Formelsammlung zu Modul 101: Statistik 1

Marina Haller und Prof. Dr. Carolin Strobl

HS 2018 und FS 2019

Inhaltsverzeichnis

1 Univariate statistische Kennwerte

1.1 Maße der zentralen Tendenz (Lagemaße)

1.1.1 Mittelwert (arithmetisches Mittel)

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

1.1.2 Median

sortierte Daten: $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$

falls n ungerade: Median $= x_{((n+1)/2)}$

falls n gerade: Median $= \dfrac{x_{(n/2)} + x_{(n/2+1)}}{2}$

1.2 Maße der Variabilität (Streuungsmaße)

1.2.1 Stichprobenvarianz

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

1.2.2 Standardabweichung

$$s = \sqrt{s^2}$$

1.3 Lineare Transformation

$$y = a + b \cdot x$$

1.3.1 Mittelwert bei linearer Transformation

$$\bar{y} = a + b \cdot \bar{x}$$

1.3.2 Stichprobenvarianz und Standardabweichung bei linearer Transformation

$$s_y^2 = b^2 \cdot s_x^2, \quad s_y = |b| \cdot s_x$$

2 Wahrscheinlichkeitstheorie

2.1 Additionstheorem

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Spezialfall wenn A und B disjunkt:

$P(A \cup B) = P(A) + P(B)$

$P(A \cup B \cup C) = P(A) + P(B) + P(C)$ etc.

2.2 Multiplikationstheorem

$P(A \cap B) = P(A) \cdot P(B|A)$

$P(A \cap B \cap C) = P(A) \cdot P(B|A) \cdot P(C|A, B)$ etc.

Spezialfall für unabhängige Ereignisse:

$P(A \cap B) = P(A) \cdot P(B)$

$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$ etc.

2.3 Bedingte Wahrscheinlichkeit

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

2.4 Satz von der totalen Wahrscheinlichkeit

$P(B) = P(B|A) \cdot P(A) + P(B|\overline{A}) \cdot P(\overline{A})$

2.5 Satz von Bayes

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\overline{A}) \cdot P(\overline{A})}$$

$P(\overline{A}) = 1 - P(A)$

$P(B|\overline{A}) = 1 - P(\overline{B}|\overline{A})$

3 Wahrscheinlichkeitsverteilungen

3.1 Diskrete Zufallsvariablen

3.1.1 Wahrscheinlichkeitsfunktion $P(x_i)$

Erwartungswert: $\mu = \sum_{i=1}^{N} x_i \cdot P(x_i)$

Varianz: $\sigma^2 = \sum_{i=1}^{N} (x_i - \mu)^2 \cdot P(x_i)$

Verteilungsfunktion $F(x_i) = \sum_{j \leq i} P(x_j)$

3.1.1.1 Binomialverteilung $P(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}$

$\binom{n}{x} = \dfrac{n!}{x!(n-x)!}$ mit $x! = 1 \cdot 2 \cdots (x-2) \cdot (x-1) \cdot x$, wobei $0! = 1$

Erwartungswert: $\mu = n \cdot \pi$

Varianz: $\sigma^2 = n \cdot \pi(1-\pi)$

3.2 Stetige Zufallsvariablen

3.2.1 Dichte $f(x)$

Erwartungswert: $\mu = \int_{-\infty}^{\infty} x \, f(x) \, \mathrm{d}x$

Varianz: $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \, f(x) \, \mathrm{d}x$

Verteilungsfunktion: $F(x_p) = P(x \leq x_p) = \int_{-\infty}^{x_p} f(t) \, \mathrm{d}t$

3.3 z-Transformation

z-**Transformation (Daten)**: $z = \dfrac{x - \bar{x}}{s}$

z-**Transformation (Verteilung)**: $z = \dfrac{x - \mu}{\sigma}$

3.4 Stichprobenverteilung des Mittelwerts

$\mu_{\bar{x}} = \mu$

$\sigma_{\bar{x}} = \sigma / \sqrt{n}$ bzw. $\sigma_{\bar{x}}^2 = \sigma^2 / n$

$\sigma_{\bar{x}}$ wird als *Standardfehler* des Mittelwerts bezeichnet

wenn $x \sim N(\mu, \sigma^2) \Rightarrow \bar{x} \sim N(\mu, \sigma_{\bar{x}}^2)$

Plug-in Schätzer für $\sigma_{\bar{x}}$:

$s_{\bar{x}} = s / \sqrt{n} = \sqrt{s^2 / n}$

4 Hypothesentests und Konfidenzintervalle

4.1 Ein-Stichproben-Tests für den Mittelwert

Hypothesen:
$H_0 : \mu = \mu_0$
a) $H_1 : \mu > \mu_0$ (einseitiger Test)
b) $H_1 : \mu < \mu_0$ (einseitiger Test)
c) $H_1 : \mu \neq \mu_0$ (zweiseitiger Test)

