6. Hypothesentests

$\hookrightarrow$ Bisher behandelte statistische Verfahren

- Punktschätzer: Aus Stichprobe wird „gute“ Schätzung für den Wert einer Kennzahl / eines Parameters einer Grundgesamtheit gewonnen.

Covid-19: theor. Reproduktionsfaktor ρ in Bevölkerung unbekannt, bestimme eine Schätzung R anhand der Infektionszahlen (zufallsabhängig, tagweise Änderung)

- Intervallschätzer: Zu „gutem“ Schätzer und Überdeckungs-WS $1-\alpha$ werden „Toleranzen nach unten und oben“ angegeben.

Covid-19: Angabe eines „möglichst schmalen“ KI $[R-\Delta, R+\Delta]$ mit 99% Überdeckungswahrscheinlichkeit für ρ.

$\hookrightarrow$ Jetzt Fragestellung: Erfüllt die Grundgesamtheit eine bestimmte Hypothese?

Covid-19: Liegt der (theor.) Reproduktionsfaktor unter 1 (unter 0,75,...)?

Als Verfahren zur Beantwortung solcher Fragen werden statistische Tests eingesetzt. Die Entscheidungen können - mit Bezug auf den wahren Parameter - falsch sein, wobei zwei mögliche Fehler (1./2. Art) auftreten können. Ein Test sollte wenigstens den Fehler 1. Art „unter Kontrolle haben“.

6.1 Testen von Hypothesen

$\hookrightarrow$ Anhand Stichproben: Prüfung von Annahmen über eine Grundgesamtheit

$\hookrightarrow$ Formulierung der Annahmen mit Bezug auf

- Verteilungsparameter
- Verteilungskennzahlen (Erwartungswert, ...)

als komplementäres Hypothesenpaar: Nullhypothese H_0/Gegenhypothese H_1.

$\hookrightarrow$ H_0 wird verworfen (statistisch widerlegt), wenn das Stichprobenergebnis „deutlich gegen H_0 spricht".

- Verwerfung bedeutet nicht, dass H_0 falsch bzw. H_1 wahr ist.
- Nichtverwerfung bedeutet nicht, dass H_0 wahr bzw. H_1 falsch ist.

$\hookrightarrow$ Verfahren werden **Hypothesentests** oder **statistische Tests** genannt.

$\hookrightarrow$ Unterschieden werden Parametertests und Verteilungstests.

$\hookrightarrow$ Tests sind nach einem einheitlichen Schema aufgebaut.

Bsp.: Verkauf von Kartoffeln in 10kg-Säcken. Wird das Sollgewicht eingehalten?
Stichprobe: $\bar{x} = 9.85$ ($n = 100$). Modell: $X_1, \ldots, X_{100}$ u.i.v. $\sim \mathcal{N}(\mu, 4)$, μ unbek.

Generelles Testschema

1. Festlegung von
   - **Nullhypothese** H_0 und **Gegen- bzw. Alternativhypothese** H_1

     $$H_0 : \mu = 10 \qquad \text{vs.} \qquad H_1 : \mu \neq 10$$
   - **Signifikanzniveau** α (= WS H_0 irrtümlich abzulehnen)

     $\alpha = 0.05$ (H_0 wird mit WS 5% irrtümlich verworfen)
2. Herleitung einer **Teststatistik** $V = V(X_1, \ldots, X_n)$ und Verteilung F_V für H_0

   $$V = \sqrt{100} \cdot \frac{\bar{X} - 10}{\sqrt{4}} \quad \sim \quad F_V = \mathcal{N}(0,1) \text{ für } \mu = 10$$
3. Aufstellen einer **Entscheidungsregel** anhand der statistischen Verteilung F_V und Festlegung des **kritischen Bereiches** K mit Ablehnung von H_0 für $V \in K$.

   H_0 verwerfen, wenn $|V| > z_{1-\alpha/2} = 1.96$ (0.975-Quantil von $\mathcal{N}(0,1)$)
4. **Entscheidung** anhand der berechneten Teststatistik $V = v$

   Aus $\bar{x} = 9.85$ folgt $|v| = \left|10 \cdot \frac{9.85-10}{2}\right| = 0.75$

   $|v| \leq 1.96 \Rightarrow H_0$ kann nicht verworfen werden.

Hypothesen (über die Grundgesamtheit)

$\hookrightarrow$ **Verteilungshypothesen:** Annahme eines bestimmten Verteilungstyps (oder einer speziellen Verteilung), auch: einer bestimmten Verteilungseigenschaft.

Bsp.: Normalverteilungsannahme $H_0 : F = \mathcal{N}$ vs. $H_1 : F \neq \mathcal{N}$

$\hookrightarrow$ **Parameterhypothesen:** beziehen sich auf einen bestimmten Parameter

$H_0 : \theta = \theta_0$	vs.	$H_1 : \theta \neq \theta_0$	Punkthypothese, zweiseitig
$H_0 : \theta \geq \theta_0$	vs.	$H_1 : \theta < \theta_0$	Bereichshypothese, rechtsseitig
$H_0 : \theta \leq \theta_0$	vs.	$H_1 : \theta > \theta_0$	Bereichshypothese, linksseitig

Beispiele für θ: Anteilswert, Erwartungswert, Varianz, ...

$\hookrightarrow$ **Beachte:**

- Sprechen in den Daten „genügend Indizien“ gegen H_0 , so wird H_0 verworfen und man entscheidet sich aktiv für H_1 („H_0 wird verworfen“)
- H_0 kann *nie (aktiv) akzeptiert*, nur mangels genügend Informationen nicht wiederlegt werden („H_0 wird beibehalten bzw. nicht verworfen“)
- Formulierung des „intendierten“ Sachverhaltes (sofern möglich) als H_1

Fehler und Signifikanzniveau eines Tests

↪ **Fehler 1. und 2. Art:**

	Testentscheidung	
Wahrer Zustand	Nichtablehnung von H_0	Ablehnung von H_0
H_0 korrekt	richtig	α-Fehler (Fehler 1. Art)
H_0 falsch	β-Fehler (Fehler 2. Art)	richtig

↪ **Signifikanzniveau α:**

- vorgegebener Maximalwert für die Wahrscheinlichkeit des Fehlers 1. Art (d.h. W.-keit H_0 abzulehnen, obwohl H_0 korrekt ist)
- muss vor der Anwendung des Tests festgelegt werden
- üblich sind $\alpha = 0.01$, $\alpha = 0.05$, $\alpha = 0.1$

Testbestimmung: Konfliktäre Ziele

- $\hookrightarrow$ Idealerweise würde man die WS des α- und β-Fehler simultan minimieren
- $\hookrightarrow$ Problem: $\alpha \downarrow \quad \Rightarrow \quad \beta \uparrow$
- $\hookrightarrow$ Wie groß sind die WS der Fehler 1./2. Art, wenn $\underline{H_0 \text{ stets abgelehnt}}$ wird?

$$P(\alpha\text{-Fehler}) = P(\text{Ablehnung} \mid H_0 \text{ korrekt}) = 1$$

$$P(\beta\text{-Fehler}) = P(\text{Nichtablehnung} \mid H_0 \text{ falsch}) = 0$$

- $\hookrightarrow$ Wie groß sind die WS der Fehler 1./2. Art, wenn $\underline{H_0 \text{ stets angenommen}}$ wird?

$$P(\alpha\text{-Fehler}) = P(\text{Ablehnung} \mid H_0 \text{ korrekt}) = 0$$

$$P(\beta\text{-Fehler}) = P(\text{Nichtablehnung} \mid H_0 \text{ falsch}) = 1$$

- $\hookrightarrow$ Ausweg: Wähle für fixes α den Test mit niedrigstem β (mächtigster Test).

Teststatistik V und ihre Verteilung F_V

- ↪ Aggregation der Stichprobenwerte zu einer Teststatistik $V = V(X)$, anhand derer die Testentscheidung (H_0 beibehalten oder verwerfen) getroffen wird
- ↪ V ist charakteristisches Merkmal eines statistischen Tests und ergibt sich oft
  - □ aus einer Schätzfunktion bei Parameterhypothesen
  - □ aus Gegenüberstellung von empirischen und theoretischen Kennzahlen.
- ↪ Als Aggregation von Zufallsvariablen ist die Teststatistik ebenfalls eine Zufallsvariable (und folgt einer Verteilung F_V)

Beispiel

- □ Nullhypothese H_0: $\mu = \mu_0$
- □ Stichprobe: $X_1, \ldots, X_n$ mit $X_i \sim \mathcal{N}(\mu_0, \sigma^2)$
- □ Teststatistik: $V = \sqrt{n} \cdot \frac{\bar{X} - \mu_0}{\sigma} \sim \mathcal{N}(0, 1)$

  $\Rightarrow$ zugehörige Verteilung der Teststatistik V: $F_V = \mathcal{N}(0, 1)$

Teststatistik V und ihre Verteilung F_V

$\hookrightarrow$ Über die statistische Verteilung F_V wird die Einhaltung des Signifikanzniveaus α geprüft.

$\hookrightarrow$ Verteilung F_V hängt von der Verteilung der Grundgesamtheit, insbesondere von der Korrektheit von H_0, ab

$\hookrightarrow$ Bei Korrektheit von H_0 muss F_V (mindestens approximativ) bestimmbar sein.

