konklusion in diesem Fall?

Hypothesentests

- Wir nehmen an, dass die Gewichte der Brote eine Normalverteilung $N(\mu; \sigma^2)$ folgen.
- Der Einfachheit halber nehmen wir zuerst an, dass auch die Standardabweichung $\sigma = 0{,}05$ kg der Verteilung bekannt ist.
- Unsere Nullhypothese ist

$$H_0 : \quad \mu = \mu_0 = 1 \text{ kg},$$

und die Alternativhypothese ist

$$H_1 : \quad \mu \neq \mu_0.$$

Hypothesentests

Im u-Test entscheidet man folgenderweise: ein Konfidenzintervall I wird für den Erwartungswert konstruiert, und

- falls $\mu_0 \in I$ gilt, dann wird die Nullhypothese angenommen,
- sonst wird H_0 verworfen.

Bei der Konstruktion des Konfidenzintervalls $[T_1(\mathbf{X}); T_2(\mathbf{X})]$ wird ein Fehlerniveau $\varepsilon \in (0;1)$ so gewählt, dass

$$\mathbb{P}_\mu(\mu \in [T_1(\mathbf{X}); T_2(\mathbf{X})]) = 1 - \varepsilon$$

für alle $\mu \in \mathbb{R}$ gilt. Also, falls H_0 richtig ist, dann wird die Wahrscheinlichkeit dafür, dass sie verworfen wird, also die Wahrscheinlichkeit des Fehlers erster Art

$$\mathbb{P}_{\mu_0}(\mu_0 \notin [T_1(\mathbf{X}); T_2(\mathbf{X})]) = \varepsilon.$$

Wir sagen, dass das *Fehlerniveau* oder *Signifikanzniveau* des Tests ε ist. (A magyar nyelvű irodalomban a szignifikanciaszint $1 - \varepsilon$.)

Hypothesentests

Das Konfidenzintervall für den Erwartungswert zum Fehlerniveau ε ist

$$\left[\overline{X} - \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}; \overline{X} + \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}\right]$$

wo $u_{\varepsilon/2} = \Phi^{-1}\left(1 - \frac{\varepsilon}{2}\right)$ das $(1 - \frac{\varepsilon}{2})$-Quantil der Standardnormalverteliung ist.

D.h.: die Nullhypothese wird genau dann angenommen, falls

$$\overline{X} - \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}} \leq \mu_0 \leq \overline{X} + \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}.$$

Hypothesentests

Im Beispiel:

$$\mu = \mu_0 = 1 \text{ kg}, \quad \sigma = \sigma_0 = 0{,}05 \text{ kg}, \quad n = 25, \quad \overline{x} = 0{,}98 \text{ kg}.$$

Typische Werte von ε sind 0,05 und 0,01. Für $\varepsilon = 0{,}05$ erhält man den Radius

$$\frac{0{,}05 \cdot u_{0{,}025}}{\sqrt{25}} = \frac{0{,}05 \cdot \Phi^{-1}(0{,}975)}{5} = 0{,}01 \cdot 1{,}96 = 0{,}0196.$$

Also wird das Konfidenzintervall

$$I = [0{,}98 - 0{,}0196; 0{,}98 + 0{,}0196] = [0{,}9604; 0{,}9996],$$

und $\mu_0 \notin I$, deshalb wird die Nullhypothese abgelehnt (verworfen).

Hypothesentests

Das Konfidenzintervall hängt von den Stichprobe ($\overline{x}$, n) und auch von unserer Wahl von ε ab. Diese werden follgenderweise getrennt:

- Eine *Teststatistik* wird definiert, die durch eine Stichprobenfunktion $T : \mathbb{R}^n \to \mathbb{R}$ angegeben wird (und desahlb nur von der Stichprobe abhängt).
- Zum Signifikanzniveau gehört ein Quantil (in diesem Fall $u_{\varepsilon/2}$), mit dem der Wert der Teststatistik verglichen wird.