4.1.1 Ein-Stichproben z-Test bei bekannter Varianz σ^2

Prüfgrösse: $z = \dfrac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \sqrt{n}\left(\dfrac{\bar{x} - \mu_0}{\sigma}\right)$
Ablehnbereich:
a) $z > z_{1-\alpha}$
b) $z < z_\alpha$
c) $z < z_{\alpha/2}$ oder $z > z_{1-\alpha/2}$ bzw. $|\,z\,| > z_{1-\alpha/2}$

4.1.1.1 Standardisierte Effektgrösse

$$\delta = \frac{\mu - \mu_0}{\sigma}$$

4.1.1.2 Bestimmung des Stichprobenumfangs

$$n = \left(\frac{z_\beta - z_{1-\alpha}}{\delta}\right)^2$$

wenn Teststärke (Power) $1 - \beta$ angegeben: $z_\beta = -z_{1-\beta}$
sonst: $z_\beta = z_{1-\alpha} - \sqrt{n} \cdot \delta$

4.1.2 Ein-Stichproben t-Test bei unbekannter Varianz

Prüfgrösse: $t = \dfrac{\bar{x} - \mu_0}{s_{\bar{x}}} = \sqrt{n}\left(\dfrac{\bar{x} - \mu_0}{s}\right)$
Ablehnbereich:
a) $t > t_{1-\alpha}(n-1)$
b) $t < t_\alpha(n-1)$
c) $t < t_{\alpha/2}(n-1)$ oder $t > t_{1-\alpha/2}(n-1)$ bzw. $|\,t\,| > t_{1-\alpha/2}(n-1)$

4.2 Konfidenzintervalle für den Mittelwert

4.2.1 Konfidenzintervall für $\bar{x}$ bei bekannter Varianz σ^2

$\bar{x} \pm z_{1-\alpha/2} \cdot \sigma_{\bar{x}}$ bzw. $\bar{x} \pm z_{1-\alpha/2} \cdot \sigma/\sqrt{n}$

4.2.2 Konfidenzintervall für $\bar{x}$ bei unbekannter Varianz

$\bar{x} \pm t_{1-\alpha/2}(n-1) \cdot s_{\bar{x}}$ bzw. $\bar{x} \pm t_{1-\alpha/2}(n-1) \cdot s/\sqrt{n}$

4.3 Zwei-Stichproben-Tests zum Vergleich von Mittelwerten

4.3.1 t-Test für unabhängige Stichproben

Hypothesen:
$H_0 : \mu_1 = \mu_2$
a) $H_1 : \mu_1 > \mu_2$ (einseitiger Test)
b) $H_1 : \mu_1 < \mu_2$ (einseitiger Test)
c) $H_1 : \mu_1 \neq \mu_2$ (zweiseitiger Test)

Prüfgrösse: $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$

$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{s_p^2 \left(\dfrac{1}{n_1} + \dfrac{1}{n_2} \right)}$ mit $s_p^2 = \dfrac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{(n_1 - 1) + (n_2 - 1)}$ bzw. falls $n_1 = n_2$: $s_p^2 = \dfrac{s_1^2 + s_2^2}{2}$

Ablehnbereich:
a) $t > t_{1-\alpha}(n_1 + n_2 - 2)$
b) $t < t_{\alpha}(n_1 + n_2 - 2)$
c) $t < t_{\alpha/2}(n_1 + n_2 - 2)$ oder $t > t_{1-\alpha/2}(n_1 + n_2 - 2)$ bzw. $|\, t \,| > t_{1-\alpha/2}(n_1 + n_2 - 2)$

wobei: n_1 = Anzahl Personen in Gruppe 1
n_2 = Anzahl Personen in Gruppe 2

4.3.2 t-Test für verbundene Stichproben

Hypothesen:
$H_0 : \mu_d = 0$
a) $H_1 : \mu_d > 0$ (einseitiger Test)
b) $H_1 : \mu_d < 0$ (einseitiger Test)
c) $H_1 : \mu_d \neq 0$ (zweiseitiger Test)

Prüfgrösse: $t = \sqrt{n} \left(\dfrac{\bar{d}}{s_d} \right)$

$$d_i = x_{i1} - x_{i2}, \quad \bar{d} = \frac{\sum_{i=1}^n d_i}{n}, \quad s_d = \sqrt{\frac{\sum_{i=1}^n \left(d_i - \bar{d}\right)^2}{n-1}}$$

Ablehnbereich:
a) $t > t_{1-\alpha}(n - 1)$
b) $t < t_{\alpha}(n - 1)$
c) $t < t_{\alpha/2}(n - 1)$ oder $t > t_{1-\alpha/2}(n - 1)$ bzw. $|\, t \,| > t_{1-\alpha/2}(n - 1)$

wobei: n = Anzahl Beobachtungspaare

4.4 χ^2-Unabhängigkeitstest

Hypothesen:
H_0: A und B unabhängig
H_1: A und B abhängig

4.4.1 Test für 2x2-Tabellen

Prüfgrösse: $\chi^2 = \dfrac{n \cdot (ad - bc)^2}{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)}$

Ablehnbereich: $\chi^2 > \chi^2_{1-\alpha}(1)$

4.4.2 Test für beliebig grosse Tabellen

Prüfgrösse: $\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \dfrac{(n_{ij} - m_{ij})^2}{m_{ij}}$

$m_{ij} = \dfrac{n_{i\cdot} \cdot n_{\cdot j}}{n}$

wobei:
$n_{i\cdot}$ = Summe der Häufigkeiten in der i-ten Zeile
$n_{\cdot j}$ = Summe der Häufigkeiten in der j-ten Spalte
n = Anzahl Beobachtungen insgesamt

Ablehnbereich: $\chi^2 > \chi^2_{1-\alpha}\left((k-1) \cdot (l-1)\right)$

wobei:
k = Anzahl Zeilen
l = Anzahl Spalten

Annahme: Alle *erwarteten* Häufigkeiten müssen > 5 sein.

