ETH zürich

ANalysis Of VAriance (ANOVA) 1/2

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

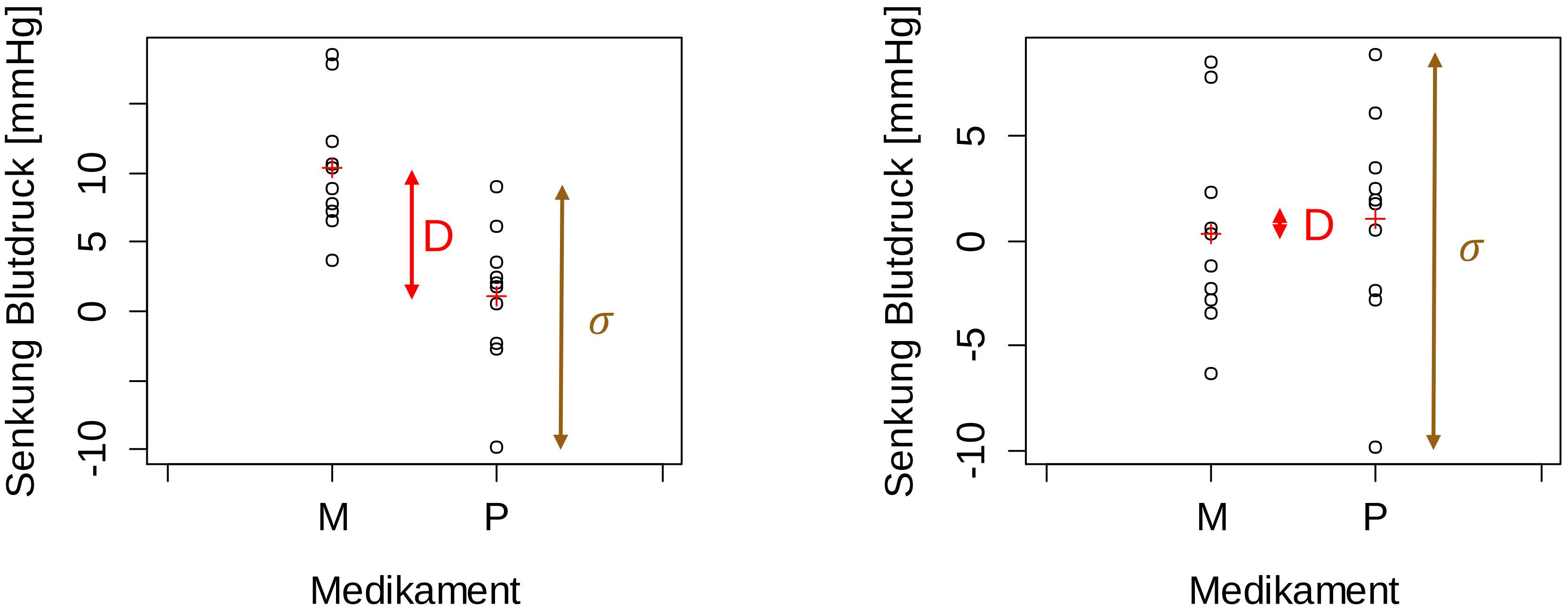
$$kont.Variable \sim \boldsymbol{Faktoren} + Fehler$$

- Verallgemeinerung des t-Test (2 Gruppen → viele Gruppen)
- Historisch: Sehr verbreitet; heute: Immer noch extrem verbreitet

Wdh: Ungepaarter t-Test

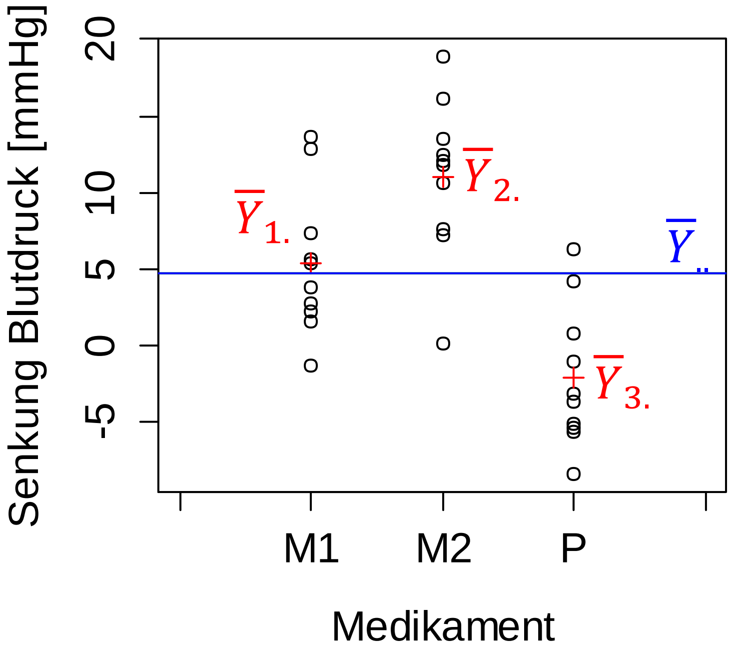
D: “Streuung” zwischen MW (“Signal”)
σ: “Streuung” um MW (“Fehler”)

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



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)