Hypothesentests

Im obigen Beispiel überprüften wir, ob

$$\overline{X} - \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}} \leq \mu_0 \leq \overline{X} + \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}.$$

Diese Ungleichungen können folgenderweise umgeornet werden:

$$-u_{\varepsilon/2} \leq \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} \leq u_{\varepsilon/2}.$$

Hypothesentests

Die Teststatistik ist also

$$\boxed{u(\mathbf{X}) = \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n}}$$

und der Annahmebereich wird durch

$$\mathfrak{X}_{\mathrm{An}} = u^{-1}([-u_{\varepsilon/2}; u_{\varepsilon/2}]) \cap \mathfrak{X}$$

definiert (und $\mathfrak{X}_{\mathrm{Ab}} = \mathfrak{X} \setminus \mathfrak{X}_{\mathrm{An}}$).

Die Nullhypothese wird also genau dann angenommen, falls

$$|u(\mathbf{X})| \leq u_{\varepsilon/2}$$

gilt, anderfalls wird sie verworfen. Wir sagen, dass $u_{\varepsilon/2}$ der *kritische Wert* des Tests ist.

Hypothesentests

Im obigen Beispiel gilt

$$u(x_1, \ldots, x_n) = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n} = \frac{0{,}98 - 1}{0{,}05}\sqrt{25} = \frac{-0{,}02}{0{,}05} \cdot 5 = -2$$

und

$$|u(x_1, \ldots, x_n)| = 2 > u_{0{,}025} = 1{,}96,$$

also wird H_0 verworfen.

Wenn wir das Fehlerniveau $\varepsilon = 0{,}01$ wählen, dann ändert sich der Testwert nicht, wir benutzen aber das Quantil $u_{0{,}005}$:

$$|u(x_1, \ldots, x_n)| = 2 < u_{0{,}005} = \Phi^{-1}(0{,}995) \approx 2{,}57,$$

also wird H_0 in diesem Fall angenommen.

Hypothesentests

Zusammenfassung - zweiseitiger u-Test für eine Stichprobe:
Test, ob der Erwartungswert μ von normalverteilten Variablen gleich einer gegebenen Zahl μ_0 ist.

Sei $(x_1, \ldots, x_n)$ eine Realisierung einer einfachen Stichprobe $\mathbf{X} = (X_1, \ldots, X_n)$, wo $X_i \sim \mathrm{N}(\mu, \sigma^2)$ gilt, und σ *bekannt* ist.

Vorgehen beim u-Test:

1. Aufstellen der Nullhypothese H_0: $\mu = \mu_0$.
2. Wahl des Fehlerniveaus $\varepsilon \in (0; 1)$.
3. Bestimmung des kritischen Wertes $u_{\varepsilon/2}$.
4. Erhebung der Stichprobenwerten $x_1, \ldots, x_n$.
5. Berechnung des Wertes der Teststatistik:

$$u(x_1, \ldots, x_n) = \frac{\overline{x} - \mu_0}{\sigma} \sqrt{n}$$

6. Entscheidung, ob der Testwert im Annahmebereich liegt, also ob $|u(x_1, \ldots, x_n)| \leq u_{\varepsilon/2}$ gilt.

Hypothesentests

Bemerkung. Das Wort "zweiseitig" bezieht sich darauf, dass die Alternativhypothese den richtigen Wert von μ auf beiden Seiten von μ_0 erlaubt (und folglich wird das Konfidenzintervall hier so gewählt, dass es auf beiden Seiten des empirischen Mittels beschränkt wird).

Hypothesentests

Zweiseitiger t-Test für eine Stichprobe: Test, ob der Erwartungswert μ von normalverteilten Variablen gleich einer gegebenen Zahl μ_0 ist.

In diesem Fall ist aber die **Standardabweichung** der Verteilung **nicht bekannt**.

Die Methode ist gleich wie im vorigen Fall: ein Konfidenzintervall wird für den Erwartungswert konstruiert und es wird überprüft, ob der Wert μ_0 in ihm liegt.

Hypothesentests

Sei also $\mathbf{X} = (X_1, \ldots, X_n)$ eine einfache Stichprobe, wo $X_i \sim \mathrm{N}(\mu, \sigma^2)$ gilt und σ nicht bekannt ist.

