



ANalysis Of VAriance (ANOVA) 1/2

Sicht Ihrer ProfessorInnen



- Planung der Vorlesung: 15 ProfessorInnen aus D-BIOL, D-HEST befragt
- Auswahl an Themen; welche sind *wichtig, nice to have, unwichtig*
- ALLE 15 ProfessorInnen haben ANOVA als *wichtig* angegeben
- ANOVA ist gemäss Ihren Profs das Thema mit der höchsten Priorität

Mögliches Wahlfach

401-0625-01L Applied Analysis of Variance and Experimental Design


Semester Herbstsemester 2017

Dozierende [L. Meier](#)

Periodizität jährlich wiederkehrende Veranstaltung

Lehrsprache Englisch

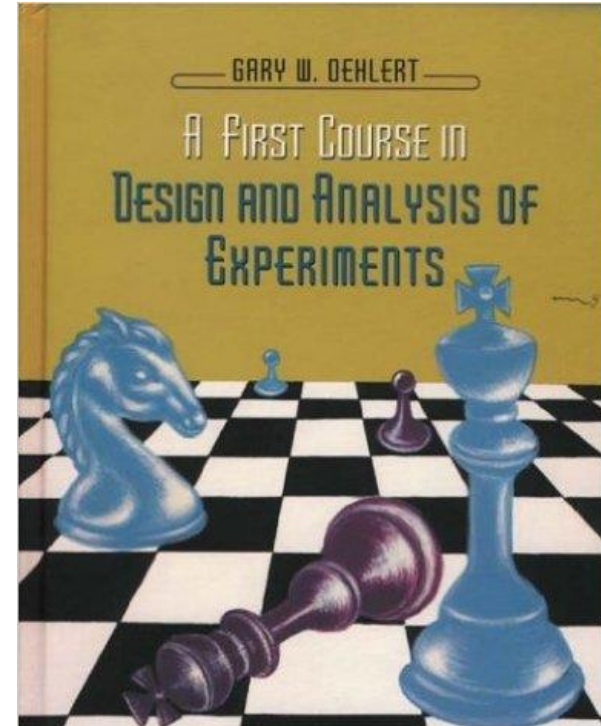
[Katalogdaten](#)[Leistungskontrolle](#)[Lernmaterialien](#)[Lehrveranstaltungen](#)[Einschränkungen](#)[Angeboten in](#)[► Übersicht](#)

Studiengang	Bereich	Typ
Biologie Master	Wahlpflicht Masterkurse	W 
⋮		



Weiterführende Literatur

- Buch aus Wahlfachvorlesung; deckt alle Grundlagen ab, geht aber noch viel weiter
- Interessante Kapitel:
Ch 3: Gründliche Einführung in ANOVA
Ch 4: Kontraste
Ch 5: Paarweise Vergleiche u.v.m.



- Buch **gratis als pdf**: <http://users.stat.umn.edu/~gary/Book.html>
- Skript aus der Wahlfachvorlesung

ANOVA - Idee

- ANOVA 1: Zwei Medikamente zur Blutdrucksenkung und Placebo (Faktor). Gibt es einen sign. Unterschied in der Wirkung (kontinuierlich)?

$$Y \sim X + \varepsilon$$

1-weg ANOVA

- ANOVA 2: Zwei Medikamente zur Blutdrucksenkung, Placebo (Faktor) und Geschlecht (Faktor). Gibt es einen sign. Unterschied in der Wirkung (kontinuierlich) (evtl. geschlechterspezifisch)?

