t-Test

= Test der Differenz der Mittelwerte zweier Gruppen auf Signifikanz:

H_0: $\mu_1 = \mu_2$

- Standardfehler der Differenz: $\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
- Testgröße: $t_{n_1 + n_2 - 2; 1-\alpha/2} = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)}}$
- Zum Vergleich - Konfidenzintervall: $x_{KI} = \bar{x} \pm \frac{\sigma}{\sqrt{n}} \cdot t_{n-1; 1-\alpha/2}$
- Voraussetzungen:
  - Normalverteilung (bei kleinen Stichproben)
  - Homogenität der Varianzen
  - Unabhängigkeit der Stichproben

Zweifaktorielle ANOVA

- $f=2$ Faktoren (A, B), p Stufen, n Wiederholungen => $N = 2 \cdot p \cdot n$
- gleichzeitige Überprüfung von **3 Hypothesen**:
  1. Es gibt keinen Effekt des Faktors A (**H_0**: $\mu_{1A} = \mu_{2A} = \mu_{3A} = ...$)
  2. Es gibt keinen Effekt des Faktors B (**H_0**: $\mu_{1B} = \mu_{2B} = \mu_{3B} = ...$)
  3. Es gibt keine Wechselwirkungen zwischen den Faktoren A und B (**H_0**: $\mu_{AB} = \mu_A + \mu_B - \mu$)

Zweifaktorielle ANOVA: Quadratsummenzerlegung

- ***2*** Faktoren (UV) ***A***, ***B***, mit jeweils ***p*** Stufen und ***n*** Wiederholungen => $N = p^2 \cdot n$
- Quadratsummen:

1-faktoriell	2-faktoriell
QS_{tot}	$QS_{tot} = QS_{Zellen} + QS_{Fehler} = QS_A + QS_B + QS_{AxB} + QS_{Fehler}$
QS_{Fehler}	QS_{Fehler}
QS_{treat}	$QS_{Zellen} = QS_A + QS_B + QS_{AxB}$
	QS_A, QS_B
	QS_{AxB}

Beispiel

		Faktor A = Medikament			
		(1) Plazebo	(2) einfache Dosis	(3) doppelte Dosis	
Faktor B = Geschlecht	(1) männlich	22 25 22 **μ_{11} = 22,4** 21 22	16 16 16 **μ_{21} = 15,6** 15 15	13 12 12 **μ_{31} = 12,4** 13 12	**μ_{B1} = 16,8**
	(2) weiblich	18 19 17 **μ_{12} = 18,8** 21 19	19 20 17 **μ_{22} = 17,6** 16 16	16 14 16 **μ_{32} = 14,6** 13 14	**μ_{B2} = 17,0**
		μ_{A1} = 20,6	**μ_{A2} = 16,6**	**μ_{A3} = 13,5**	**μ = 16,9**

Quadratsummenzerlegung

2 Faktoren (UV) ***A***, ***B***, mit jeweils ***p*** Stufen und ***n*** Wiederholungen => $N = p^2 \cdot n$

- QS_{tot}: $$QS_{tot} = \sum_{i=1}^{n \cdot p} (x_i - \bar{x})^2 \qquad QS_{tot} = QS_{Zellen} + QS_{Fehler}$$
- **Zelle**: alle Wiederholungen der gleichen Behandlung (A, B) und der gleichen Behandlungsstufe
  QS_{Zellen}: $$QS_{Zellen} = n \cdot \sum_{k=1}^{2} \sum_{I=1}^{p} (\bar{x}_{ki} - \bar{x})^2$$
- QS_{Fehler}: $$QS_{Fehler} = \sum_{k=1}^{2} \sum_{i=1}^{p} \sum_{m=1}^{n} (x_{kim} - \bar{x}_{ki})^2$$
- QS_A, QS_B: $$QS_A = p \cdot n \cdot \sum_{I=1}^{p} (\bar{x}_{Ai} - \bar{x})^2 \qquad QS_B = p \cdot n \cdot \sum_{I=1}^{p} (\bar{x}_{Bi} - \bar{x})^2$$
- QS_{AxB}: $$QS_{Zellen} = QS_A + QS_B + QS_{AxB}$$

Zweifaktorielle ANOVA: Freiheitsgrade

- *2* Faktoren (UV) ***A***, ***B***, mit jeweils ***p*** Stufen und ***n*** Wiederholungen => $N = p^2 \cdot n$

