Statistische Hypothesentests

- Allgemeines
- Tests in normalverteilten Grundgesamtheiten
- Asymptotische Tests

Allgemeines zu Tests

- Statistischer Test: Verfahren (Entscheidungsregel),
  - mit dem auf Basis einer einfachen Stichprobe $X_1, X_2, \ldots, X_n$
  - anhand einer geeigneten Stichprobenfunktion (= Teststatistik oder Prüfgröße) $T_n(X_1, \ldots, X_n)$ entschieden wird,
  - ob eine Behauptung (= statistische Hypothese) bezüglich der Verteilung der Grundgesamtheit
  - abgelehnt wird oder nicht.

- Einfache (statistische) (Verteilungs-) Hypothese:

$$H: \ X \text{ folgt der Verteilung } F_X(x;\theta) \text{ mit } \theta = \theta_0.$$

- Beispiel: Sei $X_1, \ldots, X_n$ einfache Zufallsstichprobe aus normalverteilter Grundgesamtheit mit unbekanntem Mittelwert und bekannter Varianz $\sigma^2 = 36$.

  Einfache Hypothese:

$$H: \ \mu = 1000.$$

  Mögliche Prüfgröße: Stichprobenmittel $\overline{X}_n$.

  Mögliche Entscheidungsregel: Lehne Hypothese ab, wenn Stichprobenmittel $\overline{X}_n$ kleiner ist als 950 bzw. größer ist als 1100.

- Zusammengesetzte Hypothese:

$$H: \ X \text{ folgt der Verteilung } F_X(x;\theta) \text{ mit } \theta \geq \theta_0.$$

oder

$$H: \ X \text{ folgt der Verteilung } F_X(x;\theta) \text{ mit } \theta \leq \theta_0.$$

oder

$$H: \ X \text{ folgt der Verteilung } F_X(x;\theta) \text{ mit } \theta \neq \theta_0.$$

- Beispiel: Sei $X_1, \ldots, X_n$ einfache Zufallsstichprobe aus normalverteilter Grundgesamtheit mit unbekanntem Mittelwert und bekannter Varianz $\sigma^2 = 36$.

Zusammengesetzte Hypothese:

$$H: \ \mu \geq 990.$$

Mögliche Entscheidungsregel: Lehne Hypothese ab, wenn Stichprobenmittel kleiner ist als 950.

- Fragen:
  - Warum Stichprobenmittel als Prüfgröße?
  - Wie werden kritischen Grenzen festgelegt, ab denen die Hypothese abzulehnen ist?
  - Wie gut ist eine solche Entscheidung?
  - Für welche andere Hypothese entscheidet man sich, wenn Hypothese abgelehnt wird?

Null- (H_0) und Alternativhypothese (H_1)

- Z.B. für einfache Hypothesen

$$H_0: \ F_X(x) = F_X(x;\theta_0) \ \text{ gegen } \ H_1: \ F_X(x) = F_X(x;\theta_1) \ .$$

- Ziel: Entscheidung aufgrund einer Stichprobe zwischen H_0 und H_1.
- Beispiel: X Abfüllgewicht von Zuckertüten

$$H_0: \ \mu_X = E[X] = 1000g \ \text{ gegen } \ H_1: \ \mu_X = E[X] = 1010g$$

- Beachte: Bleibt Alternativhypothese unspezifiziert, so umfasst sie alle Parameterwerte, die nicht in der Nullhypothese enthalten sind.
- Beispiel:

$$H_0: \ \mu_X \geq 990g$$

führt zu

$$H_1: \ \mu_X < 990g$$

- Bezeichnung: Θ_0 (Θ_1) = Parameterbereich, der zur Null- (Alternativ-) Hypothese gehört.

Mögliche Fehler bei Alternativentscheidungen unter Unsicherheit

- Mögliche Entscheidungen:

Zustand in der Grundgesamtheit	Testentscheidung: H_0 ablehnen	Testentscheidung: H_0 nicht ablehnen
H_0 ist richtig	Fehler 1. Art	kein Fehler
H_0 ist falsch	kein Fehler	Fehler 2. Art

- Wahrscheinlichkeiten der möglichen Entscheidungen:

Zustand in der Grundgesamtheit	Testentscheidung: H_0 ablehnen	Testentscheidung: H_0 nicht ablehnen
H_0 ist richtig	α	$1-\alpha$
H_0 ist falsch	$1-\beta$	β

- $\alpha =$ Irrtumswahrscheinlichkeit = Wahrscheinlichkeit des Fehlers 1. Art = Wahrscheinlichkeit, Nullhypothese abzulehnen, obwohl sie richtig ist.

  Bez.: Fehler 1. Art = α-Fehler

- $\beta =$ Wahrscheinscheinlichkeit des Fehlers 2. Art = Wahrscheinlichkeit, Nullhypothese nicht abzulehnen, obwohl sie falsch ist.