$$Y \sim X1 + X2 + \varepsilon$$

2-weg ANOVA

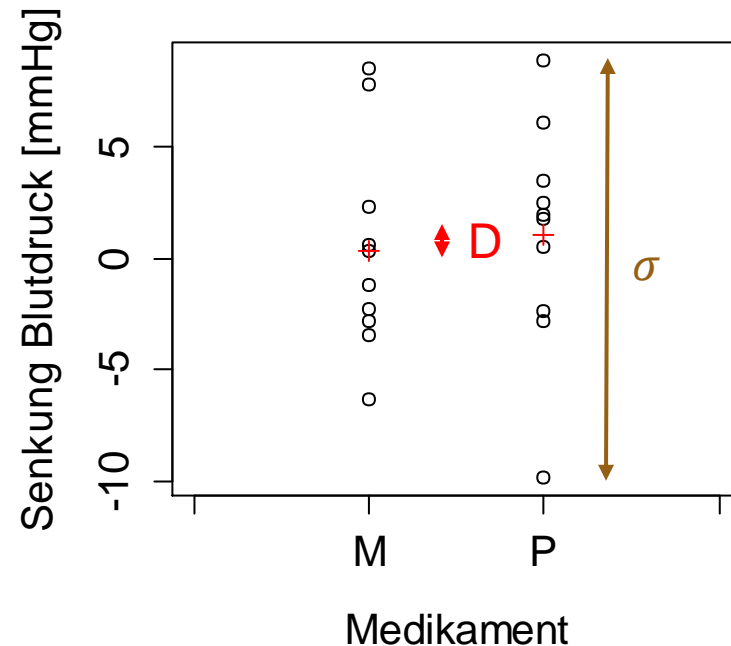
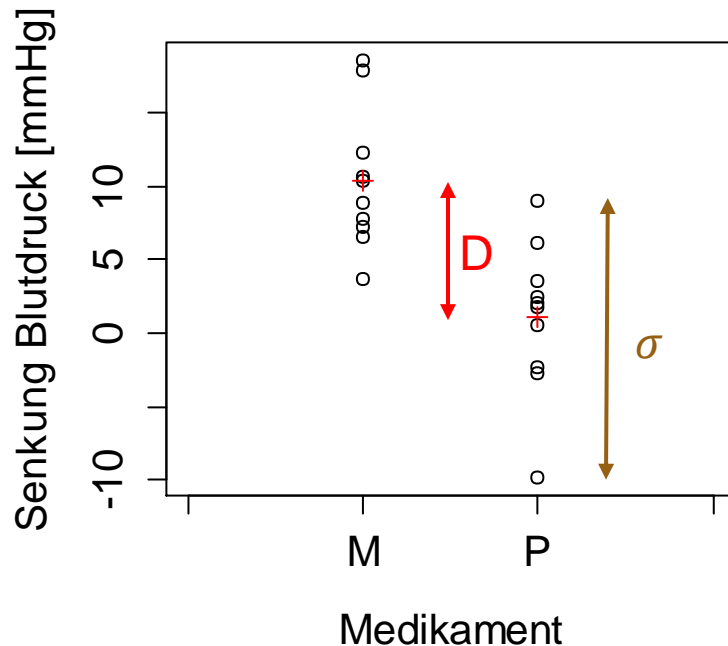
ANOVA: Mögliche Missverständnisse

- ANOVA = «Varianzanalyse»
Macht Aussagen über **Mittelwerte** (analysiert dazu Varianzen)
- ANOVA = Spezialfall einer Linearen Regression
kont. Variable ~ Faktoren + Fehler
- Verallgemeinerung des t-Test (2 Gruppen → viele Gruppen)
- Historisch: Sehr verbreitet; heute: Immer noch extrem verbreitet

Wdh: Ungepaarter t-Test

D: “Streuung” zwischen MW (“Signal”)
 σ : “Streuung” um MW (“Fehler”)

$$t \approx \frac{D}{\sigma} ; \text{falls } H_0 \text{ stimmt: } t \sim t_{n-1} \approx N(0,1)$$