5 Kovarianz und Korrelation

5.1 Stichprobenkovarianz

$$s_{xy} = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right) \cdot \left(y_i - \bar{y}\right)}{n-1}$$

5.2 Korrelationskoeffizient nach Bravais-Pearson

$$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right) \cdot \left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2} \cdot \sqrt{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}}$$

bzw. Berechnung über Summen: $r = \dfrac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \left(\sum_{i=1}^{n} x_i\right) \cdot \left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\left[n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right] \cdot \left[n \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}}$

5.3 Lineare Transformation

$u = a + b \cdot x$

5.3.1 Stichprobenkovarianz bei linearer Transformation

$s_{uy} = b \cdot s_{xy}$

5.3.2 Korrelationskoeffizient bei linearer Transformation

$r_{uy} = r_{xy}$

5.4 Test für den Korrelationskoeffizienten

Hypothesen:
$H_0 : \varrho = 0$
a) $H_1 : \varrho > 0$
b) $H_1 : \varrho < 0$
c) $H_1 : \varrho \neq 0$

Prüfgrösse: $t = \dfrac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}$

Ablehnbereich:
a) $t > t_{1-\alpha}(n-2)$
b) $t < t_{\alpha}(n-2)$
c) $t < t_{\alpha/2}(n-2)$ oder $t > t_{1-\alpha/2}(n-2)$ bzw. $|\,t\,| > t_{1-\alpha/2}(n-2)$

5.5 Rang-Korrelationskoeffizient nach Spearman

= Korrelationskoeffizient nach Bravais-Pearson angewendet auf die Ränge:

$$r_s = \frac{n \cdot \sum_{i=1}^{n} rg(x_i) \cdot rg(y_i) - \left(\sum_{i=1}^{n} rg(x_i) \right) \cdot \left(\sum_{i=1}^{n} rg(y_i) \right)}{\sqrt{\left[n \cdot \sum_{i=1}^{n} rg(x_i)^2 - \left(\sum_{i=1}^{n} rg(x_i) \right)^2 \right] \cdot \left[n \cdot \sum_{i=1}^{n} rg(y_i)^2 - \left(\sum_{i=1}^{n} rg(y_i) \right)^2 \right]}}$$

bzw. falls keine Bindungen vorliegen:

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^{n} \left(rg(x_i) - rg(y_i) \right)^2}{n \cdot (n^2 - 1)}$$

6 Lineare Einfachregression

6.1 Regressionsgleichung

Modell: $y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i$

Schätzung: $y_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_i + \hat{\varepsilon}_i$

6.2 Kleinste-Quadrate-Schätzer

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x}$

$$\hat{\beta}_1 = \frac{n \cdot \sum_{i=1}^n x_i \cdot y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}$$

Alternative Berechung:

$\hat{\beta}_1 = \frac{s_{xy}}{s_x^2}, \quad \hat{\beta}_1 = r_{xy} \cdot \frac{s_y}{s_x}$

6.3 Vorhersage

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_i$

6.4 Standardisierte Regressionskoeffizienten

$\hat{\tilde{\beta}}_0 = 0$

$\hat{\tilde{\beta}}_1 = r_{xy}$

6.5 Maße für die Güte des Regressionsmodells

6.5.1 Residuen

$\hat{\varepsilon}_i = y_i - \hat{y}_i$

6.5.2 Standardschätzfehler

$$s_{\hat{\varepsilon}} = \sqrt{\frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{n-2}}$$

6.5.3 Bestimmtheitsmaß (Determinationskoeffizient) R^2

$R^2 = \frac{QS_{\hat{y}}}{QS_y}, \quad R^2 = r_{\hat{y}\,y}^2 = r_{x\,y}^2$

mit **Quadratsummen**:

$QS_y = QS_{\hat{y}} + QS_{\hat{\varepsilon}}$

entsprechend der **Streuungszerlegung**:

$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$

6.6 Test und Konfidenzintervall für den Steigungsparameter

Hypothesen:
$H_0 : \beta = 0$
a) $H_1 : \beta > 0$
b) $H_1 : \beta < 0$
c) $H_1 : \beta \neq 0$

Prüfgrösse: $t = \dfrac{\hat{\beta}_1}{s_{\hat{\beta}_1}}$

$$s_{\hat{\beta}_1} = \frac{s_{\hat{\varepsilon}}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \quad \text{bzw.} \quad s_{\hat{\beta}_1} = \frac{s_{\hat{\varepsilon}}}{\sqrt{(n-1)\cdot s_x^2}} \quad \text{mit } s_{\hat{\varepsilon}} = \sqrt{\frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{n-2}}$$