Sei $H_0 : \mu = \mu_0$ die Nullhypothese und wählen wir ein Fehlerniveau $\varepsilon \in (0; 1)$. Das Konfidenzintervall für den Erwartungswert μ zum Fehlerniveau ε ist

$$\left[\overline{X} - \frac{S_n^* t_{n-1,\varepsilon/2}}{\sqrt{n}}; \overline{X} + \frac{S_n^* t_{n-1,\varepsilon/2}}{\sqrt{n}}\right],$$

wo $t_{n,\delta}$ das $(1-\delta)$-Quadntil der t-Verteilung mit n Freiheitsgraden ist.

Es wird überprüft, ob die Ungleichungen

$$\overline{X} - \frac{S_n^* t_{n-1,\varepsilon/2}}{\sqrt{n}} \leq \mu_0 \leq \overline{X} + \frac{S_n^* t_{n-1,\varepsilon/2}}{\sqrt{n}}$$

gelten.

Hypothesentests

Nach Umordnung ergibt sich

$$-t_{n-1,\varepsilon/2} \leq \frac{\overline{X} - \mu_0}{S_n^*}\sqrt{n} \leq t_{n-1,\varepsilon/2},$$

also muss man überprüfen, ob

$$|t(\mathbf{X})| \leq t_{n-1,\varepsilon/2}$$

für die Teststatistik

$$t(\mathbf{X}) = \frac{\overline{X} - \mu_0}{S_n^*}\sqrt{n}$$

gilt. Falls die obige Ungleichung gilt, dann wird H_0 angenommen, sont wird sie verworfen.

Hypothesentests

Zusammenfassung - zweiseitiger t-Test für eine Stichprobe:
Test, ob der Erwartungswert μ von normalverteilten Variablen gleich einer gegebenen Zahl μ_0 ist.

Sei $(x_1, \dots, x_n)$ eine Realisierung einer einfachen Stichprobe $\mathbf{X} = (X_1, \dots, X_n)$, wo $X_i \sim \mathrm{N}(\mu, \sigma^2)$ gilt, und σ *nicht bekannt* ist.

Vorgehen beim t-Test:

1. Aufstellen der Nullhypothese H_0: $\mu = \mu_0$.
2. Wahl des Fehlerniveaus $\varepsilon \in (0; 1)$.
3. Bestimmung des kritischen Wertes $t_{n-1,\varepsilon/2}$.
4. Erhebung der Stichprobenwerten $x_1, \dots, x_n$.
5. Berechnung des Wertes der Teststatistik:

$$t(x_1, \dots, x_n) = \frac{\overline{x} - \mu_0}{s_n^*}\sqrt{n}$$

6. Entscheidung, ob der Testwert im Annahmebereich liegt, also ob $|t(x_1, \dots, x_n)| \leq t_{n-1,\varepsilon/2}$ gilt.

Hypothesentests

Beispiel. Wir kauften ein Ein-Kilo-Brot täglich in den letzten 25 Tagen im Laden, und das durchschnittliches Gewicht dieser Brote ist 0,98 kg. Aus diesem Ergebnis kann man vermuten, dass der Erwartungswert des Gewichts nicht 1 kg ist, aber das könnte keinen Beweis darauf sein, dass die Brote durchschnittlich *schwerer* als 1 kg sind.

Aus unserer Sicht erscheint es sinvoller zu überprüfen, ob das Gewicht der Brote im Erwartungswert *kleiner als 1 kg* ist.

Wir wollen die Unschuldige nicht verurteilen, also wählen wir die Nullhypothese

$$H_0 : \text{ der Erwartungswert des Gewichts ist mindestens 1 kg.}$$

Hypothesentests

Unsere Stichprobe $\mathbf{X} = (X_1, \dots, X_n)$ wird wieder als Normalverteilt betrachtet mit unbekanntem Erwartungswert μ, und wir werden zuerst annehmen, dass die Standardabweichung σ bekannt ist.

Hier wird aber die Verteilung nicht durch die Nullhypothese H_0 eindeutig bestimmt, und falls ein Fehlerniveau $\varepsilon \in (0; 1)$ gewählt wird, dann kann eine Gleichung der Form

$$\mathbb{P}_\mu(\mathbf{X} \in \mathfrak{X}_{\mathrm{Ab}}) = \varepsilon$$

nicht für alle Parameter μ erreicht werden, die die Nullhypothese erfüllen. Stattdessen velangen wir, dass die Wahrscheinlichkeit eines Fehlers erster Art *höchstens* ε ist.