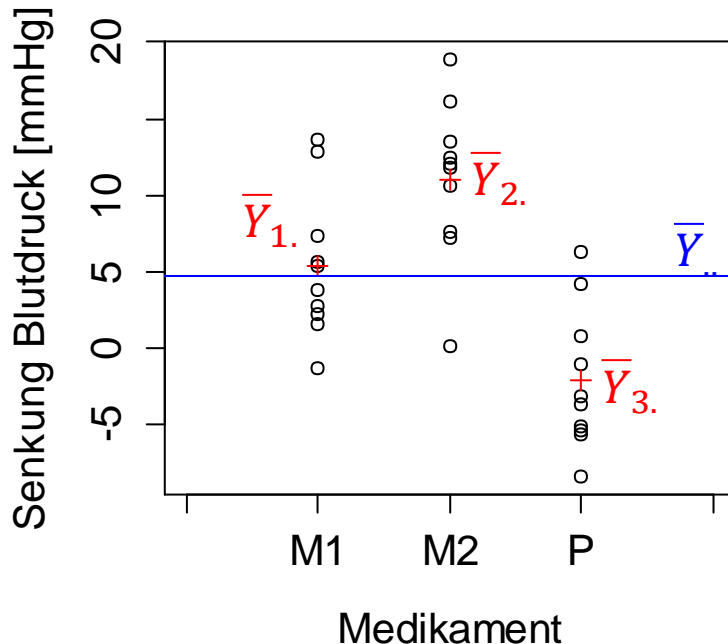


ANOVA: Idee

g : Anzahl Gruppen (3)

p : Anzahl Beob. pro Gruppe (10)

Ann: p in jeder Gruppe gleich



Streuung zwischen Gruppen:

“Between-Sum-of-Squares” (SS_B)

RSS der Gruppenmittelwerte (rote Kreuze) um den totalen Mittelwert (blaue Linie)

$$SS_B = p * \sum_{i=1}^g (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

Streuung innerhalb Gruppen:

“Within-Sum-of-Squares” (SS_W)

RSS der Einzelbeobachtungen (schwarze Kreise) um die einzelnen Mittelwerte (rote Kreuze)

$$SS_W = \sum_{i=1}^g \sum_{j=1}^p (Y_{ij} - \bar{Y}_{i.})^2$$

$$\text{Teststatistik} \approx \frac{SS_B}{SS_W}$$

ANOVA: Modell

- $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2) \text{ iid}$

Technische Nebenbedingung: $\sum_{i=1}^g \alpha_i = 0$

“Analyse der Varianzen”

- $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_g = 0$

- Teststatistik: $T = \frac{SS_B/(g-1)}{SS_W/(g*(p-1))} = \frac{MS_B}{MS_W}$

“Mean Squares”

- Theorie: Falls H_0 stimmt

“Degrees of freedom (Df)”

