Grundlagen der Mathematischen Statistik

Testen von Hypothesen
Teil I: Einführung

Uwe Menzel, 2018
uwe.menzel@matstat.org
www.matstat.org

Die "Philosophie" der Hypothesentests
Nullhypothese und alternative Hypothese

- **Ziel eines (parametrischen) Tests** ist es zu entscheiden, ob ein Verteilungs-parameter ($\mu, \Delta\mu, \sigma, \lambda, \ldots$) einen vermuteten (hypothetischen) Wert haben könnte.
  - Wir könnten z. B. vermuten, dass der Erwartungswert einer normalverteilten Population $\mu = 20$ ist.
- Die Hypothese bezüglich des Parameters wird Nullhypothese (H_0) genannt.
  - Die Nullhypothese könnte also z. B. $\boldsymbol{H_0\text{:}\ \mu = 20}$ lauten.
- Der Nullhypothese wird eine sogenannte alternative Hypothese (H_a) für den Parameter gegenübergestellt.
  - Die alternative Hypothese könnte z. B. $\boldsymbol{H_a\text{:}\ \mu > 20}$ lauten.

Die "Philosophie" der Hypothesentests
Beispiele für Hypothesen

- $X \sim N(\mu, \sigma)$ (es sei bekannt, dass Normalverteilung vorliegt)
  - Nullhypothese H_0: $\mu = 20$ (der Erwartungswert ist 20)
  - alternative Hypothese H_a: $\mu \neq 20$ (der Erwartungswert ist **nicht** 20)
- $X \sim N(\mu, \sigma)$
  - Nullhypothese H_0: $\mu = 20$ (der Erwartungswert ist 20)
  - alternative Hypothese H_a: $\mu > 20$ (der Erwartungswert ist **größer** als 20)
- $X \sim N(\mu, \sigma)$
  - Nullhypothese H_0: $\mu = 20$ (der Erwartungswert ist 20)
  - alternative Hypothese H_a: $\mu < 20$ (der Erwartungswert ist **kleiner** als 20)
- $X \sim Bin(n, p)$ (es sei bekannt, dass Binomialverteilung vorliegt)
  - Nullhypothese H_0: $p = {}^1/_2$ (die "Erfolgs"wahrscheinlichkeit ist ${}^1/_2$)
  - alternative Hypothese H_a: $p > {}^1/_2$ (die Erfolgswahrscheinlichkeit ist **größer** als ${}^1/_2$

www.matstat.org

Die "Philosophie" der Hypothesentests
Stichprobe und Teststatistik

- Um zu entscheiden, welche der beiden Hypothesen zutrifft, wird eine zufällige Stichprobe aus der Population entnommen.
- Aus der Stichprobe wird ein einzelner numerischer Wert extrahiert, welcher eine Observation (Beobachtung) einer Teststatistik darstellt.
  - Teststatistiken können z. B. sein: der (standardisierte) **Mittelwert der Messungen** oder die Anzahl der "erfolgreichen" Bernoulli-Versuche (bei Binomialverteilung) - i. A. ein Schätzer für den Parameter oder eine vom Schätzer abgeleitete Größe.
- Es wird nun festgestellt, ob der beobachtete Wert der Teststatistik im Widerspruch zur Nullhypothese steht. Wenn dies der Fall ist, wird die Nullhypothese verworfen und die alternative Hypothese akzeptiert.
  - Ein Widerspruch liegt z. B. vor, wenn der sog. ***p*-Wert** klein ist oder wenn die Teststatistik einen **kritischen Wert** überschreitet. Mehr dazu weiter unten …
- Die Ablehnung der Nullhypothese ist i. A. ein gewünschtes Resultat.
  - Dementsprechend wird i. A. auch die Nullhypothese gewählt. Will man z. B. nachweisen, dass zwei Populationen unterschiedliche Erwartungswerte haben, würde man H_0: $\Delta\mu = \mu_1 - \mu_2 = 0$ wählen und versuchen, dies zu widerlegen.

www.matstat.org

Die ”Philosophie” der Hypothesentests
Verwerfen der Nullhypothese

Es gibt verschiedene Methoden um zu prüfen, ob die Nullhypothese verworfen werden sollte.

1. Vergleich der Teststatistik mit einem kritischen Wert:
- Die Nullhypothese wird verworfen, wenn die Teststatistik einen kritischen Wert in Richtung der alternativen Hypothese überschreitet.

2. Vergleich der Teststatistik mit einem Konfidenzintervall:
- Die Nullhypothese wird verworfen, wenn der hypothetische Wert des Parameters nicht im berechneten Konfidenzintervall für den Parameter liegt.

3. Vergleich des p-Wertes mit dem Signifikanzniveau
- Die Nullhypothese wird verworfen, wenn der p-Wert kleiner ein zuvor festgelegter Wert – das sogenannte Signifikanzniveau - ist.

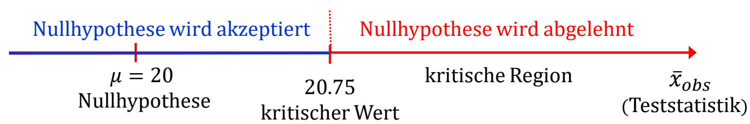
Alle drei Methoden führen zu den gleichen Schlussfolgerungen.

1. Vergleich der Teststatistik mit dem kritischen Wert

Beispiel ” Trockenzeit”:
- Ein Unternehmen stellt Farbe her und behauptet, dass die durchschnittliche Trockenzeit nur 20 Minuten beträgt.
- Eine Verbraucherorganisation misstraut dieser Behauptung und vermutet, dass die tatsächliche Trockenzeit größer sein könnte.
- Es wird angenommen, dass die Trockenzeit normalverteilt ist: $X \sim N(\mu, \sigma)$
  - Nullhypothese H_0: $\mu = 20$
  - Alternative Hypothese: H_a: $\mu > 20$
- Eine Stichprobe soll die Behauptung des Unternehmens testen:
  - es werden 36 Probenflächen bemalt ($n = 36$)
  - die Trockenzeiten werden gemessen (Trockenzeiten = Zufallsvariablen X_i)
  - Teststatistik: $\bar{X}_{obs} = {}^{1}/_{36} \cdot \sum_{i=1}^{36} X_i$ (durchschnittliche Trockenzeit in der Stichprobe = Schätzer für μ)
- Für welchen Wert der Teststatistik ($\bar{X}_{obs}$) wird H_0 verworfen ?
  - → kritischer Wert muss festgelegt werden

Wahl des kritischen Wertes

- Die Nullhypothese soll verworfen werden für $\bar{x}_{obs} \geq \omega_{krit} = 20.75$ - ein solches Resultat unterstützt die alternative Hypothese H_a: $\mu > 20$
- Die Wahl des kritischen Wertes erscheint willkürlich – wir werden später sehen, wie ω_{krit} aus anderen Forderungen hergeleitet werden kann.
- Der kritische Wert ω_{krit} definiert den Ablehnungsbereich von H_0 (die kritische Region)



