7. Konfidenzintervalle und Hypothesentests

Andreas Eberle
Institut für angewandte Mathematik

November 2008

Beispiel.

Ein Orangennektar hat 50 % Fruchtsaftgehalt. Sie bestimmen den Fruchtsaftgehalt in 50 verschiedenen Packungen, und ermitteln einen Durchschnittswert von 47 % Fruchtsaftgehalt. Können Sie schließen, daß der Fruchtsaftgehalt niedriger als angegeben ist ?

- Hier wird eine unbekannte Größe m durch den Stichprobenmittelwert $\overline{X}_n$ von $n = 50$ unabhängigen Stichproben $X_1, X_2, ..., X_n$ geschätzt.

Beispiel

Schritt 1: Wahl des statistischen Modells

Ein Orangennektar hat 50 % Fruchtsaftgehalt. Sie bestimmen den Fruchtsaftgehalt in 50 verschiedenen Packungen, und ermitteln einen Durchschnittswert von 47 % Fruchtsaftgehalt. Können Sie schließen, daß der Fruchtsaftgehalt niedriger als angegeben ist ?

- Obwohl wir die einzelnen Stichproben eigentlich ohne Zurücklegen entnehmen, können wir näherungsweie die Unabhängigkeit der Beobachtungswerte X_i voraussetzen, da die Grundgesamtheit viel größer als die Anzahl $n = 50$ der Einzelstichproben ist.
- Da n hinreichend groß ist, und sich die zufälligen Fluktuationen in der Regel aus verschiedenen unabhängigen Effekten zusammensetzen, setzen wir eine Normalverteilung $N(m, \sigma^2)$ für die einzelnen Beobachtungswerte X_i voraus.
- Hierbei ist der Erwartungswert m der gesuchte durchschnittliche Fruchtsaftgehalt in der Grundgesamtheit, also ein **unbekannter Parameter**.

- Wir betrachten zwei verschiedene Situationen:
    1. Wir kennen die Standardabweichung σ unseres Meßverfahrens, und gehen davon aus, daß die zufälligen Fluktuationen der Beobachtungswerte nur durch die Meßungenauigkeit entstehen:
    -> *Schätzung des Erwartungswerts bei bekannter Varianz*
    2. Wir kennen die Größe der Fluktuationen nicht (sondern können diese nur aus den Beobachtungswerten schätzen)
    -> *Schätzung des Erwartungswerts bei unbekannter Varianz*
- Die zweite Situation ist natürlich deutlich realistischer, und wird in der Mehrzahl der praktischen Anwendungen betrachet.
- Wir betrachten trotzdem zunächst die erste Situation, da die mathematische Analyse hier etwas einfacher ist.

7.1. Schätzung des Erwartungswerts bei bekannter Varianz

Schritt 1: Wahl des statistischen Modells

- $X_1, X_2, \ldots., X_n \sim N(m, \sigma^2)$ unabhängige Stichproben ($n = 50$)
- $m \in \mathbb{R}$ unbekannter Parameter
- $\sigma > 0$ feste Standardabweichung (vorgegeben, z.B. $\sigma = 6$)

- Beachte, daß wir hier eigentlich mehrere stochastische Modelle gleichzeitig betrachten: Für jeden möglichen Wert von m (und ggf. von σ) eines ! Ein **(parametrisches) statististisches Modell** besteht also aus einer ganzen Klasse von stochastischen Modellen, die mit einem oder mehreren unbekannten Parametern parametrisiert sind.

Schätzung des Erwartungswerts bei bekannter Varianz

Schritt 2: Wahl eines Punktschätzers

- Wir wollen die unbekannte Größe m aus den Beobachtungswerten $X_1, X_2, \ldots, X_n$ schätzen.
- Da m das arithmetische Mittelwert in der Grundgesamtheit ist, liegt es nahe als Schätzwert für m das Stichprobenmittel

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

  zu verwenden. Wir nennen die Statistik $\overline{X}_n$ auch einen **Punktschätzer** für den unbekannten Parameter m.
- Ein anderer möglicher Punktschätzer wäre der Median der Beobachtungswerte.

Schätzung des Erwartungswerts bei bekannter Varianz

Schritt 3: Festlegung eines Signifikanzniveaus

- Wir legen ein Signifikanzniveau α bzw. das Konfidenzniveau $1-\alpha$ fest, z.B.
  - $\alpha = 0.05$, $1-\alpha = 95\%$: schwach signifikant
  - $\alpha = 0.01$, $1-\alpha = 99\%$: signifikant
  - $\alpha = 0.001$, $1-\alpha = 99.9\%$: stark signifikant
  - $\alpha = 0.0001$, $1-\alpha = 99.99\%$: äußerst signifikant
- Das Signifikanzniveau gibt maximale Wahrscheinlichkeiten für bestimmte Fehlschlüsse an, s.u. Man sollte **vor** Durchführung der Untersuchung festlegen, welches Signifikanzniveau man fordert. Bei $\alpha = 0.05$ kann zum Beispiel in einer von 20 Untersuchungen ein Fehlschluß auftreten.
- Im unserem Beispiel wählen wir $\alpha = 0.01$, also $1-\alpha = 99\%$.

Schätzung des Erwartungswerts bei bekannter Varianz

Schritt 4: Wahl einer Teststatistik

- Wir wählen nun eine Statistik, die von den Beobachtungswerten $X_1, X_2, \ldots, X_n$ abhängt, und die angibt, wie stark der Schätzwert $\overline{X}_n$ vom wahren Parameter m abweicht.
- Die Verteilung dieser Statistik sollten wir explizit, oder zumindest näherungsweise, kennen.
- In unserem Fall bietet es sich an, das standardisierte Stichprobenmittel

$$Z_n = \frac{\overline{X}_n - m}{\sigma / \sqrt{n}} \qquad \left(= \frac{\overline{X}_n - E\left[\overline{X}_n\right]}{\sigma\left(\overline{X}_n\right)} \right)$$

zu verwenden, denn:

1. Z_n ist proportional zum Schätzfehler $\overline{X}_n - m$.
2. Wir wissen, daß Z_n standardnormalverteilt ist.

Schätzung des Erwartungswerts bei bekannter Varianz

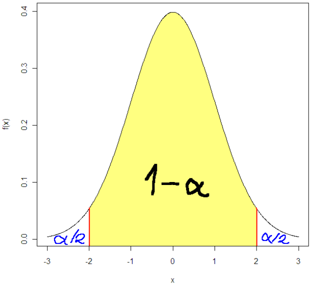
Schritt 5: Konfidenzintervall (Intervallschätzer, Bereichsschätzer)

- Wir wollen nun ein *von den Beobachtungswerten abhängendes* Intervall um den Punktschätzer $\overline{X}_n$ angeben, so daß der tatsächliche Parameter m **stets** (d.h. für alle möglichen Werte des tatsächlichen Parameters) mit Wahrscheinlichkeit $\geq 1-\alpha$ in dem Intervall liegt.

- Ein solches Intervall nennen wir ein **Konfidenzintervall** zum Konfidenzniveau $1-\alpha$ (z.B. $1-\alpha = 99\%$).

- Da die Teststatistik Z_n standardnormalverteilt ist, gilt:

$$P_{m,\sigma}\left[|Z_n| \leq z_{(1-\alpha/2)}\right] = 1 - \alpha$$



- In unserem Beispiel ist $\alpha = 0.01$, also $z_{(1-\alpha/2)} = z_{(0.995)} = 2.58$ das 99.5 % Quantil der Standardnormalverteilung.