1-faktoriell	2-faktoriell
	$df_A = p - 1$
	$df_B = p - 1$
	$df_{AxB} = p \cdot p - (2p - 1) = (p - 1) \cdot (p - 1)$
$df_{treat} = p - 1$	$df_{Zellen} = p \cdot p - 1 = df_A + df_B + df_{AxB}$
$df_{Fehler} = p \cdot (n - 1)$	$df_{Fehler} = p \cdot p \cdot (n - 1)$
$df_{tot} = (p \cdot n) - 1$	$df_{tot} = (p \cdot p \cdot n) - 1$

Zweifaktorielle ANOVA: Freiheitsgrade

- *2* Faktoren (UV) ***A*** mit ***p*** Stufen, ***B*** mit ***q*** Stufen und jeweils ***n*** Wiederholungen
  => $N = p \cdot q \cdot n$

<u>1-faktoriell</u>	<u>2-faktoriell</u>
	$df_A = p - 1$
	$df_B = q - 1$
	$df_{AxB} = p \cdot q - (p + q - 1) = (p - 1) \cdot (q - 1)$
$df_{treat} = p - 1$	$df_{Zellen} = p \cdot q - 1 = df_A + df_B + df_{AxB}$
$df_{Fehler} = p \cdot (n - 1)$	$df_{Fehler} = p \cdot q \cdot (n - 1)$
$df_{tot} = (p \cdot n) - 1$	$df_{tot} = (p \cdot q \cdot n) - 1$

Beispiel

		Faktor A = Medikament			
		(1) Plazebo	(2) einfache Dosis	(3) doppelte Dosis	
Faktor B = Geschlecht	(1) männlich	22 25 22 **μ_{11} = 22,4** 21 22	16 16 16 **μ_{21} = 15,6** 15 15	13 12 12 **μ_{31} = 12,4** 13 12	**μ_{B1} = 16,8**
	(2) weiblich	18 19 17 **μ_{12} = 18,8** 21 19	19 20 17 **μ_{22} = 17,6** 16 16	16 14 16 **μ_{32} = 14,6** 13 14	**μ_{B2} = 17,0**
		μ_{A1} = 20,6	**μ_{A2} = 16,6**	**μ_{A3} = 13,5**	**μ = 16,9**

$$\hat{\sigma}^2_{Fehler} = \frac{QS_{Fehler}}{df_{Fehler}} = 1{,}70 \qquad F_A = \frac{QS_A}{df_A} \Big/ \hat{\sigma}^2_{Fehler} = 74{,}53^{**}$$

$$F_{AxB} = \frac{QS_{AxB}}{df_{AxB}} \Big/ \hat{\sigma}^2_{Fehler} = 15{,}94^{**} \qquad F_B = \frac{QS_B}{df_B} \Big/ \hat{\sigma}^2_{Fehler} = 0{,}18$$

Post-hoc-Test

- **ANOVA**: Gibt es generell signifikante Effekte der einzelnen Faktoren?
- **t-Test**: Gibt es signifikante Unterschiede zwischen ***zwei*** Gruppen?
- **Post-hoc-Test** der ANOVA: Welche Zellen (Stufen der Faktoren) unterscheiden sich signifikant?

  → im Gegensatz zum t-Test Berücksichtigung der Tatsache, dass >2 Stichproben vorliegen

  Beispiel:

  - *20-maliges Ziehen von jeweils 10 Zufallszahlen*
  - *Berechnung der Mittelwerte dieser 20 Stichproben*
  - *Vergleich der Gruppe mit dem höchsten Mittelwert mit der Gruppe mit dem niedrigsten Mittelwert ergibt gemäß t-Test mit hoher Wahrscheinlichkeit einen signifikanten Unterschied*

Aufgabe

1. Berechnen Sie „zu Fuß“ QS_{tot}, QS_{Fehler}, QS_A, QS_B, QS_{AxB}, die entsprechenden Freiheitsgrade und die F-Werte für einen der Parameter (Cl, DOC, pH oder NO_3) des Datensatzes.
2. Führen Sie in Statistica eine mehrfaktorielle ANOVA durch und vergleichen Sie die Ergebnisse.
3. Stellen Sie zum Vergleich die Mittelwerte und Konfidenzintervalle der Variablen grafisch dar. Kennzeichnen Sie signifikante Unterschiede mittels unterschiedlicher Buchstaben.