Wichtig: Die Formulierung "Nullhypothese akzeptiert" bedeutet **nicht**, dass diese damit bewiesen wäre. Es kann lediglich gesagt werden, dass die aus der aktuellen Stichprobe gewonnene Observation der Teststatistik **nicht gegen** die Nullhypothese spricht.

www.matstat.org

Fehlertypen

Fehlertypen: Wenn ein kritischer Wert festgelegt wird, ist es unausweichlich, dass zwei Arten von Fehlern riskiert werden:

Fehler vom Typ I: es ist möglich, dass die Stichprobe $\bar{x}_{obs} > 20.75$ ergibt, selbst wenn die Nullhypothese $\mu = 20$ richtig ist:

→ die Nullhypothese wird verworfen, obwohl sie wahr ist

Fehler vom Typ II: es ist möglich, dass die Stichprobe $\bar{x}_{obs} < 20.75$ ergibt, selbst wenn die Nullhypothese $\mu = 20$ falsch ist:

→ die Nullhypothese wird akzeptiert, obwohl sie falsch ist

Diese Fehler sind unausweichlich, da wir unsere Schlussfolgerungen auf der Grundlage einer Stichprobe treffen, die vom Zufall beeinflusst wird.

	H_0 wird akzeptiert	H_0 wird verworfen
H_0 ist richtig	korrekt	Fehler Typ I
H_0 ist falsch	Fehler Typ II	korrekt

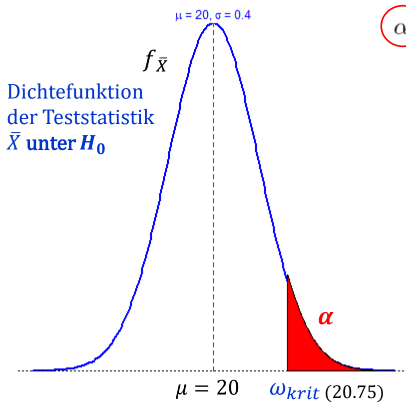
www.matstat.org

Wahrscheinlichkeit für Fehler Typ I: α

Fehler Typ I: Die Nullhypothese wird verworfen, obwohl sie wahr ist.

Annahmen: Normalverteilung, $\sigma = 2.4$ bekannt. Unter H_0 gilt $\boldsymbol{\mu = 20}$.

$$X_i \sim N(\mu, \sigma) \Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow \bar{X} \sim N(\underline{\underline{20}}, 0.4) \quad \text{(wenn } H_0 \text{ wahr)}$$



$$\begin{aligned} \alpha &= P(\text{Fehler Typ I}) \\ &= P(H_0 \text{ wird verworfen} \mid H_0 \text{ wahr}) \\ &= P(\bar{X} > 20.75 \mid \mu = 20) \\ &= 1 - P(\bar{X} \leq 20.75 \mid \mu = 20) \\ &= 1 - F_{\bar{X}}(20.75) \quad \text{für } \mu = 20 \\ &= 1 - \Phi\left(\frac{20.75 - 20}{0.4}\right) \\ &= 1 - \Phi(1.875) = 0.03 \end{aligned}$$

3% Risiko für Fehler Typ I

Man kann α verkleinern, indem man ω_{krit} **vergrößert**.

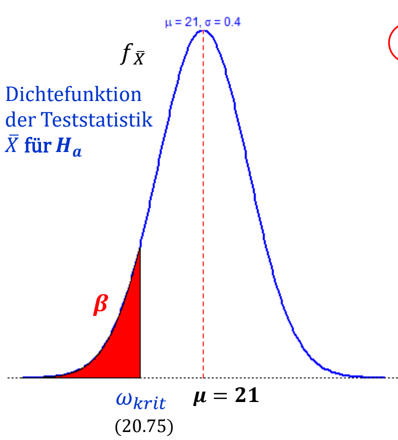
uwe.menzel@matstat.org

Wahrscheinlichkeit für Fehler Typ II: β

Fehler Typ II: Die Nullhypothese wird akzeptiert, obwohl sie falsch ist

Annahme: $\sigma = 2.4$ bekannt; H_0 falsch, z. B. $\boldsymbol{\mu = 21}$

$$X_i \sim N(\mu, \sigma) \Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow \bar{X} \sim N(\underline{\underline{21}}, 0.4) \quad (H_0 \text{ falsch})$$



$$\begin{aligned} \beta &= P(\text{Fehler Typ II}) \\ &= P(H_0 \text{ wird akzeptiert} \mid H_0 \text{ falsch}) \\ &= P(\bar{X} \leq 20.75 \mid \mu = 21) \\ &= F_{\bar{X}}(20.75) \quad \text{für } \mu = 21 \\ &= \Phi\left(\frac{20.75 - 21}{0.4}\right) \\ &= \Phi(-0.625) \\ &= 1 - \Phi(0.625) = 0.266 \end{aligned}$$

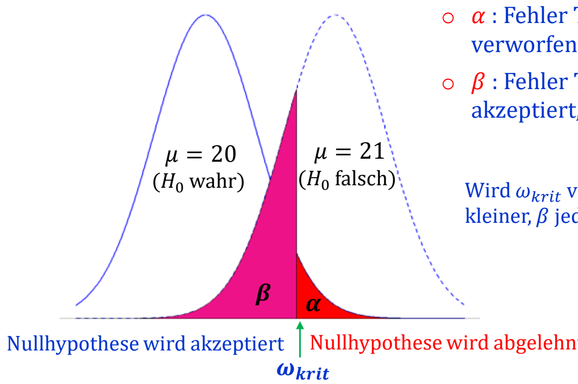
ca. 27% Risiko für Fehler Typ II

Man kann β verkleinern, indem man ω_{krit} **verkleinert**.

uwe.menzel@matstat.org

Kompromiss zwischen α und β

- Man kann α (Fehler Typ I) verkleinern, indem man ω_{krit} **vergrößert**.
- Man kann β (Fehler Typ II) verkleinern, indem man ω_{krit} **verkleinert**.
- → Kompromiss zwischen α und β notwendig!