- Wegen

$$Z_n = \frac{\overline{X}_n - m}{\sigma/\sqrt{n}}$$

erhalten wir:

$$0.99 = P_{m,\sigma}\left[|Z_n| \leq z_{(0.995)}\right] = P_{m,\sigma}\left[|\overline{X}_n - m| \leq z_{(0.995)} \cdot \sigma/\sqrt{n}\right]$$

- Das gesuchte 99% Konfidenzintervall ist also

$$\overline{X}_n \pm z_{(0.995)} \cdot \sigma/\sqrt{n}$$

- Speziell für $\sigma = 6$ und $n = 50$ ergibt sich

$$z_{(0.995)} \cdot \sigma/\sqrt{n} = 2.58 \cdot 6/\sqrt{50} = 2.2$$

Schätzung des Erwartungswerts bei bekannter Varianz

Schritt 5: Konfidenzintervall

- **Fazit:** In unserem Beispiel ist $\overline{X}_n \pm 2.2 = 47 \pm 2.2$ ein 99% Konfidenzintervall für den wahren Fruchtsaftgehalt m.
- **Warnung**: Dies bedeutet **nicht**, daß der tatsächliche Fruchtsaftgehalt mit Wahrscheinlichkeit 99% zwischen 44.8 und 49.2 liegt, sondern...
- **Richtige Interpretation:** Wenn wir dieselbe Untersuchung sehr oft wiederholen würden, dann würde der tatsächliche Fruchtsaftgehalt in 99% der Fälle in dem von uns ermittelten Intervall liegen. Dieses Intervall ist aber für jede Untersuchung ein anderes, da es aus den jeweiligen Beobachtungswerten berechnet wird !
- Allgemein erhalten wir bei bekanntem σ mit völlig analoger Rechnung das Konfidenzintervall

$$\overline{X}_n \pm z_{(1-\alpha/2)} \cdot \sigma / \sqrt{n}$$

zum Konfidenzniveau $1 - \alpha$ für den Erwartungswert m.

Schätzung des Erwartungswerts bei bekannter Varianz

Schritt 6: Formulierung einer Nullhypothese und Alternative

Die Konfidenzintervalle enthalten im Prinzip schon alle wesentlichen Informationen. In Anwendungen ist es aber üblich, die Problemstellung als Hypothesentest umzuformulieren.

- Dazu formulieren wir zunächst eine Nullhypothese H_0 und eine Alternative H_1:
  - Die **Nullhypothese** beschreibt den zu erwartenden Normalfall/Regelfall.
  - Die **Alternative** beschreibt eine Abweichung vom Normalfall, oder einen in einem Experiment zu beobachtenden Effekt.
- In unserem Beipiel:
  - $H_0 : \quad m = 50$
  - $H_1 : \quad m \neq 50$
- Dies ist ein **beidseitiges Testproblem**. Alternativ könnten wir auch das **einseitige Testproblem** $H_0\colon\ m = 50$ vs. $H_1\colon\ m < 50$ betrachten.

- Die **Nullhypothese** beschreibt den zu erwartenden Normalfall/Regelfall.
- Die **Alternative** beschreibt eine Abweichung vom Normalfall, oder einen in einem Experiment zu beobachtenden Effekt.

- Die Nullhypothese und Alternative sind nicht gleichwertig:
  - Wir wollen unbedingt vermeiden, daß wir die Nullhypothese verwerfen, obwohl sie wahr ist ("peinlicher Irrtum", **Fehler 1. Art**).
  - Wenn unsere Untersuchung nicht genau genug ist (z.B. aufgrund einer zu geringen Stichprobengröße), kann es aber durchaus passieren, daß wir einen vorhandenen Effekt nicht entdecken, also die Nullhypothese belassen,obwohl sie falsch ist (**Fehler 2. Art**)

Schätzung des Erwartungswerts bei bekannter Varianz

Schritt 7: Hypothesentest

- Ein **Hypothesentest** für eine Nullhypothese H_0 gegen eine Alternative H_1 basierend auf den Beobachtungswerten $x_1, x_2, \dots, x_n$ besteht nun aus der Angabe eines **kritischen Bereichs** K (Verwerfungsbereich) zusammen mit der **Entscheidungsregel**:
  - **Verwerfe** die Nullhypothese, falls die Beobachtungswerte im kritischen Bereich liegen.
  - **Belasse** die Nullhypothese, falls nicht.

- Ein statistischer Nachweis/Beweis gelingt bloß dann, wenn die Nullhypothese verworfen wird. *Belassen der Nullhypothese bedeutet dagegen* **nicht***, daß* H_0 *damit statistisch bewiesen ist!*

Schätzung des Erwartungswerts bei bekannter Varianz

Signifikanzniveau

- Wie oben legen wir ein Signifikanzniveau α fest, z.B.
  - $\alpha = 0.05$, $1 - \alpha = 95\%$: schwach signifikant
  - $\alpha = 0.01$, $1 - \alpha = 99\%$: signifikant
  - $\alpha = 0.001$, $1 - \alpha = 99.9\%$: stark signifikant
  - $\alpha = 0.0001$, $1 - \alpha = 99.99\%$: äußerst signifikant
- Bei einem Hypothesentest beschreibt das Signifikanzniveau die **(maximale) Wahrscheinlichkeit für den Fehler 1. Art** (Verwerfen der Nullhypothese, obwohl diese wahr ist), die wir zulassen wollen. Für unseren Test soll also auf jeden Fall gelten:

$$P_{H_0}\left[(X_1, X_2, \ldots, X_n) \text{ liegt im Verwerfungsbereich}\right] \leq \alpha$$

- Im unserem Beispiel wählen wir wieder $\alpha = 0.01$, also $1 - \alpha = 99\%$.

Schätzung des Erwartungswerts bei bekannter Varianz

Der z-Test

- Ein **Hypothesentest** für die Nullhypothese H_0: $m = m_0$ (in unserem Beispiel ist $m_0 = 50$) gegen die Alternative H_1: $m \neq m_0$ basierend auf den Beobachtungswerten $x_1, x_2, \ldots, x_n$ besteht aus der Angabe eines **kritischen Bereichs** K (Verwerfungsbereich) zusammen mit der **Entscheidungsregel**:
  - **Verwerfe** die Nullhypothese, falls die Beobachtungswerte im kritischen Bereich liegen.
  - **Belasse** die Nullhypothese, falls nicht.

- Um im konkreten Fall einen Hypothesentest zu konstruieren, betrachten wir wieder die Teststatistik

$$Z_n = \frac{\overline{X}_n - m_0}{\sigma / \sqrt{n}} \quad .$$

- Beachte, daß wir jetzt $m = m_0$ $(= 50)$ setzen, da uns zunächst die Verwerfungswahrscheinlichkeit unter der Nullhypothese interessiert.

Schätzung des Erwartungswerts bei bekannter Varianz

Der z-Test

- Da die Statistik

$$Z_n = \frac{\overline{X}_n - m_0}{\sigma / \sqrt{n}}$$

unter der Nullhypothese standardnormalverteilt ist, erhalten wir:

$$0.99 = P_{H_0}\left[|Z_n| \leq z_{(0.995)} \right] = P_{H_0}\left[\left|\overline{X}_n - m_0\right| \leq z_{(0.995)} \cdot \sigma / \sqrt{n} \right]$$

- **Wir verwerfen daher die Nullhypothese** $m = m_0$**, falls**

$$\boxed{\left|\overline{x}_n - m_0\right| > z_{(0.995)} \cdot \sigma / \sqrt{n}}$$

gilt - die Verwerfungswahrscheinlichkeit unter der Nullhypothese beträgt dann gerade 1%.