Ablehnbereich:
a) $t > t_{1-\alpha}(n-2)$
b) $t < t_{\alpha}(n-2)$
c) $t < t_{\alpha/2}(n-2)$ oder $t > t_{1-\alpha/2}(n-2)$ bzw. $|\, t \,| > t_{1-\alpha/2}(n-2)$

Konfidenzintervall:
$\hat{\beta}_1 \pm t_{1-\alpha/2}(n-2) \cdot s_{\hat{\beta}_1}$

7 Partielle Korrelation

$r_{x_0 x_1 . x_2} = r_{\hat{\varepsilon}_0 \hat{\varepsilon}_1}$

Berechnung über Korrelationen:

$$r_{x_0 x_1 . x_2} = \frac{r_{x_0 x_1} - r_{x_0 x_2} \cdot r_{x_1 x_2}}{\sqrt{1 - r_{x_0 x_2}^2} \cdot \sqrt{1 - r_{x_1 x_2}^2}}$$

8 Multiple lineare Regression

8.1 Regressionsgleichung

Modell: $y_i = \beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \cdots + \beta_p \cdot x_{ip} + \varepsilon_i$

Schätzung: $y_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_{i1} + \hat{\beta}_2 \cdot x_{i2} + \cdots + \hat{\beta}_p \cdot x_{ip} + \hat{\varepsilon}_i$

8.2 Kleinste-Quadrate-Schätzer für zwei Prädiktoren

$$\hat{\beta}_1 = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{1 - r^2_{x_1x_2}} \cdot \frac{s_y}{s_{x_1}}$$

$$\hat{\beta}_2 = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{1 - r^2_{x_1x_2}} \cdot \frac{s_y}{s_{x_2}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x}_1 - \hat{\beta}_2 \cdot \bar{x}_2$$

8.3 Standardisierte Regressionskoeffizienten

$$\hat{\hat{\beta}}_0 = 0$$

$$\hat{\hat{\beta}}_j = \hat{\beta}_j \cdot \frac{s_{x_j}}{s_y}$$

8.4 Maße für die Güte des multiplen Regressionsmodells

8.4.1 Residuen

$$\hat{\varepsilon}_i = y_i - \hat{y}_i$$

8.4.2 Standardschätzfehler

$$s_{\hat{\varepsilon}} = \sqrt{\frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{n - p - 1}}$$

8.4.3 Bestimmtheitsmaß R^2

$$R^2 = \frac{QS_{\hat{y}}}{QS_y}, \quad R^2 = r^2_{\hat{y}y}$$

mit **Quadratsummen**:
$QS_y = QS_{\hat{y}} + QS_{\hat{\varepsilon}}$

entsprechend der **Streuungszerlegung**:
$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$

8.4.4 Korrigiertes Bestimmtheitsmaß R^2_{korr}

$$R^2_{korr} = 1 - (1 - R^2) \cdot \frac{n - 1}{n - p - 1}$$

8.5 Tests im multiplen Regressionsmodell

8.5.1 F-Test

Hypothesen:
H_0: $\beta_1 = \cdots = \beta_p = 0$
H_1: $\beta_j \neq 0$ für mindestens ein j

Prüfgrösse:
$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - p - 1}{p}$$

Ablehnbereich:
$F > F_{1-\alpha}(p, n - p - 1)$

8.5.2 t-Test für eine einzelne Steigung

Hypothesen:
$H_0 : \beta_j = 0$
a) $H_1 : \beta_j > 0$
b) $H_1 : \beta_j < 0$
c) $H_1 : \beta_j \neq 0$

Prüfgrösse:
$$t = \frac{\hat{\beta}_j}{s_{\hat{\beta}_j}} \quad \text{mit } s_{\hat{\beta}_j} = \text{Standardfehler von } \hat{\beta}_j$$

Ablehnbereich:
a) $t > t_{1-\alpha}(n - p - 1)$
b) $t < t_{\alpha}(n - p - 1)$
c) $t < t_{\alpha/2}(n - p - 1)$ oder $t > t_{1-\alpha/2}(n - p - 1)$ bzw. $|t| > t_{1-\alpha/2}(n - p - 1)$

9 Varianzanalyse

9.1 Einfaktorielle Varianzanalyse

9.1.1 Modell mit festen Effekten

$y_{im} = \mu_i + \varepsilon_{im}$ (Grundmodell)

$y_{im} = \mu + \alpha_i + \varepsilon_{im}$ (Modell in Effektdarstellung)

wobei $\varepsilon_{im} \sim N(0, \sigma_e^2)$ unabhängig, $\sum_{i=1}^{p} \alpha_i = 0$ und

Faktorstufen $i = 1, \ldots, p$
Personen innerhalb jeder Faktorstufe $m = 1, \ldots, n_i$

Hypothese:

$$\mathrm{H}_0\colon \mu_1 = \ldots = \mu_p \text{ bzw. } \mathrm{H}_0\colon \alpha_1 = \ldots = \alpha_p = 0$$