- α : Fehler Typ I : H_0 wird verworfen, obwohl wahr
- β : Fehler Typ II : H_0 wird akzeptiert, obwohl falsch

Wird ω_{krit} vergrößert, wird α kleiner, β jedoch größer

www.matstat.org

Justierung der Wahrscheinlichkeit des Fehlers vom Typ I (α)

- In der Praxis wird α vor dem Test festgelegt.
- Dadurch entscheidet der Anwender, wie groß der Fehler vom Typ I sein soll.
- Der festgelegte Wert von α wird Signifikanzniveau genannt.
- Oft wird $\alpha = 0.05$ oder $\alpha = 0.01$ gewählt.
- Nach Festlegung von α kann ω_{krit} nun wie folgt berechnet werden:

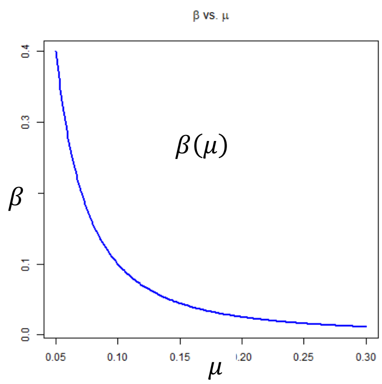
$$\begin{aligned}
\alpha &= P\left(H_0 \text{ wird verworfen} \mid H_0 \text{ wahr}\right) && \text{Fehler vom Typ I} \\
&= P\left(\bar{X} > \omega_{krit} \mid \mu = 20\right) && \omega_{krit} \text{ muss bestimmt werden} \\
&= 1 - P\left(\bar{X} \le \omega_{krit} \mid \mu = 20\right) && \text{immer unter } H_0\text{: } \mu = 20 \\
&= 1 - F_{\bar{X}}\left(\omega_{krit}\right) && F_{\bar{X}} = \text{Verteilungsfunktion von } \bar{X} \\
&= 1 - \Phi\left(\frac{\omega_{krit} - 20}{0.4}\right) && \text{Bestimmungsgleichung für } \omega_{krit} \text{ für vorgegebenes } \alpha \text{ (Tabelle für } \Phi \text{ nutzen)}
\end{aligned}$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

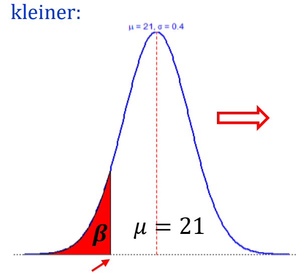
uwe.menzel@matstat.org

Justierung der Wahrscheinlichkeit des Fehlers vom Typ II (β) ?

- Der Fehler Typ II (β) kann nicht auf diese Weise justiert werden.
- Im Beispiel wurde für H_a mit $\mu = 21$ gerechnet. Diesen Wert kennen wir jedoch nicht (die alternative Hypothese war H_a: $\mu > 20$).
- Man kann eine Funktion $\beta(\mu)$ in einem gewissen Intervall berechnen:



Wenn μ größer wird, wird die Dichtefunktion nach rechts verschoben. Bei gleichem kritischen Wert wird damit β kleiner:



www.matstat.org

Wahl des Signifikanzniveaus α

- In der Praxis wird das Signifikanzniveau α vor dem Test festgelegt.
- Det kritische Wert ω_{krit} wird dann unter H_0 berechnet
- Oft werden $\alpha = 0.05$ oder $\alpha = 0.01$ gewählt.
  - $\alpha = 0.05$: "auf lange Sicht" wird H_0 in 1 von 20 Tests fälschlich verworfen
  - $\alpha = 0.01$: "auf lange Sicht" wird H_0 in 1 von 100 Tests fälschlich verworfen

Konsequenzen der Wahl von α :

... für das Beispiel " Trockenzeit":

Ein kleineres α, z. B. $\alpha = 0.01$ hat folgende Konsequenzen:

- ω_{krit} wird zu höheren Werten verschoben (20.75 → 20.93, siehe unten)
- Es wird unwahrscheinlicher, dass die Nullhypothese verworfen wird
- besser für Unternehmen – eventuell schlechter für Konsumenten

$$\alpha = 0.01 = 1 - \Phi\left(\frac{\omega_{krit} - 20}{0.4}\right) = 1 - \Phi(x)$$

$$x = \Phi^{-1}(0.99) = 2.326$$

$$\omega_{krit} = x \cdot 0.4 + 20 = \underline{\underline{20.93}}$$

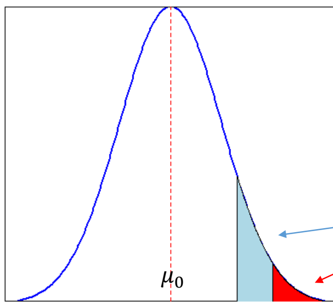
uwe.menzel@matstat.org

Wahl des Signifikanzniveaus α

Ein kleineres α bedeutet:

- Das Risiko für den Fehler Typ I wird vermindert, das Risiko für den Fehler Typ II (β) erhöht sich.
- Man kann sicherer sein, dass H_0 wirklich falsch ist, wenn H_0 verworfen wird.

Dichtefunktion der Teststatistik $\bar{X}$ **unter H_0**



wenn $\alpha = \alpha_2$ wird H_0 erst verworfen, wenn $\bar{x}_{obs}$ noch weiter von μ_0 abweicht (in Richtung H_a).

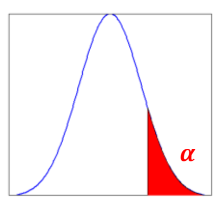
$\alpha_1 = 0.05$

$\alpha_2 = 0.01$

www.matstat.org

Verschiedene alternative Hypotesen (H_a)

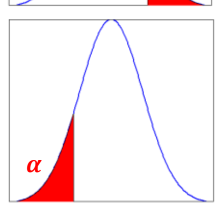
H_0	H_a	Test
$\mu = \mu_0$	$\mu > \mu_0$	einseitig



verwerfe H_0 für $\bar{x}_{obs} > \omega_\alpha$

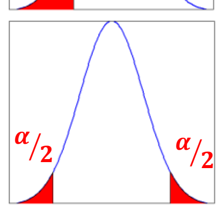
ω_α = kritischer Wert

H_0	H_a	Test
$\mu = \mu_0$	$\mu < \mu_0$	einseitig



verwerfe H_0 für $\bar{x}_{obs} < \omega_\alpha$

H_0	H_a	Test
$\mu = \mu_0$	$\mu \neq \mu_0$	zweiseitig



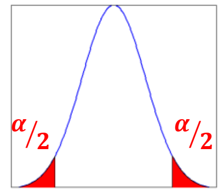
verwerfe H_0 für $\bar{x}_{obs} < -\,\omega_{\alpha/2}$ oder $\bar{x}_{obs} > +\,\omega_{\alpha/2}$ (für symmetrische Dichtefunktion)

www.matstat.org

Verschiedene alternative Hypothesen

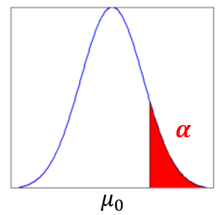
Beispiel, zweiseitiger Test:

- Ein Bauelement muss einen Durchmesser von 30 cm haben, dieser darf weder zu groß noch zu klein sein
- H_0: $\mu = 30$; verwerfe H_0 wenn
  - $\bar{x}_{obs} > 30.5$ oder
  - $\bar{x}_{obs} < 29.5$



Beispiel, einseitiger Test:

