Spiekbriefjes bij Conclusies trekken

Hypothese toetsen, z-toetsen

Hypothese toetsen is een manier om een bewering over een populatie te controleren.

Stel X is normaal verdeeld met gemiddelde μ_X en standaardafwijking σ_X.
Iemand denkt dat het gemiddelde het lager is. Je hebt dan een **linkszijdige toets**:

- de **nulhypothese** is H_0: $\mu = \mu_X$ (gemiddelde klopt)
- de **alternatieve hypothese** is H_1: $\mu < \mu_X$ (gemiddelde lager)

Je neemt een steekproef van grootte n en berekent $\overline{X}$.

$\overline{X}$ is normaal verdeeld met $\overline{X} \approx \mu_X$ en $S_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}$.

Er hoort een **kritiek gebied** $\overline{X} < g$ bij dat aangeeft wanneer je H_0 verwerpt.

De grens g ervan bepaal uit $\text{P}\left(\overline{X} < g\right) = \text{P}\left(z < \frac{g-\mu_X}{\frac{\sigma_X}{\sqrt{n}}}\right) = \alpha$.

Het **significantieniveau** α kies je voordat je de toets uitvoert, bijvoorbeeld $\alpha = 5\%$ of $\alpha = 1\%$.

Je kunt ook een **rechtszijdige toets** (H_0 tegen $\text{H}_1 : \mu > \mu_X$ of een **tweezijdige toets** (H_0 tegen $\text{H}_1 : \mu \neq \mu_X$) uitvoeren.

meer info

t-toetsen

Stel X is normaal verdeeld met gemiddelde μ_X en standaardafwijking σ_X.
Iemand toetst H_0: $\mu = \mu_X$ tegen H_1: $\mu \neq \mu_X$ met significantieniveau α, dus met een betrouwbaarheid van $(1-\alpha) \cdot 100\%$ en met een steekproef van grootte n:

- Is de populatiestandaarddeviatie bekend, gebruik je de z-verdeling.
- Is de populatiestandaarddeviatie niet bekend gebruik je **Student's**
  ***t*-verdeling** $t = \frac{\overline{X}-\mu_X}{\frac{s_X}{\sqrt{n}}}$
  met de steekproefstandaarddeviatie s_X en **vrijheidsgraad** $v = n - 1$.

Bij het bepalen van betrouwbaarheidsintervallen heb je deze twee mogelijkheden:

- De populatiestandaarddeviatie is bekend.
  Het betrouwbaarheidsinterval is $\overline{X} - z \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z \cdot \frac{\sigma}{\sqrt{n}}$.
- De populatiestandaarddeviatie is niet bekend.
  Het betrouwbaarheidsinterval is $\overline{X} - t \cdot \frac{s_X}{\sqrt{n}} < \mu < \overline{X} + t \cdot \frac{s_X}{\sqrt{n}}$.

meer info

f-toetsen

Je wilt door twee series metingen A en B een bepaalde grootheid X vaststellen. Bij A met steekproefgrootte n, krijg je gemiddelde $\overline{X_1}$ en standaardafwijking s_{X_1}. Bij B met steekproefgrootte m, krijg je gemiddelde $\overline{X_2}$ en standaardafwijking s_{X_2}.

Met een ***f*-toets** kun je de precisie van beide metingen vergelijken.
Neem aan $s_{X_1} \geq s_{X_2}$, dan toets je

- H_0: $s_{X_1} = s_{X_2}$;
- H_1: $s_{X_1} \neq s_{X_2}$ (dubbelzijdige toets) of H_1: $s_{X_1} > s_{X_2}$ (enkelzijdige toets).

Daarbij gebruik je de f-verdeling $f = \frac{s^2_{X_1}}{s^2_{X_2}}$ omdat $s_{X_1} > s_{X_2}$.

Meestal neem je een betrouwbaarheid van 95% en daarbij een ***f*-tabel**.
Je kunt ook de juistheid van beide series metingen vergelijken. Daartoe gebruik je een t-toets.

meer info

Regressielijnen

In een **spreidingsdiagram** van de variabelen x en y zet je alle combinaties (x, y) als een **puntenwolk** in een assenstelsel. Of er een **statistisch verband** bestaat tussen x en y wordt bepaald door de **correlatiecoëfficiënt** r_{xy}.

- $r_{xy} = 1$: perfecte positieve correlatie tussen x en y; de punten liggen op een stijgende lijn.
- $r_{xy} = 0$: geen enkele correlatie tussen x en y.
- $r_{xy} = -1$: perfecte negatieve correlatie tussen x en y; de punten liggen op een dalende lijn.

De correlatie wordt beter naarmate r_{xy} dichter bij 1 of -1 ligt.

Gebruik de **r-tabel**. In Excel wordt vaak de determinatiecoëfficiënt r^2_{xy} gegeven.

Een verband waarbij de toename (of afname) van de éne variabele een gevolg is van een toename (of afname) van de andere heet een **causaal verband**: er is dan sprake van oorzaak en gevolg. Een **statistisch verband** tussen twee variabelen hoeft niet causaal te zijn.
Bij voldoende correlatie kun je een formule van de vorm $y = ax + b$ opstellen, de **regressielijn** van y op x. Zo'n regressielijn gaat door het punt $(\overline{x}, \overline{y})$ en heeft als richtingscoëfficiënt (hellingsgetal): $a = r_{xy} \cdot \frac{s_y}{s_x}$

meer info