Streuungszerlegung:

$$\mathrm{QS}_{tot} = \mathrm{QS}_A + \mathrm{QS}_e$$

mit

$$\mathrm{QS}_A = n \cdot \sum_{i=1}^{p} (\overline{A}_i - \overline{G})^2 \quad \text{bzw.} \quad \sum_{i=1}^{p} n_i \cdot (\overline{A}_i - \overline{G})^2$$

$$\mathrm{QS}_e = \sum_{i=1}^{p} \sum_{m=1}^{n_i} (y_{im} - \overline{A}_i)^2$$

$$\mathrm{QS}_{tot} = \sum_{i=1}^{p} \sum_{m=1}^{n_i} (y_{im} - \overline{G})^2$$

wobei

$$\overline{A}_i = \frac{1}{n_i} \sum_{m=1}^{n_i} y_{im} \qquad \overline{G} = \frac{1}{N} \sum_{i=1}^{p} \sum_{m=1}^{n_i} y_{im} \qquad N = n \cdot p \text{ bzw. } \sum_{i=1}^{p} n_i$$

Varianztabelle:

Quelle	QS	df	MQ	F
A	QS_A	$p-1$	$\mathrm{MQ}_A = \dfrac{\mathrm{QS}_A}{p-1}$	$F = \dfrac{\mathrm{MQ}_A}{\mathrm{MQ}_e}$
Fehler	QS_e	$N-p$	$\mathrm{MQ}_e = \dfrac{\mathrm{QS}_e}{N-p}$	
Total	QS_{tot}	$N-1$		

Ablehnbereich: $F > F_{1-\alpha}(\mathrm{df}_A,\ \mathrm{df}_e) = F_{1-\alpha}(p-1,\ N-p)$

9.1.2 Modell mit zufälligen Effekten

$y_{im} = \mu + \alpha_i + \varepsilon_{im}$

wobei: $\alpha_i \sim N(0, \sigma_A^2)$, $\varepsilon_{im} \sim N(0, \sigma_e^2)$,
alle Zufallsvariablen sind voneinander unabhängig

Hypothese:

H$_0$: $\sigma_A^2 = 0$

Varianztabelle:

Quelle	QS	df	MQ	F
A	QS_A	$p-1$	$\mathrm{MQ}_A = \dfrac{\mathrm{QS}_A}{p-1}$	$F = \dfrac{\mathrm{MQ}_A}{\mathrm{MQ}_e}$
Fehler	QS_e	$N-p$	$\mathrm{MQ}_e = \dfrac{\mathrm{QS}_e}{N-p}$	
Total	QS_{tot}	$N-1$		

Ablehnbereich: $F > F_{1-\alpha}(\mathrm{df}_A,\ \mathrm{df}_e) = F_{1-\alpha}(p-1,\ N-p)$

Intraklassenkorrelation:

$$\mathrm{ICC} = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_A^2 + \hat{\sigma}_e^2} \quad \text{mit} \quad \hat{\sigma}_A^2 = \frac{\mathrm{MQ}_A - \mathrm{MQ}_e}{n} \quad \text{und} \quad \hat{\sigma}_e^2 = \mathrm{MQ}_e$$

9.2 Post-hoc Tests und multiples Testen

9.2.1 Lineare Kontraste

$$D = \sum_{i=1}^{p} c_i \cdot \overline{A}_i \quad \text{mit} \quad \sum_{i=1}^{p} c_i = 0$$

zwei Kontraste j und k sind orthogonal wenn: $\sum_{i=1}^{p} c_{ij} \cdot c_{ik} = 0$

Hypothese:

$$\mathrm{H}_0:\ \sum_{i=1}^{p} c_i \cdot \mu_i = 0$$

Prüfgröße:

$$F = QS_D / MQ_e \quad \text{mit } QS_D = \frac{n \cdot D^2}{\sum_{i=1}^{p} c_i^2}$$

Ablehnbereich: $F > F_{1-\alpha}(1,\ N-p)$

9.2.2 Kontrolle der familywise error rate

Korrektur nach Šidák:

$$\alpha = 1 - (1 - \alpha_{gesamt})^{1/m}$$

Approximation nach Bonferroni:

$$\alpha = \alpha_{gesamt}/m$$

mit m = Anzahl Tests

9.3 Zweifaktorielle Varianzanalyse

9.3.1 Modell mit festen Effekten für balanciertes Design

$y_{ijm} = \mu_{ij} + \varepsilon_{ijm}$ (Grundmodell)

$y_{ijm} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijm}$ (Modell in Effektdarstellung)

wobei $\varepsilon_{ijm} \sim N(0, \sigma_e^2)$ unabhängig,
$\sum_{i=1}^{p} \alpha_i = 0$, $\sum_{j=1}^{q} \beta_i = 0$ und $\sum_{i=1}^{p} \sum_{j=1}^{q} (\alpha\beta)_{ij} = 0$
Faktorstufen $i = 1, \dots, p$ und $j = 1, \dots, q$
Personen innerhalb jeder Faktorstufe $m = 1, \dots, n$

Hypothesen:

H$_0$: $\alpha_i = 0$ für alle i
H$_0$: $\beta_j = 0$ für alle j
H$_0$: $(\alpha\beta)_{ij} = 0$ für alle i, j

Streuungszerlegung:

$$\text{QS}_{tot} = \text{QS}_A + \text{QS}_B + \text{QS}_{AB} + \text{QS}_e$$

Varianztabelle:

Quelle	QS	df	MQ	F
A	QS_A	$p-1$	$\text{MQ}_A = \dfrac{\text{QS}_A}{p-1}$	$F_A = \dfrac{\text{MQ}_A}{\text{MQ}_e}$
B	QS_B	$q-1$	$\text{MQ}_B = \dfrac{\text{QS}_B}{q-1}$	$F_B = \dfrac{\text{MQ}_B}{\text{MQ}_e}$
AB	QS_{AB}	$(p-1)\cdot(q-1)$	$\text{MQ}_{AB} = \dfrac{\text{QS}_{AB}}{(p-1)\cdot(q-1)}$	$F_{AB} = \dfrac{\text{MQ}_{AB}}{\text{MQ}_e}$
Fehler	QS_e	$N - p\cdot q$	$\text{MQ}_e = \dfrac{\text{QS}_e}{N - p\cdot q}$	
Total	QS_{tot}	$N-1$		

Ablehnbereich: H_0 ablehnen, falls

$$F_A > F_{1-\alpha}(\mathrm{df}_A,\ \mathrm{df}_e) = F_{1-\alpha}(p-1,\ N-p\cdot q)$$
$$F_B > F_{1-\alpha}(\mathrm{df}_B,\ \mathrm{df}_e) = F_{1-\alpha}(q-1,\ N-p\cdot q)$$
$$F_{AB} > F_{1-\alpha}(\mathrm{df}_{AB},\ \mathrm{df}_e) = F_{1-\alpha}((p-1)\cdot(q-1),\ N-p\cdot q)$$

9.3.2 Quadratsummen für unbalanciertes Design

Effekt	Quadratsummen Typ	Vergleich von	gegen
	I	A	–
A	II	A + B	B
	III	A + B + AB	B + AB
	I	A + B	A
B	II	A + B	A
	III	A + B + AB	A + AB
	I	A + B + AB	A + B
AB	II	A + B + AB	A + B
	III	A + B + AB	A + B

9.3.3 Modell mit zufälligen Effekten

$y_{ijm} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijm}$
wobei $\varepsilon_{ijm} \sim N(0, \sigma_e^2)$, $\alpha_i \sim N(0, \sigma_A^2)$, $\beta_j \sim N(0, \sigma_B^2)$ und $(\alpha\beta)_{ij} \sim N(0, \sigma_{AB}^2)$,
alle Zufallsvariablen sind voneinander unabhängig

Hypothesen:

H_0: $\sigma_A^2 = 0$
H_0: $\sigma_B^2 = 0$
H_0: $\sigma_{AB}^2 = 0$

Prüfgrößen: $F_A = \dfrac{\mathrm{MQ}_A}{\mathrm{MQ}_{AB}},\ F_B = \dfrac{\mathrm{MQ}_B}{\mathrm{MQ}_{AB}},\ F_{AB} = \dfrac{\mathrm{MQ}_{AB}}{\mathrm{MQ}_e}$

9.3.4 Gemischtes Modell

$y_{ijm} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijm}$
wobei $\varepsilon_{ijm} \sim N(0, \sigma_e^2)$, $\sum_{i=1}^{p} \alpha_i = 0$, $\beta_j \sim N(0, \sigma_B^2)$ und $(\alpha\beta)_{ij} \sim N(0, \sigma_{AB}^2)$,
alle Zufallsvariablen sind voneinander unabhängig

Hypothesen:

H_0: $\alpha_i = 0$ für alle i
H_0: $\sigma_B^2 = 0$
H_0: $\sigma_{AB}^2 = 0$

Prüfgrößen: $F_A = \dfrac{\mathrm{MQ}_A}{\mathrm{MQ}_{AB}},\ F_B = \dfrac{\mathrm{MQ}_B}{\mathrm{MQ}_{AB}},\ F_{AB} = \dfrac{\mathrm{MQ}_{AB}}{\mathrm{MQ}_e}$

9.4 Varianzanalyse mit Messwiederholungen

$y_{im} = \mu + \alpha_i + \pi_m + \varepsilon_{im}$

wobei $\varepsilon_{im} \sim N(0, \sigma_e^2)$, $\pi_m \sim N(0, \sigma_P^2)$, $\sum_{i=1}^{p} \alpha_i = 0$,
alle Zufallsvariablen sind voneinander unabhängig und es gilt Sphärizität

Hypothese:

H_0: $\mu_1 = ... = \mu_p$ bzw. H_0: $\alpha_1 = ... = \alpha_p = 0$

Prüfgröße: $F_A = \dfrac{\mathrm{MQ}_A}{\mathrm{MQ}_e}$

10 Tabellen

10.1 Vereinfachte Normalverteilungstabelle

Tabelliert sind einige Quantile z_p und die entsprechenden Werte der Verteilungsfunktion $F(z_p)$ für $p \geq 0.5$. Für das Quantil z_p gilt $F(z_p) = P(z \leq z_p) = p$.