- Eine alte Maschine produziert durchschnittlich 2500 Komponenten pro Tag: $\mu_0 = 2500$ (normalverteilt) .
- Neue, teurere Maschine kaufen ?
  - nur wenn μ für die neue Maschine größer ist!
  - wähle H_a: $\mu > \mu_0$
  - nur kaufen wenn H_0 verworfen werden kann



www.matstat.org

Berechnung des kritischen Wertes aus α

wenn σ unbekannt ist

Wenn die Nullhypothese $\mu = \mu_0$ wahr ist, dann gilt $X_i \sim N(\mu_0, \sigma)$ und daher

$$\bar{X} \sim N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right) \Longrightarrow Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

eine $N(0,1)$-Variable wird oft mit Z bezeichnet

Die Standardabweichung σ ist jedoch unbekannt, deshalb kann Z nicht als Teststatistik benutzt werden. (Es kann kein numerischer Wert für eine Observation von Z berechnet werden). Wir wissen jedoch (Vorlesung **F9**), dass wir eine t-verteilte Variable erhalten, wenn σ im obigen Ausdruck durch den Schätzer S ersetzt wird:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$$

$t(n-1)$: Students t–Verteilung für $n-1$ Freiheitsgrade

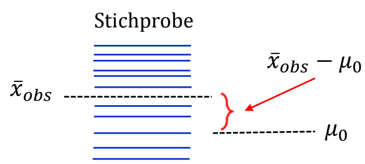
Die Teststatistik T ist indikativ für die Frage, ob der Erwartungswert einer Population den hypothetischen Wert μ_0 haben könnte. Weicht der beobachtete Mittelwert $\bar{x}_{obs}$ stark von μ_0 ab und ist gleichzeitig die beobachtete Standardabweichung s der Stichprobe nicht allzu groß, nimmt eine Observation von T große positive oder negative Werte an. Liegen diese in Richtung der alternativen Hypothese sind wir geneigt, diese anzuerkennen.

uwe.menzel@matstat.org

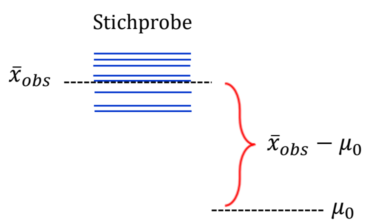
Die Teststatistik T ist indikativ für den Test

Eine Observation von T soll im Folgenden mit t_{obs} bezeichnet werden:

$$t_{obs} = \frac{\bar{x}_{obs} - \mu_0}{s/\sqrt{n}}$$



Der beobachtete Mittelwert $\bar{x}_{obs}$ weicht nicht stark von μ_0 ab, außerdem ist die beobachtete Standardabweichung s der Stichprobe groß. Eine Observation von T nimmt nicht allzu große positive oder negative Werte an. Die Nullhypothese wird **nicht** verworfen.

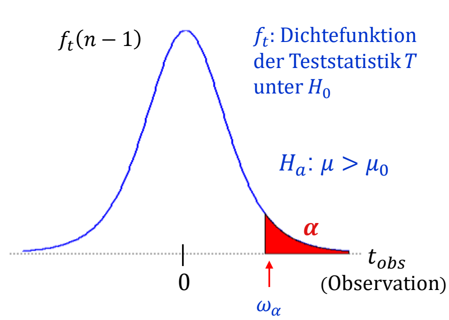


Der beobachtete Mittelwert $\bar{x}_{obs}$ weicht stark von μ_0 ab, außerdem ist die beobachtete Standardabweichung s der Stichprobe klein. Eine Observation von T nimmt große positive oder negative Werte an. Wenn diese Werte in Richtung der alternativen Hypothese liegen, kann die Nullhypothese verworfen werden.

www.matstat.org

Berechnung des kritischen Wertes aus α für H_a: $\mu > \mu_0$

Die t-Verteilung ist symmetrisch um den Nullpunkt. Wenn die Nullhypothese wahr ist, sollte die Observation t_{obs} nicht zu weit vom Nullpunkt entfernt liegen, da dann $\bar{x}_{obs} \approx \mu_0$. Wir sind geneigt, z. B. die alternative Hypothese H_a: $\mu > \mu_0$ anzunehmen, wenn $\bar{x}_{obs} \gg \mu_0$ und wenn gleichzeitig s nicht allzu groß ist. In diesem Fall nimmt t_{obs} große positive Werte an. Wenn schließlich $t_{obs} > \omega_a$ wird H_0 zugunsten der alternativen Hypothese verworfen.



$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t\,(n-1)$$

$$t_{obs} = \frac{\bar{x}_{obs} - \mu_0}{s/\sqrt{n}}$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

ω_α: kritischer Wert

www.matstat.org

Kritischer Wert für H_a: $\mu > \mu_0$

Für H_a: $\mu > \mu_0$ und ein zuvor festgelegtes Signifikanzniveau α kann der kritische Wert ω_α für die Teststatistik T aus der Forderung berechnet werden, dass die Wahrscheinlichkeit für den Fehler vom Typ I gleich α sein soll:

$$\boxed{P\left(\underbrace{\frac{\bar{X}-\mu_0}{S/\sqrt{n}} > \omega_\alpha}_{H_0 \text{ wird verworfen}} \mid H_0\ wahr\right) = \alpha}$$

Fehler Typ I

Bestimmungsgleichung für ω_α für vorgegebenes α (= Wahrscheinlichkeit für Fehler Typ I)

Unter H_0 ist der linke Term in der Klammer $t(n-1)$-verteilt, wir können also schreiben (womit die Bedingung "| H_0 *wahr*""verarbeitet" ist):
(Der Ausdruck ist dann und nur dann t-verteilt, wenn H_0 wahr ist, also wenn $\mu = \mu_0$)

$$P\left(T > \omega_\alpha\right) = \alpha \quad \text{mit} \quad T \sim t(n-1)$$

Allgemein gilt: $P\left(T > t_\alpha(n-1)\right) = \alpha$

$t_\alpha(n-1)$ = Quantil füt t-Verteilung mit $f = n-1$ und Signifikanzniveau α

Durch Vergleich der beiden letzten Beziehungen erhalten wir für den kritischen Wert $\omega_\alpha = t_\alpha(n-1)$. H_0 wird also verworfen, wenn eine Observation von T größer als $t_\alpha(n-1)$ wird.