$$T \sim F_{g-1, g*(p-1)}$$

- Damit kann ein Hypothesentest mit den üblichen 6 Schritten durchgeführt werden

Exkurs: Verteilungen

- Angenommen: $Z_i \sim N(0,1)$, $i = 1, \dots, n$ alle unabhängig

$$A = \sum_{i=1}^n Z_i^2$$

Chi-Quadrat-Verteilung mit n Freiheitsgraden: $A \sim X_n$

- Angenommen: $A \sim X_n, B \sim X_m$ unabhängig

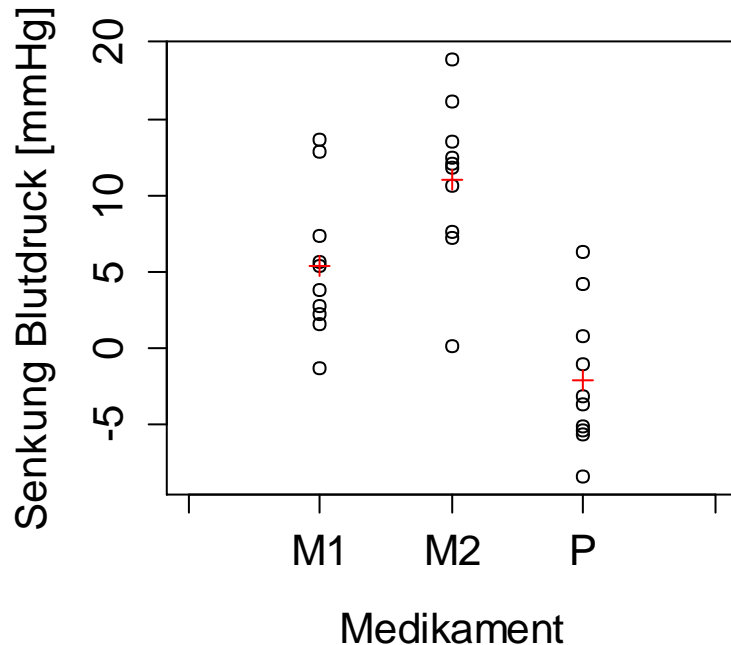
$$Q = \frac{A/n}{B/m}$$

- **F-Verteilung** mit n und m Freiheitsgraden $Q \sim F_{n;m}$



Beispiel in R: ANOVA-Tabelle

$g = 3, p = 10$



$$SS_B = 872.3$$

$$SS_W = 642.1$$

$$F = \frac{436.1}{23.8} = 18.34$$

```
> fm <- aov(y ~ g, data = df)
> summary(fm)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
g	2	872.3	436.1	18.34	9.32e-06 ***
Residuals	27	642.1	23.8		

$$g - 1 = 2$$

$$g^*(p-1) = 27$$

$$MS_B = \frac{872.3}{2} = 436.1$$

$$MS_W = \frac{642.1}{27} = 23.8$$

Wo ist der Unterschied ?

Teil 1: Paarweise Tests

- Falls ANOVA signifikant: Zwischen welchen Gruppen sind signifikante Unterschiede ?
→ t-Tests für alle Gruppenpaare
- Problem: Multiples Testen
- Lösung: t-Test korrigieren (z.B. Bonferroni, ...)
→ “Bonferroni significant difference (BSD)”

Wdh: Multiples Testen

- Microarray Test mit $m=1000$ Genen
- 1000 t-Tests auf dem 5% Sign.niveau
- Angenommen, kein Gen hat einen Effekt:

Ca. 50 Tests (5% von 1000) werden trotzdem ein signifikantes Ergebnis liefern !

50 Gene werden als “wichtig” angegeben, obwohl sie gar nicht “wichtig” sind.

- Wie erzeugt man eine Liste von “wirklich wichtigen” Genen?

Multiples Testen: Bonferroni Korrektur

- Wollen eine “konservative” Liste mit der Eigenschaft:
 $P(\text{mind. ein Fehler 1. Art}) = \text{Family Wise Error Rate} = \text{FWER}$
 $\text{FWER} \leq \alpha$
- Bonferroni Korrektur: Teste jedes Gen mit Sign.niveau $\frac{\alpha}{m}$, statt α ; m ist die Anzahl Gene
- Nachteil: Liste ist evtl. “extrem” konservativ (z.B. gar kein Gen enthalten)