Entscheidungsregel (I)

$\hookrightarrow$ anhand einer konkreten Stichprobe x wird die Realisation $v = v(x)$ bestimmt

$\hookrightarrow$ Bestimmung eines **kritischen Bereiches** K: Ablehnung von H_0 für $v \in K$

H_0 wird abgelehnt, falls $|v| > z_{1-\alpha/2}$, d.h. $K = (-\infty, z_{\alpha/2}) \cup (z_{1-\alpha/2}, \infty)$

- $\hookrightarrow$ Sicherstellung, dass $P(\text{Fehler 1. Art}) \leq \alpha$.
- $\hookrightarrow$ endgültige Festlegung von K so, dass $P(\text{Fehler 2. Art})$ minimal ist.
- $\hookrightarrow$ oft (z.B. stetiges F_V) realisiert durch $P(\text{Fehler 1. Art}) = \alpha$ für „ungünstigsten Fall der Hypothese“.

Faustregel zur Festlegung des Hypothesenpaars

Statistische Tests können nur zur Widerlegung von H_0 verwendet werden – nicht zur Bestätigung.

$\rightsquigarrow$ Die zu bestätigende Hypothese muss als Gegenhypothese formuliert werden.

Entscheidungsregel (II)

$v \in K$:

$\hookrightarrow$ Stichprobe steht in signifikantem Widerspruch zu H_0

$\hookrightarrow$ H_0 **kann abgelehnt werden**

$\hookrightarrow$ H_1 ist statistisch bestätigt mit Signifikanzniveau α.

$v \notin K$:

$\hookrightarrow$ Stichprobe steht nicht in signifikantem Widerspruch zu H_0

$\hookrightarrow$ H_0 **kann nicht abgelehnt werden** anhand der konkreten Stichprobe

$\hookrightarrow$ H_0 ist dadurch nicht bestätigt!

Klassifizierungen von Tests anhand der Anzahl einbezogener Stichproben

Einstichprobentest

$\hookrightarrow$ Grundlage bildet eine u.i.v. Stichprobe $X_1, \ldots, X_n$ zu einer Grundgesamtheit

$\hookrightarrow$ Bsp: Tests für den Anteilswert p oder den Erwartungswert μ

Zweistichprobentest

$\hookrightarrow$ für den Vergleich verschiedener Grundgesamtheiten

Unterscheidet sich das mittlere Einkommen der 20- bis 30-Jährigen signifikant vom mittleren Einkommen der 30- bis 40-Jährigen?

$\hookrightarrow$ Stichproben $X_{11}, \ldots, X_{1n_1}$ und $X_{21}, \ldots, X_{2n_2}$, gezogen aus 2 Grundgesamtheiten

$\hookrightarrow$ Annahme: unabhängige Stichproben

$\hookrightarrow$ Bsp: Tests für die Differenz zweier Erwartungswerte oder zweier Anteilswerte

Spezialfall: Zweistichprobentest für verbundene Stichproben

- ↪ Ziehen einer zweidimensionalen Stichprobe aus einer einzelnen Grundgesamtheit.
- ↪ pro Merkmalsträger werden mehrere Merkmale erhoben
- ↪ Bsp: Differenzentests für das arithmetische Mittel oder Korrelations- oder Kontingenztest

I) Besteht eine Abhängigkeit in den Leistungen der Studierenden in den Fächern Mathematik und Physik?

II) Blutdruck von n Patienten vor und nach der Einnahme eines blutdrucksenkenden Medikamentes. Hat das Medikament einen Einfluss?

III) Einkommen von n Erwerbslosen vor und nach einer Arbeitsbeschaffungsmaßnahme. Hat diese einen Einfluss?

Gemeinsame Struktur aller behandelten Tests

Abhängig von Hypothese (und Signifikanzniveau α) haben sämtliche Tests folgende Struktur mit spezifischer Teststatistik $V = V(X)$:

	Nullhypothese H_0	**kritischer Bereich**	p-**value**
a)	zweiseitig: $\theta = \theta_0$	$V \notin \left[q_{\alpha/2}; q_{1-\alpha/2}\right]$	$2 \cdot \min(F(v), 1 - F(v))$
	F symmetrisch	$\lvert V\rvert > q_{1-\alpha/2}$	$2 \cdot (1 - F(\lvert v\rvert)$
b)	rechtsseitig: $\theta \geq \theta_0$	$V < q_\alpha$	$F(v)$
c)	linksseitig: $\theta \leq \theta_0$	$V > q_{1-\alpha}$	$1 - F(v)$

$\hookrightarrow$ Dabei ist q_α das α-Quantil der Stichprobenverteilung $F = F_V$ von V

$\hookrightarrow$ F und q_α werden i.d.R. auch bei b) und c) für $\theta = \theta_0$ bestimmt, weil die WS eines Fehlers 1. Art dann besonders hoch ist.

$\hookrightarrow$ p-value: für alternative Darstellung der Test-Entscheidung (s.u.)

6.2 Einstichproben-Erwartungswert-Tests

Konstruktionsschema (s.o.)

Zunächst werden die Modellannahmen weiter spezifiziert, danach:

1. Aufstellen der **Hypothesen** H_0 und H_1, sowie des Signifikanzniveaus α.
2. Bestimmung der **Teststatistik** $V = V(X_1, \ldots, X_n)$ mitsamt zugehöriger Stichprobenverteilung F_V
3. Aufstellen der **Entscheidungsregel** bzgl. der Ablehnung von H_0

Anhand der konkreten Stichprobe $x_1, \ldots, x_n$ wird dann mit dem berechneten Wert $v(x_1, \ldots, x_n)$ der Teststatistik $V(X_1, \ldots, X_n)$ die Entscheidung (H_0 beibehalten oder verwerfen) getroffen.

Einstichprobentests für den Erwartungswert

↪ Stichprobe $X_1, \ldots, X_n$ von u.i.v. ZVen mit existierendem (unbekanntem) Erwartungswert μ und existierender (bekannter/unbekannter) Varianz σ^2

↪ Die Vorgehensweise ist jeweils ähnlich, unterscheidet sich im Detail nach

- □ Verteilung der Grundgesamtheit: $X_i \sim \mathcal{N}(\mu, \sigma^2)$ bzw. $\mathcal{L}(X)$ unspezifiziert
- □ Varianz: bekannt bzw. (in der Praxis meist) unbekannt

1. Hypothesen aufstellen

Wähle aus den folgenden Möglichkeiten das zum Fall passende Hypothesenpaar:

a) $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$

b) $H_0 : \mu \geq \mu_0$ vs. $H_1 : \mu < \mu_0$

c) $H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$

Einkommen von niedergelassenen Ärzten

Köln – *Niedergelassene Ärzte haben die Erträge ihrer Praxen zwischen 2007 und 2011 deutlich steigern können. Das geht aus Erhebungen des Statistischen Bundesamts hervor. Danach stieg der Reinertrag je Arztpraxis, der dem Überschuss nach Abzug der Praxisaufwendungen vor Steuern und sonstigen Abgaben entspricht, in diesen vier Jahren um 21 Prozent auf durchschnittlich 234.000 Euro. Nach Angaben des Statistischen Bundesamts gab es bei den Arzteinkommen je nach Fachrichtung große Unterschiede. Allgemeinarztpraxen verbuchten 2011 Reinerträge von durchschnittlich 181.000 Euro, Kinderarztpraxen brachten es auf 191.000 Euro, Orthopädie-Praxen auf 293.000 Euro und Praxen für Haut- und Geschlechtskrankheiten auf 249.000 Euro.* *Quelle:* `https://www.aerzteblatt.de/nachrichten/55525/`

Über alle Sparten hinweg beträgt demnach das mittlere Einkommen[a] 181 T€ oder mehr. Gilt dies auch für Ärzte in unserer Region? D.h. Hypothesenpaar:

$$H_0 : \mu \geq 181 \quad \text{vs.} \quad H_1 : \mu < 181$$

[a] Dass die Angaben (näherungsweise) Grundgesamtheitswerte sind, ist allerdings kritisch, da scheinbar nur freiwillige Angaben von ca. 4000 (6%) der deutschen Arztpraxen zugrunde lagen.

Festlegung: „Grenzfall“ $\mu = \mu_0$ (hier $\mu = 181$) gehört zu H_0, nicht zu H_1.

Fiktives Datenbeispiel: In einer Stichprobe von $n = 200$ Arztpraxen unserer Region wurde ein Durchschnittseinkommen von $\bar{X} = 168$ T€ ermittelt.

Durchschnittliche Praktikumsgehälter, Clevis Praktikantenspiegel 2016

Öffentlicher Sektor	1194	Pharma	1280
TK, IT und Internet	1216	Baugewerbe, -industrie	1321
Transport, Logistik	1128	Personaldienstleistungen	1334
Medien, Unterhaltung	1231	Konsum-, Gebrauchsgüter	1380
Finanzen, Banken, Versicherungen	1260	Consulting, WP & Recht	1383

Werte gerundet, Quelle: `http://www.clevis.de/CLEVIS_Praktikantenspiegel/2016.pdf`

Das Studienergebnis soll für Praktikanten in der **Versicherungsbranche** geprüft werden, d.h. Hypothesen[a]:

$$H_0 : \mu = 1260 \quad \text{vs.} \quad H_1 : \mu \neq 1260$$

[a] Auch hier ist die Annahme, dass $\mu = 1260$ eine Angabe zur Grundgesamtheit ist, kritisch zu beurteilen.

