Beispiel

Beispiel: Parfum wird in 100 ml Flaschen abgefüllt. Durch eine Zufallsstichprobe vom Umfang $n = 36$ möchte das Management überprüfen, ob nicht etwa die Füllmenge zu hoch ist. Die Messungen können als normalverteilt angesehen werden. Die Fehlerwahrscheinlichkeit, fälschlicherweise aus den Daten auf eine zu hohe Füllmenge zu schlieflen, wird vom Management mit $\alpha = 1\%$ festgelegt.

$$\overline{x} = 101.78, \quad \sigma = 2.34, \quad n = 36$$

Modell: $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, μ: wahre Füllmenge, $\sigma^2 = 2.34^2$

1. Formulierung des Testproblems:

$H_0 : \mu \leq 100$ versus $H_1 : \mu > 100$

2. Testverfahren:

Einseitiger 1-SP-Gaufltest (nach oben) mit $\mu_0 = 100$ durchgeführt.

3. Testdurchführung:

3a. Berechne $t = \sqrt{n}\frac{\overline{x}-\mu_0}{\sigma}$.

3b. Berechne kritischen Wert $c_{krit} = q_{0.99}$

3c. Lehne H_0 ab, falls $t > c_{krit}$. Behalte H_0 bei, falls $t \leq c_{krit}$

Beispiel

Datenanalyse und -interpretation:

$$\overline{x} = 101.78, \quad \sigma = 2.34, \quad n = 36$$

Berechnung der Teststatistik:

$$t = T_{obs} = \sqrt{36}\frac{101.78 - 100}{2.34} = 4.564...$$

Berechnung des kritischen Werts (z.B. Raussuchen Tabelle Buch S. 304):

$$c_{krit} = q_{1-0.01} = q_{0.99} \approx 2.33$$

Testdurchführung: Da $t = 4.465 > 2.33 = c_{krit}$, wird H_0 auf dem Niveau $\alpha = 0.01$ abgelehnt.
Antwortsatz: Basierend auf einer normalverteilten Zufallsstichprobe vom Umfang $n = 36$ konnte durch Anwendung eines einseitigen Gaußtests statistisch auf einem Signifikanzniveau von $\alpha = 1\%$ nachgewiesen werden, dass die Füllmenge größer als der Sollwert $100[ml]$ ist.

Der t-Test (für eine Stichprobe)

Gegeben: $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ mit *unbekannter* Varianz σ^2

Idee: Ersetze die unbekannte Varianz σ^2 durch den (erwartungstreuen und konsistenten) Schätzer

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 \ .$$

Teststatistik: $T = \sqrt{n}\, \frac{\overline{X}_n - \mu_0}{S}$ ($\mu_0 \in \mathbb{R}$ vorgegebener Sollwert)

Verteilung der Teststatistik: $T \sim t(n-1)$ für $\mu = \mu_0$

Der t-Test (für eine Stichprobe)

Einseitiger t-Test (1)

Der einseitige t-Test verwirft die Nullhypothese $H_0 : \mu \leq \mu_0$ auf dem Signifikanzniveau α zugunsten von $H_1 : \mu > \mu_0$, wenn $T > t(n-1)_{1-\alpha}$.

Einseitiger t-Test (2)

Der einseitige t-Test verwirft die Nullhypothese $H_0 : \mu \geq \mu_0$ auf dem Signifikanzniveau α zugunsten von $H_1 : \mu < \mu_0$, wenn $T < -t(n-1)_{1-\alpha} = t(n-1)_{\alpha}$.

Zweiseitiger t-Test

Der zweiseitige t-Test verwirft die Nullhypothese $H_0 : \mu = \mu_0$ auf dem Signifikanzniveau α zugunsten von $H_1 : \mu \neq \mu_0$, wenn $|T| > t(n-1)_{1-\alpha/2}$.

(Hierbei bezeichnet $t(n-1)_p$ das p-Quantil zur t-Verteilung mit $n-1$ Freiheitsgraden für $p \in (0,1)$.)

Der t-Test (für eine Stichprobe)

Beispiel

Die Schätzung der mittleren Ozonkonzentration während der Sommermonate ergaben für eine Großstadt anhand von $n = 26$ Messungen den Mittelwert $\overline{x}_n = 244$ und die Stichproben-Standardabweichung $s = 5.1$ (jeweils in $\mu g/m^3$).

Der im Ozongesetz von 1995 festgelegte verbindliche Alarmwert beträgt 240 $\mu g/m^3$. Kann das gemessene Ergebnis als signifikante Überschreitung des Warnwerts gewertet werden zum Signifikanzniveau $\alpha = 0.01$?