- In unserem Beipiel ist $m_0 = 50$ und $z_{(0.995)} \cdot \sigma/\sqrt{n} = 2.2$, siehe oben. Wir verwerfen also die Nullhypothese $m = 50$ genau dann, wenn
$$|\overline{x}_n - 50| > 2.2 \qquad \text{gilt,}$$
also wenn das Stichprobenmittel $\overline{x}_n$ nicht in dem Intervall $[47.8, 52.2]$ liegt.

- Beobachtet wurde $\overline{x}_n = 47$, so daß wir die Nullhypothese also in diesem Fall tatsächlich verwerfen können. Weil wir das Signifikanzniveau $\alpha = 1\%$ gewählt haben, ist unser Testergebnis im Sinne der Notation von oben **signifikant** (d.h. ein Fehler 1. Art tritt nur bei einer von 100 Untersuchungen auf).

- Hätte unsere Untersuchung hingegen nur auf $n = 25$ Einzelstichproben basiert, dann wäre $z_{(0.995)} \cdot \sigma/\sqrt{n} = 3.1$, und wir würden nur verwerfen, wenn
$$|\overline{x}_n - 50| > 3.1 \qquad \text{gilt.}$$

- In diesem Fall könnten wir die Nullhypothese also bei einem Signifikanzniveau $\alpha = 1\%$ nicht verwerfen.

Schätzung des Erwartungswerts bei bekannter Varianz

Der z-Test

Allgemein erhalten wir auf analoge Weise folgenden Test:

- Voraussetzung: Gaußmodell mit bekannter Varianz σ^2.
- Nullhypothese: $m = m_0$, Alternative: $m \neq m_0$
- Signifikanzniveau: α
- Teststatistik:
$$Z_n = \frac{\overline{X}_n - m_0}{\sigma/\sqrt{n}} \sim N(0,1)$$
- **Verwerfungsbereich:** Die Nullhypothese $m = m_0$ wird verworfen, falls $|z_n| > z_{(1-\alpha/2)}$ bzw.
$$\boxed{|\overline{x}_n - m_0| > z_{(1-\alpha/2)} \cdot \sigma/\sqrt{n}}$$

Schätzung des Erwartungswerts bei bekannter Varianz

- Unser Testergebnis sagt uns nur, ob wir bei einem *vorgegebenen* Signifikanzniveau die Nullhypothese verwerfen oder nicht. Wenn wir das Signifikanzniveau ändern, müssen wir den Verwerfungsbereich wieder neu berechnen.
- Um zu messen, wie eindeutig eine Nullhypothese verworfen wird (oder nicht verworfen wird), können wir zu gegebenen Beobachtungswerten dasjenige Signifikanzniveau p berechnen, bei dem H_0 gerade noch verworfen wird:

Definition

Der p-**Wert** eines Hypothesentests ist das kleinste Signifikanzniveau α, bei dem die Nullhypothese verworfen wird.

- Wenn wir den p-Wert kennen, sieht unsere Testentscheidung bei vorgegebenem Signifikanzniveau α so aus:
  - Verwerfe H_0 falls $p \leq \alpha$ — Belasse H_0 falls $p > \alpha$.

Schätzung des Erwartungswerts bei bekannter Varianz

Der p-Wert beim z-Test

- Der z-Test basiert auf der Teststatistik

$$Z_n = \frac{\overline{X}_n - m_0}{\sigma / \sqrt{n}} \sim N(0, 1)$$

- Wir verwerfen , wenn der Betrag des beobachteten Wertes

$$z_n = \frac{\overline{x}_n - m_0}{\sigma / \sqrt{n}}$$

  der Teststatistik hinreichend groß ist.

Theorem

Der p-Wert beim z-Test mit Beobachtungswert z_n ist

$$p = P_{H_0}\left[|Z_n| \geq |z_n|\right] = P_{H_0}\left[\left|\overline{X}_n - m_0\right| \geq |\overline{x}_n - m_0|\right]$$

Schätzung des Erwartungswerts bei bekannter Varianz

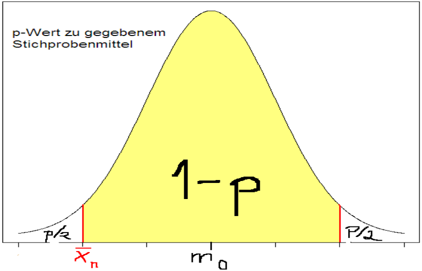
Berechnung des p-Werts

Theorem

Der p-Wert beim z-Test mit Beobachtungswert z_n ist

$$p = P_{H_0}\left[|Z_n| \geq |z_n|\right] = P_{H_0}\left[\left|\overline{X}_n - m_0\right| \geq \left|\overline{x}_n - m_0\right|\right]$$

- **Der p-Wert gibt also an, mit welcher Wahrscheinlichkeit das Stichprobenmittel noch weiter von m_0 entfernt ist, als tatsächlich beobachtet wurde.**



- **Berechnung des p-Werts im z-Test**:

$$\begin{aligned} p &= P_{H_0}\left[|Z_n| \geq |z_n|\right] = 2 \cdot P_{H_0}\left[Z_n \geq |z_n|\right] \\ &= 2 \cdot (1 - \Phi(|z_n|)) \end{aligned}$$

- In unserem Beispiel ist

$$|z_n| = \frac{|\overline{x}_n - m_0|}{\sigma/\sqrt{n}} = \frac{|47 - 50|}{6/\sqrt{50}} = 3.53$$

- Wir erhalten also

$$p = 2 \cdot (1 - \Phi(3.53)) \approx 0.0004 = 0.04\%$$

- Wir würden die Nullhypothese also sogar beim Signifikanzniveau $\alpha = 0.1\%$ verwerfen, d.h. das Testergebnis ist sogar "**stark signifikant**".
- Beim Signifikanzniveau $\alpha = 0.01\%$ könnten wir H_0 dagegen nicht verwerfen - das Ergebnis ist also nicht "**äußerst signifikant**".

Schätzung des Erwartungswerts bei bekannter Varianz

Schritt 9: Die Gütefunktion

- Bisher haben wir nur den Fehler 1. Art beachtet, und den Test so konstruiert, das dessen W'keit unter dem Signifikanzniveau liegt.
- Um die Qualität verschiedener Tests zu demselben Signifikanzniveau zu vergleichen, untersucht man, mit welcher W'keit die Tests die Nullhypothese verwerfen, wenn $m \neq m_0$ gilt.

Definition

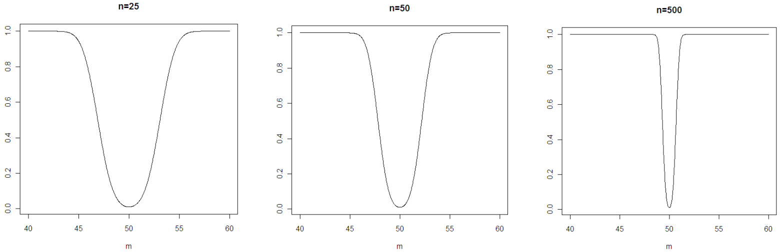
Die **Gütefunktion (Macht, power)** des z-Tests ist

$$g(m) = P_{m,\sigma}\left[H_0 \text{ wird verworfen}\right]$$

- Die **Gütefunktion (Macht, power)** des z-Tests ist

$$g(m) = P_{m,\sigma}\left[H_0 \text{ wird verworfen}\right]$$

- Die Gütefunktion des z-Tests kann man explizit berechnen. Die folgenden Grafiken zeigen Gütefunktionen für verschiedene Stichprobengrößen:



7.2. Schätzung des Erwartungswerts bei unbekannter Varianz

- Wenn wir die Standardabweichung σ nicht kennen, dann können wir die Teststatistiken
$$Z_n = \frac{\overline{X}_n - m}{\sigma/\sqrt{n}}$$
von oben nicht verwenden.
- Stattdessen **schätzen** wir in diesem Fall die Standardabweichung σ durch die renormierte empirische Standardabweichung
$$s_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2}$$
der Meßwerte, und verwenden die Teststatistiken
$$T = \frac{\overline{X}_n - m}{S_n/\sqrt{n}}$$
(**Studentsche t-Statistik mit $\mathbf{n-1}$ Freiheitsgraden**).