Fiktives Datenbeispiel: Von $n = 40$ Studierenden, die 2015 ein Praktikum in der Versicherungsbranche absolvierten, wurde das Praktikumsgehalt erfragt:
1346,1187,1353,1347,1261,1066,1127,1191,1219,1460,1296,1140,1105,1191,1190,
1179,1245,1131,1264,1096,1198,1258,1233,1300,1214,1270,1329,1151,1092,1225,
1196,1166,1177,1155,1293,1335,1319,1177,1344,1192 Mittelwert: 1225,45 €

2. Bestimmung der Teststatistik V und zugehöriger Verteilung F_V

V und F_V hängen davon ab, ob σ^2 bekannt ist und welches WS-Modell für die u.i.v.-Stichprobe vorliegt.

Fall	σ^2	V	Modell für X_i	Verteilung von V	Testbezeichnung
i)	bekannt	$\sqrt{n} \cdot \frac{\bar{X} - \mu_0}{\sigma}$	Normalvtlg.	$\mathcal{N}(0, 1)$	Gaußtest
ii)			beliebig	approx. $\mathcal{N}(0, 1)$	approx. Gaußtest
iii)	unbekannt	$\sqrt{n} \cdot \frac{\bar{X} - \mu_0}{S}$	Normalvtlg.	t_{n-1}	t-Test
iv)			beliebig	approx. $\mathcal{N}(0, 1)$	approx. Gaußtest

$\hookrightarrow$ Dabei $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

$\hookrightarrow$ in R: Gaußtest: `compositions::Gauss.test(...)`, t-Test: `ttest(...)`

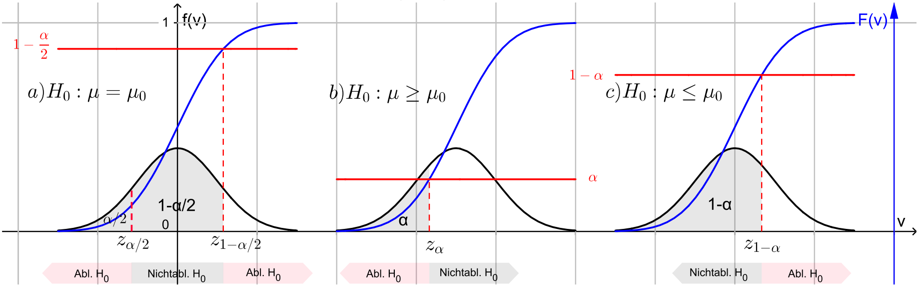
Ärzteeinkommen ($n = 200, \mu_0 = 181, \bar{X} = 168$): Bestimmen Sie den Wert v der Teststatistik $V = \sqrt{n} \cdot \frac{\bar{X}-\mu_0}{\sigma}$ die Verteilung von V für $\mu = \mu_0$ ($\sigma = 100$ bekannt)

$v = \sqrt{200}\frac{168-181}{100} = -1.838478$, V ist für $\mu = 181$ $\mathcal{N}(0,1)$-verteilt.

Praktikumsbeispiel ($n = 40, \mu_0 = 1260, \bar{X} = 1225,45$):
Bestimmen Sie $V = \sqrt{n} \cdot \frac{\bar{X}-\mu_0}{\sigma}$ (Ann.: $\sigma = 120$ bekannt) bzw. $V = \sqrt{n} \cdot \frac{\bar{X}-\mu_0}{S}$ (Ann.: σ unbekannt) und jeweils die Verteilung von V unter $\mu = \mu_0$.

Varianz bekannt: $v = \sqrt{40}\frac{1225.45-1260}{120} = -1.820945$
V ist für $\mu = 1260$ $\mathcal{N}(0,1)$-verteilt.
Varianz unbekannt: $S = 87.23354$ (z.B. mit Taschenrechner),
damit $v = \sqrt{40}\frac{1225.45-1260}{87.23354} = -2.504924$
V ist für $\mu = 1260$ $t(39)$-verteilt.

Bezeichne z_α das α-Quantil von $\mathcal{N}(0,1)$. In R: `qnorm(alpha)`



https://ggbm.at/u3kbgjby

3. Entscheidungsregel für die Fälle i), ii) und iv)

a)	$H_0 : \mu = \mu_0$	vs.	$H_1 : \mu \neq \mu_0$:	Ablehnung von H_0, falls	$\lvert v \rvert > z_{1-\alpha/2}$
b)	$H_0 : \mu \geq \mu_0$	vs.	$H_1 : \mu < \mu_0$:	Ablehnung von H_0, falls	$v < z_\alpha$
c)	$H_0 : \mu \leq \mu_0$	vs.	$H_1 : \mu > \mu_0$:	Ablehnung von H_0, falls	$v > z_{1-\alpha}$

Durchschnittseinkommen von niedergelassenen Ärzten

Stichprobe von $n = 200$ Arztpraxen, $\bar{X} = 168$T€. Laut einer statistischen Bundesamt liegt das „wahre Durchschnittseinkommen" bei mindestens 181T€.
Wird das „wahre" Durchschnittseinkommen (bei einem Signifikanzniveau von 5%) durch die Stichprobe gestützt?
Annahme: $X \sim \mathcal{N}(\mu, \sigma^2), \sigma = 100$ bekannt.

$\hookrightarrow$ Hypothesen: $H_0 : \mu \geq 181$, vs. $H_1 : \mu < 181$

$\hookrightarrow$ Teststatistik: $v = \sqrt{n} \cdot \frac{\bar{x} - \mu_0}{\sigma} \approx -1.83$

Ermitteln Sie die Testentscheidung.

Ablehnungsbereich (Fall (b)) ist $v < z_\alpha$ mit $\alpha = 0.05$
Testentscheidung : $v = -1.83, z_{0.05} = -1.64$, $v < -z_{0.05}$, d.h. H_0 wird abgelehnt.
Das Einkommen von Ärzten in unserer Region ist signifikant niedriger als im Bundesdurchschnitt.

Tabelle der α-Quantile der Standardnormalverteilung:

α	0.001	0.005	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990	0.995	0.999
z_α	-3.090	-2.576	-2.326	-1.960	-1.645	-1.282	1.282	1.645	1.960	2.326	2.576	3.090

Durchschnittseinkommen von Praktikanten

Stichprobe von $n = 40$ Studenten mit durchschnittlichem Einkommen $\bar{X} = 1225.45$.
Laut einer Studie liegt das „wahre Durchschnittseinkommen" bei 1260 €.
Wird die Angabe der Studie (bei einem Signifikanzniveau von $\alpha = 5\%$) durch die Stichprobe gestützt?

Annahme: $X \sim \mathcal{N}(\mu, \sigma^2), \sigma = 120$ als bekannt angenommen.

Wie lautet, ausgehend von Ihren bisherigen Berechnungen, Ihre Testentscheidung?

Testentscheidung: Zweiseitiges Testproblem $H_0 : \mu = \mu_0 = 1260$

$$v = -1.820945 \quad \Rightarrow \quad |v| = 1.820945$$

$$1 - \frac{\alpha}{2} = 0.975 \quad \Rightarrow \quad z_{0.975} = 1.96$$

$$|v| = 1.820945 \leq 1.96 = z_{0.975} \quad \Rightarrow \quad H_0 \text{ wird nicht abgelehnt}$$

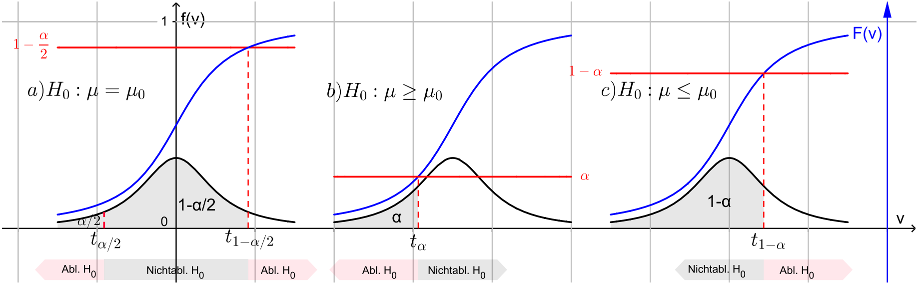
Tabelle der α-Quantile der Standardnormalverteilung:

α	0.001	0.005	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990	0.995	0.999
z_α	-3.090	-2.576	-2.326	-1.960	-1.645	-1.282	1.282	1.645	1.960	2.326	2.576	3.090

Bezeichne $t_\alpha = t_{\alpha;\, n-1}$ das α-Quantil von t_{n-1}. In R: `qt(alpha, df = n - 1)`

3. Entscheidungsregel bei iii), d.h. Normalvtlg. mit unbekanntem σ^2

a) $H_0 : \mu = \mu_0$	vs.	$H_1 : \mu \neq \mu_0$:	Ablehnung von H_0, falls	$\lvert v\rvert > t_{1-\alpha/2}$
b) $H_0 : \mu \geq \mu_0$	vs.	$H_1 : \mu < \mu_0$:	Ablehnung von H_0, falls	$v < t_\alpha$
c) $H_0 : \mu \leq \mu_0$	vs.	$H_1 : \mu > \mu_0$:	Ablehnung von H_0, falls	$v > t_{1-\alpha}$



Für $n > 30$: Approximation der t-Verteilung durch die Standardnormalverteilung.