Lösung:

• Als beobachtete Teststatistik erhalten wir aus den Daten:

$$t = T_{obs} = \sqrt{26}\frac{244 - 240}{5.1} = 3.999,$$

Die Statistik T ist $t(n-1=25)$-verteilt, wenn $\mu = \mu_0$, also am Rand der Nullhypothese.

• Bestimmung des kritischen Werts: $c_{krit} = t(25)_{0.99} = 2.485$.
• Testentscheidung: Da $t > 2.485$ wird die Nullhypothese $H_0 : \mu \leq 240$ zu Gunsten von $H_1 : \mu > 240$ verworfen.
• Antwortsatz: Durch einen Signifikanztest (1-Stichproben $t-$Test) konnte basierend auf einer Zufallsstichprobe vom Umfang $n = 26$ auf einem Signifikanzniveau von 1% statistisch nachgewiesen werden, dass von einer Überschreitung des Alarmwert 240 ausgegangen werden kann.

Wichtige Angaben: n, α, verwendetes Testverfahren, die Hypothesenformulierung muss ersichtlich sein.

Zusammenhang Test $\leftrightarrow$ Konfidenzintervall

Der **zweiseitige** t**-Test** akzeptiert H_0 auf dem Niveau α, wenn

$$\left| \sqrt{n} \frac{\overline{X} - \mu_0}{S} \right| \le t(n-1)_{1-\alpha/2}$$

(sonst wird H_0 abgelehnt). Dies ist äquivalent zur Ungleichungskette

$$\mu_0 - t(n-1)_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} \ \le \ \overline{X} \ \le \ \mu_0 + t(n-1)_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}.$$

Man kann also auch $\overline{X}$ mit $\mu_0 \pm t(n-1)_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$ vergleichen.
Weiteres Umformen liefert:

$$\overline{X} - t(n-1)_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} \ \le \ \mu_0 \ \le \ \overline{X} + t(n-1)_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}.$$

$H_0 : \mu = \mu_0$ wird somit genau dann akzeptiert, wenn der Sollwert μ_0 vom $(1-\alpha)$-Konfidenzintervall für μ überdeckt wird.

Zusammenhang Test $\leftrightarrow$ Konfidenzintervall

Merke: (Erfahrene Data Analysts kennen solche Zusammenhänge...)

Das (zweiseitige) Konfidenzintervall für μ liefert die wertvolle Information, welche Nullhypothesen vom vorliegenden Datenmaterial durch den (zweiseitigen) t-Test abgelehnt werden.

p-Wert

Durchführung eines statistischen Tests:

1. Formuliere H_0 und H_1.
2. Wähle Signifikanzniveau α.
3. Bestimme kritischen Wert c_{krit}.
4. Berechne t_{obs}.
5. Vergleiche t_{obs} mit c_{krit}.

Nachteile:

Bei Änderung von α müssen 3. bis 5. neu durchgeführt werden.

In der Praxis geht man so vor:

Angabe einer Zahl p, so dass folgende Regel gilt:

$$p < \alpha \quad \Leftrightarrow \quad H_0 \text{ ablehnen}$$

p-Wert

Frage

Wie wahrscheinlich ist es, bei einer (gedanklichen) Wiederholung des Experiments, einen Teststatistik-Wert zu beobachten, der noch deutlicher gegen H_0 spricht als t_{obs}?

p-Wert

Einseitige Tests

Testproblem: $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$

$$p = P_{\mu_0}(T > t_{obs})$$

Testproblem: $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$

$$p = P_{\mu_0}(T < t_{obs})$$

Lehne H_0 genau dann ab, wenn $p < \alpha$.

Zweiseitiger Test

Testproblem: $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$

$$p_{zweis} = P_{\mu_0}(|T| > |t_{obs}|)$$

Lehne H_0 genau dann ab, wenn $p_{zweis} < \alpha$.

p-Wert

Einseitige Tests

Gegeben: t_{obs} und p_{zweis}.

1. Lehne $H_0 : \mu \leq \mu_0$ zugunsten von $H_1 : \mu > \mu_0$ ab, falls

$$t_{obs} \geq 0 \quad \text{und} \quad \frac{p_{zweis}}{2} \overset{t_{obs} \geq 0}{=} P_{\mu_0}(T > t_{obs}) < \alpha \ .$$

2. Lehne $H_0 : \mu \geq \mu_0$ zugunsten von $H_1 : \mu < \mu_0$ ab, falls

$$t_{obs} \leq 0 \quad \text{und} \quad \frac{p_{zweis}}{2} \overset{t_{obs} \leq 0}{=} P_{\mu_0}(T < t_{obs}) < \alpha \ .$$

Gütefunktion

Frage: Wie wahrscheinlich ist es, dass die Alternative H_1 tatsächlich aufgedeckt wird?