  Bez.: Fehler 2. Art = β-Fehler

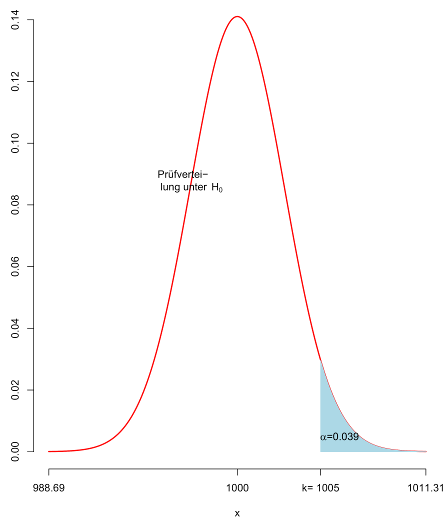
- Beispiel: Berechnung der Wahrscheinlichkeit des Fehlers 1. Art α

  Einfache Nullhypothese $H_0: \ \mu = 1000$.

  Entscheidungsregel: ,,Lehne H_0 ab, wenn das Stichprobenmittel $\overline{X}_n$ größer als $k = 1005$g ist."

  Seien $n = 18$ und $\sigma = 12$.

$$\begin{aligned}\alpha &= P(\overline{X}_n > 1005|H_0) = P(\overline{X}_n > 1005|\mu = 1000)\\ &= 1 - \Phi\left(\frac{\sqrt{n}(1005-\mu)}{\sigma}\right)\\ &= 1 - \Phi\left(\frac{\sqrt{18}(1005-1000)}{12}\right) = 1 - \Phi(1.768) = 0.039.\end{aligned}$$



Festlegung des kritischen Bereiches

- Ablehnungsbereich K (kritischer Bereich) = Menge aller Stichprobenbefunde, für die die Nullhypothese abzulehnen ist.
- Strategie: Festlegung des kritischen Bereiches durch Vorgabe der Irrtumswahrscheinlichkeit α:

  D.h. K wird durch

  $$P((X_1, \ldots, X_n) \in K | H_0) = \alpha.$$

  festgelegt.

Beispiel: Mittelwerttest für eine normalverteilte Grundgesamtheit, d.h. $X \sim \mathcal{N}(\mu, \sigma^2)$ mit bekanntem σ^2

- Entscheidung zwischen zwei einfachen Hypothesen:

  $$H_0 : \mu = \mu_0 \text{ gegen } H_1 : \mu = \mu_1 \text{ mit } \mu_1 > \mu_0.$$

- Plausible Entscheidungsregel: Ablehnung der Nullhypothese, wenn Stichprobenmittel größer als eine kritische Schranke k ist.
- Verteilung der Prüfgröße unter H_0:

  $$\overline{X}_n \sim \mathcal{N}(\mu_0, \sigma^2/n).$$

- Wahrscheinlichkeit α für Fehler 1. Art:

  $$\begin{aligned} \alpha &= P(H_0 \text{ ablehnen} | H_0 \text{ gilt}) = P((X_1, \ldots, X_n) \in K | \mu = \mu_0) \\ &= P(\overline{X}_n > k | \mu = \mu_0) \\ &= 1 - F_N(k; \mu_0, \sigma^2/n) = 1 - \Phi\left(\sqrt{n}\frac{k - \mu_0}{\sigma}\right). \end{aligned}$$

  D.h.

  $$\Phi\left(\sqrt{n}\frac{k - \mu_0}{\sigma}\right) = 1 - \alpha$$

  und

  $$\sqrt{n}\frac{k - \mu_0}{\sigma} = \lambda_{1-\alpha} \Longrightarrow k = \mu_0 + \lambda_{1-\alpha}\sigma/\sqrt{n}.$$

- Kritischer Bereich: Lehne H_0 mit Irrtumswahrscheinlichkeit α ab, wenn Stichprobenbefund $(x_1, \ldots, x_n)$ in kritischen Bereich

$$K = \left\{ (x_1, \ldots, x_n) | \sqrt{n} \frac{\overline{X}_n - \mu_0}{\sigma} > \lambda_{1-\alpha} \right\}$$

  fällt.

- Kritischer Wert:

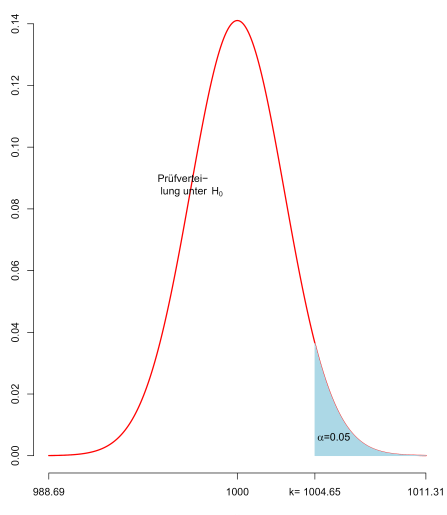
$$k = \mu_0 + \lambda_{1-\alpha} \frac{\sigma}{\sqrt{n}}.$$

- Zahlenbeispiel: $\mu_0 = 1000$, $\mu_1 = 1010$, $\sigma = 12$, $n = 18$, $\alpha = 0.05$

$$k = 1000 + \frac{12}{\sqrt{18}} \lambda_{0.95} = 1004.625.$$

  D.h.

$$K = \{(x_1, \ldots, x_{18}) | \overline{x}_{18} > 1004.652\}.$$



Kann man H_0 annehmen?