Hypothesentests

Formalisierung der Nullhypothese:

- Die Nullhypothese H_0 teilt die Menge Θ der möglichen Parameter in zwei Teile auf:

$$\mathbb{R} = \Theta = \Theta_0 \cup \Theta_1,$$

  wo $\Theta_0 = \{\mu \in \Theta : \mu \geq 1\}$ und $\Theta_1 = \{\mu \in \Theta : \mu < 1\}$.

- Im vorigen Fällen hatten wir auch eine solche (aber einfachere) Aufteilung: die Nullhypothese $\{\mu = 1\}$ ergab die Mengen $\Theta_0 = \{1\}$ und $\Theta_1 = \mathbb{R} \setminus \{1\}$.
- Alle Nullhypothesen, die sich auf die Parameter beziehen, können durch eine solche Aufteilung $\Theta = \Theta_0 \cup \Theta_1$ bestimmt werden, wobei $\Theta_0 \cap \Theta_1 = \emptyset$ und die Nullhypothese H_0 kann durch die formale Aussage

$$\vartheta_{\mathrm{r}} \in \Theta_0$$

  angegeben werden, wo ϑ_{r} der richtige Parameter der Verteilung ist.

Hypothesentests

Ein Hypothesentest wird dann formal durch eine Aufteilung

$$\mathfrak{X} = \mathfrak{X}_{\mathrm{An}} \cup \mathfrak{X}_{\mathrm{Ab}}$$

beschreibt (H_0 wird genau dann angenommen, falls die Realisierung der Stichprobe in $\mathfrak{X}_{\mathrm{An}}$ liegt).

Definition. Wir sagen, dass ein Hypothesentest das *Signifikanzniveau* oder *Fehlerniveau* ε hat, falls

$$\mathbb{P}_\vartheta(\mathbf{X} \in \mathfrak{X}_{\mathrm{Ab}}) \leq \varepsilon$$

für alle $\vartheta \in \Theta_0$ gilt, also falls die Wahrscheinlchkeit des Fehlers erster Art höchstens ε ist.

Hypothesentests

In unserem Beispiel werden wir sogar

$$\sup_{\mu \in \Theta_0} \mathbb{P}(\mathbf{X} \in \mathfrak{X}_{\mathrm{Ab}}) = \varepsilon$$

erreichen, also werden wir einen Test konstruieren, der für eine früher festgelegte Zahl ε das Fehlerniveau ε hat, und für den ε die kleinste solche Zahl ist.

Hypothesentests

Wir stellen die Nullhypothese und die Alternativhypothese folgenderweise auf:

$$H_0 : \mu_\mathrm{r} \geq \mu_0 \text{ vs. } H_1 : \mu_\mathrm{r} < \mu_0,$$

wo μ_r der richtige Parameter ist. Wir wählen ein Fehlerniveau $\varepsilon \in (0; 1)$ und verwerfen die Nullhypothese, falls das empirische Mittel zu klein ist, also falls $\overline{X} + r_\varepsilon < \mu_0$ für eine bestimme Zahl r_ε gilt. Wir werden

$$\mathbb{P}_\mu(\overline{X} + r_\varepsilon < \mu_0) \leq \varepsilon$$

für alle $\mu \in \Theta_0 = \{\mu \in \mathbb{R} : \mu \geq \mu_0\}$ garantieren, dann wird die Nullhypothese höchstens mit Wahrscheinlichkeit ε verworfen, wenn $\mu_\mathrm{r} \in \Theta_0$ gilt.

Hier ist es genügend, den Grenzfall $\mu = \mu_0$ zu handeln. Nämlich, je größer μ_r ist, desto kleiner ist die Wahrscheinlichkeit dafür, dass $\overline{X}$ klein wird.