Gesucht: $P_{H_1}(„H_1“)$.

Gütefunktion

Die Funktion

$$G(\mu) = P_\mu(„H_1“) = P(„H_1“|\mu, \sigma^2), \qquad \mu \in \mathbb{R},$$

heißt Gütefunktion (an der Stelle μ).

Beispiel

Beispiel

Sei $\mu_0 = 150$ *und* $\sigma = 10$. *Betrachte das Testproblem*

$$H_0 : \mu \leq 150, \qquad \textit{versus} \qquad H_1 : \mu > 150.$$

Wähle $\alpha = 0.01$. *Der einseitige Gauß-Test verwirft* H_0*, falls*

$$T > z_{0.99} = 2.3263$$

Bestimme $G(\mu)$ *für* $\mu \in \{155, 160\}$.

Fallzahlplanung:

Lösung: Berechnung der Gütefunktion

$$G(\mu) = P_\mu(T > 2.3263).$$

Ist μ der wahre Erwartungswert, dann ist T nicht korrekt zentriert. Korrektur:

$$\frac{\overline{X} - 150}{10/\sqrt{n}} = \underbrace{\frac{\overline{X} - \mu}{10/\sqrt{n}}}_{\sim N(0,1)} + \frac{\mu - 150}{10/\sqrt{n}}.$$

Schreibe $G(\mu)$ um:

$$\begin{aligned} G(\mu) &= P_\mu\left(\frac{\overline{X} - 150}{10/\sqrt{n}} > 2.3263\right) \\ &= P_\mu\left(\frac{\overline{X} - \mu}{10/\sqrt{n}} + \frac{\mu - 150}{10/\sqrt{n}} > 2.3263\right) \\ &= P_\mu\left(\frac{\overline{X} - \mu}{10/\sqrt{n}} > 2.3263 - \frac{\mu - 150}{10/\sqrt{n}}\right) \\ &= 1 - \Phi\left(2.3263 - \frac{\mu - 150}{10/\sqrt{n}}\right) \end{aligned}$$

Fallzahlplanung:

Für $n = 25$ und $\mu = 155$ erhalten wir

$$G(155) = \Phi(-2.3263 + 2.5) = \Phi(0.1737) \approx 0.569.$$

Genauso berechnet man $G(160) = \Phi(2.6737) \approx 0.9962$.

Eine Abweichung von 10 Einheiten wird also mit sehr hoher Wahrscheinlichkeit entdeckt, 5 Einheiten jedoch lediglich mit Wahrscheinlichkeit ≈ 0.57.

Ersetzt man in der obigen Herleitung 2.3263 durch $z_{1-\alpha}$, 150 durch μ und 10 durch σ, so erhält man die allgemeine Formel für die Güte des einseitigen Gaußtests:

Gütefunktion

Formel für die Güte des einseitigen Gaußtests:

$$G(\mu) = \Phi\left(-z_{1-\alpha} + \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right)$$

Analog für den zweiseitigen Test:

$$G_{\text{zweis.}}(\mu) = 2\Phi\left(-z_{1-\alpha/2} + \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right)$$

Hinweis: In der Praxis wird σ aus Trainingsdaten (historischen Daten) durch S geschätzt.

Fallzahlplanung

Ziel: Bestimme den Stichproben-Umfang n so, dass eine vorgegebene Lageänderung d $(= \mu - \mu_0)$ von μ_0 mit einer Mindestwahrscheinlichkeit von $1 - \beta$ aufgedeckt wird.

Hierdurch wird auch der Fehler 2. Art kontrolliert: Die Fehlerwahrscheinlichkeit 2. Art ist ein Abweichung d (oder schlimmer) höchstens β.