Durchschnittseinkommen von Praktikanten

Stichprobe von $n = 40$ Studenten mit durchschnittlichem Einkommen $\bar{X} = 1225.45$.
Laut einer Studie liegt das „wahre Durchschnittseinkommen" bei 1260 €.
Wird die Angabe der Studie (bei einem Signifikanzniveau von $\alpha = 5\%$) durch die Stichprobe gestützt?

Annahme: $X \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 unbekannt (wird durch $S^2 = 87.23354^2$ geschätzt).

Wie lautet, ausgehend von Ihren bisherigen Berechnungen, Ihre Testentscheidung?

Testentscheidung: Zweiseitiger t-Test von H_0: $\mu = 1260$

$$v = -2.504924 \quad \Rightarrow \quad |v| = 2.504924$$

$$1 - \frac{\alpha}{2} = 0.975 \quad \Rightarrow \quad t_{0.975,\,40-1} = 2.02$$

$$|v| = 2.504924 > 2.02 = t_{0.975,\,39} \quad \Rightarrow \quad H_0 \text{ wird verworfen}$$

(Tabelle der t-Quantile: siehe nächste Folie)

Tabelle: Quantile von Standardnormal- und t-Verteilung

	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
∞	1.28	1.64	1.96	2.33	2.58	3.09	3.29
1	3.08	6.31	12.71	31.82	63.66	318.31	636.62
2	1.89	2.92	4.30	6.96	9.92	22.33	31.60
3	1.64	2.35	3.18	4.54	5.84	10.21	12.92
4	1.53	2.13	2.78	3.75	4.60	7.17	8.61
5	1.48	2.02	2.57	3.36	4.03	5.89	6.87
6	1.44	1.94	2.45	3.14	3.71	5.21	5.96
7	1.41	1.89	2.36	3.00	3.50	4.79	5.41
8	1.40	1.86	2.31	2.90	3.36	4.50	5.04
9	1.38	1.83	2.26	2.82	3.25	4.30	4.78
10	1.37	1.81	2.23	2.76	3.17	4.14	4.59
11	1.36	1.80	2.20	2.72	3.11	4.02	4.44
12	1.36	1.78	2.18	2.68	3.05	3.93	4.32
13	1.35	1.77	2.16	2.65	3.01	3.85	4.22
14	1.35	1.76	2.14	2.62	2.98	3.79	4.14
15	1.34	1.75	2.13	2.60	2.95	3.73	4.07
16	1.34	1.75	2.12	2.58	2.92	3.69	4.01
17	1.33	1.74	2.11	2.57	2.90	3.65	3.97
18	1.33	1.73	2.10	2.55	2.88	3.61	3.92
19	1.33	1.73	2.09	2.54	2.86	3.58	3.88
20	1.33	1.72	2.09	2.53	2.85	3.55	3.85
21	1.32	1.72	2.08	2.52	2.83	3.53	3.82
22	1.32	1.72	2.07	2.51	2.82	3.50	3.79
23	1.32	1.71	2.07	2.50	2.81	3.48	3.77
24	1.32	1.71	2.06	2.49	2.80	3.47	3.75
25	1.32	1.71	2.06	2.49	2.79	3.45	3.73
26	1.31	1.71	2.06	2.48	2.78	3.43	3.71
27	1.31	1.70	2.05	2.47	2.77	3.42	3.69
28	1.31	1.70	2.05	2.47	2.76	3.41	3.67
29	1.31	1.70	2.05	2.46	2.76	3.40	3.66
30	1.31	1.70	2.04	2.46	2.75	3.39	3.65
31	1.31	1.70	2.04	2.45	2.74	3.37	3.63
32	1.31	1.69	2.04	2.45	2.74	3.37	3.62
33	1.31	1.69	2.03	2.44	2.73	3.36	3.61
34	1.31	1.69	2.03	2.44	2.73	3.35	3.60
35	1.31	1.69	2.03	2.44	2.72	3.34	3.59
36	1.31	1.69	2.03	2.43	2.72	3.33	3.58
37	1.30	1.69	2.03	2.43	2.72	3.33	3.57
38	1.30	1.69	2.02	2.43	2.71	3.32	3.57
39	1.30	1.68	2.02	2.43	2.71	3.31	3.56
40	1.30	1.68	2.02	2.42	2.70	3.31	3.55

	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
41	1.30	1.68	2.02	2.42	2.70	3.30	3.54
43	1.30	1.68	2.02	2.42	2.70	3.29	3.53
44	1.30	1.68	2.02	2.41	2.69	3.29	3.53
45	1.30	1.68	2.01	2.41	2.69	3.28	3.52
46	1.30	1.68	2.01	2.41	2.69	3.28	3.51
47	1.30	1.68	2.01	2.41	2.68	3.27	3.51
49	1.30	1.68	2.01	2.40	2.68	3.27	3.50
50	1.30	1.68	2.01	2.40	2.68	3.26	3.50
51	1.30	1.68	2.01	2.40	2.68	3.26	3.49
52	1.30	1.67	2.01	2.40	2.67	3.25	3.49
53	1.30	1.67	2.01	2.40	2.67	3.25	3.48
54	1.30	1.67	2.00	2.40	2.67	3.25	3.48
56	1.30	1.67	2.00	2.39	2.67	3.24	3.47
57	1.30	1.67	2.00	2.39	2.66	3.24	3.47
59	1.30	1.67	2.00	2.39	2.66	3.23	3.46
62	1.30	1.67	2.00	2.39	2.66	3.23	3.45
63	1.30	1.67	2.00	2.39	2.66	3.22	3.45
64	1.29	1.67	2.00	2.39	2.65	3.22	3.45
66	1.29	1.67	2.00	2.38	2.65	3.22	3.44
68	1.29	1.67	2.00	2.38	2.65	3.21	3.44
69	1.29	1.67	1.99	2.38	2.65	3.21	3.44
71	1.29	1.67	1.99	2.38	2.65	3.21	3.43
73	1.29	1.67	1.99	2.38	2.64	3.21	3.43
74	1.29	1.67	1.99	2.38	2.64	3.20	3.43
76	1.29	1.67	1.99	2.38	2.64	3.20	3.42
77	1.29	1.66	1.99	2.38	2.64	3.20	3.42
79	1.29	1.66	1.99	2.37	2.64	3.20	3.42
81	1.29	1.66	1.99	2.37	2.64	3.19	3.41
85	1.29	1.66	1.99	2.37	2.63	3.19	3.41
88	1.29	1.66	1.99	2.37	2.63	3.19	3.40
89	1.29	1.66	1.99	2.37	2.63	3.18	3.40
96	1.29	1.66	1.98	2.37	2.63	3.18	3.39
99	1.29	1.66	1.98	2.36	2.63	3.17	3.39
102	1.29	1.66	1.98	2.36	2.62	3.17	3.39
106	1.29	1.66	1.98	2.36	2.62	3.17	3.38
112	1.29	1.66	1.98	2.36	2.62	3.16	3.38
118	1.29	1.66	1.98	2.36	2.62	3.16	3.37
128	1.29	1.66	1.98	2.36	2.61	3.16	3.37
129	1.29	1.66	1.98	2.36	2.61	3.15	3.37
132	1.29	1.66	1.98	2.35	2.61	3.15	3.37

	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
134	1.29	1.66	1.98	2.35	2.61	3.15	3.36
152	1.29	1.65	1.98	2.35	2.61	3.14	3.36
154	1.29	1.65	1.98	2.35	2.61	3.14	3.35
159	1.29	1.65	1.97	2.35	2.61	3.14	3.35
171	1.29	1.65	1.97	2.35	2.60	3.14	3.35
182	1.29	1.65	1.97	2.35	2.60	3.14	3.34
185	1.29	1.65	1.97	2.35	2.60	3.13	3.34
202	1.29	1.65	1.97	2.34	2.60	3.13	3.34
222	1.29	1.65	1.97	2.34	2.60	3.13	3.33
237	1.29	1.65	1.97	2.34	2.60	3.12	3.33
247	1.28	1.65	1.97	2.34	2.60	3.12	3.33
259	1.28	1.65	1.97	2.34	2.59	3.12	3.33
285	1.28	1.65	1.97	2.34	2.59	3.12	3.32
332	1.28	1.65	1.97	2.34	2.59	3.11	3.32
401	1.28	1.65	1.97	2.34	2.59	3.11	3.31
433	1.28	1.65	1.97	2.33	2.59	3.11	3.31
473	1.28	1.65	1.96	2.33	2.59	3.11	3.31
538	1.28	1.65	1.96	2.33	2.58	3.11	3.31
555	1.28	1.65	1.96	2.33	2.58	3.10	3.31
675	1.28	1.65	1.96	2.33	2.58	3.10	3.30
1712	1.28	1.65	1.96	2.33	2.58	3.09	3.30
∞	1.28	1.64	1.96	2.33	2.58	3.09	3.29

Beispiele und Nutzungshinweise:

- $t_{0.95}(15) \approx 1.75$
- Für nicht aufgeführte n nächstkleineres gelistetes n nutzen, z.B. $t_{0.9}(250) \approx t_{0.9}(247)$
- $z_{0.975} = t_{0.975}(\infty) = 1.96$
- Andere α:
  - $t_\alpha(n) = -t_{1-\alpha}(n)$
  - $z_\alpha = -z_{1-\alpha}$
- Für $n > 2000$ gilt $t_\alpha(n) \approx t_\alpha(\infty)$

6.3 Der p-value

↪ für praktische Durchführung: Bringe den Ablehnungsbereich auf eine einheitliche Form:

$$v \in K \Leftrightarrow p^* < \alpha$$

Dabei ist $p^* = p^*(v)$ eine (geeignete) Transformation der Teststatistik.