- Wenn

$$(x_1, \ldots, x_n) \notin K$$

  wird Nullhypothese nicht abgelehnt.

- Problem: Annahme von H_0 ist nicht möglich, da die Wahrscheinlichkeit, H_0 nicht abzulehnen, obwohl H_0 falsch ist ($=\beta$-Fehler) nicht kontrolliert werden kann.

  Sprachgebrauch statt Annahme von H_0: H_0 kann nicht abgelehnt werden.

- Lösung für einfache Hypothesen: Berechnung der Wahrscheinlichkeit des β-Fehlers

$$\beta = P((X_1, \ldots, X_n) \notin K | H_1)$$

Sprachgebrauch: Wenn β relativ klein (z.B. kleiner als α), dann kann H_0 angenommen werden.

Beispiel: Mittelwerttest für eine normalverteilte Grundgesamtheit, d.h. $X \sim \mathcal{N}(\mu, \sigma^2)$ mit bekanntem σ^2

- Entscheidung zwischen zwei einfachen Hypothesen:

$$H_0 : \mu = \mu_0 \text{ gegen } H_1 : \mu = \mu_1 \text{ mit } \mu_1 > \mu_0.$$

- Verteilung der Prüfgröße unter H_1:

$$\overline{X}_n \sim \mathcal{N}(\mu_1, \sigma^2/n).$$

- Wahrscheinlichkeit des Fehlers 2. Art β:

$$\begin{aligned}
\beta &= P(H_0 \text{ nicht ablehnen} | H_1) \\
&= P((X_1, \ldots, X_n) \notin K | \mu = \mu_1) \\
&= P(\overline{X}_n \leq k | \mu = \mu_1) \\
&= F_N(k; \mu_1, \sigma^2/n) = \Phi\left(\sqrt{n}\frac{k - \mu_1}{\sigma}\right) \\
&= \Phi\left(\lambda_{1-\alpha} + \sqrt{n}\frac{\mu_0 - \mu_1}{\sigma}\right)
\end{aligned}$$

wegen

$$k = \mu_0 + \lambda_{1-\alpha}\frac{\sigma}{\sqrt{n}}.$$

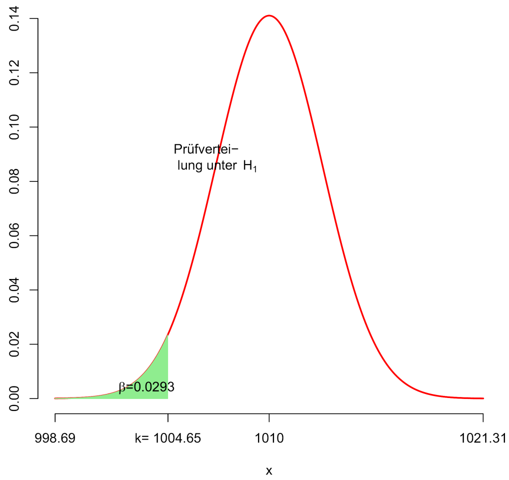
- Zahlenbeispiel: $\mu_0 = 1000$, $\mu_1 = 1010$, $\sigma = 12$, $n = 18$, $\alpha = 0.05$

$$k = 1000 + \lambda_{0.95}\frac{12}{\sqrt{18}} = 1004.652.$$

D.h.

$$\beta = \Phi\left(\lambda_{0.95} + \sqrt{18}\frac{1000 - 1010}{12}\right) = \Phi(-1.891) = 0.0293.$$

Da $\beta < \alpha = 0.05$, kann H_0 nicht nur nicht abgelehnt, sondern sogar mit einer Wahrscheinlichkeit des β-Fehlers von weniger als 5% angenommen werden, wenn der Stichprobenbefund einen empirischen Wert des Prüfmaßes liefert, der nicht in kritischen Bereich fällt.

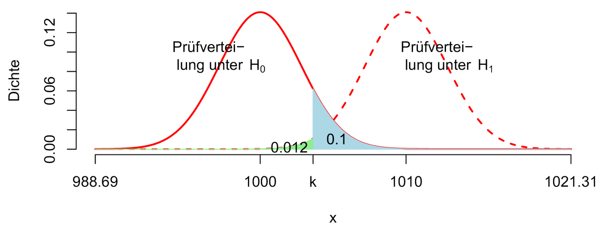


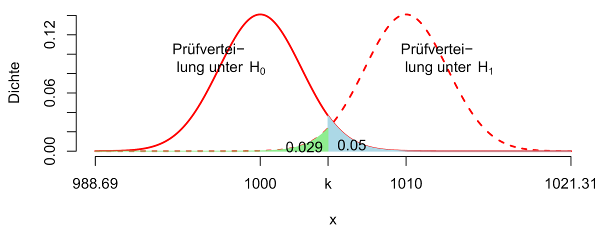
Warum kann man α nicht beliebig klein wählen?