Ablesebeispiel: $z_{0.975} = 1.96$

Werte der Verteilungsfunktion für $z_p < 0$: $F(-z_p) = 1 - F(z_p)$

Quantile für $0 < p < 0.5$: $z_p = -z_{1-p}$

z_p	$F(z_p)$
0	0.50
0.68	0.75
1.28	0.90
1.65	0.95
1.96	0.975
2.33	0.99
2.58	0.995

10.2 χ^2-Verteilung

Tabelliert sind die Quantile $\chi^2_p(df)$ für df Freiheitsgrade und einige Werte der Verteilungsfunktion. Für das Quantil $\chi^2_p(df)$ gilt $F(\chi^2_p(df)) = p$.

Ablesebeispiel: $\chi^2_{0.95}(10) = 18.31$

df	0.01	0.025	0.05	0.1	0.5	0.9	0.95	0.975	0.99
1	0.00	0.00	0.00	0.02	0.45	2.71	3.84	5.02	6.63
2	0.02	0.05	0.10	0.21	1.39	4.61	5.99	7.38	9.21
3	0.11	0.22	0.35	0.58	2.37	6.25	7.81	9.35	11.35
4	0.30	0.48	0.71	1.06	3.36	7.78	9.49	11.14	13.28
5	0.55	0.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09
6	0.87	1.24	1.64	2.20	5.35	10.64	12.59	14.45	16.81
7	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48
8	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.54	20.09
9	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67
10	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21
11	3.05	3.82	4.57	5.58	10.34	17.27	19.68	21.92	24.73
12	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22
13	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69
14	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14
15	5.23	6.26	7.26	8.55	14.34	22.31	25.00	27.49	30.58
16	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.84	32.00
17	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41
18	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.80
19	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19
20	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57
21	8.90	10.28	11.59	13.24	20.34	29.61	32.67	35.48	38.93
22	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29
23	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64
24	10.86	12.40	13.85	15.66	23.34	33.20	36.41	39.36	42.98
25	11.52	13.12	14.61	16.47	24.34	34.38	37.65	40.65	44.31
26	12.20	13.84	15.38	17.29	25.34	35.56	38.88	41.92	45.64
27	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.20	46.96
28	13.56	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28
29	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59
30	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89

10.3 Students t-Verteilung

Tabelliert sind die Quantile $t_p(df)$ für df Freiheitsgrade und einige Werte der Verteilungsfunktion für $p > 0.5$. Für das Quantil $t_p(df)$ gilt $F(t_p(df)) = p$.

Ablesebeispiel: $t_{0.99}(20) = 2.528$

Quantile für $0 < p < 0.5$: $t_p(df) = -t_{1-p}(df)$

Approximation für $df > 30$: $t_p(df) \approx z_p$ (z_p ist das p-Quantil der Standardnormalverteilung)

df	0.6	0.8	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
1	0.325	1.376	3.078	6.314	12.706	31.820	63.657	318.309	636.619
2	0.289	1.061	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.277	0.978	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.271	0.941	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.267	0.920	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.265	0.906	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.263	0.896	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.262	0.889	1.397	1.859	2.306	2.897	3.355	4.501	5.041
9	0.261	0.883	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.260	0.879	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.260	0.876	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.259	0.873	1.356	1.782	2.179	2.681	3.054	3.930	4.318
13	0.259	0.870	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.258	0.868	1.345	1.761	2.145	2.624	2.977	3.787	4.141
15	0.258	0.866	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.258	0.865	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.257	0.863	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.257	0.862	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.257	0.861	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.257	0.860	1.325	1.725	2.086	2.528	2.845	3.552	3.849
21	0.257	0.859	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.256	0.858	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.256	0.858	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.256	0.857	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.256	0.856	1.316	1.708	2.059	2.485	2.787	3.450	3.725
26	0.256	0.856	1.315	1.706	2.055	2.479	2.779	3.435	3.707
27	0.256	0.855	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.256	0.855	1.312	1.701	2.048	2.467	2.763	3.408	3.674
29	0.256	0.854	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.256	0.854	1.310	1.697	2.042	2.457	2.750	3.385	3.646
∞	0.25	0.84	1.28	1.65	1.96	2.33	2.58	3.09	3.29

10.4 F-Verteilung

Tabelliert sind die Quantile $F_p(df_1, df_2)$ für df_1 und df_2 Freiheitsgrade und $p = 0.95$ bzw. $p = 0.99$. Für das Quantil $F_p(df_1, df_2)$ gilt $F(F_p(df_1, df_2)) = p$.