↪ p^* wird als p-**value** (p-Wert, Signifikanz) bezeichnet.

↪ Der p-Wert entspricht der geringsten vorgegebenen Irrtumswahrscheinlichkeit, zu der die Nullhypothese bei vorliegender Stichprobe abgelehnt werden kann. aber (!): Der p-Wert ist keine Wahrscheinlichkeit, sondern zufallsabhängig.

↪ Vorteile:

- Einheitlichkeit der Test-Darstellung,
- Nutzung in statistischer Software.
- Nur noch p^* wird berechnet, α kommt vom Anwender.

↪ Nachteile:

- p^* muss i.d.R. mit Software berechnet werden.
- Verschiedene Hypothesenpaare $\rightsquigarrow$ verschiedene Formeln für p^*.

↪ Klassische Testdarstellung: Teststatistik $V = v$, Schwellenwert q_α (Quantil)
↪ Für Umschreibung auf p-value: Transformation mittels Verteilungsfunktion:

Memo DuW: Verteilungsfunktion versus Quantilfunktion

↪ Die Verteilungsfunktion einer ZV ist die Funktion $x \mapsto F(x) = P(X \leq x)$
Symmetrische Verteilung: $1 - F(-x) = F(x) \quad \forall x$ (z.B. $\mathcal{N}(0,1)$, t_n)

↪ Die Quantilfunktion einer ZV X mit VF F ist die Funktion

$$\alpha \mapsto F^{-1}(\alpha) := q_\alpha := \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$$

↪ Für alle $x \in \mathbb{R}, \alpha \in]0;1[$ gilt:

$$F(x) \geq \alpha \Leftrightarrow x \geq F^{-1}(\alpha)$$

↪ Falls F stetig, streng monoton, so ist F^{-1} Umkehrfunktion von F,d.h. es gilt

$$F(q_\alpha) = \alpha \quad \text{für alle } \alpha \in]0;1[\qquad (*)$$

↪ Die Quantilfunktion ist Ersatz für die Umkehrfunktion, wenn F nicht umkehrbar ist. Daher auch Bezeichung: „verallgemeinerte Umkehrfunktion“.

$p - value$ beim Mittelwerttest (σ bekannt) für (b) $H_0 : \mu \geq \mu_0$, $H_1 : \mu < \mu_0$

$\hookrightarrow$ Ablehnung von H_0 für $\sqrt{n}\frac{\bar{x}-\mu_0}{\sigma} = v < z_\alpha$ $\qquad (z_\alpha = \Phi^{-1}(\alpha))$.

$\hookrightarrow$ Äquivalent $\Phi(v) < \Phi(z_\alpha) \stackrel{(*)}{=} \alpha$

$\hookrightarrow$ d.h. der p-value ist $p^*(v) = F(v)$ mit $F = \Phi$ $\qquad$ (Φ VF zu $\mathcal{N}(0,1)$)

Durchschnittseinkommen von niedergelassenen Ärzten

Stichprobe von $n = 200$ Arztpraxen, $\bar{X} = 168$T€. Laut einer statistischen Bundesamt liegt das „wahre Durchschnittseinkommen“ bei mindestens 181T€.
Annahme: $X \sim \mathcal{N}(\mu, \sigma^2), \sigma = 100$ bekannt.

$\hookrightarrow$ Hypothesen: $H_0 : \mu \geq 181$, vs. $H_1 : \mu < 181$, $v = \sqrt{n} \cdot \frac{\bar{x}-\mu_0}{\sigma} \approx -1.83$

Ermitteln Sie den p-value und die Testentscheidung ($\alpha = 0.05$).

$v = -1.83$, $p^*(v) = F(-1.83) = \Phi(-1.83) = 1 - \Phi(1.83) \approx 1 - 0.966 = 0.034$

$0.034 < 0.05$, d.h. die Hypothese wird abgelehnt.

Man kann am p-value ablesen, zu welchem Signifikanzniveau die Hypothese „so eben“ noch abgelehnt werden kann.

Hier z.B. $\alpha = 0.01$: Hypothese wird nicht abgelehnt.

Tabelle der VF zu $\mathcal{N}(0,1)$ auf folgender Folie:

	000	005	010	015	020	025	030	035	040	045	050	055	060	065	070	075	080	085	090	095
0.0	500	502	504	506	508	510	512	514	516	518	520	522	524	526	528	530	532	534	536	538
0.1	540	542	544	546	548	550	552	554	556	558	560	562	564	566	567	569	571	573	575	577
0.2	579	581	583	585	587	589	591	593	595	597	599	601	603	604	606	608	610	612	614	616
0.3	618	620	622	624	626	627	629	631	633	635	637	639	641	642	644	646	648	650	652	654
0.4	655	657	659	661	663	665	666	668	670	672	674	675	677	679	681	683	684	686	688	690
0.5	691	693	695	697	698	700	702	704	705	707	709	711	712	714	716	717	719	721	722	724
0.6	726	727	729	731	732	734	736	737	739	741	742	744	745	747	749	750	752	753	755	756
0.7	758	760	761	763	764	766	767	769	770	772	773	775	776	778	779	781	782	784	785	787
0.8	788	790	791	792	794	795	797	798	800	801	802	804	805	806	808	809	811	812	813	815
0.9	816	817	819	820	821	823	824	825	826	828	829	830	831	833	834	835	836	838	839	840
1.0	841	843	844	845	846	847	848	850	851	852	853	854	855	857	858	859	860	861	862	863
1.1	864	865	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884
1.2	885	886	887	888	889	890	891	892	893	893	894	895	896	897	898	899	900	901	901	902
1.3	903	904	905	906	907	907	908	909	910	911	911	912	913	914	915	915	916	917	918	918
1.4	919	920	921	921	922	923	924	924	925	926	926	927	928	929	929	930	931	931	932	933
1.5	933	934	934	935	936	936	937	938	938	939	939	940	941	941	942	942	943	944	944	945
1.6	945	946	946	947	947	948	948	949	949	950	951	951	952	952	953	953	954	954	954	955
1.7	955	956	956	957	957	958	958	959	959	960	960	960	961	961	962	962	962	963	963	964
1.8	964	964	965	965	966	966	966	967	967	967	968	968	969	969	969	970	970	970	971	971
1.9	971	972	972	972	973	973	973	974	974	974	974	975	975	975	976	976	976	976	977	977
2.0	977	978	978	978	978	979	979	979	979	980	980	980	980	981	981	981	981	981	982	982
2.1	982	982	983	983	983	983	983	984	984	984	984	984	985	985	985	985	985	986	986	986
2.2	986	986	986	987	987	987	987	987	987	988	988	988	988	988	988	989	989	989	989	989
2.3	989	989	990	990	990	990	990	990	990	990	991	991	991	991	991	991	991	991	992	992
2.4	992	992	992	992	992	992	992	993	993	993	993	993	993	993	993	993	993	994	994	994
2.5	994	994	994	994	994	994	994	994	994	995	995	995	995	995	995	995	995	995	995	995
2.6	995	995	995	996	996	996	996	996	996	996	996	996	996	996	996	996	996	996	996	996
2.7	997	997	997	997	997	997	997	997	997	997	997	997	997	997	997	997	997	997	997	997
2.8	997	997	998	998	998	998	998	998	998	998	998	998	998	998	998	998	998	998	998	998
2.9	998	998	998	998	998	998	998	998	998	998	998	998	998	998	999	999	999	999	999	999
3.0	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999
3.1	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999
3.2	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999	≈ 1

Beispiel:
$\Phi(1.240) = \Phi(1.\overset{2}{\cancel{3}} + .040) \approx 0.893.$
Für nicht aufgeführte x:

- $x \geq 3.2$:
  $\Phi(x) \approx 1$ für $x \geq 3.2$.
- Für $x < 0$:
  $\Phi(x) = 1 - \Phi(-x)$
- Interpolation:
  für vertafelte x, y und
  $t = \lambda x + (1-\lambda)y$, $\lambda \in]0; 1[$
  $\Phi(t) \approx \lambda\Phi(x) + (1-\lambda)\Phi(y)$

Berechnung des p-values in den anderen Fällen (σ bekannt, mit $F = \Phi$):

c) $H_0 : \mu \leq \mu_0$, $H_1 : \mu > \mu_0$: $\quad v \in K \Leftrightarrow v > z_{1-\alpha}$