- Problem: Je kleiner α, desto größer β.
- Via Formel im Beispiel des Mittelwerttestes für Normalverteilung für $\mu_0 < \mu_1$: Sinkt α, steigt $\lambda_{1-\alpha}$, steigt

$$\beta = \Phi\left(\lambda_{1-\alpha} + \sqrt{n}\frac{\mu_0 - \mu_1}{\sigma}\right)$$

- Als Grafik: β (=grüne Fläche) in Abhängigkeit von α (blaue Fläche)





Welchen Test nehmen?

- Existieren zwei Tests (Entscheidungsregeln, kritische Bereiche) für dieselbe Testaufgabe (einfache Hypothesen) und dieselbe Irrtumswahrscheinlichkeit $\Longrightarrow$ Entscheidung für Test mit kleinerer Wahrscheinlichkeit des Fehlers 2. Art.
- Beispiel: Mittelwerttest für Normalverteilung mit $\mu_0 < \mu_1$

  Alternativer kritischer Bereich

$$K' = \left\{(x_1, \ldots, x_n) | \sqrt{n}\frac{\overline{x}_n - \mu_0}{\sigma} < \lambda_\alpha\right\}$$

  liefert ebenfalls Test mit der Wahrscheinlichkeit α.

Aber:

$$\begin{aligned}
\beta' &= P\left(\sqrt{n}\frac{\overline{X}_n-\mu_1}{\sigma} > \lambda_\alpha + \sqrt{n}\frac{\mu_0-\mu_1}{\sigma}\right)\\
&= \Phi\left(\lambda_{1-\alpha} - \sqrt{n}\frac{\mu_0-\mu_1}{\sigma}\right)\\
&> \Phi\left(\lambda_{1-\alpha} + \sqrt{n}\frac{\mu_0-\mu_1}{\sigma}\right) = \beta
\end{aligned}$$

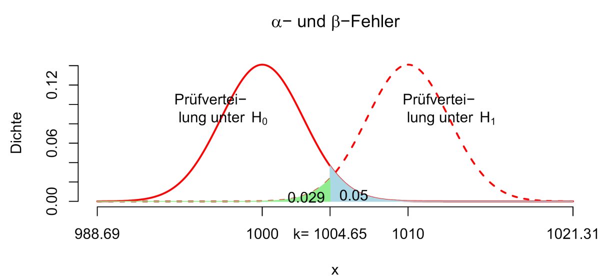
wegen $\mu_0 < \mu_1$.

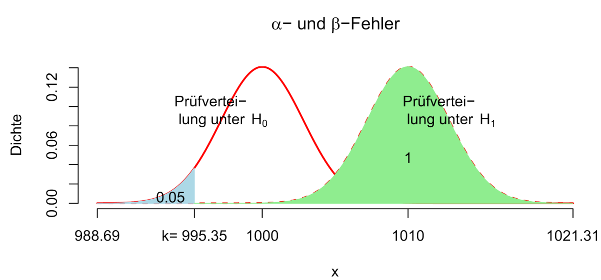
- D.h. Test mit K' hat bei derselben Irrtumswahrscheinlichkeit α eine größere Wahrscheinlichkeit für den Fehler 2. Art $\Longrightarrow$ Entscheidung für Test mit kritischem Bereich K.

- Zahlenbeispiel: $\mu_0 = 1000$, $\mu_1 = 1010$, $\sigma = 12$, $n = 18$, $\alpha = 0.05$

$$\beta' = \Phi\left(\lambda_{0.95} - \sqrt{18}\frac{1000-1010}{12}\right) = \Phi(6.736) = 1 > 0.029 = \beta.$$

- Grafik: Vergleich der Wahrscheinlichkeiten des β-Fehlers für K und K'





Wie misst man die Güte eines Testes für einfache Hypothesen?

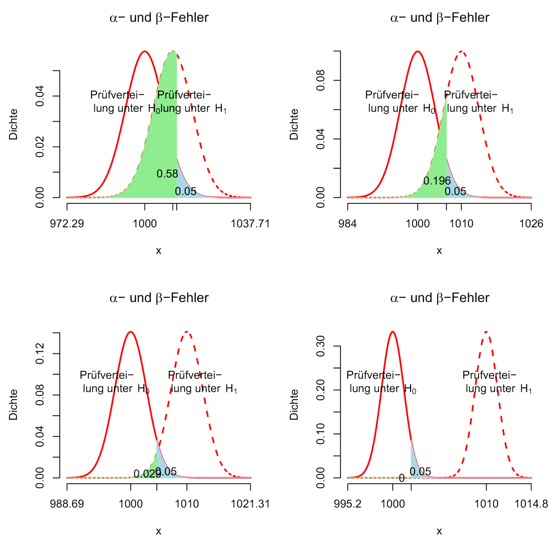
- Wähle denjenigen Test (= kritischen Bereich), der die Fehlerwahrscheinlichkeit β bei gegebener Wahrscheinlichkeit α für einen Fehler erster Art (zumeist 1%, 5% oder 10%) minimiert.
- Test mit minimalem β bei gegebenem α wird bester, mächtigster oder trennscharfer Test genannt.
- Güte (= Power = Macht) eines Testes $= 1 - \beta$.