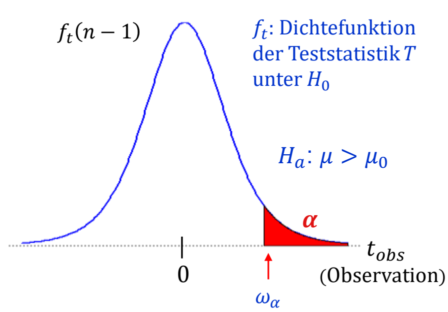
Kritischer Wert für H_a: $\mu > \mu_0$

Die Nullhypothese wird verworfen, wenn $t_{obs} > t_\alpha(n-1)$. Diese Werte bilden die kritische Region für t_{obs} zum Signifikanzniveau α. Die kritische Region wird mit Ω_α bezeichnet:

$$\Omega_\alpha = \{t_{obs} > t_\alpha(n-1)\}$$

$$T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t\,(n-1)$$

$$t_{obs} = \frac{\bar{x}_{obs}-\mu_0}{s/\sqrt{n}}$$



f_t: Dichtefunktion der Teststatistik T unter H_0

H_0 wird verworfen, wenn die Observation t_{obs} in der kritischen Region (rot) landet.

$t_\alpha(n-1)$ = Quantil für t-Verteilung mit $f = n-1$ für Signifikanzniveau α

Kritischer Wert für $\bar{X}$ für H_a: $\mu > \mu_0$

Es ist üblich, den Test mit der Teststatistik T durchzuführen. Aus dem oben berechneten kritischen Wert für t_{obs} lässt sich aber auch leicht ein kritischer Wert für $\bar{x}_{obs}$ herleiten. Wir hatten festgestellt, dass die Nullhypothese verworfen wird, wenn $t_{obs} > t_\alpha(n-1)$ also wenn

$$t_{obs} = \frac{\bar{x}_{obs} - \mu_0}{s/\sqrt{n}} > t_\alpha(n-1)$$

H_0 wird dann verworfen wenn $\bar{x}_{obs} > \mu_0 + t_\alpha(n-1) \cdot \frac{s}{\sqrt{n}}$

Der auf $\bar{X}$ bezogene kritische Wert ist daher

$$\omega_{\bar{X}} = \mu_0 + t_\alpha(n-1) \cdot \frac{s}{\sqrt{n}}$$

Die Nullhypothese wird also verworfen für $\bar{x}_{obs} > \omega_{\bar{X}}$.

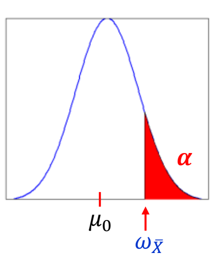
Die kritischen Werte für t_{obs} bzw. $\bar{x}_{obs}$ für die anderen alternativen Hypothesen können analog hergeleitet werden (siehe Vorlesung **F12**). Im Folgenden werden die kritischen Regionen für $\bar{x}_{obs}$ gelistet:

www.matstat.org

Kritische Werte für $\bar{x}_{obs}$ für verschiedene alternative Hypothesen

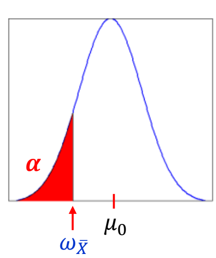
Die Lokalisation der kritischen Region (rot) für $\bar{x}_{obs}$ hängt von der Wahl der alternativen Hypothese ab.

H_a: $\mu > \mu_0$



$\omega_{\bar{X}} = \mu_0 + t_\alpha(n-1) \cdot \frac{s}{\sqrt{n}}$

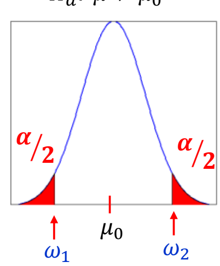
H_a: $\mu < \mu_0$



$\omega_{\bar{X}} = \mu_0 - t_\alpha(n-1) \cdot \frac{s}{\sqrt{n}}$

einseitige Tests

H_a: $\mu \neq \mu_0$



$\omega_1 = \mu_0 - t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$

$\omega_2 = \mu_0 + t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$

zweiseitiger Test

uwe.menzel@matstat.org

Zusammenfassung

Vergleich der Teststatistik mit einem kritischen Wert

1. Definiere die Nullhypothese (H_0) und die alternative Hypothese (H_a)
2. Setze das Signifikanzniveau fest: $\alpha = P(Fehler\ Typ\ I \mid H_0)$
3. Berechne den kritischen Wert ω_{krit} der Teststatistik. Die Verteilung der Teststatistik unter H_0 muss bekannt sein.
4. Berechne die Stichproben-Teststatistik ($\bar{x}_{obs}$, t_{obs},...).
5. Verwerfe H_0 auf Signifikanzniveau α wenn die Stichproben-Teststatistik in der kritischen Region liegt, sonst "akzeptiere" H_0

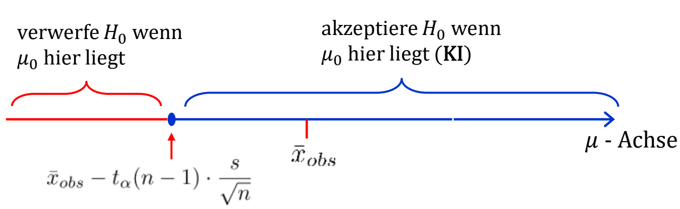
www.matstat.org

2. Vergleich Teststatistik ⇄ Konfidenzintervall

Einseitiger Test H_0: $\mu = \mu_0$ H_a: $\mu > \mu_0$

Wir hatten: verwerfe H_0 wenn $\bar{x}_{obs} > \mu_0 + t_\alpha(n-1) \cdot \frac{s}{\sqrt{n}}$

Das bedeutet: verwerfe H_0 wenn $\mu_0 < \bar{x}_{obs} - t_\alpha(n-1) \cdot \frac{s}{\sqrt{n}}$



Zur Erinnerung: einseitiges Konfidenzintervall für μ (Vorlesung **F9**):

$$I_\mu = \left(\bar{x}_{obs} - t_\alpha(n-1) \cdot \frac{s}{\sqrt{n}}\ ;\ +\infty \right)$$

dies ist genau der oben blau markierte Bereich

Die Nullhypothese wird **nicht** verworfen ("akzeptiert"), wenn der hypothetische Parameter innerhalb des Konfidenzintervalles liegt.

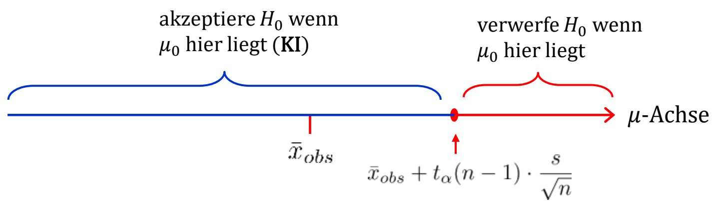
uwe.menzel@matstat.org

Vergleich Teststatistik ⇄ Konfidenzintervall

Einseitiger Test $H_0\colon \mu = \mu_0 \quad H_a\colon \mu < \mu_0$

Wir hatten: verwerfe H_0 wenn $\bar{x}_{obs} < \mu_0 - t_\alpha(n-1) \cdot \frac{s}{\sqrt{n}}$

Das bedeutet: verwerfe H_0 wenn $\mu_0 > \bar{x}_{obs} + t_\alpha(n-1) \cdot \frac{s}{\sqrt{n}}$



Zur Erinnerung: einseitiges Konfidenzintervall für μ (Vorlesung **F9**):

$$I_\mu = \left(-\infty\ ;\ \bar{x}_{obs} + t_\alpha(n-1) \cdot \frac{s}{\sqrt{n}}\right)$$

dies ist genau der oben blau markierte Bereich

Die Nullhypothese wird **nicht** verworfen ("akzeptiert"), wenn der hypothetische Parameter innerhalb des Konfidenzintervalles liegt.