Schätzung des Erwartungswerts bei unbekannter Varianz

Schritt 1: Wahl des statistischen Modells

- $X_1, X_2, \ldots, X_n \sim N(m, \sigma^2)$ unabhängige Stichproben ($n = 50$)
- $m \in \mathbb{R}, \quad \sigma > 0$ unbekannte Parameter

Schätzung des Erwartungswerts bei unbekannter Varianz

Schritt 2: Wahl eines Punktschätzers

- Wir wollen die unbekannte Größe m aus den Beobachtungswerten $X_1, X_2, ..., X_n$ schätzen.
- Da m das arithmetische Mittelwert in der Grundgesamtheit ist, verwenden wir als Schätzwert für m wieder das Stichprobenmittel

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Exkurs: Eigenschaften von Punktschätzern

Definition

Ein Punktschätzer $\widehat{\theta}$ für einen unbekannten Parameter θ heißt **erwartungstreu (unbiassed)**, falls der Erwartungswert des Schätzers stets gleich dem unbekannten Parameter ist:

$$E_\theta\left[\widehat{\theta}\right] = \theta \qquad \text{für alle möglichen Werte des Parameters } \theta.$$

- **Beipiel 1:** Das Stichprobenmittel $\overline{X}_n$ ist ein erwartungstreuer Schätzer für den Mittelwert m in der Grundgesamtheit, denn es gilt stets

$$\begin{aligned} E_{m,\sigma}\left[\overline{X}_n\right] &= E_{m,\sigma}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} \underbrace{E_{m,\sigma}\left[X_i\right]}_{=m} = \frac{n \cdot m}{n} = m. \end{aligned}$$

Exkurs: Eigenschaften von Punktschätzern

- **Beipiel 2:** Die empirische Varianz σ_n^2 einer Stichprobe ist **kein erwartungstreuer Schätzer** für die Varianz σ^2 der zugrundeliegenden Verteilung, denn es gilt

$$\begin{aligned} E_{m,\sigma}\left[\sigma_n^2\right] &= E_{m,\sigma}\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2\right] = \frac{1}{n}\sum_{i=1}^{n} E_{m,\sigma}\left[\left(X_i - \overline{X}_n\right)^2\right] \\ &= \; ..... = \frac{n-1}{n}\cdot\sigma^2 \neq \sigma^2 \end{aligned}$$

- Dagegen ist die **renormierte Stichprobenvarianz**

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$$

  ein **erwartungstreuer Schätzer** für σ^2.

- Dies ist der Grund, warum man bei Stichproben (aber nicht in der Grundgesamtheit) in der Regel S_n^2 statt σ_n^2 verwendet !

Exkurs: Eigenschaften von Punktschätzern

Definition

Ein (von der Stichprobengröße n abhängender) Punktschätzer $\widehat{\theta}_n$ für einen unbekannten Parameter θ heißt **konsistent**, falls der Schätzwert sich für große n dem tatsächlichen Wert θ annähert (mit W'keit 1):

$$\lim_{n\to\infty} \widehat{\theta}_n = \theta \qquad \text{mit W'keit 1.}$$

- Alle von uns hier betrachteten Schätzer (z.B. $\overline{X}_n, s_n^2$, σ_n^2, ...) sind konsistent. Dies folgt aus dem Gesetz der großen Zahlen.

Definition

Der **mittlere quadratische Fehler (MSE=mean square error)** eines Schätzers $\widehat{\theta}$ ist die mittlere quadratische Abweichung vom Schätzwert:

$$MSE\left(\widehat{\theta}\right) = E\left[\left|\widehat{\theta} - \theta\right|^2\right]$$

Schätzung des Erwartungswerts bei unbekannter Varianz

Schritt 3: Festlegung eines Signifikanzniveaus

- Wie oben legen wir ein Signifikanzniveau α bzw. das Konfidenzniveau $1 - \alpha$ fest, z.B.
  - $\alpha = 0.05,\ \ 1 - \alpha = 95\%$: schwach signifikant
  - $\alpha = 0.01,\ \ 1 - \alpha = 99\%$: signifikant
  - $\alpha = 0.001,\ \ 1 - \alpha = 99.9\%$: stark signifikant
  - $\alpha = 0.0001,\ \ 1 - \alpha = 99.99\%$: äußerst signifikant
- Beispielsweise wählen wir $\alpha = 0.01$, also $1 - \alpha = 99\%$.

Schätzung des Erwartungswerts bei unbekannter Varianz

Schritt 4: Wahl einer Teststatistik

- Wir wählen nun wieder eine Statistik, die von den Beobachtungswerten $X_1, X_2, \ldots, X_n$ abhängt, und die angibt, wie stark der Schätzwert $\overline{X}_n$ vom wahren Parameter m abweicht.
- Die Verteilung dieser Statistik sollten wir explizit, oder zumindest näherungsweise, kennen.

- Da wir die Standardabweichung σ nicht kennen, also schätzen müssen, verwenden wir nun (im Gegensatz zu oben) die **Studentsche t-Statistik**

$$T = \frac{\overline{X}_n - m}{S_n / \sqrt{n}}$$

Exkurs:Studentsche t-Verteilung

Theorem

Die Verteilung der t-Statistik

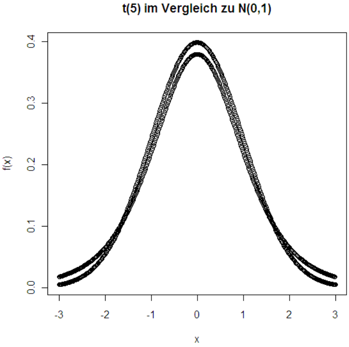
$$T = \frac{\overline{X}_n - m}{S_n / \sqrt{n}}$$

im Gaußmodell mit Mittelwert m und Varianz σ^2 hängt nicht von σ ab, und läßt sich explizit berechnen.

Definition

Die Verteilung von T heißt Studentsche t-Verteilung mit $n - 1$ Freiheitsgraden.

- Die Studentschen t-Verteilungen und ihre Quantile sind an vielen Stellen tabelliert. $t(n-1)$ ist eine **symmetrische** Verteilung um 0, die **langschwänziger** als die Standardnormalverteilung ist.



- Für große n ist die t-Verteilung **der Standardnormalverteilung sehr ähnlich**.