- Beispiel des Mittelwerttestes für normalverteilte Grundgesamtheiten für $\mu_0 < \mu_1$:

$$\beta = \Phi\left(\lambda_{1-\alpha} + \sqrt{n}\frac{\mu_0 - \mu_1}{\sigma}\right)$$

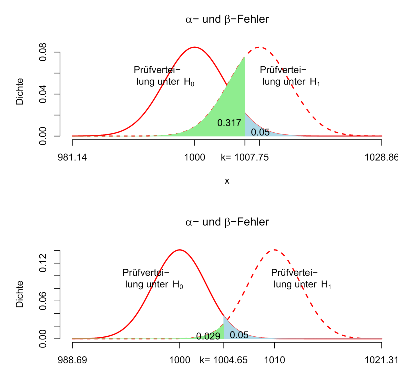
  – Wirkung des Stichprobenumfanges: Wenn n steigt, wird $\lambda_{1-\alpha} + \sqrt{n}\frac{\mu_0-\mu_1}{\sigma}$ kleiner und damit
    * β kleiner bzw.
    * $1 - \beta$ größer.

Für $n = 3, 12, 18, 100$:



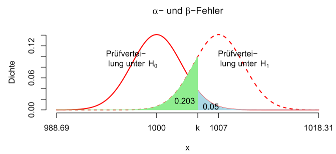
– Wirkung der Varianz der Grundgesamtheit: Je größer σ, desto kleiner $\lambda_{1-\alpha} + \sqrt{n}\frac{\mu_0-\mu_1}{\sigma}$ und desto größer β.

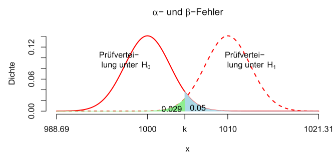
Für $\sigma = 12, 20$:



– Wirkung von Differenz zwischen Null- und Alternativhypothese: Je negativer $\mu_0 - \mu_1$, desto kleiner $\lambda_{1-\alpha} + \sqrt{n}\frac{\mu_0 - \mu_1}{\sigma}$ und desto kleiner β.

Für $\mu_1 = 1007, 1010$ und $\mu_0 = 1000$.





Zusammengesetzte Nullhypothese und Testniveau

- Sei Θ_0 der durch die Nullhypothese festgelegte Parameterbereich.
- Problem: Wahrscheinlichkeiten des Fehlers 1. Art

$$P((X_1, \ldots, X_n) \in K|\theta) \ \text{für alle } \theta \in \Theta_0$$

- Größe (= Niveau) eines Testes: Maximale Wahrscheinlichkeit des Fehlers 1. Art

$$\alpha = \sup_{\Theta_0} P((X_1, \ldots, X_n) \in K|\theta)$$

- Beispiel des Mittelwertestes für normalverteilte Grundgesamtheiten für $\mu_0 < \mu_1$ und zusammengesetzte Nullhypothese

$$H_0: \ \mu \leq \mu_0.$$

Da

$$P((X_1, \ldots, X_n) \in K|\mu) = P(\overline{X}_n > k|\mu) = 1 - \Phi\left(\sqrt{n}\frac{k - \mu}{\sigma}\right).$$

monoton zunehmend in μ ist, gilt

$$\alpha = \sup_{\mu \leq \mu_0} \left(1 - \Phi\left(\sqrt{n}\frac{k - \mu_0}{\sigma}\right)\right).$$

D.h. die Größe (=Niveau) des Testes wird an der Grenze μ_0 des Parameterbereiches der Nullhypothese berechnet.

Konsequenz: Bestimmung des kritischen Wertes k wie im Fall einer einfachen Nullhypothese als

$$k = \mu_0 + \lambda_{1-\alpha}\sigma/\sqrt{n}.$$

Zusammengesetzte Alternativhypothese und Gütefunktion

- Sei $\Theta_1 \subseteq \Theta$ der durch die Nullhypothese festgelegte Parameterbereich.
- Problem: Wahrscheinlichkeiten des Fehlers 2. Art

$$P((X_1, \ldots, X_n) \notin K|\theta) \text{ für alle } \theta \in \Theta_1$$

- Lösung: Gütefunktion

$$G(\theta) = P((X_1, \ldots, X_n) \in K|\theta)$$

gibt für alle Parameterwerte die Wahrscheinlichkeit an, die Nullhypothese abzulehnen.

- Für $\theta \in \Theta_0$ misst die Gütefunktion die Wahrscheinlichkeit, die Nullhypothese abzulehnen, wenn sie richtig ist.
- Für $\theta \in \Theta_1$ misst die Gütefunktion die Wahrscheinlichkeit, die Nullhypothese abzulehnen, wenn sie falsch ist.
- D.h. je niedriger (größer) die Werte der Gütefunktion für Θ_0 (Θ_1), desto besser ist ein Test.

