7. Hypothesentests

Lernziele:

- Sie können eine Testentscheidung sinnvoll als Hypothesentest formulieren.
- Sie können Folgerungen aus dem Testergebnis zu einer konkreten Stichprobe ziehen.
- Sie unterscheiden zwischen dem Fehler 1. Art (dem Signifikanzniveau) und dem Fehler 2. Art.
- Sie kennen den p-Wert und können auch anhand des p-Wertes Testentscheidungen treffen.
- Sie verstehen den Zusammenhang zwischen einem Konfidenzintervall, einem klassischen Parametertest und dem p-Wert.

Literatur:

- Teschl Band 2, Kap. 30.4
- Zucchini, Kap. 8.1 + 8.3 + 8.5
- Arens et al., Kap 40.5

Situation:

Basierend auf n i.i.d. Zufallsvariablen $X_1, \ldots, X_n$ (Messungen) soll eine Entscheidung getroffen werden, ob eine Hypothese für einen unbekannten Parameter der Verteilung, z. B. den Erwartungswert μ, gültig ist oder nicht.

Beispiel: Unterschiedliche Fragestellungen beim **Schätzen** bzw. **Testen** eines unbekannten Parameters

- Schätzen:
  Wie groß ist die durchschnittliche Abfüllmenge von 0.5-Liter Flaschen?
- Testen:
  Kommt es zu Verbraucherklagen, weil die angegebene Abfüllmenge unterschritten wird?

7.1 Nullhypothese und Gegenhypothese

Vor dem Test:
Formulierung des Modells, der Nullhypothese H_0 und Gegenhypothese H_1

- **Modell**: Verteilung der Grundgesamtheit bzw. einer Testgröße TG (häufig: Mittelwert) ist bekannt bis auf einen Parameter, z. B. den Erwartungswert μ, für den eine Hypothese aufgestellt wird.
  z. B. $TG \sim N_{\mu,\sigma^2}$
- **Nullhypothese** H_0: Angezweifelte Aussage, der widersprochen werden kann, wenn die Stichprobe einen Gegenbeweis liefert
  z. B. $H_0 : \mu = \mu_0$
- **Gegenhypothese** H_1: Gegenteil von H_0
  z. B. $H_1 : \mu \neq \mu_0$

7.2 Signifikanzniveau, Ablehnungsbereich, Fehler 1. + 2. Art

Treffen der Testentscheidung
basierend auf einer konkreten Stichprobe $\{x_1, \ldots, x_n\}$

- Berechnung der Realisation $tg = TG(x_1, \ldots, x_n)$ der **Testgröße** TG
- **Ablehnungsbereich / Kritischer Bereich** C: Werte der Testgröße, die für H_1 sprechen und bei Gültigkeit von H_0 mit Wahrscheinlichkeit $\leq \alpha$ (meist: 0.1, 0.05 oder 0.01), dem sog. **Signifikanzniveau** auftreten. α ist also die Wahrscheinlichkeit, dass H_0 verworfen wird, obwohl sie richtig ist, der sog. **Fehler 1. Art**.
- **Annahmebereich**: Komplement $\bar{C}$ des Ablehnungsbereichs
H_0 kann nicht abgelehnt werden, falls $tg \in \bar{C}$ $(P(tg \in \bar{C}) \geq 1 - \alpha)$.
Die Wahrscheinlichkeit, dass H_0 nicht abgelehnt wird, obwohl sie falsch ist, ist der sog. **Fehler 2. Art**.

Testszenarien

	Testentscheidung	
Realität	H_0 wird nicht abgelehnt.	H_0 wird abgelehnt.
H_0 ist wahr.	richtig	falsch Fehler 1. Art ist die entsprechende Wsk $\leq \alpha$
H_0 ist falsch.	falsch Fehler 2. Art ist die entsprechende Wsk	richtig

Wenn α klein, dann hat man eine signifikante Aussage, falls H_0 abgelehnt wird.