- Daher können wir die Quantile der t-Verteilung bei **großen Stichproben** (Faustregel: $n \geq 30$) auch näherungsweise durch die Quantile der Standardnormalverteilung ersetzen. Die Konfidenzintervalle und Tests bei unbekannter Varianz basieren dann also auf genau denselben Quantilen wie bei bekannter Varianz.
- Bei **kleinen Stichproben** ($n < 30$) muß dagegen berücksichtigt werden, daß die Schätzung der Varianz einen zusätzlichen Schätzfehler verursacht. Die Quantile der Standardnormalverteilung müssen dann durch die Quantile der $t(n-1)$-Verteilung ersetzt werden, die etwas größer sind.
- Für $T \sim t(n-1)$ gilt

$$E[T] = 0 \qquad \text{und} \qquad Var(T) = \frac{n-1}{n-3} > 1.$$

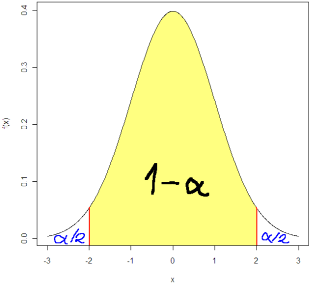
Schätzung des Erwartungswerts bei unbekannter Varianz

Schritt 5: Konfidenzintervall (Intervallschätzer, Bereichsschätzer)

- Wir wollen nun ein *von den Beobachtungswerten abhängendes* Konfidenzintervall um den Punktschätzer $\overline{X}_n$ angeben, so daß der tatsächliche Parameter m **stets** (d.h. für alle möglichen Werte der Parameter m und σ^2) mit Wahrscheinlichkeit $\geq 1 - \alpha$ $(= 99\%)$ in dem Intervall liegt.
- Sei dazu $t_{n-1,1-\alpha/2}$ das $1 - \alpha/2$ Quantil der t-Verteilung mit $n - 1$ Freiheitsgraden.

- Wegen $T \sim t_{n-1}$ gilt:

$$P_{m,\sigma}\left[|T| \leq t_{n-1,1-\alpha/2}\right] = 1-\alpha$$



- In unserem Beispiel ist $\alpha = 0.01$ und $n = 50$, also $t_{n-1,1-\alpha/2} = t_{49,0.995} = 2.68$ das 99.5 % Quantil der $t-$verteilung mit 49 Freiheitsgeraden.

- Wegen
$$T = \frac{\overline{X}_n - m}{S_n/\sqrt{n}}$$
erhalten wir:
$$0.99 = P_{m,\sigma}\left[\left|T\right| \leq t_{49,0.995}\right] = P_{m,\sigma}\left[\left|\overline{X}_n - m\right| \leq t_{49,0.995} \cdot S_n/\sqrt{n}\right]$$

- Das gesuchte 99% Konfidenzintervall ist also
$$\overline{X}_n \pm t_{49,0.995} \cdot S_n/\sqrt{n}$$

- Im Gegensatz zu oben hängt die Breite des Konfidenzintervalls nun von dem Beobachtungswert S_n ab!

- Haben wir zum Beipiel die empirische Standardabweichung $s_n = 8.32$ beobachtet, dann ergibt sich
$$t_{49,0.995} \cdot s_n/\sqrt{n} = 2.68 \cdot 8.32/\sqrt{50} = 3.15$$

Schätzung des Erwartungswerts bei unbekannter Varianz

Schritt 5: Konfidenzintervall

- **Fazit:** In unserem Beispiel mit $\overline{x}_n = 47$ und $s_n = 8.32$ ist $\overline{x}_n \pm 2.2 = 47 \pm 3.15$ ein 99% Konfidenzintervall für m.
- **Warnung**: Dies bedeutet **nicht**, daß der tatsächliche Fruchtsaftgehalt mit Wahrscheinlichkeit 99% zwischen 43.85 und 50.15 liegt, sondern...
- **Richtige Interpretation:** Wenn wir dieselbe Untersuchung sehr oft wiederholen würden, dann würde der tatsächliche Fruchtsaftgehalt in 99% der Fälle in dem von uns ermittelten Intervall liegen. Dieses Intervall ist aber für jede Untersuchung ein anderes, da es aus den jeweiligen Beobachtungswerten berechnet wird (sogar die Breite hängt jetzt von den Meßwerten ab !!!)
- Allgemein erhalten wir mit völlig analoger Rechnung das Konfidenzintervall

$$\overline{X}_n \pm t_{n-1,1-\alpha/2} \cdot S_n / \sqrt{n}$$

zum Konfidenzniveau $1 - \alpha$ für den Erwartungswert m.

Schätzung des Erwartungswerts bei unbekannter Varianz

Schritt 6: Formulierung einer Nullhypothese und Alternative

- Wir formulieren nun wieder eine Nullhypothese H_0 und eine Alternative H_1:
  - Die **Nullhypothese** beschreibt den zu erwartenden Normalfall/Regelfall.
  - Die **Alternative** beschreibt eine Abweichung vom Normalfall, oder einen in einem Experiment zu beobachtenden Effekt.
- Wir könnten wie oben das beidseitige Testproblem mit der Nullhypothese $m = 50$ und der Alternative $m \neq 50$ betrachten. Die Testentscheidung würde dann genau wie oben beim z-Test verlaufen - wir müssten nur die Quantile der Standardnormalverteilung durch die Quantile der t(n-1)-Verteilung ersetzen, und die unbekannte Standardabweichung durch den Schätzwert s_n.
- Stattdessen betrachten wir zur Illustration diesmal das einseitige Testproblem $H_0 : \; m = 50$ vs. $H_1 : \; m < 50$.

Schätzung des Erwartungswerts bei unbekannter Varianz

- Nullhypothese $H_0 : \ m = 50$
  (oder auch $m \geq 50$, das macht hier keinen Unterschied)
- Alternative $H_1 : \ m < 50$

- Dies ist ein **einseitiges Testproblem**.
- Die **Nullhypothese** beschreibt den zu erwartenden Normalfall/Regelfall.
- Die **Alternative** beschreibt eine Abweichung vom Normalfall, oder einen in einem Experiment zu beobachtenden Effekt.

Schätzung des Erwartungswerts bei unbekannter Varianz

Schritt 7: Hypothesentest

- Ein **Hypothesentest** für die Nullhypothese H_0 gegen die Alternative H_1 basierend auf den Beobachtungswerten $x_1, x_2, \ldots, x_n$ besteht nun aus der Angabe eines **kritischen Bereichs** K (Verwerfungsbereich) zusammen mit der **Entscheidungsregel**:
  - **Verwerfe** die Nullhypothese, falls die Beobachtungswerte im kritischen Bereich liegen.
  - **Belasse** die Nullhypothese, falls nicht.
- Ein statistischer Nachweis/Beweis gelingt bloß dann, wenn die Nullhypothese verworfen wird. *Belassen der Nullhypothese bedeutet dagegen* **nicht***, daß H_0 damit statistisch bewiesen ist!*

- Die Nullhypothese und Alternative sind nicht gleichwertig:
  - Wir wollen unbedingt vermeiden, daß wir die Nullhypothese verwerfen, obwohl sie wahr ist
    ("peinlicher Irrtum", **Fehler 1. Art**).
  - Wenn unsere Untersuchung nicht genau genug ist (z.B. aufgrund einer zu geringen Stichprobengröße), kann es aber durchaus passieren, daß wir einen vorhandenen Effekt nicht entdecken, also die Nullhypothese belassen,obwohl sie falsch ist (**Fehler 2. Art**)
- Wie oben legen wir deshalb ein **Signifikanzniveau** α fest, z.B. $\alpha = 0.01$, also $1 - \alpha = 99\%$.
- Bei einem Hypothesentest beschreibt das Signifikanzniveau die **(maximale) Wahrscheinlichkeit für den Fehler 1. Art**, die wir zulassen wollen. Für unseren Test soll also auf jeden Fall gelten:

$$P_{H_0}\left[(X_1, X_2, \ldots, X_n) \text{ liegt im Verwerfungsbereich}\right] \leq \alpha$$

Schätzung des Erwartungswerts bei unbekannter Varianz

Der einseitige t-Test

- Um im diesem Fall einen Hypothesentest zu konstruieren, betrachten wir wieder die Teststatistik
$$T = \frac{\overline{X}_n - m_0}{S_n / \sqrt{n}} \quad .$$
- Beachte, daß wir jetzt $m = m_0$ $(= 50)$ setzen, da uns zunächst die Verwerfungswahrscheinlichkeit unter der Nullhypothese interessiert.