$p^* = 1 - \Phi(v)$

denn $v > z_{1-\alpha} \quad \Leftrightarrow \quad F(v) > F(z_{1-\alpha}) \overset{(*)}{=} 1 - \alpha \quad \Leftrightarrow \quad 1 - F(v) < \alpha$

a) $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$: $\quad v \in K \Leftrightarrow |v| > z_{1-\alpha/2}$

$p^* = 2\min(\Phi(v), 1 - \Phi(v)) = 2\min(1 - \Phi(-v), 1 - \Phi(v)) = 2(1 - \Phi(|v|))$

$$
\begin{aligned}
|v| > z_{1-\alpha/2} \quad &\Leftrightarrow \quad v < -z_{1-\alpha/2} \quad \text{oder} \quad v > z_{1-\alpha/2} \\
&\Leftrightarrow \quad -v > z_{1-\alpha/2} \quad \text{oder} \quad v > z_{1-\alpha/2} \\
&\Leftrightarrow \quad F(-v) > 1 - \tfrac{\alpha}{2} \quad \text{oder} \quad F(v) > 1 - \tfrac{\alpha}{2} \\
&\Leftrightarrow \quad 1 - F(-v) < \tfrac{\alpha}{2} \quad \text{oder} \quad 1 - F(v) < \tfrac{\alpha}{2} \\
&\Leftrightarrow \quad F(v) < \tfrac{\alpha}{2} \quad \text{oder} \quad 1 - F(v) < \tfrac{\alpha}{2} \\
&\Leftrightarrow \quad \min(F(v), 1 - F(v)) < \tfrac{\alpha}{2} \quad \Leftrightarrow \quad 2\min(F(v), 1 - F(v)) < \alpha
\end{aligned}
$$

Durchschnittseinkommen von Praktikanten

Stichprobe von $n = 40$ Studenten mit durchschnittlichem Einkommen $\bar{X} = 1225.45$.
Laut einer Studie liegt das „wahre Durchschnittseinkommen" bei 1260 €.
Wird die Angabe der Studie (bei einem Signifikanzniveau von $\alpha = 5\%$) durch die Stichprobe gestützt? Berechnen Sie den p-value.

Annahme: $X \sim \mathcal{N}(\mu, \sigma^2), \sigma = 120$ als bekannt angenommen.

Wie lautet, ausgehend von Ihren bisherigen Berechnungen, Ihre Testentscheidung?

Testentscheidung: Zweiseitiges Testproblem $H_0 : \mu = \mu_0 = 1260$

$$\begin{aligned} v = -1.820945 \quad &\Rightarrow \quad |v| = 1.820945 \\ p^* = 2(1 - \Phi(|v|) \quad &= \quad 2(1 - \Phi(1.82)) = 2(1 - 0.966) = 0.068 \\ 0.068 > 0.05 \end{aligned}$$

H_0 wird nicht abgelehnt.

Berechnung des p-Wertes

a) bei zweiseitiger Hypothese $H_0 : \mu = \mu_0$ (zweiseitigem Test):

$$p^* \quad = \quad 2 \cdot \min(P(V \leq v)\ ,\ 1 - P(V \leq v))$$

b) bei rechtssseitiger Hypothese $H_0 : \mu \geq \mu_0$ (rechtsseitigem Test):

$$p^* \quad = \quad P(V \leq v)$$

c) bei linksseitiger Hypothese $H_0 : \mu \leq \mu_0$ (linkssseitigem Test):

$$p^* \quad = \quad 1 - P(V \leq v)$$

$\hookrightarrow$ In allen Fällen wird p^* jeweils mit $\mu = \mu_0$ berechnet.

p-Werte p^* bei Mittelwert-Tests (Fälle i),ii),iv), d.h. Normalverteilung mit σ^2 bekannt, bzw. approx. Normalverteilung s.o.)

$$
\begin{array}{lllcl}
a) & H_0 : \mu = \mu_0 & p^* & = & 2 \cdot (1 - P(V \leq |v|)) = 2 \cdot (1 - \Phi(|v|)) \\
b) & H_0 : \mu \geq \mu_0 & p^* & = & P(V \leq v) = \Phi(v) \\
c) & H_0 : \mu \leq \mu_0 & p^* & = & (1 - P(V \leq v)) = 1 - \Phi(v)
\end{array}
$$

Fall iv): Für Normalverteilung und unbekanntes σ^2 ist die Nullverteilung die t_{n-1}-Verteilung, dann analog:

$$
\begin{array}{lllcl}
a) & H_0 : \mu = \mu_0 & p^* & = & 2 \cdot (1 - P(V \leq |v|)) = 2 \cdot (1 - F_{t_{n-1}}(|v|)) \\
b) & H_0 : \mu \geq \mu_0 & p^* & = & P(V \leq v) = F_{t_{n-1}}(v) \\
c) & H_0 : \mu \leq \mu_0 & p^* & = & (1 - P(V \leq v)) = 1 - F_{t_{n-1}}(v)
\end{array}
$$

Berechnung mittels Software (R) oder (für $n > 30$ approximativ) durch $F_{t_{n-1}} \approx \Phi$

Durchschnittseinkommen von Praktikanten

Stichprobe von $n = 40$ Studenten mit durchschnittlichem Einkommen $\bar{X} = 1225.45$.
Laut einer Studie liegt das „wahre Durchschnittseinkommen" bei 1260 €.
Wird die Angabe der Studie (bei einem Signifikanzniveau von $\alpha = 5\%$) durch die Stichprobe gestützt? Berechnen Sie den p-value.
Annahme: $X \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 unbekannt (wird durch $S^2 = 87.23354^2$ geschätzt).

Testentscheidung: Zweiseitiger t-Test von H_0: $\mu = 1260$

$$v = -2.504924 \quad \Rightarrow \quad |v| = 2.504924$$

$$p^* = 2(1 - F_{t_{n-1}}(|v|)) \quad = \quad 2(1 - F_{t_{39}}(2.505)) = 2(1 - 0.9917) \approx 0.0165$$

$$0.0165 < 0.05$$

H_0 wird verworfen.

Der p-value muss mit R berechnet werden:

```
praktdata=c(1346,1187,1353,1347,1261,1066,1127,1191,1219,1460,1296,1140,
1105,1191,1190,1179,1245,1131,1264,1096,1198,1258,1233,1300,1214,1270,1329,

1151,1092,1225,1196,1166,1177,1155,1293,1335,1319,1177,1344,1192)
n=length(praktdata)
mu0=1260
t.test(praktdata,mu=mu0)
2*(1-pt(abs(sqrt(n)*(mean(praktdata)-mu0)/sd(praktdata)),df=39))
```

$\hookrightarrow$ Formeln für p-Werte auf viele stetige(!) Modelle übertragbar:

- $H: \theta = \theta_0$: $p^* = 2 \cdot \min(F(v), 1 - F(v))$
- $H: \theta \geq \theta_0$: $p^* = F(v)$
- $H: \theta \leq \theta_0$: $p^* = 1 - F(v)$

Geeignet bei unimodaler Nullverteilung F.

$\hookrightarrow$ Bei diskreten Modellen Formeln nicht einfach übertragbar, weil die VF nicht invertierbar ist. Grundsätzlich aber weiterhin:

$p^* \cong$ min. Signifikanzniveau, zu dem H_0 bei $V = v$ abgelehnt wird.

Vor allem bei zweiseitigen Hypothesen dann diverse Ansätze, z.B.:

„Principle of minimum likelihood“ (z.T. in R verwendet)

$p^* \cong$ Summe aller WS $P(V = k)$ mit $P(V = k) \leq P(V = v)$.

Vgl.: *Gibbons/Pratt: P-values: Interpretation and Methodology, Amer. Statist. 29 (1975), 20-25.*

6.4 Einstichproben-Anteilswert-Tests

Modell: $X_i \sim \mathcal{B}(1, p)$

1. Hypothesensysteme

a) $H_0 : p = p_0$ vs. $H_1 : p \neq p_0$

b) $H_0 : p \geq p_0$ vs. $H_1 : p < p_0$

c) $H_0 : p \leq p_0$ vs. $H_1 : p > p_0$

Beispiele:

a) Der Anteil iPad-Besitzer unter den Studenten in Münster liegt bei p_0.
Eine (verbogene) Münze ist fair ($p = \frac{1}{2}$).

b) Der Anteil Kunden, die einem Vertreterbesuch zustimmen, ist mindestens p_0.

c) Der Anteil fehlerhafter Artikel einer Tagesproduktion ist höchstens p_0.

Anteilswert-Tests sind Mittelwert-Tests für dichotome X_i. ($E(X_i) = p$). Sie lassen sich exakt oder approximativ ausführen.

Exakter Test

$\hookrightarrow$ Teststatistik: $V = X_1 + \cdots + X_n \sim Bin(n, p_0)$ für $p = p_0$

a) $H_0 : p = p_0$: Ablehnung wenn $v \notin [q_{\alpha/2}; q_{1-\alpha/2}]$, p^* s.u.

b) $H_0 : p \geq p_0$: Ablehnung wenn $v < q_\alpha$, $p^* = F(v)$.

c) $H_0 : p \leq p_0$: Ablehnung, wenn $v > q_{1-\alpha}$, $p^* = 1 - F(v-1)$.