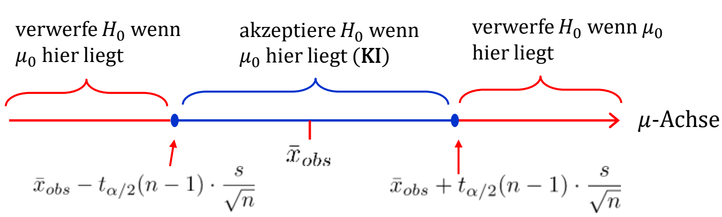
uwe.menzel@matstat.org

Vergleich Teststatistik ⇄ Konfidenzintervall

Zweiseitiger Test $H_0\colon \mu = \mu_0 \quad H_a\colon \mu \neq \mu_0$

Verwerfe H_0 wenn $\bar{x}_{obs} < \mu_0 - t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$ oder $\bar{x}_{obs} > \mu_0 + t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$

→ verwerfe H_0 wenn $\mu_0 > \bar{x}_{obs} + t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$ oder $\mu_0 < \bar{x}_{obs} - t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$



Zur Erinnerung: zweiseitiges Konfidenzintervall für μ (Vorlesung **F9**):

$$I_\mu = \bar{x}_{obs} \pm t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$$

dies ist genau der oben blau markierte Bereich

Die Nullhypothese wird **nicht** verworfen ("akzeptiert"), wenn der hypothetische Parameter innerhalb des Konfidenzintervalles liegt.

uwe.menzel@matstat.org

Zusammenfassung

Vergleich der Teststatistik mit dem Konfidenzintervall

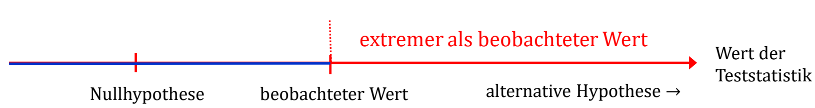
1. Auf der Grundlage einer Stichprobe wird ein Konfidenzintervall mit dem Konfidenzniveau $1-\alpha$ für den zu testenden Parameter θ berechnet.
2. Akzeptiere $H_0: \theta = \theta_0$ auf dem Signifikanzniveau α wenn der hypothetische Parameter θ_0 innerhalb des berechneten Konfidenzintervalles liegt.
3. Verwerfe $H_0: \theta = \theta_0$ auf dem Signifikanzniveau α wenn der hypothetische Parameter θ_0 außerhalb des berechneten Konfidenzintervalles liegt.

- Die Methode des kritischen Wertes und die Konfidenzmethode führen zu den gleichen Schlussfolgerungen.

3. Vergleich p-Wert ⇄ Signifikanzniveau

Definition p-Wert: Der p-Wert ist die Wahrscheinlichkeit für mindestens so extreme Testresultate wie das aktuell beobachte, berechnet unter der Annahme dass die Nullhypothese wahr ist.

Das Testresultat ist hier der erhaltene Wert der Teststatistik. Werte der Teststatistik werden als ”extremer” angesehen, wenn sie noch weiter von der Nullhypothese entfernt in Richtung der alternativen Hypothese liegen als der beobachtete Wert; sie sind unter H_0 noch unwahrscheinlicher als der beobachtete Wert.



Zur Berechnung der obigen Wahrscheinlichkeit muss man eine Teststatistik wählen, deren Verteilung unter H_0 bekannt ist - den Schätzer für den Parameter oder eine vom Schätzer abgeleitete Größe, z. B. den standardisierten Schätzer.

p-Wert für Normalverteilung

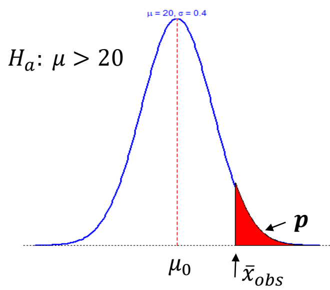
Beispiel 1: Es soll getestet werden, ob der Erwartungswert einer normalverteilten Population $\mu = 20$ oder größer ist:

Hypothesen: H_0: $\mu = \mu_0 = 20$; H_a: $\mu > 20$

Teststatistik: Schätzer für μ, also Mittelwert $\bar{X}$

Verteilung der Teststatistik "unter H_0": $X_i \sim N(\mu_0, \sigma) \Longrightarrow \bar{X} \sim N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)$

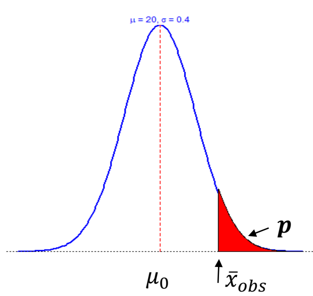
Wir nehmen hier der Einfachheit halber an, dass die Standardabweichung σ bekannt ist. Dann ist die Verteilung von $\bar{X}$ unter H_0 vollständig bekannt.



Die blaue Kurve sei die Dichtefunktion für $\bar{X}$ unter der Annahme, dass die Nullhypothese wahr ist: $f_{\bar{X}}(x \mid H_0)$. Der beobachtete Wert der Teststatistik sei $\bar{x}_{obs}$. Dann entspricht die Fläche des rot markierten Bereiches dem p-Wert = Wahrscheinlickeit der Beobachtung $\bar{x}_{obs}$ oder noch extremerer Werte in Richtung der alternativen Hypothese.

www.matstat.org

p-Wert für Normalverteilung



Die blaue Kurve sei die Dichtefunktion für $\bar{X}$ unter der Annahme, dass die Nullhypothese wahr ist: $f_{\bar{X}}(x \mid H_0)$ Der beobachtete Wert der Teststatistik sei $\bar{x}_{obs}$. Dann entspricht die Fläche des rot markierten Bereiches dem p-Wert.

$$P_{wert} = \int_{\bar{x}_{obs}}^{\infty} f_{\bar{X}}(x \mid H_0)\, dx$$

für H_a: $\mu > 20$

Da das Integral nicht immer leicht zu berechnen ist, ist es günstiger, den standardisierten Mittelwert als Testvariable zu wählen, dazu mehr später ...