Ablesebeispiel: $F_{0.95}(4, 8) = 3.84$

	$F_{0.95}(df_1, df_2)$															
df_1 \ df_2	1	2	3	4	5	6	7	8	9	10	15	20	30	40	50	100
1	161	18.5	10.1	7.71	6.61	5.99	5.59	5.32	5.12	4.96	4.54	4.35	4.17	4.08	4.03	3.94
2	200	19.0	9.55	6.94	5.79	5.14	4.74	4.46	4.26	4.10	3.68	3.49	3.32	3.23	3.18	3.09
3	216	19.2	9.28	6.59	5.41	4.76	4.35	4.07	3.86	3.71	3.29	3.10	2.92	2.84	2.79	2.70
4	225	19.2	9.12	6.39	5.19	4.53	4.12	3.84	3.63	3.48	3.06	2.87	2.69	2.61	2.56	2.46
5	230	19.3	9.01	6.26	5.05	4.39	3.97	3.69	3.48	3.33	2.90	2.71	2.53	2.45	2.40	2.31
6	234	19.3	8.94	6.16	4.95	4.28	3.87	3.58	3.37	3.22	2.79	2.60	2.42	2.34	2.29	2.19
7	237	19.4	8.89	6.09	4.88	4.21	3.79	3.50	3.29	3.14	2.71	2.51	2.33	2.25	2.20	2.10
8	239	19.4	8.85	6.04	4.82	4.15	3.73	3.44	3.23	3.07	2.64	2.45	2.27	2.18	2.13	2.03
9	241	19.4	8.81	6.00	4.77	4.10	3.68	3.39	3.18	3.02	2.59	2.39	2.21	2.12	2.07	1.97
10	242	19.4	8.79	5.96	4.74	4.06	3.64	3.35	3.14	2.98	2.54	2.35	2.16	2.08	2.03	1.93
15	246	19.4	8.70	5.86	4.62	3.94	3.51	3.22	3.01	2.85	2.40	2.20	2.01	1.92	1.87	1.77
20	248	19.4	8.66	5.80	4.56	3.87	3.44	3.15	2.94	2.77	2.33	2.12	1.93	1.84	1.78	1.68
30	250	19.5	8.62	5.75	4.50	3.81	3.38	3.08	2.86	2.70	2.25	2.04	1.84	1.74	1.69	1.57
40	251	19.5	8.59	5.72	4.46	3.77	3.34	3.04	2.83	2.66	2.20	1.99	1.79	1.69	1.63	1.52
50	252	19.5	8.58	5.70	4.44	3.75	3.32	3.02	2.80	2.64	2.18	1.97	1.76	1.66	1.60	1.48
100	253	19.5	8.55	5.66	4.41	3.71	3.27	2.97	2.76	2.59	2.12	1.91	1.70	1.59	1.52	1.39
	$F_{0.99}(df_1, df_2)$															
1	4052	98.5	34.1	21.2	16.3	13.7	12.2	11.3	10.6	10.0	8.68	8.10	7.56	7.31	7.17	6.90
2	4999	99.0	30.8	18.0	13.3	10.9	9.55	8.65	8.02	7.56	6.36	5.85	5.39	5.18	5.06	4.82
3	5403	99.2	29.5	16.7	12.1	9.78	8.45	7.59	6.99	6.55	5.42	4.94	4.51	4.31	4.20	3.98
4	5625	99.3	28.7	16.0	11.4	9.15	7.85	7.01	6.42	5.99	4.89	4.43	4.02	3.83	3.72	3.51
5	5764	99.3	28.2	15.5	11.0	8.75	7.46	6.63	6.06	5.64	4.56	4.10	3.70	3.51	3.41	3.21
6	5859	99.3	27.9	15.2	10.7	8.47	7.19	6.37	5.80	5.39	4.32	3.87	3.47	3.29	3.19	2.99
7	5928	99.4	27.7	15.0	10.5	8.26	6.99	6.18	5.61	5.20	4.14	3.70	3.30	3.12	3.02	2.82
8	5981	99.4	27.5	14.8	10.3	8.10	6.84	6.03	5.47	5.06	4.00	3.56	3.17	2.99	2.89	2.69
9	6022	99.4	27.3	14.7	10.2	7.98	6.72	5.91	5.35	4.94	3.89	3.46	3.07	2.89	2.78	2.59
10	6056	99.4	27.2	14.5	10.1	7.87	6.62	5.81	5.26	4.85	3.80	3.37	2.98	2.80	2.70	2.50
15	6157	99.4	26.9	14.2	9.72	7.56	6.31	5.52	4.96	4.56	3.52	3.09	2.70	2.52	2.42	2.22
20	6209	99.5	26.7	14.0	9.55	7.40	6.16	5.36	4.81	4.41	3.37	2.94	2.55	2.37	2.27	2.07
30	6261	99.5	26.5	13.8	9.38	7.23	5.99	5.20	4.65	4.25	3.21	2.78	2.39	2.20	2.10	1.89
40	6287	99.5	26.4	13.7	9.29	7.14	5.91	5.12	4.57	4.17	3.13	2.69	2.30	2.11	2.01	1.80
50	6303	99.5	26.4	13.7	9.24	7.09	5.86	5.07	4.52	4.12	3.08	2.64	2.25	2.06	1.95	1.74
100	6334	99.5	26.2	13.6	9.13	6.99	5.75	4.96	4.41	4.01	2.98	2.54	2.13	1.94	1.82	1.60
df_1 / df_2	1	2	3	4	5	6	7	8	9	10	15	20	30	40	50	100