Hypothesentests

Wir wählen also r_ε so, dass die Gleichung

$$\mathbb{P}_{\mu_0}(\overline{X} + r_\varepsilon < \mu_0) = \varepsilon$$

gilt. Diese ist äquivalent zu

$$\begin{aligned}\mathbb{P}_{\mu_0}\left(\frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} < \frac{-r_\varepsilon\sqrt{n}}{\sigma}\right) = \\ = \Phi\left(\frac{-r_\varepsilon\sqrt{n}}{\sigma}\right) = 1 - \Phi\left(\frac{r_\varepsilon\sqrt{n}}{\sigma}\right) = \varepsilon,\end{aligned}$$

also zu

$$r_\varepsilon = \frac{\sigma\Phi^{-1}(1-\varepsilon)}{\sqrt{n}} = \frac{\sigma u_\varepsilon}{\sqrt{n}}.$$

Hypothesentests

Wir müssen noch zeigen, dass

$$\mathbb{P}_\mu\left(\overline{X} + \frac{\sigma u_\varepsilon}{\sqrt{n}} < \mu_0\right) \leq \varepsilon$$

für alle $\mu \geq \mu_0$ gilt:

$$\begin{aligned}
\mathbb{P}_\mu\left(\overline{X} + \frac{\sigma u_\varepsilon}{\sqrt{n}} < \mu_0\right) &= \mathbb{P}_\mu\left(\frac{\overline{X} - \mu}{\sigma}\sqrt{n} < -u_\varepsilon + \frac{(\mu_0 - \mu)\sqrt{n}}{\sigma}\right) \\
&\leq \mathbb{P}_\mu\left(\frac{\overline{X} - \mu}{\sigma}\sqrt{n} < -u_\varepsilon\right) \\
&= \Phi(-u_\varepsilon) = 1 - \Phi(u_\varepsilon) \\
&= 1 - \Phi(\Phi^{-1}(1-\varepsilon)) = 1 - (1-\varepsilon) = \varepsilon.
\end{aligned}$$

Hypothesentests

Wir verwerfen also H_0 genau dann, falls

$$\overline{X} + \frac{\sigma u_\varepsilon}{\sqrt{n}} < \mu_0 \iff \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} < -u_\varepsilon = -\Phi^{-1}(1-\varepsilon)$$

gilt.

Die Nullhypothese $\mu \leq \mu_0$ wird ähnlicherweise behandelt, in diesem Fall verwerfen wir H_0 genau dann, falls

$$\frac{\overline{X} - \mu_0}{\sigma}\sqrt{n} > u_\varepsilon$$

gilt.

Hypothesentests

Zusammenfassung - einseitiger u-Test für eine Stichprobe:
Test, ob der Erwartungswert μ von normalverteilten Variablen mindestens/höchstens einer gegebenen Zahl μ_0 ist.
Sei $(x_1, \ldots, x_n)$ eine Realisierung einer einfachen Stichprobe $\mathbf{X} = (X_1, \ldots, X_n)$, wo $X_i \sim \mathrm{N}(\mu, \sigma^2)$ gilt, und σ *bekannt* ist.

Vorgehen beim u-Test:

1. Aufstellen der Nullhypothese H_0: $\mu \geq \mu_0$ (bzw. $\mu \leq \mu_0$).
2. Wahl des Fehlerniveaus $\varepsilon \in (0; 1)$.
3. Bestimmung des kritischen Wertes u_ε.
4. Erhebung der Stichprobenwerten $x_1, \ldots, x_n$.
5. Berechnung des Wertes der Teststatistik:

$$u(x_1, \ldots, x_n) = \frac{\overline{x} - \mu_0}{\sigma} \sqrt{n}$$

6. Entscheidung, ob der Testwert im Annahmebereich liegt, also ob $u(x_1, \ldots, x_n) \geq -u_\varepsilon$ (bzw. $u(x_1, \ldots, x_n) \leq u_\varepsilon$) gilt.

Hypothesentests

Der Fall der unbekannten Standardabweichung ist analogisch, wir geben nur den Test an:

Hypothesentests

Zusammenfassung - einseitiger t-Test für eine Stichprobe:
Test, ob der Erwartungswert μ von normalverteilten Variablen mindestens/höchstens einer gegebenen Zahl μ_0 ist.
Sei $(x_1, \ldots, x_n)$ eine Realisierung einer einfachen Stichprobe $\mathbf{X} = (X_1, \ldots, X_n)$, wo $X_i \sim \mathrm{N}(\mu, \sigma^2)$ gilt, und σ *unbekannt* ist.