Schätzung des Erwartungswerts bei unbekannter Varianz

Der einseitige t-Test

- Da die Statistik

$$T = \frac{\overline{X}_n - m_0}{S_n / \sqrt{n}}$$

unter der Nullhypothese $t(n-1)$ verteilt ist, erhalten wir:

$$\alpha = P_{H_0}\left[T \leq t_{n-1,\alpha}\right] = P_{H_0}\left[\overline{X}_n \leq m_0 + t_{n-1,\alpha} \cdot S_n / \sqrt{n}\right]$$

- **Wir verwerfen daher die Nullhypothese $m = m_0$, falls**

$$\boxed{\overline{x}_n \leq m_0 + t_{n-1,\alpha} \cdot S_n / \sqrt{n}}$$

gilt - die Verwerfungswahrscheinlichkeit unter der Nullhypothese beträgt dann gerade α.

- Beachte, daß das Quantil $t_{n-1,\alpha}$ wegen $\alpha < 0.5$ negativ ist. In unserem Beipiel ist $m_0 = 50$ und

$$t_{n-1,\alpha} = t_{49,0.01} = -t_{49,0.99} \approx -2.4$$

- Im Fall $S_n = 8.3$ verwerfen wir die Nullhypothese $m = 50$ (bzw. $m \geq 50$) also genau dann, wenn

$$\overline{x}_n \leq 50 - 2.4 \cdot 8.3/\sqrt{50} \approx 47.4 \qquad \text{gilt.}$$

- Beobachtet wurde $\overline{x}_n = 47$, so daß wir die Nullhypothese also in diesem Fall tatsächlich verwerfen können. Weil wir das Signifikanzniveau $\alpha = 1\%$ gewählt haben, ist unser Testergebnis im Sinne der Notation von oben **signifikant** (d.h. ein Fehler 1. Art tritt nur bei einer von 100 Untersuchungen auf).

- Wäre die beobachtete Standardabweichung s_n der Meßwerte hingegen deutlich größer als 8.3 gewesen, dann hätten wir die Hypothese zum Signifikanzniveau $\alpha = 1\%$ nicht mehr verwerfen können, evtl. aber zum Signifikanzniveau $\alpha = 5\%$.

Schätzung des Erwartungswerts bei unbekannter Varianz

Der einseitige t-Test

Allgemein erhalten wir auf analoge Weise folgenden Test:

- Voraussetzung: Gaußmodell mit unbekanntem Mittel und Varianz.
- Nullhypothese: $m \geq m_0$ (bzw. $m \leq m_0$)
  Alternative: $m < m_0$ (bzw.$m > m_0$)
- Signifikanzniveau: α
- Teststatistik:
$$T = \frac{\overline{X}_n - m_0}{S_n/\sqrt{n}} \sim t(n-1)$$
- **Verwerfungsbereich:** Die Nullhypothese wird verworfen, falls $T \leq t_{n-1,\alpha}$ (bzw. $T \geq t_{n-1,1-\alpha}$) gilt, also falls

$$\boxed{\overline{x}_n \leq m_0 + t_{n-1,\alpha} \cdot S_n/\sqrt{n} \quad (\text{bzw. } \overline{x}_n \geq m_0 + t_{n-1,1-\alpha} \cdot S_n/\sqrt{n})}$$

Schätzung des Erwartungswerts bei unbekannter Varianz

Der beidseitige t-Test

Analog zu oben erhalten wir auch einen beidseitigen t-Test:

- Voraussetzung: Gaußmodell mit unbekanntem Mittel und Varianz.
- Nullhypothese: $m = m_0$, Alternative: $m \neq m_0$
- Signifikanzniveau: α
- Teststatistik:
$$T = \frac{\overline{X}_n - m_0}{S_n / \sqrt{n}} \sim t(n-1)$$
- **Verwerfungsbereich:** Die Nullhypothese wird verworfen, falls $|T| \leq t_{n-1,1-\alpha/2}$ gilt, also falls

$$\boxed{|\overline{x}_n - m_0| \geq t_{n-1,1-\alpha/2} \cdot S_n / \sqrt{n}}$$

Schätzung des Erwartungswerts bei unbekannter Varianz

- Wie schon beim z-Test können wir auch beim t-Test leicht den p-Wert berechnen. Zur Erinnerung:

Definition

Der p-**Wert** eines Hypothesentests ist das kleinste Signifikanzniveau α, bei dem die Nullhypothese verworfen wird.

- Der p-Wert hängt von den beobachteten Daten ab.
- Wenn wir den p-Wert kennen, sieht unsere Testentscheidung bei vorgegebenem Signifikanzniveau α so aus:
  - Verwerfe H_0 falls $p \leq \alpha$ — Belasse H_0 falls $p > \alpha$.

Schätzung des Erwartungswerts bei unbekannter Varianz

Der p-Wert beim t-Test

- Der t-Test basiert auf der Teststatistik

$$T = \frac{\overline{X}_n - m_0}{S_n / \sqrt{n}} \sim t(n-1)$$

- Wir verwerfen, wenn der beobachtete Wert der t-Statistik

$$t = \frac{\overline{x}_n - m_0}{s_n / \sqrt{n}}$$

einen gewissen Wert unter- bzw. überschreitet.

Theorem

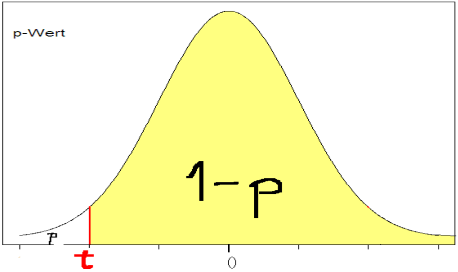
Der p-Wert beim t-Test mit Beobachtungswert t ist

$$\begin{aligned} p &= P_{H_0}\left[|T| \geq |t|\right] && \textit{bei beidseitiger Alternative,} \\ p &= P_{H_0}\left[T \leq t\right] && \textit{bei linksseitiger Alternative, und} \\ p &= P_{H_0}\left[T \geq t\right] && \textit{bei rechtsseitiger Alternative.} \end{aligned}$$

Schätzung des Erwartungswerts bei unbekannter Varianz

Berechnung des p-Werts

- **Der p-Wert gibt also an, mit welcher Wahrscheinlichkeit die t-Statistik einen Wert annimmt, der noch stärker zur Alternative hinzeigt, als der tatsächlich beobachtete Wert.**
- Die Grafik zeigt den p-Wert beim t-Test mit linksseitiger Alternative:



- **Berechnung des p-Werts in unserem Beispiel**:

$$p = P_{H_0}\left[T \leq t\right] = F_T(t)$$

- Im Beispiel ist

$$t = \frac{\overline{x}_n - m_0}{s_n/\sqrt{n}} = \frac{47 - 50}{8.3/\sqrt{50}} = -2.56$$

- Wir erhalten also

$$p = F_T(-2.56) = 1 - F_T(2.56) \approx 1 - 0.993 = 0.7\%$$

- Wir würden die Nullhypothese also beim Signifikanzniveau $\alpha = 0.1\%$ nicht verwerfen, d.h. das Testergebnis ist nicht "**stark signifikant"**.

7.3. Die Normalverteilungsannahme

- Bisher haben wir vorausgesetzt, daß die Beobachtungswerte normalverteilt sind. In diesem Fall kann man beweisen, daß die t-Tests in gewissem Sinn optimal sind.
- Bei großen Stichproben kann der t-Test aufgrund des zentralen Grenzwertsatzes auch ohne diese Voraussetzung angewendet werden, falls die zugrundeliegenden Verteilungen nicht zu langschwänzig sind.
- Bei kleinen Stichproben oder vielen Ausreißern in den Daten ist der t-Test aber nicht mehr optimal - und die im t-Test berechneten Wahrscheinlichkeiten sind nicht korrekt.