Beispiel

Ansatz: Bestimme n so, dass eine Abweichung von 5 mit Wahrscheinlichkeit von mindestens 90% aufgedeckt wird.
Mit $\mu(= 155)$ ist n so zu bestimmen, dass

$$\Phi\left(-2.3263 + \frac{\mu - 150}{10/\sqrt{n}}\right) \geq 0.9.$$

Bezeichne das Argument von Φ mit z. Zu Lösen ist also $\Phi(z) \geq 0.9$. Da $\Phi(z)$ und die Inverse $\Phi^{-1}(p)$ streng monton wachsend sind, ist

$$\Phi(z) \geq 0.9 \Leftrightarrow z \geq z_{0.9}$$

(allg.: $\Phi(z) \geq 1 - \beta \Leftrightarrow z \geq z_{1-\beta}$). Also:

$$z = -2.3263 + \sqrt{n}\frac{\mu - 150}{10} \geq z_{0.9}$$

Formales Auflösen nach n liefert für $\mu = 155$ und $z_{0.9} = 1.12816..$:

$$n \geq \frac{10^2}{5^2}(2.3263 + 1.2816)^2 = 52.068$$

$\rightarrow$ **Die gewünschte Schärfe des Tests erfordert $n \geq 53$.**

Beispiel

Hatten:

$$G(\mu) = \Phi\left(-2.3263 + \frac{\mu - 150}{10/\sqrt{n}}\right), \qquad \mu \in \mathbb{R}.$$

Finde den minimalen Stichprobenumfang $n \in \mathbb{N}$*, so dass eine Abweichung von* $d = 5$ *mit einer Wahrscheinlichkeit von mindestens* 90% *aufgedeckt wird.*

Fallzahlplanung

Mindestfallzahl

$$n \geq \frac{\sigma^2}{|\mu - \mu_0|^2}(z_{1-\alpha} + z_{1-\beta})^2.$$

Für den zweiseitigen Fall ergibt sich die Forderung

$$n \geq \frac{\sigma^2}{|\mu - \mu_0|^2}(z_{1-\alpha/2} + z_{1-\beta})^2,$$

damit Abweichungen größer oder gleich $\Delta = |\mu - \mu_0|$ mit einer Mindestwahrscheinlichkeit von $1 - \beta$ aufgedeckt werden.

2-Stichproben-Tests

Zwei Grundsituationen

1. Verbundenes Design
2. Unverbundenes Design

Verbundenes Design

Typische Anwendungssituation:
'Vorher-Nachher-Design' zur Analyse von zeitlichen Effekten bzw. Effekten nach Änderung der Versuchsbedingungen.

Beispiele:

- Änderung Druck/Temperatur/Spannung/...,
- neue Marketing-Maßnahme
- Schulung
- Umstrukturierung
- Medikament

Verbundenes Design

Für $i = 1, \dots, n$ erhebe

X_i : Messung an der i-ten Versuchseinheit vorher,

Y_i : Messung an der i-ten Versuchseinheit nachher.

Modell: Bivariate einfache Stichprobe

$$(X_1, Y_1), \dots, (X_n, Y_n)$$

von normalverteilten Zufallsvektoren mit

$$\mu_X = E(X_i) \qquad \text{und} \qquad \mu_Y = E(Y_i)$$

Betrachte die Differenzen (nachher - vorher):

$$D_i = Y_i - X_i, \qquad i = 1, \dots, n.$$

Erwartungswert der Differenzen:

$$E(D_i) = E(Y_i) - E(X_i) = \mu_Y - \mu_X = \delta.$$

Verbundenes Design

Es sei $\sigma_D^2 = Var(D_1) = \ldots = Var(D_n)$ unbekannt.

Verwerfe dann

$$H_0 : \delta = 0 \Leftrightarrow \mu_X = \mu_Y \qquad \text{(kein Effekt)}$$

zugunsten von

$$H_1 : \delta \neq 0 \Leftrightarrow \mu_X \neq \mu_Y \qquad \text{(Effekt vorhanden)}$$

falls

$$|T| > t(n-1)_{1-\alpha/2},$$

wobei

$$T = \sqrt{n}\,\frac{\overline{D}}{S_D} \quad \text{mit} \quad \overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i\ ,\ S_D = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})^2}\ .$$

Analog konstruiert man einseitige Tests, s. Buch.

Unverbundes 2-Stichprobenproblem

Modell: Zwei unabhängige Stichproben

$$X_{11}, \ldots, X_{1n_1} \overset{i.i.d.}{\sim} N(\mu_1, \sigma_1^2)$$
$$X_{21}, \ldots, X_{2n_2} \overset{i.i.d.}{\sim} N(\mu_2, \sigma_2^2)$$

Schritte:

1. Test auf Varianzhomogenität: Gilt $\sigma_1^2 = \sigma_2^2$?
2. Test auf Lageunterschied: Gilt $\mu_1 = \mu_2$?

Test auf Varianzhomogenität

Testproblem

$$H_0 : \sigma_1^2 = \sigma_2^2 \qquad \text{versus} \qquad H_1 : \sigma_1^2 \neq \sigma_2^2$$

Varianzschätzungen:

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (X_{1j} - \overline{X}_1)^2, \quad S_2^2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (X_{2j} - \overline{X}_2)^2$$

Teststatistik: $F = \frac{S_1^2}{S_2^2}$. Unter $H_0 : \sigma_1^2 = \sigma_2^2$ ist F F-verteilt!