Vorgehen beim t-Test:

1. Aufstellen der Nullhypothese H_0: $\mu \geq \mu_0$ (bzw. $\mu \leq \mu_0$).
2. Wahl des Fehlerniveaus $\varepsilon \in (0; 1)$.
3. Bestimmung des kritischen Wertes $t_{n-1,\varepsilon}$.
4. Erhebung der Stichprobenwerten $x_1, \ldots, x_n$.
5. Berechnung des Wertes der Teststatistik:

$$t(x_1, \ldots, x_n) = \frac{\overline{x} - \mu_0}{s_n^*}\sqrt{n}$$

6. Entscheidung, ob der Testwert im Annahmebereich liegt, also ob $t(x_1, \ldots, x_n) \geq -t_{n-1,\varepsilon}$ (bzw. $t(x_1, \ldots, x_n) \leq t_{n-1,\varepsilon}$) gilt.

Hypothesentests

Bemerkung.

- Wenn der Umfang der Stichprobe "groß" ist, dann ist die empirische Varianz eine gute (erwartungswerte und stark konsistente) Schätzung für die Varianz.
- Also betrachtet man die empirische Standardabweichung bei einer "großen" Stichprobe oft als der richtige Wert der Standardabweichung, und dementsprechend benutzt man u-Tests statt t-Tests.
- In der Praxis kann eine Stichprobe vom Umfang 30 oft als "groß" betrachtet werden.

Hypothesentests

Wie behandelt man solche Beispiele, wo die Stichprobenvariablen nicht Normaverteilt sind?

Beispiel. Betrachten wir eine auf der Straße gefundene Münze, mit der man mit Wahrscheinlichkeit $\vartheta \in (0; 1)$ einen Kopf erhält.

- Werfen wir die Münze (unabhängig) vielmal, und betrachten wir die Anzahl X der Köpfe.
- Wir sahen früher, dass die Standardisierte von X approximativ standardnormalverteilt ist.
- Eine einfache Rechnung zeigte auch, dass man zu einer sehr präzisen Schätzung des Parameters ϑ sogar mehrere Millionen Experimenten machen muss.
- Wenn aber weniger präzise Ergebnisse genügen, so kann der zentrale Grenzwertsatz bei einer ziemlich großen Stichprobe (z.B. mit etwa 100 Würfe) angewandt werden.

Hypothesentests

Nehmen wir an, dass wir entscheiden möchten, ob die Münze regulär ist, also stellen wir die Hypothesen

$$H_0 : \vartheta = \frac{1}{2}, \qquad H_1 : \vartheta \neq \frac{1}{2}$$

auf.

- Nehmen wir an, dass die Münze 100 Mal geworfen wurde. Sei X_i die Indikatorvariable des Ereignisses, dass der i-te Wurf ein Kopf ist.
- Falls H_0 richtig ist, dann ist die Variable

$$\frac{\sum_{i=1}^{n} X_i - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} = \frac{\overline{X} - 1/2}{1/2}\sqrt{n}$$

  nach dem zentralen Grenzwertsatz approximativ standardnormalverteilt.
- Also funktioniert die Argumentation, die bei dem u-Test durchgeführt wurde, auch in diesem Fall.

Hypothesentests

- Wählen wir das Fehlerniveau $\varepsilon = 0{,}05$.
- Angenommen, 60 Köpfe und 40 Zahlen werden geworfen. Der Wert der Teststatistik ist

$$\frac{\overline{x} - 1/2}{1/2}\sqrt{n} = \frac{\frac{6}{10} - \frac{1}{2}}{\frac{1}{2}} \cdot \sqrt{100} = \frac{1}{5} \cdot 10 = 2.$$

- Falls H_0 richtig ist, dann wird der Wert $1/2$ im Konfidenzintervall genau dann, falls der Wert der Testsatistik kleiner als $u_{0{,}025} = 1{,}96$ ist, also verwerfen wir die Nullhypothese.