Beispiel: Mittelwerttest für eine normalverteilte Grundgesamtheit, d.h. $X \sim \mathcal{N}(\mu, \sigma^2)$ mit bekanntem σ^2

- Entscheidung zwischen zwei zusammengesetzten Hypothesen:

$$H_0 : \mu \leq \mu_0 \text{ gegen } H_1 : \mu > \mu_0$$

- Verteilung der Prüfgröße unter H_1:

$$\overline{X}_n \sim \mathcal{N}(\mu_1, \sigma^2/n).$$

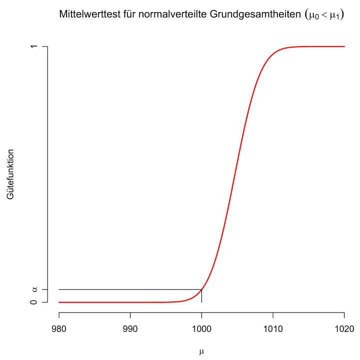
- Für $\mu \in \mathbb{R}$ ist

$$\begin{aligned} G(\mu) &= P((X_1, \ldots, X_n) \in K|\mu) \\ &= P(\overline{X}_n > k|\mu) \\ &= 1 - F_N(k; \mu, \sigma^2/n) = 1 - \Phi\left(\sqrt{n}\frac{k-\mu}{\sigma/\sqrt{n}}\right) \\ &= 1 - \Phi\left(\lambda_{1-\alpha} + \sqrt{n}\frac{\mu_0 - \mu}{\sigma}\right) \end{aligned}$$

wegen $k = \mu_0 + \lambda_{1-\alpha}\sigma/\sqrt{n}$.

- Zahlenbeispiel: $\mu_0 = 1000$, $\sigma = 12$, $n = 18$

$$G(\mu) = 1 - \Phi\left(\lambda_{0.95} + \sqrt{18}\frac{1000-\mu}{12}\right).$$



Links-, rechtsseitige und zweiseitige Tests

- Es lassen sich für eine Prüfgröße T_n folgende Entscheidungsregeln unterscheiden:
  - Rechtsseitige Entscheidungsregel: Lehne H_0 ab, wenn $T_n > k$ ist.
  - Linksseitige Entscheidungsregel: Lehne H_0 ab, wenn $T_n < k$ ist.
  - Zweiseitige Entscheidungsregel: Lehne H_0 ab, wenn $T_n < k_1$ oder $T_n > k_2$ ist.

- Sinnvolle Festlegung des kritischen Bereiches im Beispiel des Mittelwerttestes für normalverteilte Grundgesamtheiten:
  - $\mu_0 < \mu_1 \Longrightarrow$ rechtsseitiger Test (s.o.)
  - $\mu_0 > \mu_1 \Longrightarrow$ linksseitiger Test
  - $\mu_0 \neq \mu_1 \Longrightarrow$ zweiseitiger Test

Beispiel für linksseitigen Test: Lebensdauer von Glühbirnen

- Lebensdauer einer Glühbirne in Tagen: $X \sim \mathcal{N}(\mu, \sigma^2 = 900)$.

  Einfache Zufallsstichprobe des Umfangs $n = 16$.

  Stichprobenergebnis: $\overline{x}_{16} = 780$Tage.

  Entscheidung zwischen zwei zusammengesetzten Hypothesen

$$H_0:\ \mu \geq \mu_0 = 800 \quad \text{gegen} \quad H_1:\ \mu < \mu_0.$$

- Verteilung der Prüfgröße an der Grenze des Parameterbereiches μ_0 unter H_0:

$$\overline{X}_{16} \sim \mathcal{N}\left(800, \frac{900}{16}\right)$$

bzw.

$$\frac{\sqrt{16}(\overline{X}_{16} - 800)}{\sqrt{900}} \sim \mathcal{N}(0,1).$$

- Vorgegebenes Niveau des Testes: $\alpha = 0.05$.

- Festlegung des kritischen Bereiches (der kritischen Schranke):

$$\begin{aligned} 0.05 &= \alpha = P\left(\overline{X}_{16} < k | \mu_0 = 800\right) \\ &= P\left(\frac{\sqrt{16}(\overline{X}_{16} - 800)}{\sqrt{900}} < \frac{\sqrt{16}(k - 800)}{\sqrt{900}} | \mu_0 = 800\right) \\ &= \Phi(\lambda_\alpha), \end{aligned}$$

so dass der kritische Wert

$$k = 800 + \lambda_{0.05}\frac{30}{4} = 800 - 1.645 \cdot 7.5 = 787.66$$

ist.

- Entscheidungsregel (Linksseitiger Test): Lehne Nullhypothese mit 5%-Irrtumswahrscheinlichkeit ab, wenn der Stichprobenbefund $(x_1, \ldots, x_{16})$ in den kritischen Bereich K fällt mit

$$K = \left\{(x_1, \ldots, x_n) | \overline{x}_{16} < 787.66\right\}.$$

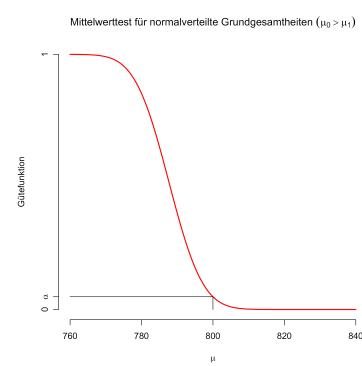
- Testentscheidung: Einsetzen des Stichprobenbefundes $\overline{x}_{16} = 780$ in die Prüfgröße ergibt Stichprobenwert (= empirischer Wert = realisierter Wert) der Prüfgröße.