Die Normalverteilungsannahme

Damit stellen sich zwei wichtige Fragen:

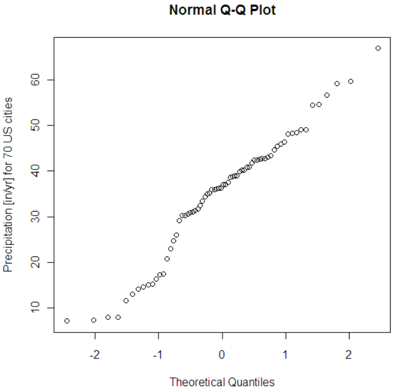
1. Wie können wir feststellen, ob eine Normalverteilungsannahme aufgrund der beobachteten Daten realistisch ist ?
2. Durch was können wir den t-Test ersetzen, wenn wir nicht von einer Normalverteilung ausgehen können, und nur wenig über die zugrundeliegende Verteilung wissen ?

- Wir gehen zunächst auf die erste Frage ein, und behandeln dann die zweite Frage in Abschnitt 7.4.

Die Normalverteilungsannahme

QQ-Plot/Normalplot

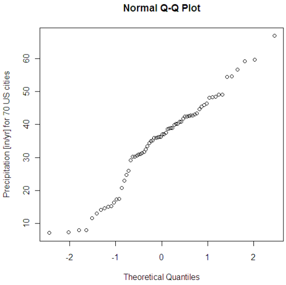
Um sich rasch einen Überblick zu veschaffen, ob die Daten durch eine Normalverteilung angemessen beschrieben werden, kann man einen **Quantil-Quantil-Plot (QQ-Plot)** der empirischen Verteilung der Beobachtungswerte $x_1, x_2, \ldots, x_n$ gegen die Standardnormalverteilung erstellen.



Die Normalverteilungsannahme

QQ-Plot/Normalplot

- In einem **Normal-QQ-Plot** werden die empirischen Quantile $x_{(\alpha)}$ auf der y-Achse gegen die theoretischen Quantile z_α unter der Standardnormalverteilung auf der x-Achse aufgetragen.
- Genauer werden die Punkte mit den Koordinaten $\left(z_\alpha, x_{(\alpha)}\right)$ für $\alpha = \frac{0.5}{n}, \frac{1.5}{n}, \frac{2.5}{n}, \ldots, 1 - \frac{0.5}{n}$ aufgetragen.
- **Beispiel:** Normalplot der Niederschlagsdaten aus 70 amerikanischen Städten:



Die Normalverteilungsannahme

QQ-Plot/Normalplot

- Sind die Daten Realisierungen einer normalverteilten Zufallsgröße

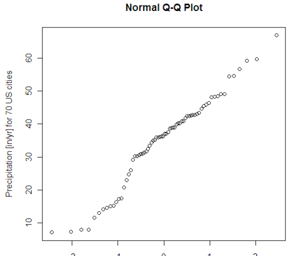
$$X \sim N(m, \sigma^2) \quad ,$$

dann gilt für die empirischen Quantile $x_{(\alpha)}$ näherungsweise

$$x_{(\alpha)} \approx m + \sigma \cdot z_\alpha \quad .$$

- Die Punkte im Normalplot liegen also etwa auf der Gerade

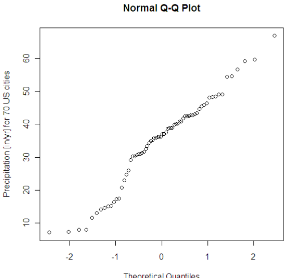
$$y = m + \sigma \cdot x$$



Durch Erstellen des **Normalplots** können wir also einschätzen, ob eine Normalverteilungsannahme gerechtfertigt ist, und Näherungswerte für die Parameter m und σ ablesen:

- Liegt der QQ-Plot in guter Näherung auf einer Gerade, dann können wir von einer Normalverteilung ausgehen.
- In diesem Fall liefert die Steigung der Geraden einen Schätzwert für die Standardabweichung, und der Achsenabschnitt auf der y-Achse einen Schätzwert für den Mittelwert.
- Normalplots und allgemeinere QQ-Plots kann man mit gängiger Statistiksoftware leicht erstellen.

Beispiel:



Die Normalverteilungsannahme

Chiquadrat-Anpassungstest

- Wenn man eine Normalverteilungsannahme mit bestimmten Parametern (die z.B. auf einem QQ-Plot basiert) noch weitergehend empirisch bestätigen möchte, kann man einen Chiquadrat-Anpassungstest durchführen.
- Dazu werden die Beobachtungswerte in Klassen eingeteilt, und dann eine Chiquadrat-Statistik basierend auf der Normalverteilung berechnet. Überschreitet diese Statistik einen gewissen kritischen Wert (der vom Signifikanzniveau abhängt), dann wird die Normalverteilungsannahme verworfen.

Die Normalverteilungsannahme

Chiquadrat-Anpassungstest

- Wollen wir beispielsweise testen, ob die Beobachtungswerte $x_1, x_2, \ldots, x_n$ Realisierungen einer standardnormalverteilten Zufallsvariable sind, dann teilen wir die reellen Zahlen zunächst in die 10 unter N(0,1) gleich wahrscheinlichen Klassen

$$\begin{aligned} K_1 &= (-\infty, z_{0.1}], \quad K_2 = (z_{0.1}, z_{0.2}], \\ K_3 &= (z_{0.2}, z_{0.3}], \quad \ldots \quad , \ K_{10} = (z_{0.9}, \infty) \qquad \text{ein.} \end{aligned}$$

- Wir bezeichnen mit $\widehat{n}_l$ die Anzahl der Beobachtungwerte, die in der l-ten Klasse K_l liegen, und mit $n_l = n/10$ die bei Standardnormalverteilung zu erwartende durchschnittliche Anzahl von Werten in K_l.
- Wir berechnen dann die **Chiquadrat-Statistik**

$$\chi^2 = \sum_{l=1}^{10} \frac{(n_l - \widehat{n}_l)^2}{\widehat{n}_l}$$

Die Normalverteilungsannahme

Chiquadrat-Anpassungstest

- Wir betrachten nun die Nullhypothese

$$H_0 : \quad X_i \sim N(0,1)$$

und die Alternative

$$H_1 : \quad X_i \nsim N(0,1).$$

- Unter der Nullhypothese hat die Teststatistik χ^2 näherungsweise eine Chiquadratverteilung mit $10 - 1 = 9$ Freiheitsgraden. Die Quantile dieser Verteilung sind tabelliert.
- Damit ergibt sich folgender **Test zum Signifikanzniveau** α :
  - Verwerfe H_0, falls

$$\chi^2 = \sum_{l=1}^{10} \frac{(n_l - \widehat{n}_l)^2}{\widehat{n}_l} \geq u_{1-\alpha}$$

wobei $u_{1-\alpha}$ das $1 - \alpha$ Quantil der $\chi^2(9)$-Verteilung bezeichnet.

Die Normalverteilungsannahme

Chiquadrat-Anpassungstest

- Mit dem Chiquadrat-Anpassungstest kann man auch analog auf andere Verteilungen testen (z.B. Test auf gleiche Wahrscheinlichkeiten aller Augenzahlen beim Würfeln).
- Die asymptotische Verteilung der Teststatistik unter der Nullhypothese bleibt dabei unverändert eine Chiquadratverteilung.