Beispiel:

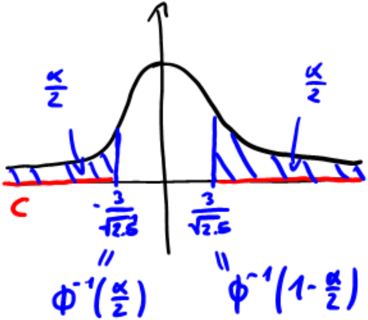
Für die normalverteilte Abfüllmenge $X \sim N_{\mu,25}$ von 500 ml-Flaschen soll basierend auf einer Stichprobe vom Umfang $n = 10$ getestet werden:

$$H_0 : \mu = 500 \quad \text{gegen} \quad H_1 : \mu \neq 500$$

Der kritische Bereich C für die Testgröße $TG = \bar{X}$ sei $C = \mathbb{R} \setminus [497, 503]$.

- ~~Wahrscheinlichkeit für den~~ Fehler 1. Art:
  Unter H_0 gilt: $\bar{X} \sim N_{500,\frac{25}{10}}$ bzw. $\frac{\bar{X}-500}{\sqrt{2.5}} \sim N_{0,1}$ und damit

  $P(\bar{X} \in C) = P\left(\frac{\bar{X}-500}{\sqrt{2.5}} < -\frac{3}{\sqrt{2.5}}\right) + P\left(\frac{\bar{X}-500}{\sqrt{2.5}} > \frac{3}{\sqrt{2.5}}\right) =$
  $2\Phi(-\frac{3}{\sqrt{2.5}}) = 0.058 = \alpha$ Wahrscheinlichkeit, dass H_0 abgelehnt wird, obwohl richtig
- ~~Wahrscheinlichkeit für den~~ Fehler 2. Art:
  Lässt sich unter H_1 nur für einen festen Wert von μ berechnen, z. B. für $\mu = 504$



Berechnung des Fehlers 2. Art (β) für $\mu = 504$:

$$\beta_{504} = P_{\mu=504}(\bar{X} \in \bar{C}) = P_{\mu=504}(497 \le \bar{X} \le 503) =$$

$$= P_{\mu=504}\left(\frac{497-504}{\sqrt{2.5}} \le \underbrace{\frac{\bar{X}-504}{\sqrt{2.5}}}_{\sim N_{0,1}} \le \frac{503-504}{\sqrt{2.5}}\right) = \Phi\left(-\frac{1}{\sqrt{2.5}}\right) - \Phi\left(-\frac{7}{\sqrt{2.5}}\right)$$

$$\approx 26.4\,\%$$

Für $\mu = 501$: $\beta_{501} \approx 89\%$

(Der Fehler 2. Art ist umso größer, je näher μ an μ_0 (hier 500) liegt.)

7.3 Klassische Parametertests

Testprobleme:

- Zweiseitiger Test:
  $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$
- Einseitige Tests:
  $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$ bzw.
  $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$

Testentscheidung basierend auf Stichprobe $\{x_1, \ldots, x_n\}$:

- H_0 wird abgelehnt, falls $tg = TG(x_1, \ldots, x_n) \in C$
- H_0 kann nicht abgelehnt werden, falls $tg = TG(x_1, \ldots, x_n) \in \bar{C}$

Der kritische Bereich ergibt sich analog zu den Konfidenzintervallen durch die Vorgabe eines kleinen Signifikanzniveaus α, d. h. der maximalen Wahrscheinlichkeit für einen Fehler 1. Art. Mit der standardisierten Testgröße TG^* gilt:

$P(TG \in C) < \alpha \iff TG^* \in]-\infty; \Phi^{-1}(\frac{\alpha}{2})[\cup]\Phi^{-1}(1-\frac{\alpha}{2}); \infty[$

$$P(TG \in \bar{C}) \geq 1-\alpha \quad \Longleftrightarrow \quad TG^* \in [\Phi^{-1}\left(\frac{\alpha}{2}\right), \Phi^{-1}\left(1-\frac{\alpha}{2}\right)]$$

Wird dann H_0 verworfen,so spricht man von einer **signifikanten** Schlußfolgerung.
Kann allerdings H_0 nicht verworfen werden, dann lässt sich keine Aussage über den Fehler 2. Art treffen und man spricht von einer **schwachen** Schlußfolgerung.