- Bonferroni significant difference (BSD): t-Test & Bonferroni-Korrektur

Beliebte Alternative bei ANOVA: Tukey's Honestly Significant Difference (HSD) Test

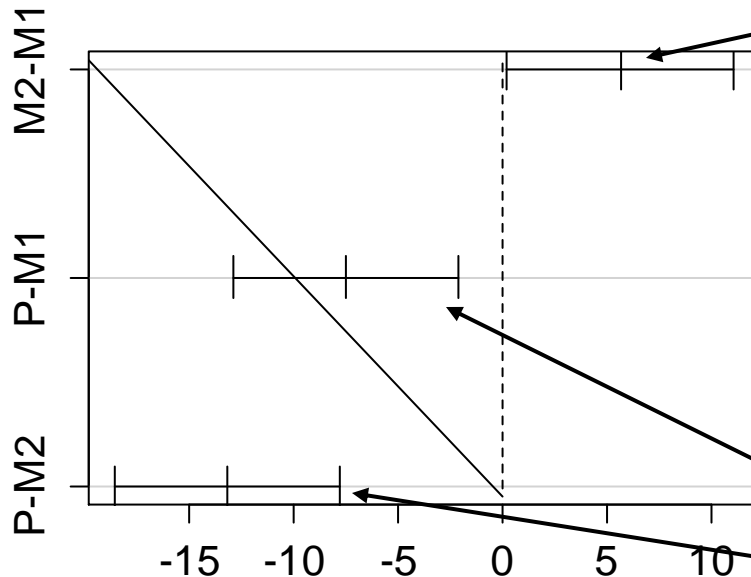
- Vorteil:
 - Vertrauensintervalle für Differenzen der Gruppenmittelwerte
 - Wa., dass **alle** wahren Differenzen in den Vertrauensintervallen liegen: 95%
- Alternative zu BSD;
Tukey HSD liefert kürzere Vertrauensintervalle als BSD, wenn wir alle Gruppenmittelwerte paarweise vergleichen
- Empfehlung: Tukey HSD verwenden



Beispiel in R: TukeyHSD

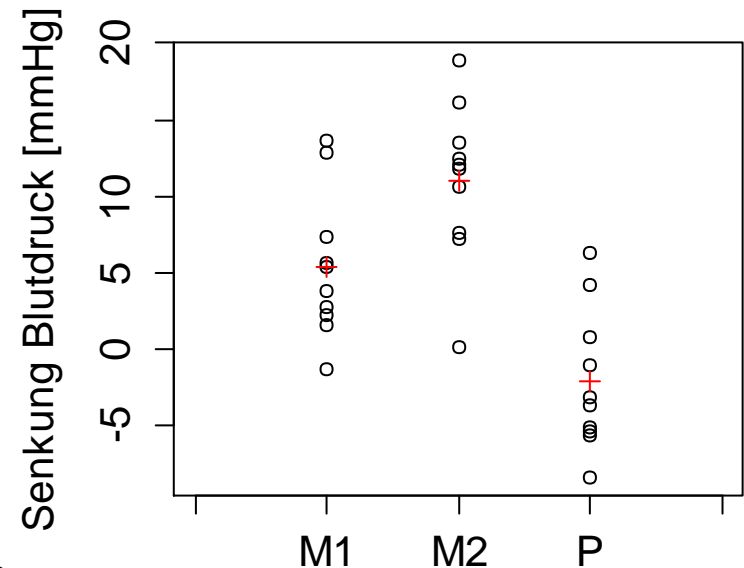
```
TukeyHSD(fm)
plot(TukeyHSD(fm))
```

95% family-wise confidence level



Differences in mean levels of g

M2 ist deutlich wirksamer als M1



M1 und M2 sind deutlich wirksamer als Placebo

Medikament

Wo ist der Unterschied ?

Teil 2: Allgemeine Kontraste

- Bisher: Differenz von zwei Gruppen
- Jetzt: Linearkombination von beliebigen Gruppen
- Bsp: Sind die beiden Medikamente im Mittel besser als das Placebo ?

Kontraste: Notation

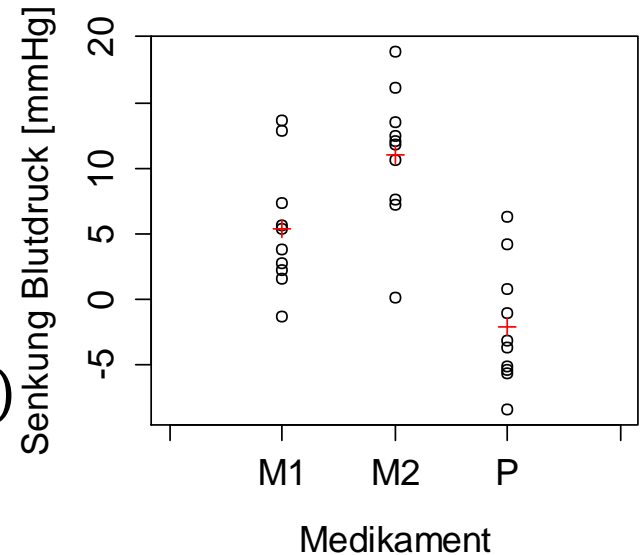
- Vektor mit wahren Gruppenmittelwerten:

$$\mu = (\mu_{M1}, \mu_{M2}, \mu_P)$$

- Kontraste-Matrix K
- Parameter-Vektor m

- $H_0: K * \mu = m$

- Praxis: Benutzer definiert K und m ; Computer berechnet p-Werte für Hypothesen und korrigiert für mult. Testen



Kontraste – Bsp 1: Paarweise Vergleiche

(Alternative zu TukeyHSD)

$$\begin{array}{c} \text{K} \\ \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \end{array} * \begin{array}{c} \mu \\ \begin{pmatrix} \mu_{M1} \\ \mu_{M2} \\ \mu_P \end{pmatrix} \end{array} = \begin{array}{c} \text{m} \\ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \end{array} \quad \longleftrightarrow \quad \begin{array}{l} \mu_{M2} - \mu_{M1} = 0 \\ \mu_P - \mu_{M1} = 0 \\ \mu_P - \mu_{M2} = 0 \end{array}$$



Kontraste – Bsp 1: R

- Funktion 'glht' (General Linear Hypotheses Test) im package 'multcomp'