7.4. Nichtparametrische Tests

- Bisher haben wir vorausgesetzt, daß die Beobachtungswerte normalverteilt sind. In diesem Fall kann man beweisen, daß die t-Tests in gewissem Sinn optimal sind.
- Bei großen Stichproben kann der t-Test aufgrund des zentralen Grenzwertsatzes auch ohne diese Voraussetzung angewendet werden, falls die zugrundeliegenden Verteilungen nicht zu langschwänzig sind.
- Bei kleinen Stichproben oder vielen Ausreißern in den Daten ist der t-Test aber nicht mehr optimal - und die im t-Test berechneten Wahrscheinlichkeiten sind nicht korrekt.

- Daher wollen wir nun zwei Tests betrachten, die **unabhängig von der zugrundeliegenden Verteilung** anwendbar, und relativ **stabil unter Ausreißern** sind. Diese Tests basieren auf der **Schätzung des Medians** statt des Mittelwerts.

- **Voraussetzung:** $X_1, X_2, ..., X_n$ sind unabhängige Stichproben einer stetigen W'keitsverteilung mit Median μ.
- **Nullhypothese:** $H_0 : \mu = \mu_0$, **Alternative:** $H_1 : \mu \neq \mu_0$
.
- **Teststatistik:**

$$\begin{aligned} V &= \text{Anzahl der } X_i\text{'s mit } X_i > \mu_0 \\ &= \text{Anzahl der positiven Vorzeichen von } X_i - \mu_0 \end{aligned}$$

- Verteilung der Teststatistik unter der Nullhypothese:

$$V \sim Bin(n, 1/2) \quad (\approx N(n/2, n/4) \text{ für große } n)$$

$$
\begin{aligned}
V &= \text{Anzahl der } X_i\text{'s mit } X_i > \mu_0 \\
V &\sim Bin(n, 1/2) \qquad (\approx N(n/2, n/4) \text{ für große } n)
\end{aligned}
$$

- **Testentscheidung für große n ($n \geq 30$):**
  - $H_0 : \mu = \mu_0$ wird verworfen, falls

$$\left| V - \frac{n}{2} \right| \geq \sqrt{\frac{n}{4}} \cdot z_{1-\alpha/2}$$

- **Testentscheidung für kleine n ($n < 30$):**
  - $H_0 : \mu = \mu_0$ wird verworfen, falls

$$V \geq b_{1-\alpha/2}(n/2, n/4) \quad \text{oder} \quad V \leq b_{\alpha/2}(n/2, n/4)$$

  - Hierbei bezeichnet $b_\alpha(n, p)$ das α-Quantil der Bin(n,p)-Verteilung.

- Der Vorzeichentest kann IMMER (für alle Verteilungen der Einzelstichproben) angewendet werden.
- Bei normalverteilten Daten ist die W'keit, daß die Nullhypothese beim Vorzeichentest verworfen wird, deutlich kleiner als z.B. beim t-Test.
- Bei nicht in guter Approximation normalverteilten Daten ist der Vorzeichentest dagegen dem t-Test oft überlegen.
- Ein Nachteil beim Vorzeichentest ist, daß er die Information, um wieviel die Beobachtungswerte von μ_0 abweichen, nicht benutzt.

Der Wilcoxon-Test ist ein Kompromiss, der keine Normalverteilung voraussetzt wie der t-Test, aber die Information der Daten besser ausnützt als der Vorzeichentest.

- **Voraussetzung:** $X_1, X_2, \ldots, X_n$ sind unabhängige Stichproben einer stetigen **symmetrischen** W'keitsverteilung mit Median (=Mittelwert) μ.
- **Nullhypothese:** $H_0 : \mu = \mu_0$, **Alternative:** $H_1 : \mu \neq \mu_0$.
- **Teststatistik:** W_n berechnet sich aus den Rängen von den der Größe (im Absolutbetrag) nach angeordneten Abweichungen $X_i - \mu_0$ der Daten von μ_0.
- Den p-Wert des Tests kann man mit den gängigen Statistiksoftwarepaketen berechnen.

Nichtparametrische Tests

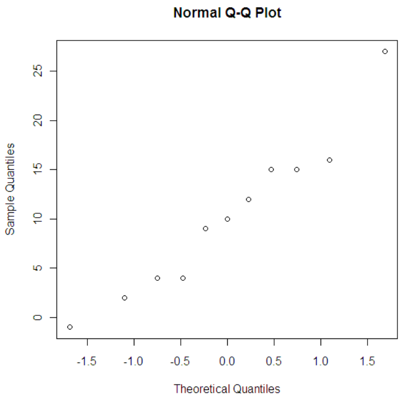
- **Der Wilcoxon-Test ist in vielen Fällen vorzuziehen. Er hat oftmals eine bessere Macht (Gütefunktion) als der t-Test und der Vorzeichentest.**
- **Nur falls die Daten sehr gut mit einer Normalverteilung beschrieben werden, ist der t-Test für eine gute Datenanalyse "in vollem Umfang tauglich" !**

Beispiel. (Blutplättchen-Aggregation)

Bei 11 Individuen wurde die Aggregation von Blutplättchen vor und nach dem Rauchen einer Zigarette gemessen. Die folgenden Daten geben den Anteil aggregierter Blutplättchen (in Prozent) nach einer Stimulation an:

Testperson	1	2	3	4	5	6	7	8	9	10	11
vorher	25	25	27	44	30	67	53	53	52	60	28
nachher	27	29	37	56	46	82	57	80	61	59	43
Differenz x_i	2	4	10	12	16	15	4	27	9	-1	15

- Als erstes erstellen wir einen **Normalplot** der Daten:



- Die Punkte lassen sich recht gut durch eine Gerade approximieren, sodass wir zunächst von einer Normalverteilung ausgehen können.

- Wir fassen daher die Differenzen

$$x_i = \text{Aggregation nachher} \; - \; \text{Aggregation vorher}$$

  als unabhängige Realisierungen einer $N(m, \sigma^2)$ verteilten Zufallsgröße auf, und führen einen **t-Test** durch.
- $H_0 : m = 0 \,, \qquad H_1 : m > 0.$
- Die realisierte Teststatistik ist

$$t = \frac{\overline{x}_n - 0}{s_n / \sqrt{n}} = 4.27$$

- Als p-Wert ergibt sich

$$p = P_{H_0}\left[T > 4.27\right] = 1 - F_{t(10)}(4.27) = 0.00082$$

- Der Einfluß des Rauchens einer Zigarette ist also bzgl. der Blutplättchenaggregation **hochsignifikant**.

- Hätten wir stattdessen den t-Test für die **beidseitige Alternative** $m \neq 0$ durchgeführt, dann hätten wir den p-Wert

$$p = 0.0016$$

erhalten.

- Die Testentscheidung für einen Einfluss des Rauchens auf die Blutplättchenaggregation wäre jetzt signifikant, aber nicht hochsignifikant.

- Zum Vergleich führen wir auch den Vorzeichen- und den Wilcoxontest für das einseitige Testproblem

$$H_0 : \mu = 0 \quad \text{vs.} \quad H_1 : \mu > 0$$

durch.
- Beim **Vorzeichentest** ist die realisierte Teststatistik

$$v = 10$$

Als p-Wert erhalten wir

$$p = 0.005$$

- Beim **Wilcoxon-Test** ergibt sich der p-Wert

$$p = 0.0025$$

- In diesem Fall ist der t-Test also mächtiger, was an der approximativen Normalverteilung der Daten liegt. Die Testentscheidung beim Wilcoxon-Test ist aber zuverlässiger, weil sie keine Normalverteilung voraussetzt.