$p^* \cong$ min. Sig.-niveau, mit dem H_0 abgelehnt wird.
Davon abweichend in R: „minimum likelihood principle“

Approx. Test für $np_0(1-p_0) \geq 9$ (Fall iv) der Mittelwerttests)

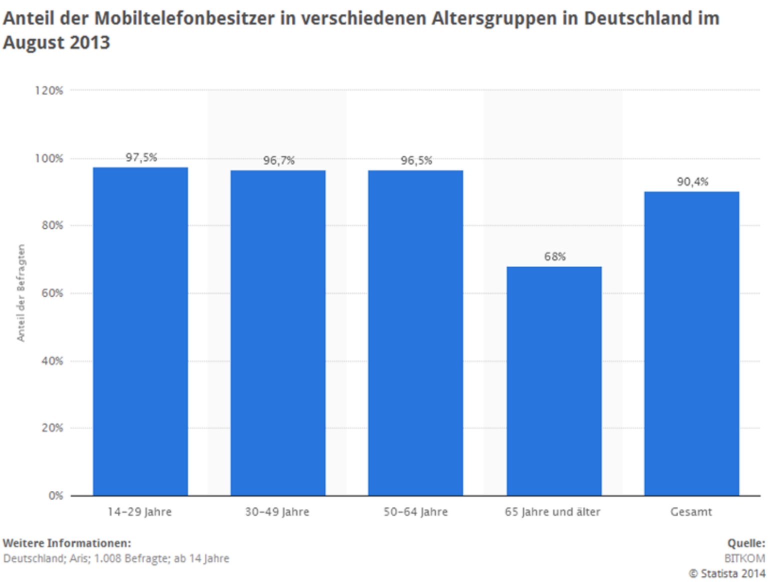
$\hookrightarrow$ Teststatistik $V = \sqrt{n}\frac{\bar{X}-p_0}{\sqrt{p_0(1-p_0)}}$, näherungsweise $\sim \mathcal{N}(0,1)$ für $p = p_0$.

a) $H_0 : p = p_0$: Ablehnung wenn $|v| > z_{1-\alpha/2}$, $p^* = 2(1 - \Phi(|v|))$.

b) $H_0 : p \geq p_0$: Ablehnung wenn $v < z_\alpha$, $p^* = \Phi(v)$.

c) $H_0 : p \leq p_0$: Ablehnung, wenn $v > z_{1-\alpha}$, $p^* = 1 - \Phi(v)$.

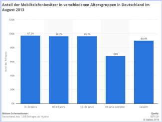
Exakter und approx. Test ergeben nicht immer die gleiche Entscheidung (Übung!)

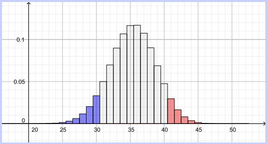

Anteil der Mobiltelefonbesitzer in verschiedenen Altersgruppen in Deutschland im August 2013
120%
100%
80%
60%
40%
20%
0%
Anteil der Befragten
97,5%
96,7%
96,5%
68%
90,4%
14-29 Jahre
30-49 Jahre
50-64 Jahre
65 Jahre und älter
Gesamt
Weitere Informationen:
Deutschland; Aris; 1.008 Befragte; ab 14 Jahre
Quelle:
BITKOM
© Statista 2014

Beispiel: Handynutzung unter Rentnern

Laut einer Studie besitzen 68% der Rentner ein Handy. Wird dies durch eine Stichprobe von $n = 52$ Rentnern, von denen 30 ein Handy besitzen, gestützt ($\alpha = 0.05$)?

Hypothesen: $H_0 : p = p_0 = 0.68$, $H_1 : p \neq 0.68$





https://ggbm.at/rqrpwked

↪ exakter Test:

- ☐ $V = v = 30$
- ☐ $q_{0.025} = 29 \leq 30 \leq q_{0.975} = 42$:
  H_0 wird nicht verworfen.
- ☐ $p^* = P(V \leq 30) + P(V \geq 41) \approx 0.136$

in R: `binom.test(30,52,0.68)$p.value` („minimum likelihood principle“)

↪ approximativer Test:

- ☐ $V = v = \sqrt{n}\frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}} = \sqrt{52}\frac{30/52 - 0.68}{\sqrt{0.68 \cdot 0.32}} = -1.59$
- ☐ $|v| = 1.59 < z_{0.975} = 1.96$: H_0 wird nicht verworfen.
- ☐ $p^* = 2 \cdot (1 - \Phi(1.59)) = 0.11 > 0.05$

Gültigkeit der Normalapproximation: $52 \cdot 0.68 \cdot 0.32 = 11.32 > 9$


Anteil der Tablet-Nutzer in Deutschland in den Jahren 2012 bis 2017
Anteil an der Bevölkerung
60%
50%
40%
30%
20%
10%
0%
13%
15%
28%
40%
41%
47%
2012
2013
2014
2015
2016
2017*
Quelle
Bitkom
© Statista 2019
Weitere Informationen:
Deutschland; Bitkom Research; ab 14 Jahre
statista

Umfrage unter Studierenden

Von $n = 40$ Befragten teilten 25 mit, dass sie im Jahr 2018 Tablet-Nutzer gewesen seien. Lässt sich hieraus statistisch schließen, dass der Anteil der Tablet-Nutzer in 2018 auf über 47% gestiegen ist ($\alpha = 5\%$)?

Fiktives Umfragebeispiel:
Anteilswerttest mit $n = 40$, $v = 25$
und $H_0 : p \leq 0.47$
exakt: $p^* = 1 - F(24) \approx 0.035$
Die Nullhypothese wird verworfen.
appr.: $p^* = 1 - \Phi(\sqrt{40}\frac{25/40-0.47}{\sqrt{0.47 \cdot 0.53}}) =$
$1 - \Phi(1.964) \approx 0.025$
Die Nullhypothese wird approximativ verworfen.
Insgesamt lässt sich die Steigung des Anteils statistisch bestätigen.

Umfrage im Hörsaal:
Anteilswerttest mit $n =$, $v =$
und $H_0 : p \leq 0.47$

Anmerkung: Eine verallgemeinernde Aussage ist aufgrund der speziellen Auswahl der Befragten eher nicht möglich („selection bias“).

Zusammenhang von Test und Konfidenzintervall

$H_0 : p = p_0$ wird nicht abgelehnt, falls

$$
\begin{aligned}
|V| &= \left|\frac{\hat{p} - p_0}{\sigma_{\hat{p}}}\right| \leq z_{1-\alpha/2} \\
\Leftrightarrow \quad -z_{1-\alpha/2} &\leq \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} \leq z_{1-\alpha/2} \\
\Leftrightarrow \quad \hat{p} - z_{1-\alpha/2}\sigma_{\hat{p}} &\leq p_0 \leq \hat{p} + z_{1-\alpha/2}\sigma_{\hat{p}}
\end{aligned}
$$

$\hookrightarrow$ entspricht dem (approximativen) Konfidenzintervall für den Anteilswert (s.o.)

$\hookrightarrow$ H_0 **wird abgelehnt, wenn der hypothetische Wert** p_0 **außerhalb der Realisation des Konfidenzintervalles liegt.**

$\hookrightarrow$ übertragbar auf einseitige Tests/KI und (z.B.) Mittelwert-Tests

6.5 Chi-Quadrat-Test für die Varianz bei $X_i \sim \mathcal{N}(\mu, \sigma^2)$

1. Hypothesensystem

a) $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 \neq \sigma_0^2$

b) $H_0 : \sigma^2 \geq \sigma_0^2$ vs. $H_1 : \sigma^2 < \sigma_0^2$

c) $H_0 : \sigma^2 \leq \sigma_0^2$ vs. $H_1 : \sigma^2 > \sigma_0^2$

2. Teststatistik und ihre Verteilung

Für $\sigma = \sigma_0$ gilt (s.o.) mit $\hat{\sigma}^2 = S^2$:

$$V = \frac{(n-1)\hat{\sigma}^2}{\sigma_0^2} = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2(n-1) \quad (\mu \text{ unbekannt}),$$

$$V = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i - \mu)^2 \sim \chi^2(n) \quad (\mu \text{ bekannt})$$

3. Entscheidungsregel

Für unbekanntes μ wird H_0 abgelehnt, wenn:

a) $H_0: \sigma^2 = \sigma_0^2, H_1: \sigma^2 \neq \sigma_0^2$ $\qquad v < \chi^2_{\alpha/2}(n-1)$ oder $v > \chi^2_{1-\alpha/2}(n-1)$

b) $H_0: \sigma^2 \geq \sigma_0^2, H_1: \sigma^2 < \sigma_0^2$ $\qquad 0 < v < \chi^2_{\alpha}(n-1)$

c) $H_0: \sigma^2 \leq \sigma_0^2, H_1: \sigma^2 > \sigma_0^2$ $\qquad v > \chi^2_{1-\alpha}(n-1)$

$\hookrightarrow$ die χ^2-Verteilung ist nicht symmetrisch, deshalb unsymmetrische Festlegung bei zweiseitiger Hypothese.

p-Werte

a) $H_0, \sigma^2 = \sigma_0^2$: $\qquad p^* = 2 \cdot \min(F_{\chi^2(n-1)}(v), 1 - F_{\chi^2(n-1)}(v))$

b) $H_0, \sigma^2 \geq \sigma_0^2$: $\qquad p^* = F_{\chi^2(n-1)}(v)$

c) $H_0, \sigma^2 \leq \sigma_0^2$: $\qquad p^* = 1 - F_{\chi^2(n-1)}(v)$

$\hookrightarrow$ Für bekanntes μ werden Quantile und Verteilungsfunktion der χ^2-Verteilung mit n Freiheitsgraden verwendet (aber eher unrealistisch).