Test

H_0 ablehnen, falls

$$F < F(n_1 - 1, n_2 - 1)_{\alpha/2} \qquad \text{oder} \qquad F > F(n_1 - 1, n_2 - 1)_{1-\alpha/2}$$

Äquivalent: Nummeriere so, dass $S_1^2 \leq S_2^2$ und lehne H_0 ab, falls $F < F(n_1 - 1, n_2 - 1)_{\alpha/2}$.

2-Stichproben-t-Test auf Lageunterschied

Annahme: $\sigma_1 = \sigma_2 =: \sigma^2$ (Varianzhomogenität).

Testproblem (zweiseitig):

$$H_0 : \mu_1 = \mu_2 \qquad \text{(kein Lageunterschied)}$$

versus

$$H_1 : \mu_1 \neq \mu_2 \quad \text{(Lageunterschied)}$$

Testprobleme (einseitig):

$$H_0 : \mu_1 \geq \mu_2 \qquad \text{versus} \qquad H_1 : \mu_1 < \mu_2.$$

bzw.

$$H_0 : \mu_1 \leq \mu_2 \qquad \text{versus} \qquad H_1 : \mu_1 > \mu_2.$$

2-Stichproben-t-Test auf Lageunterschied

Teststatistik: $$T = \frac{\overline{X}_2 - \overline{X}_1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, S}$$ mit

$$\begin{aligned} S^2 &= \frac{n_1 - 1}{n_1 + n_2 - 2} S_1^2 + \frac{n_2 - 1}{n_1 + n_2 - 2} S_2^2 \\ &= \frac{1}{n_1 + n_2 - 2} \left(\sum_{i=1}^{n_1} (X_{1i} - \overline{X}_1)^2 + \sum_{j=1}^{n_2} (X_{1j} - \overline{X}_2)^2 \right) . \end{aligned}$$

2-Stichproben-t-Test

1. Lehne $H_0 : \mu_1 = \mu_2$ zugunsten von $H_1 : \mu_1 \neq \mu_2$ ab, wenn $|T| > t(n_1 + n_2 - 2)_{1-\alpha/2}$.
2. Lehne $H_0 : \mu_1 \geq \mu_2$ zugunsten von $H_1 : \mu_1 < \mu_2$ ab, wenn $T > t(n_1 + n_2 - 2)_{\alpha}$.
3. Lehne $H_0 : \mu_1 \leq \mu_2$ zugunsten von $H_1 : \mu_1 > \mu_2$, falls $T < t(n_1 + n_2 - 2)_{1-\alpha}$.

Welch-Test auf Lageunterschied

Bei Varianzinhomogenität $\sigma_1^2 \neq \sigma_2^2$ verwende man den Welch-Test.

Teststatistik:

$$T = \frac{\overline{X}_2 - \overline{X}_1}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}.$$

Lehne $H_0 : \mu_1 = \mu_2$ auf dem Niveau α ab, wenn $|T| > t(df)_{1-\alpha/2}$, wobei

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\left(\frac{S_1^2}{n_1}\right)^2 \frac{1}{n_1-1} + \left(\frac{S_2^2}{n_2}\right)^2 \frac{1}{n_2-1}}$$

Falls $df \notin \mathbb{N}$, dann vorher auf nächste ganze Zahl abrunden.

Fallzahlplanung

Für $n = n_1 = n_2$ kann man folgende Näherungen verwenden:
Zweiseitiger Test: Wähle

$$n \geq \frac{\sigma^2}{\Delta^2}(z_{1-\alpha/2} + z_{1-\beta})^2,$$

um eine Schärfe von $1 - \beta$ bei einer Abweichung von $\Delta = |\mu_A - \mu_B|$ näherungsweise zu erzielen.
Einseitiger Test: Wähle

$$n \geq \frac{\sigma^2}{\Delta^2}(z_{1-\alpha} + z_{1-\beta})^2,$$

um eine Schärfe von $1 - \beta$ bei einer Abweichung von $\Delta = |\mu_A - \mu_B|$ näherungsweise zu erzielen.