  Wert ist kleiner als die kritische Schranke $k = 787.66$.

  D.h. Stichprobenergebnis fällt in kritischen Bereich und Nullhypothese ist mit 5%-Irrtumswahrscheinlichkeit abzulehnen.

- Gütefunktion:

$$G(\mu) = P(\overline{X}_{16} < k | \mu) = \Phi\left(\sqrt{16}\frac{787.66 - \mu}{\sqrt{900}}\right)$$



Testentscheidung mittels p-Wert

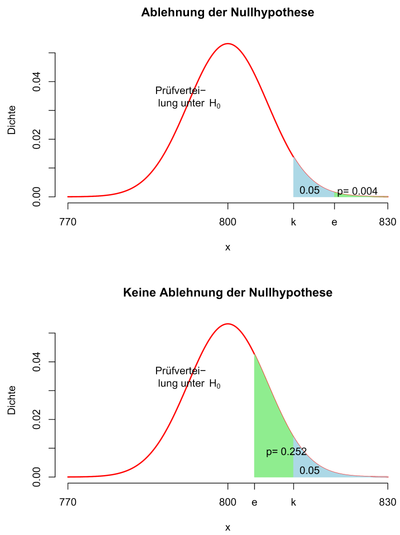
- Strategie 1:
  - Vorgabe der Irrtumswahrscheinlichkeit $\alpha = 0.05$,
  - Berechnung der zugehörigen kritische Schranke k
  - Prüfung, ob empirischer Wert der Prüfgröße e in den kritischen Bereich fällt.

- Strategie 2 für linksseitigen Test:
  - Berechnung des empirischen Wertes der Prüfgröße e,
  - Bestimmung der Wahrscheinlichkeit ($=$ p-Wert), dass dieser Wert "unterschritten" wird bei Gültigkeit der Nullhypothese,
  - Vergleich p-Wert mit gewünschter Irrtumswahrscheinlichkeit (z.B. 0.05)
  - und Ablehnung der Nullhypothese, wenn p-Wert kleiner oder gleich der Irrtumswahrscheinlichkeit ist.

- Strategie 2 für rechtsseitigen Test:
  - Berechnung des empirischen Wertes der Prüfgröße e,
  - Bestimmung der Wahrscheinlichkeit ($=$ p-Wert), dass dieser Wert "überschritten" wird bei Gültigkeit der Nullhypothese,
  - Vergleich p-Wert mit gewünschter Irrtumswahrscheinlichkeit (z.B. 0.05)
  - und Ablehnung der Nullhypothese, wenn p-Wert kleiner oder gleich der Irrtumswahrscheinlichkeit ist.

- Vorteil des p-Wertes: Nur eine Wahrscheinlichkeitsberechnung und Standardvergleich (mit z.B. 0.05) nötig.

- Nachteil: Keine Vorstellung über die Größe, die Prüfgröße besitzen muss, damit H_0 abgelehnt werden kann (=kritischer Wert) und keine Güteberechnung mittels p-Wert möglich.



p-Wert am Beispiel des rechtsseitigen Mittelwertestes bei Normalverteilung und bekannter Varianz (Zuckertüten)

- Allgemein für Stichprobenergebnis e:

$$p = P(\overline{X}_n > e|H_0) = 1 - \Phi\left(\sqrt{n}\frac{e-\mu_0}{\sigma}\right).$$

- Wenn Stichprobe $e = \overline{x}_{18} = 1002$ liefert, ist

$$p = 1 - \Phi\left(\sqrt{18}\frac{1002-1000}{12}\right) = 0.240.$$

  D.h. $p > \alpha = 0.05 \Longrightarrow H_0 : \mu \leq 1000$ kann mit einer Irrtumswahrscheinlichkeit von 5% nicht abgelehnt werden.

p-Wert am Beispiel des linksseitigen Mittelwertestes bei Normalverteilung und bekannter Varianz (Glühbirnen)

- Allgemein für Stichprobenergebnis e:

$$p = P(\overline{X}_n < e|H_0) = \Phi\left(\sqrt{n}\frac{e-\mu_0}{\sigma}\right).$$

- Stichprobe hat $e = \overline{x}_{18} = 780$ geliefert, womit

$$p = \Phi\left(\sqrt{16}\frac{780-800}{\sqrt{900}}\right) = 0.00384.$$

  D.h. $p < \alpha = 0.05 \Longrightarrow H_0 : \mu \geq 800$ kann mit einer Irrtumswahrscheinlichkeit von 5% abgelehnt werden.

Auswahl der zusammengesetzten Hypothesen

- Einfache Alternativhypothese: β lässt sich u.U. berechnen. Wenn β hinreichend klein ist und H_1 abgelehnt wird, kann H_0 als ,,statistisch abgesichert" (signifikant) gelten.

- Problem für zusammengesetzte Alternativhypothese: Welches β, wenn β vom unbekannten Wert des Parameters θ abhängt?

- D.h. H_0 lässt sich (bei einer zusammengesetzten Alternativhypothese) nicht mit einem quantifizierbaren Fehlerniveau statistisch absichern.