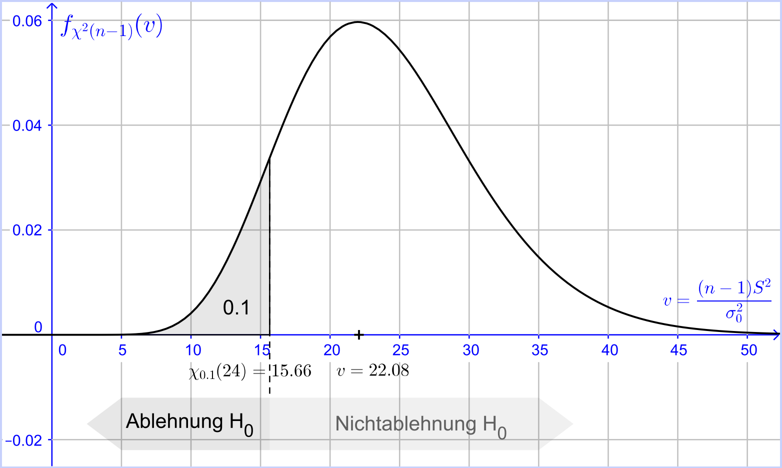
Lebensdauer eines Verschleißteils

Annahme: Lebensdauer X eines Verschleißteils sei normalverteilt. In der Vergangenheit galt $\sigma^2 = 2.5$ Jahre^2. Nach Umstellung des Produktionsverfahrens soll geprüft werden, ob die Variabilität zurückgegangen ist ($\alpha = 0.1$). Eine Stichprobe bei unbekanntem μ ergab: $n = 25, \hat{\sigma}^2 = 2.3$.

Hypothesen	$H_0 : \sigma^2 \geq 2.5 \quad \text{vs.} \quad H_1 : \sigma^2 < 2.5$
Teststatistik	$v = \dfrac{(n-1)\hat{\sigma}^2}{\sigma_0^2} = \dfrac{24 \cdot 2.3}{2.5} = 22.08$
Testentscheidung	$v = 22.08 \geq \chi^2_{0.1}(24) = 15.66$
$\Rightarrow$	H_0 kann nicht verworfen werden
	$p^* = F_{\chi^2(24)}(22.08) = 0.42 > 0.1.$

In R: `qchisq(0.1,df=c(24,25))` ergibt c(15.65868,16.47341)

Illustration



6.6 Gütefunktionen von Parametertests

↪ Partition des Parameterraums gemäß Hypothesenpaar H_0, H_1 in $\Theta = \Theta_0 \cup \Theta_1$.

↪ Bestimmung der Teststatistik $v = V = V(X)$. Damit Testentscheidungen:

- $d(v) = d_1$ (Ablehnung von H_0) für $v \in K$
- $d(v) = d_0$ (Nichtablehnung von H_0) für $v \notin K$

↪ Risiko[2]: $R(\theta, V) = \begin{cases} P_\theta(V \in K) & \theta \in \Theta_0 \text{ (Fehler 1. Art)} \\ P_\theta(V \notin K) & \theta \in \Theta_1 \text{ (Fehler 2. Art)} \end{cases}$

↪ Tests sollten ein möglichst kleines Risiko haben $\rightsquigarrow$ 2 konfliktäre Ziele:

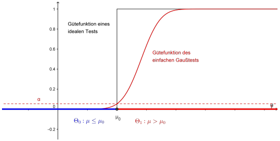
- Für $\theta \in \Theta_0$: $P_\theta(V \in K)$ möglichst klein
- Für $\theta \in \Theta_1$: $P_\theta(V \notin K)$ möglichst klein, d.h. $P_\theta(V \in K)$ möglichst groß

Gütefunktion

$$g : \Theta \to [0; \infty], g(\theta) \quad = \quad P_\theta(\text{„Ablehnung von } H_0\text{“}) = P_\theta(V \in K)$$

[2] $R(\theta, V) = E_\theta(L(\theta, d(V)))$ m. Verlustfunktion $L(\theta, d) = \begin{cases} 1 & \text{für } \theta \in \Theta_i, d = d_j,\ i \neq j \\ 0 & \text{sonst} \end{cases}$

Beispiel: linksseitiger Gaußtest : $\Theta_0 :] -\infty; \mu_0] \quad \Theta_1 :]\mu_0; \infty[$



https://ggbm.at/cedn5dkr

$$\begin{aligned}
g(\mu) &= P_\mu\left(\sqrt{n}\frac{\bar{X}-\mu_0}{\sigma} > z_{1-\alpha}\right) = P_\mu\left(\sqrt{n}\frac{\bar{X}-\mu+\mu-\mu_0}{\sigma} > z_{1-\alpha}\right) \\
&= P_\mu\left(\sqrt{n}\frac{\bar{X}-\mu}{\sigma} + \sqrt{n}\frac{\mu-\mu_0}{\sigma} > z_{1-\alpha}\right) \\
&= P_\mu\left(\sqrt{n}\frac{\bar{X}-\mu}{\sigma} > z_{1-\alpha} - \sqrt{n}\frac{\mu-\mu_0}{\sigma}\right) = 1 - \Phi\left(z_{1-\alpha} - \sqrt{n}\frac{\mu-\mu_0}{\sigma}\right) \\
&= \Phi\left(\sqrt{n}\frac{\mu-\mu_0}{\sigma} - z_{1-\alpha}\right) = \Phi\left(\frac{\sqrt{n}}{\sigma}\cdot\left(\mu - \left(\mu_0 + \frac{\sigma}{\sqrt{n}}z_{1-\alpha}\right)\right)\right) = \Phi(b(\mu-a))
\end{aligned}$$

Die Gütefunktion entsteht aus der VF Φ zu $\mathcal{N}(0,1)$ durch lineare Transformation:

- Skalenfaktor $b = \frac{\sqrt{n}}{\sigma}$ legt „Steilheit“ fest. Je größer n/kleiner σ, desto steiler.
- $a = \mu_0 + \frac{\sigma}{\sqrt{n}}z_{1-\alpha}$ ist Wendestelle (Übung).

Ein Test mit steiler verlaufender Gütefunktion stellt einen besseren Kompromiss zwischen Fehler-WS dar.

Einfluss des Stichprobenumfangs n auf die Testgüte

Durch Erhöhung von n können Vorgaben an beide Fehlertypen erfüllt werden.

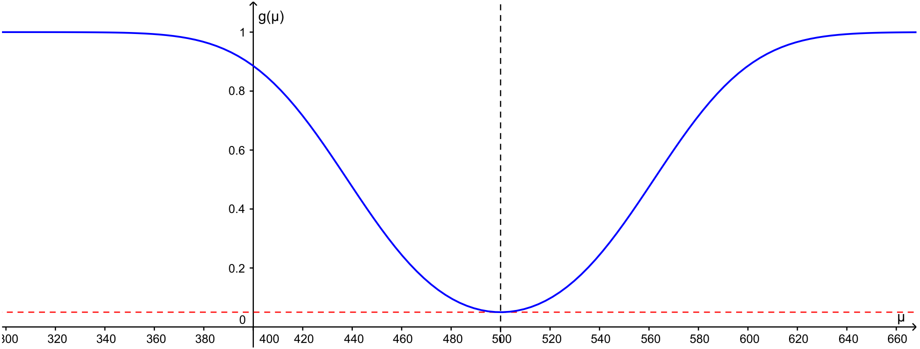
Konkret: Wie groß muss n sein, damit die Gütefunktion an der Stelle μ^* einen vorgegebenen Wert β überschreitet?

$$\begin{aligned}
\Phi\left(\frac{\mu^* - \mu_0}{\sigma}\sqrt{n} - z_{1-\alpha}\right) \geq \beta \quad &\Leftrightarrow \quad \frac{\mu^* - \mu_0}{\sigma}\sqrt{n} - z_{1-\alpha} \geq z_\beta \\
&\Leftrightarrow \quad \frac{\mu^* - \mu_0}{\sigma}\sqrt{n} \geq z_{1-\alpha} + z_\beta \\
&\Leftrightarrow \quad \sqrt{n} \geq \frac{z_{1-\alpha} + z_\beta}{\mu^* - \mu_0}\sigma \\
&\Leftrightarrow \quad n \geq \frac{(z_{1-\alpha} + z_\beta)^2}{(\mu^* - \mu_0)^2}\sigma^2
\end{aligned}$$

Für festes n wird jede Abweichung $d = \mu^* - \mu_0 > 0$ und höher mindestens mit Wahrscheinlichkeit β aufgedeckt.

Gütefunktion im zweiseitigen Testproblem

$$g(\mu) \quad = \quad \Phi(\frac{\mu - \mu_0}{\sigma/\sqrt{n}} - z_{1-\alpha/2}) + \Phi(-\frac{\mu - \mu_0}{\sigma/\sqrt{n}} - z_{1-\alpha/2})$$



Rechnung dazu: Übung