Je größer α, desto kleiner ist der Annahmebereich.

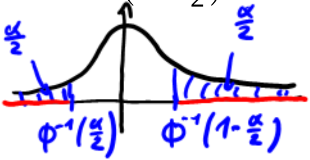
7.3.1 Gauß-Test

Test für Erwartungswert einer Normalverteilung bei bekannter Varianz σ_0^2

(1) Zweiseitiger Gauß-Test $\quad H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$

$$\bar{X} \sim N_{\mu_0, \sigma_0^2/n} \quad \Longrightarrow \frac{\bar{X}-\mu_0}{\sigma_0}\sqrt{n} \sim N_{0,1}$$

$$P_{\mu_0}(\bar{X} \in C) \leq \alpha \quad \Longleftrightarrow \quad |TG| = \frac{|\bar{X}-\mu_0|}{\sigma_0}\sqrt{n} > \overbrace{\Phi^{-1}\left(1-\frac{\alpha}{2}\right)}^{= -\Phi^{-1}\left(\frac{\alpha}{2}\right)}$$



Testentscheidung:

- H_0 wird abgelehnt, falls $|TG| > \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$
- H_0 kann nicht abgelehnt werden, falls $|TG| \leq \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$

Beispiel 7.3.1:
Das Verpackungsgewicht von Schokoladetafeln sei normalverteilt mit unbekanntem EW μ und Standardabweichung $\sigma_0 = 4$ [g]. Es wird vermutet, dass die Verpackungsanlage nicht richtig funktioniert und deshalb der EW für das Gewicht einer Tafel $\neq 100$ [g] ist.
Lässt sich anhand einer Stichprobe vom Umfang $n = 25$ mit $\bar{x} = 98$ [g] die Hypothese $\mu = \underbrace{100}_{\mu_0}$ [g] zum Signifikanzniveau $\alpha = 5\%$ signifikant widerlegen?

Testproblem: $H_0: \mu = 100$ gegen $H_1: \mu \neq 100$

Standardisierte Testgröße: $TG = \frac{\bar{x} - \mu_0}{\sigma_0}\sqrt{n} = \frac{98-100}{4} \cdot 5 = -2.5$

H_0 kann abgelehnt werden, wenn $|TG| = 2.5 > \Phi^{-1}(\underbrace{1-\frac{\alpha}{2}}_{0.975}) \approx 1.96$

Das ist der Fall, d.h. H_0 kann signifikant widerlegt werden.

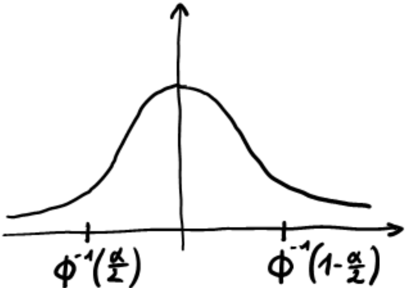
Annahmebereich $\bar{C}$ für $\alpha = 5\%$: $[-1.96; 1.96]$

für $\alpha = 1\%$: $[-\Phi^{-1}(0.995); \underbrace{\Phi^{-1}(0.995)}_{2.58}]$

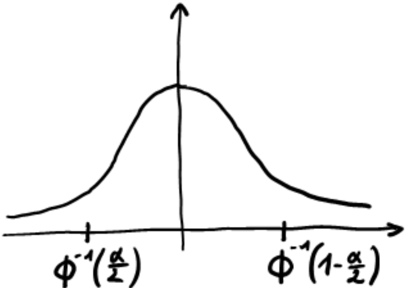
Welche Testentscheidung wird für $\alpha = 1\%$ getroffen?

Jetzt kann H_0 nicht verworfen werden.