Hypothesentests

Zweiseitiger u-Test für zwei Stichproben.

Seien $\mathbf{X} = (X_1, \ldots, X_{n_1})$ und $\mathbf{Y} = (Y_1, \ldots, Y_{n_2})$ einfache Stichproben, wo *alle* Variablen unabhängig sind. Nehmen wir an, dass $X_i \sim \mathrm{N}(\mu_1; \sigma_1^2)$ und $Y_j \sim \mathrm{N}(\mu_2; \sigma_2^2)$ gelten, wo σ_1 und σ_2 bekannt sind.

Wir werden die folgende Nullhypothese betrachten:

$$H_0 : \mu_1 = \mu_2, \qquad (H_1 : \mu_1 \neq \mu_2).$$

Hypothesentests

Behauptung. Seien $\mathbf{X} = (X_1, \ldots, X_{n_1})$ und $\mathbf{Y} = (Y_1, \ldots, Y_{n_2})$ einfache Stichproben wie oben. Falls $\mu_1 = \mu_2$ gilt, so ist die Variable

$$u(\mathbf{X}, \mathbf{Y}) = \frac{\overline{X_{n_1}} - \overline{Y_{n_2}}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

standardnormalverteilt.

Beweis. Übung.

Hypothesentests

Der Test geht wie früher vor:

- Ein Fehlerniveau $\varepsilon \in (0;1)$ wird gewählt.
- Falls H_0 richtig ist, dann wird der Wert der Teststatistik $u(\mathbf{X},\mathbf{Y})$ im Intervall $[-u_{\varepsilon/2}; u_{\varepsilon/2}]$ mit Wahrscheinlichkeit $1-\varepsilon$ enthalten.
- Also wird H_0 im Fall $|u(\mathbf{X},\mathbf{Y})| \leq u_{\varepsilon/2}$ angenommen, sonst wird sie verworfen.
- Die Wahrscheinlichkeit des Fehlers erster Art ist ε.

Hypothesentests

Zweiseitiger t-Test für zwei Stichproben.

Seien $\mathbf{X} = (X_1, \ldots, X_{n_1})$ und $\mathbf{Y} = (Y_1, \ldots, Y_{n_2})$ einfache Stichproben, wo alle Variablen unabhängig sind. Nehmen wir an, dass $X_i \sim \mathrm{N}(\mu_1; \sigma^2)$ und $Y_j \sim \mathrm{N}(\mu_2; \sigma^2)$ gelten, also **die Varianzen gleich sind**.

Wir werden wieder die folgende Nullhypothese betrachten:

$$H_0 : \mu_1 = \mu_2, \qquad (H_1 : \mu_1 \neq \mu_2).$$

Hypothesentests

Behauptung. Seien $\mathbf{X} = (X_1, \ldots, X_{n_1})$ und $\mathbf{Y} = (Y_1, \ldots, Y_{n_2})$ einfache Stichproben wie oben. Falls $\mu_1 = \mu_2$ gilt, so ist die Variable

$$t(\mathbf{X}, \mathbf{Y}) = \frac{\overline{X_{n_1}} - \overline{Y_{n_2}}}{\sqrt{\frac{(n_1 - 1)(S_X^*)^2 + (n_2 - 1)(S_Y^*)^2}{n_1 + n_2 - 2}}} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

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

Beweis. Übung.

Hypothesentests

Der Test geht wie früher vor:

- Ein Fehlerniveau $\varepsilon \in (0; 1)$ wird gewählt.
- Falls H_0 richtig ist, dann wird der Wert der Teststatistik $t(\mathbf{X}, \mathbf{Y})$ im Intervall $[-t_{n_1+n_2-2;\varepsilon/2}; t_{n_1+n_2-2;\varepsilon/2}]$ mit Wahrscheinlichkeit $1 - \varepsilon$ enthalten.
- Also wird H_0 im Fall $|t(\mathbf{X}, \mathbf{Y})| \leq t_{n_1+n_2-2;\varepsilon/2}$ angenommen, sonst wird sie verworfen.
- Die Wahrscheinlichkeit des Fehlers erster Art ist ε.