Ein sehr kleiner p-Wert bedeutet:

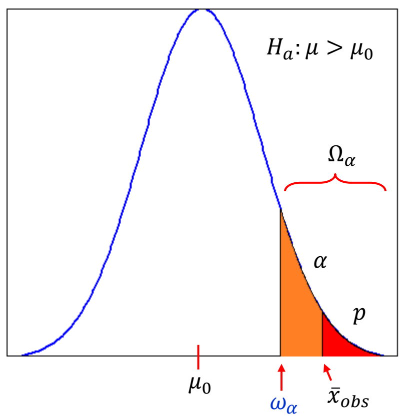
- Der beobachtete Wert der Teststatistik (= "Observation") ist sehr unwahrscheinlich, wenn man annimmt, dass die Nullhypothese H_0 zutrifft.
- Die Nullhypothese wird daher verworfen.

www.matstat.org

p–Wert und kritische Region

- Wenn die Observation $\bar{x}_{obs}$ in der kritischen Region für das Signifikanzniveau α liegt, ist der p – Wert kleiner als α und umgekehrt.
- Die Nullhypothese wird also für $p < \alpha$ zurückgewiesen (auf Signifikanzniveau α).

Dichtefunktion der Teststatistik $\bar{X}$ **unter H_0**



$\alpha = P(Fehler\ Typ\ I)$

Ω_α = kritische Region

ω_α = kritischer Wert

rote Fläche = p - Wert

$\boxed{H_0 \text{ wird verworfen, wenn } p < \alpha}$

www.matstat.org

p-Wert für Binomialverteilung

Beispiel 2: Wir haben einen Behälter mit mehreren Würfeln, von denen die meisten faire Würfel sind, einige jedoch gefälscht sind und mehr als eine Sechs enthalten. Wir ziehen zufällig einen der Würfel und sollen nun anhand der Ergebnisse von 10 Würfen – ohne den Würfel selbst zu untersuchen! – feststellen, ob wir einen der gefälschten Würfel gezogen haben.

Zufallsvariable X : Augenzahl bei einem Wurf. Sei $p = P(X = 6)$.

Hypothesen: $H_0: p = 1/6$; $H_a: p > 1/6$ (Nullhypothese = fairer Würfel)

Teststatistik (T): Anzahl der gewürfelten Sechsen bei insgesamt 10 Würfen. Unter H_0 ist T binomialverteilt mit $T \sim Bin(10, 1/6)$, denn für einen fairen Würfel beträgt die Wahrscheinlichkeit, eine Sechs zu würfeln gerade 1/6.

Stichprobe: Wir nehmen an, dass die Sechs 9 Mal gewürfelt wurde, d.h. $T = 9$

p-Wert: Die Observation ist $T = 9$. Noch extremer in Richtung von H_a wäre $T = 10$. Der p-Wert ist also laut der Definition die Summe der Wahrscheinlichkeiten für $T = 9$ und $T = 10$.

www.matstat.org

p-Wert für Binomialverteilung

Hypothesen: $H_0: p = 1/6$; $H_a: p > 1/6$ (Nullhypothese = fairer Würfel)

Teststatistik (T): $T \sim Bin(10, 1/6)$ unter H_0

Stichprobe: $T = 9$

p-Wert: $P_{Wert} = P(T \geq 9 \mid H_0)$

Zur Erinnerung: $P(T = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$ wenn $T \sim Bin(n, p)$

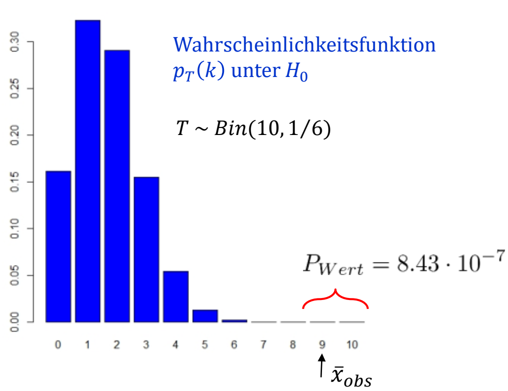
$$\begin{aligned} P_{Wert} &= P(T \geq 9 \mid H_0) = P(T = 9 \mid H_0) + P(T = 10 \mid H_0) \\ &= \binom{10}{9} \cdot (1/6)^9 \cdot (5/6)^1 + \binom{10}{10} \cdot (1/6)^{10} \cdot (5/6)^0 \\ &= 8.27 \cdot 10^{-7} + 1.65 \cdot 10^{-8} = 8.43 \cdot 10^{-7} \end{aligned}$$

Ergebnis: Der p-Wert ist sehr klein. Die erhaltene Beobachtung ist also sehr unwahrscheinlich, wenn man annimmt, dass die Nullhypothese zutrifft. Die Nullhypothese (fairer Würfel) wird daher – sogar auf Signifikanzniveau $\alpha = 10^{-6}$ - verworfen! (Das Signifikanzniveau muss **vor** dem Test festgelegt werden!)

Anmerkung: Bei Vorliegen eines solchen Testergebnisses hätte wohl auch der "gesunde Menschenverstand" die Hypothese des fairen Würfels zurückgewiesen.

www.matstat.org

p-Wert für Binomialverteilung



Ergebnis: $p_{wert} < \alpha = 10^{-6}$. Die Nullhypothese kann also sogar auf Signifikanzniveau $\alpha = 10^{-6}$ verworfen werden. Die Wahrscheinlichkeit, einen Fehler vom Typ I zu begehen ist sehr klein.

www.matstat.org

p-Wert für Normalverteilung

Beispiel 3: Oberflächen von bestimmten elektronischen Komponenten müssen eine 30 µm dicke Kupferschicht haben. Die Schicht darf weder zu dünn noch zu dick sein. Die Oberflächen von $n = 10$ Komponenten wurden gemessen, der Mittelwert der Messungen war $\bar{x}_{obs} = 30.91$.
Aus langer Erfahrung ist außerdem bekannt, dass die Dicke normalverteilt mit $\sigma = 0.788$ ist. (Wir werden später Probleme mit unbekanntem σ rechnen ...).