Signifikanzniveau



Annahmebereich



Nullhypothese

Fehler 2. Art

bei bekannter Varianz

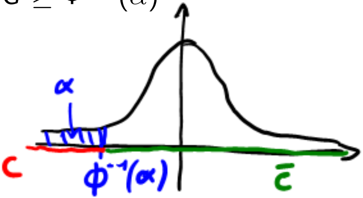
(2) Einseitiger Gauß-Test $\boxed{H_0 : \mu \geq \mu_0 \text{ gegen } H_1 : \mu < \mu_0}$

bzw. $\mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$

$$\underbrace{P_{\mu_0}(\bar{X} \in C)}_{\text{Fehler 1. Art}} \leq \alpha \quad \Longleftrightarrow \quad TG = \frac{\bar{X} - \mu_0}{\sigma_0}\sqrt{n} < \Phi^{-1}(\alpha)$$

Testentscheidung:

- H_0 wird abgelehnt, falls $TG < \Phi^{-1}(\alpha)$
- H_0 kann nicht abgelehnt werden, falls $TG \geq \Phi^{-1}(\alpha)$



Beispiel 7.3.2:

Die Anzahl der Personen in einer U-Bahn, die nominell 175 Personen fasst, sei normalverteilt. Die aus früheren Messungen bekannte Varianz der Anzahl der Fahrgäste beträgt 225. $= \sigma_0^2$

Es wird vermutet, dass die U-Bahnen teilweise überladen fahren. Deshalb soll in den Stoßzeiten eine Stichprobe durchgeführt werden, die die Nullhypothese testet, dass die Bahnen tatsächlich überfüllt sind.

Mit einem Stichprobenumfang von $n = 30$ wurde ein Mittelwert $\overline{x} = 172$ errechnet.

a) Es soll ein Konfidenzintervall für den Erwartungswert μ zum Konfidenzniveau $1-\alpha =$ 92 % geschätzt werden.
(Hinweis: `qnorm(0.96,0,1)` ≈ 1.7507)

b) Formulieren Sie das Testproblem, um die Vermutung, dass die Bahnen überfüllt sind, signifikant zu widerlegen. Zu welcher Entscheidung kommen Sie auf einem Signifikanzniveau von $\alpha = 0.04$?

c) Wie wäre die Entscheidung in Teilaufgabe b) ausgefallen, wenn die Stichprobe einen Mittelwert $\overline{x} = 170$ ergeben hätte?

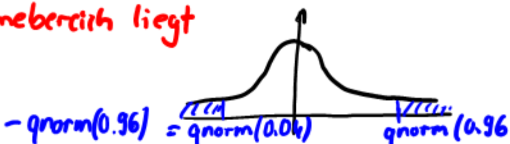
a) $I = \left]\, \bar{x} - \underbrace{\Phi^{-1}\left(1-\frac{\alpha}{2}\right)}_{qnorm(0.96) \approx 1.75} \cdot \frac{\sigma_0}{\sqrt{n}} \,,\, \bar{x} + \Phi^{-1}\left(1-\frac{\alpha}{2}\right) \cdot \frac{\sigma_0}{\sqrt{n}} \,\right[$

$I = \left]\, 167.2 \,,\, 176.8 \,\right[$

b) $\boxed{H_0 : \mu > \underbrace{175}_{\mu_0} \quad \text{gegen} \quad H_1 : \mu \leq 175}$ Testproblem

$TG = \frac{\bar{x} - 175}{15} \sqrt{30} \approx -1.095 \;\underset{>}{\cancel{\overset{?}{<}}}\; \underbrace{\Phi^{-1}(0.04)}_{qnorm(0.04)} \approx -1.75$

$\Rightarrow$ H_0 kann nicht verworfen werden, da TG im Annahmebereich liegt



c) Für $\bar{x} = 170$:

$$TG = \frac{170 - 175}{15}\sqrt{30} \approx -1.83 < \phi^{-1}(0.04) \approx -1.75$$