- Konsequenz für die Hypothesenformulierung:
  Soll Gültigkeit einer Hypothese statistisch abgesichert werden, so muß diese als Alternativhypothese H_1 formulieren werden.

  Bei einer Ablehnung von H_0 gilt H_1 zum Niveau α signifikant bestätigt.

Eigenschaften von Hypothesentests

- Null- und Alternativhypothese lassen sich wie folgt verallgemeinern:

$$H_0 : \theta \in \Theta_0 \text{ gegen } H_1 : \theta \in \Theta_1$$

  mit

$$\Theta_0 \subset \Theta \,, \quad \Theta_1 \subset \Theta \,, \quad \Theta_0 \cap \Theta_1 = \emptyset.$$

- Eigenschaften eines Tests mit kritischem Bereich K = Eigenschaften der Gütefunktion

$$G(\theta) = P((X_1, \ldots, X_n) \in K|\theta).$$

- Unter H_0 (d.h. für $\theta \in \Theta_0$)

$$G(\theta) = P((X_1, \ldots, X_n) \in K|\theta)$$

  gibt Gütefunktion $\theta \in \Theta_0$ Wahrscheinlichkeit des Fehlers 1. Art an.

- Unter H_1 (d.h. für $\theta \in \Theta_1$) beschreibt $G(\theta)$ die Wahrscheinlichkeit, die Nullhypothese abzulehnen, wenn sie falsch ist (= richtige Entscheidung).

- Idealer Test besitzt Gütefunktion G^*:

$$G^*(\theta) = \begin{cases} 0 & \text{falls } \theta \in \Theta_0 \\ 1 & \text{falls } \theta \in \Theta_1 \end{cases}$$

- Problem: Ideale Tests für endliche Stichprobenumfänge?

- Lösung: $n \to \infty$.

- Konsistenter Test:

$$\lim_{n \to \infty} P((X_1, \ldots, X_n) \in K|\theta \in \Theta_1) = 1.$$

  D.h.: Ein konsistenter Test wird mit zunehmendem Stichprobenumfang „trennschärfer“.

Abschließendes Beispiel: Mittelwerttest für normalverteilte Grundgesamtheit mit bekanntem σ (zweiseitig)

- Sei X die Abweichung des Durchmessers einer Lagerbuchse vom Plansoll in mm.
- Annahme: $X \sim N(0, 0.01^2)$
- Einfache Stichprobe des Umfangs $n = 100$: $e = \overline{x}_{100} = 0.003$.
- Zu testen:

$$H_0: \ \mu = \mu_0 = 0 \ \text{ gegen } \ H_1: \ \mu \neq \mu_0 = 0.$$

- Kritischer Bereich:

$$K = \{(x_1, \ldots, x_n) | |\overline{x}_n| > k\}$$

- Bestimmung von k:

$$\begin{aligned} \alpha &= P\left(|\overline{X}_n| > k | H_0\right) = P\left(|\overline{X}_n| > k | \mu_0 = 0\right) \\ &= 1 - P(\overline{X}_n \leq k | \mu_0 = 0) + P(\overline{X}_n \leq -k | \mu_0 = 0) \\ &= 1 - \Phi\left(\sqrt{n}\frac{k}{\sigma}\right) + \Phi\left(\sqrt{n}\frac{-k}{\sigma}\right) \\ &= 2\left(1 - \Phi\left(\sqrt{n}\frac{k}{\sigma}\right)\right). \end{aligned}$$

Damit ist

$$\Phi\left(\sqrt{n}\frac{k}{\sigma}\right) = 1 - \frac{\alpha}{2},$$

so dass

$$k = \lambda_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

D.h. für $\alpha = 0.05$ ist

$$k = \lambda_{0.975}\frac{0.01}{\sqrt{100}} = 0.00196.$$

- Testentscheidung: $e = 0.003 > k = 0.00196 \Longrightarrow$ Die mittlere Abweichung des Durchmessers ist mit einer Irrtumswahrscheinlichkeit von 5% signifikant von 0 verschieden.

- p-Wert:

$$\begin{aligned} p &= P\left(|\overline{X}_n| > e | H_0\right) = P(|\overline{X}_n| > e | \mu_0 = 0) \\ &= 1 - P(\overline{X}_n \leq e | \mu_0 = 0) + P(\overline{X}_n \leq -e | \mu_0 = 0) \\ &= 2\left(1 - \Phi\left(\sqrt{n}\frac{e}{\sigma}\right)\right) \\ &= 0.0027. \end{aligned}$$

D.h. $p = 0.0027 < \alpha = 0.05 \Longrightarrow$ Die mittlere Abweichung des Durchmessers ist mit einer Irrtumswahrscheinlichkeit von 5% signifikant von 0 verschieden.

- Gütefunktion:

$$G(\mu) = P(|\overline{X}_n| > k | \mu) = 2 - \Phi\left(\sqrt{n}\frac{k-\mu}{\sigma}\right) - \Phi\left(\sqrt{n}\frac{k+\mu}{\sigma}\right)$$

Gütefunktion für $n = 10, 50, 100, 1000$ („Konsistenz"):