```
Fit: aov(formula = y ~ g, data = df)
```

```
Linear Hypotheses:
```

	Estimate	Std. Error	t value	Pr(> t)	
M2-M1 == 0	5.670	2.181	2.600	0.03853	*
P-M1 == 0	-7.496	2.181	-3.437	0.00533	**
P-M2 == 0	-13.166	2.181	-6.037	< 0.001	***

Approx. 95%-VI für Unterschied M1 vs. M2:
 $5.67 \pm 2 * 2.181$

```
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Adjusted p values reported -- single-step method)
```

Kontraste – Bsp 2:

Gruppe der Medikamente vs. Placebo

$$\begin{pmatrix} 0.5 & 0.5 & -1 \\ -1 & 1 & 0 \end{pmatrix} * \begin{pmatrix} \mu_{M1} \\ \mu_{M2} \\ \mu_P \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$



Medikamente vs. Placebo

$$0.5 * \mu_{M1} + 0.5 * \mu_{M2} - \mu_P = 0$$
$$\mu_{M2} - \mu_{M1} = 0$$

Medikamente untereinander



Kontraste – Bsp 2: R

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: User-defined Contrasts

Fit: `aov(formula = y ~ g, data = df)`

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t)
M-P == 0	10.331	1.889	5.47	1.73e-05 ***
M2-M1 == 0	5.670	2.181	2.60	0.0294 **

Die Medikamente sind deutlich wirksamer als Placebo

M2 ist deutlich wirksamer als M1

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

Grundregeln für Kontraste

- Wenige Kontraste → viel Macht
- Software: Korrektur für multiples Testen innerhalb von einem Funktionsaufruf (aber nicht bei mehreren Funktionsaufrufen mit verschiedenen Kontrasten)
- Deshalb: **Einen** Satz von Kontrasten definieren, dann auswerten; anschliessend keinen neuen Satz von Kontrasten mehr untersuchen

Residuenanalyse bei ANOVA

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2) \text{ iid}$$

1. Daten in jeder Gruppe normalverteilt
2. Gleiche Varianz in Gruppen
3. Unabhängige Fehler ε_{ij}

In R: Funktion “plot” wie bei Linearer Regression

Vorteil: “Balanciertes Experiment” (gleiche Anzahl pro Gruppe):
ANOVA ist robuster gegen Abweichungen obiger Annahmen

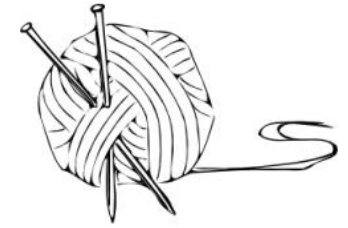
Ausblick: Alternativen um mit Multiplem Testen umzugehen (FWER vs. FDR)

Es gibt verschiedene Kriterien, mit denen man dem Problem vom multiplen Testen begegnen kann.

	H_0 wahr	H_A wahr	Total
H_0 verworfen	V	S	R
H_0 nicht verworfen	U	T	$m - R$
Total	m_0	$m - m_0$	m

- **Familywise Error Rate (FWER):** $P(V \geq 1)$ wird kontrolliert
Methoden stellen sicher, dass FWER nicht grösser als α (z.B. 0.05) ist;
wird verwendet bei: Bonferroni Korrektur, Tukey HSD
- **False Discovery Rate (FDR):** $E\left(\frac{V}{R}\right)$ wird kontrolliert
Methoden stellen sicher, dass FDR nicht grösser als α (z.B. 0.05) ist;
wird verwendet bei: Benjamini-Hochberg Korrektur

Exkurs: Reproduzierbare Auswertung mit knitr



- R-Paket “knitr”: R code und Text in einem einzigen File zusammenfügen
- Keine Copy-Paste Fehler; Sicherheit, dass Ergebnisse im pdf reproduziert werden können
- File > New File > R Markdown...
Help > Cheat Sheets > R Markdown Cheat Sheet
- Nicht prüfungsrelevant, aber praktisch 😊 !