$\Longrightarrow$ H_0 wird verworfen

Varianten des Gauß-Tests: Testgröße $tg = \frac{\bar{X} - \mu_0}{\sigma_0} \sqrt{n}$

bekannte Standardabweichung

tg liegt in C

H_0	H_1	H_0 ablehnen, falls	p-Wert
zweiseitig $\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert tg \rvert > \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$	$2(1 - \Phi(\lvert tg \rvert))$
einseitig $\mu \leq \mu_0$	$\mu > \mu_0$	$tg > \Phi^{-1}(1 - \alpha)$	$1 - \Phi(tg)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$tg < \Phi^{-1}(\alpha)$	$\Phi(tg)$

7.3.2 t-Test

Test für Erwartungswert einer Normalverteilung bei unbekannter Varianz

Testgröße $tg = \frac{\bar{X}-\mu_0}{S}\sqrt{n} \sim t_{n-1}$

Stichprobenstandardabweichung

Varianten:

H_0	H_1	H_0 ablehnen, falls	p-Wert
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert tg\rvert > t_{n-1}^{-1}(1-\frac{\alpha}{2})$	$2(1-t_{n-1}(\lvert tg\rvert))$
$\mu \leq \mu_0$	$\mu > \mu_0$	$tg > t_{n-1}^{-1}(1-\alpha)$	$1-t_{n-1}(tg)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$tg < t_{n-1}^{-1}(\alpha)$	$t_{n-1}(tg)$

erhält man aus Testentscheidung durch Auflösen nach α

7.4 p-Wert

p-Wert: "beobachtetes Signifikanzniveau" bzw. Fehler 1. Art, wenn tg Grenze des kritischen Bereichs

Wahrscheinlichkeit, bei Zutreffen von H_0 den beobachteten Wert tg der Testgröße oder einen noch stärker von μ_0 abweichenden Wert zu bekommen.

Falls $\alpha < $ p-Wert, dann kann H_0 nicht verworfen werden

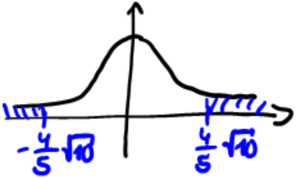
Falls p-Wert $< \alpha$, dann wird H_0 verworfen

Beispiel:

Berechnung des p-Werts für das Testproblem zur Abfüllmenge von 500 ml Flaschen für $n = 10$ und $\bar{x} = 504$ $(\sigma = 5)$

$$\text{p-Wert}: P\left(\frac{|\bar{x} - 500|}{5}\sqrt{10} > \frac{4}{5}\sqrt{10}\right) = 2 \cdot \text{pnorm}\left(-\frac{4}{5}\sqrt{10}\right)$$

$$\approx 0.0114$$



Für $\alpha = 5\%$ bzw. 10% kann H_0 verworfen werden, für $\alpha = 1\%$ nicht.

Zusammenhang p-Wert – Hypothesentest

Der p-Wert zu einer Hypothese H_0 ist der größte Wert von α, für den H_0 nicht abgelehnt wird.

Vorteil:
Anhand des p-Werts kann man für beliebige Werte von α eine Testentscheidung treffen.

R-Funktion: t.test(x, mu = μ_0, conf.level = $1-\alpha$) Default: zweiseitiger Test

Stichprobe

Bemerkung:

Falls p-Wert < 1% : sehr hohe Signifikanz

1% ≤ p-Wert < 5% : hohe Signifikanz

5% ≤ p-Wert ≤ 10% : Signifikanz

10% < p-Wert : keine Signifikanz

Zusammenhang Konfidenzintervall I – Hypothesentest

Testentscheidung: !!! Gilt nur bei zweiseitigem Test !!!

- H_0 wird abgelehnt, falls $\mu_0 \notin I$
- H_0 kann nicht abgelehnt werden, falls $\mu_0 \in I$

(I: Konfidenzintervall zum Konfidenzniveau $1-\alpha$ für unbekannten Erwartungswert μ)

Zusammenhang: I ist der Annahmebereich für $H_0: \mu = \mu_0$ gegen $H_1: \mu \neq \mu_0$ zum Signifikanzniveau α