Hypothesen: $H_0: \mu = 30$; $H_a: \mu \neq 30$

Teststatistik: Schätzer für μ, also $\bar{X}$

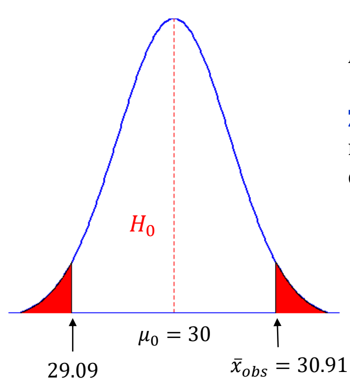
Wenn H_0 wahr ist, gilt $\mu = \mu_0 = 30$, also:

$$\bar{X} \sim N\left(30, \frac{0.788}{\sqrt{10}}\right) \Rightarrow \bar{X} \sim N(30, 0.249)$$

Die alternative Hypothese ist $H_a: \mu \neq 30$, μ darf weder zu groß noch zu klein sein. Die kritische Region liegt auf **beiden** Seiten der Dichtefunktion für die Nullhypothese.

p-Wert für Normalverteilung

- Alternative Hypoth.: $H_a: \mu \neq 30$ $\rightarrow \mu$ darf weder zu groß noch zu klein sein.
- Der p-Wert setzt sich nun aus zwei Flächen zusammen (die im Falle einer symmetrischen Dichtefunktion um μ_0 den gleichen Anstand von μ_0 haben).
- $\rightarrow$ zweiseitiger Test

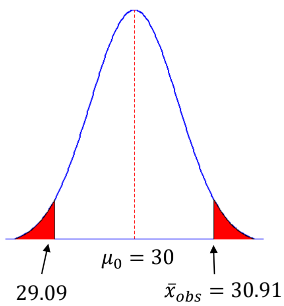


$H_0: \mu = 30$; $H_a: \mu \neq 30$

Zweiseitiger Test: beide rot markierte Regionen sind extremer als $\bar{x}_{obs}$!

p-Wert für Normalverteilung

$H_0: \mu = 30$; $H_a: \mu \neq 30$ (zweiseitiger Test) $\bar{x}_{obs} = 30.91$



Zweiseitiger Test: beide markierte Regionen sind extremer als $\bar{x}_{obs}$!

Zur Berechnung des p-Wertes können wir die Symmetrie der Dichtefunktion ausnutzen – beide rot markierte Flächen sind gleich groß:

$$
\begin{aligned}
p &= 2 \cdot P\left(\bar{X} > \bar{x}_{obs} \mid H_0\right) \\
&= 2 \cdot P\left(\bar{X} > 30.91 \mid H_0\right) \\
&= 2 \cdot \left[1 - P\left(\bar{X} \leq 30.91\right)\right] \\
&= 2 \cdot \left[1 - F_{\bar{X}}\left(30.91\right)\right] \\
&= 2 \cdot \left[1 - \Phi\left(\frac{30.91 - 30}{0.249}\right)\right] \\
&= 2 \cdot \left[1 - \Phi\left(3.65\right)\right] \\
&= 2 \cdot \left[1 - 0.99987\right] = 2.6 \cdot 10^{-4}
\end{aligned}
$$

H_0

Ergebnis: Der p-Wert ist kleiner als $10^{-3} \rightarrow H_0$ kann auf Signifikanzniveau $\alpha = 10^{-3}$ verworfen werden. Die Abweichungen von der vorgeschriebenen Dicke sind zu groß.

www.matstat.org

p-Wert für Binomialverteilung

Beispiel 4: (Alm, Britton: **Stokastik**, Liber AB, Stockholm 2008)

Eine Person behauptet so fingerfertig zu sein, dass sie Einfluss darauf hat, welche Seite bei einem Münzwurf nach oben kommt, so dass sie ”Zahl” öfter als ”Wappen” werfen kann, d. h. $p = P(Zahl) > {}^1/_2$. Wir sind skeptisch und glauben, dass $p = {}^1/_2$. Um die Behauptung der Person zu testen, soll diese 10 Würfe machen. Wir sind bereit der Person zu glauben, wenn sie ausreichend oft ”Zahl” werfen kann, genauer: wenn der p-Wert kleiner als $\alpha = 0.01$ ist. (d.h. wir nehmen eine Wahrscheinlichkeit von 0.01 für einen Fehler Typ I in Kauf).

Hypothesen: H_0: $p = {}^1/_2$; $H_a: p > {}^1/_2$ (einseitiger Test)

Teststatistik (T): Anzahl geworfener ”Zahl” $T \sim Bin(10, 1/2)$ unter H_0

Stichprobe: $T = 8$ Es wurde acht Mal ”Zahl” geworfen.

p-Wert: $P_{Wert} = P\left(T \geq 8 \mid H_0\right)$

Signifikanzniveau: $\alpha = 0.01$

www.matstat.org

p-Wert für Binomialverteilung

$T \sim Bin(10, 1/2)$ unter H_0 $\quad$ p-Wert: $\quad P_{Wert} = P(T \geq 8 \mid H_0)$

Observation $\quad$ extremer als Observation

$$P_{wert} = P(X \geq 8 \mid H_0) = p_X(8) + p_X(9) + p_X(10)$$
$$= \binom{10}{8}(1/2)^{10} + \binom{10}{9}(1/2)^{10} + \binom{10}{10}(1/2)^{10} = \underline{\underline{0.0986}}$$

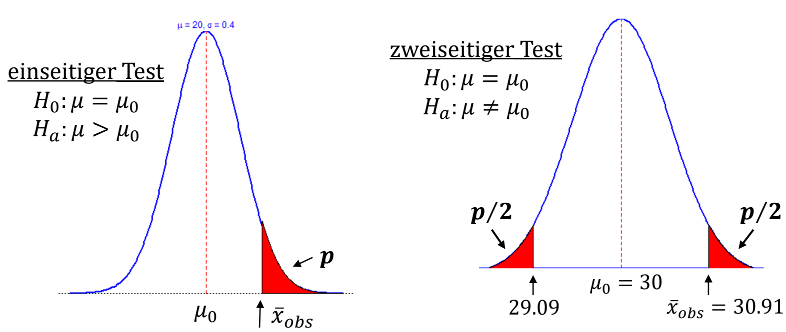
Ergebnis: Das Signifikanzniveau α wurde mit 0.01 festgelegt. Der berechnete p-Wert ist jedoch größer: $p > \alpha$. Die Nullhypothese wird deshalb **nicht** verworfen ("auf Signifikanzniveau 0.01"). Wir dürfen also weiterhin zweifeln, dass diese Person außergewöhnlich geschickt ist, denn wir halten es nicht für so unwahrscheinlich, rein zufällig in 8 von 10 Würfen "Zahl" zu bekommen.

Anmerkung: Damit ist die Nullhypothese jedoch **nicht** "bewiesen"! Vielleicht reichte einfach die Anzahl der Würfe nicht – hätte die Person 80 von 100 mal "Zahl" geworfen, sähe das Ergebnis ganz anders aus. Eine Nullhypothese kann abgelehnt, aber nicht "bewiesen" werden!

www.matstat.org

Zusammenfassung: p-Wert-Methode

1. Definiere die Nullhypothese (H_0) und die alternative Hypothese (H_a)
2. Setze das Signifikanzniveau fest: $\alpha = P(Fehler\ Typ\ I \mid H_0)$
3. Lege die Teststatistik ($\bar{X}_{obs}, T, \ldots$) fest. Die Verteilung der Teststatistik unter H_0 muss bekannt sein.
4. Berechne den p – Wert (laut Definition).
5. Verwerfe H_0 auf Signifikanzniveau α wenn $p < \alpha$, sonst "akzeptiere" H_0



www.matstat.org