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

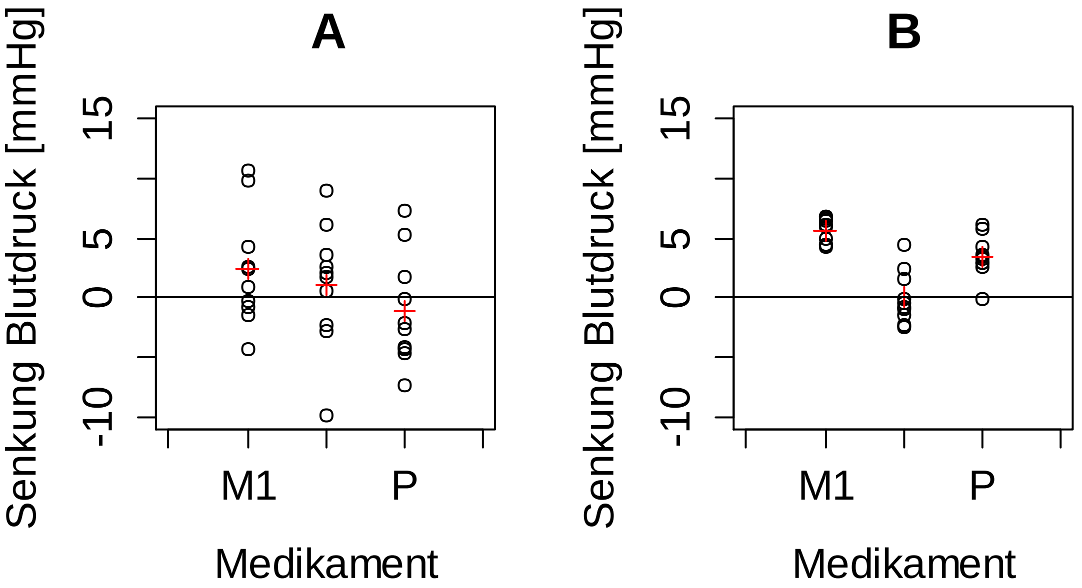
Streuung innerhalb Gruppen:
"Within-Sum-of-Squares" (SS_W)
RSS der Einzelbeobachtungen (schwarze Kreise) um die einzelnen Mittelwerte (rote Kreuze)

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

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

ANOVA: Teststatistik

In welchem Bild ist die Teststatistik der ANOVA **grösser** ?



ANOVA: Modell

- $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2) \;\; iid$
  Technische Nebenbedingung: $\sum_{i=1}^{g} \alpha_i = 0$
- $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_g = 0$
- Teststatistik: $T = \frac{SS_B/(g-1)}{SS_W/(g*(p-1))} = \frac{MS_B}{MS_W}$
- Theorie: Falls H_0 stimmt

“Analyse der Varianzen”

“Mean Squares”

“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: $A \sim X_n$

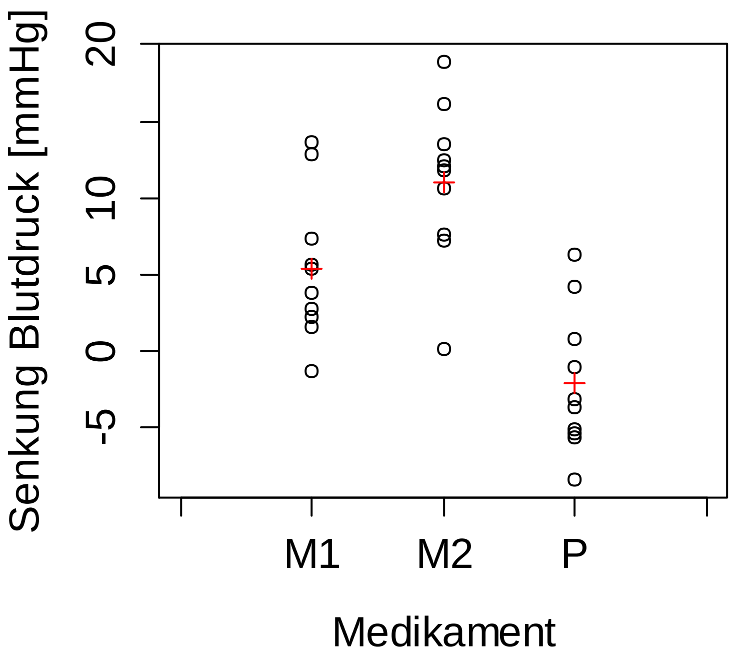
- Angenommen: $A \sim \mathrm{X}_n, B \sim \mathrm{X}_m$ unabhängig

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

- **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
```

$$g - 1 = 2$$
$$g^*(p\text{-}1)=27$$

$$MS_B = \frac{872.3}{2} = 436.1$$
$$MS_W = \frac{642.1}{27} = 23.8$$

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
  Bei n Gruppen gibt es $\binom{n}{2} = \frac{n(n-1)}{2}$ t-Tests
  Bsp: $n = 20 \rightarrow 190$ Tests auf 5%-Niveau
  Könnten etwa $0.05 * 190 \approx 10$ falsch positive Tests haben
- Lösung: t-Test korrigieren (z.B. Bonferroni, …)

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

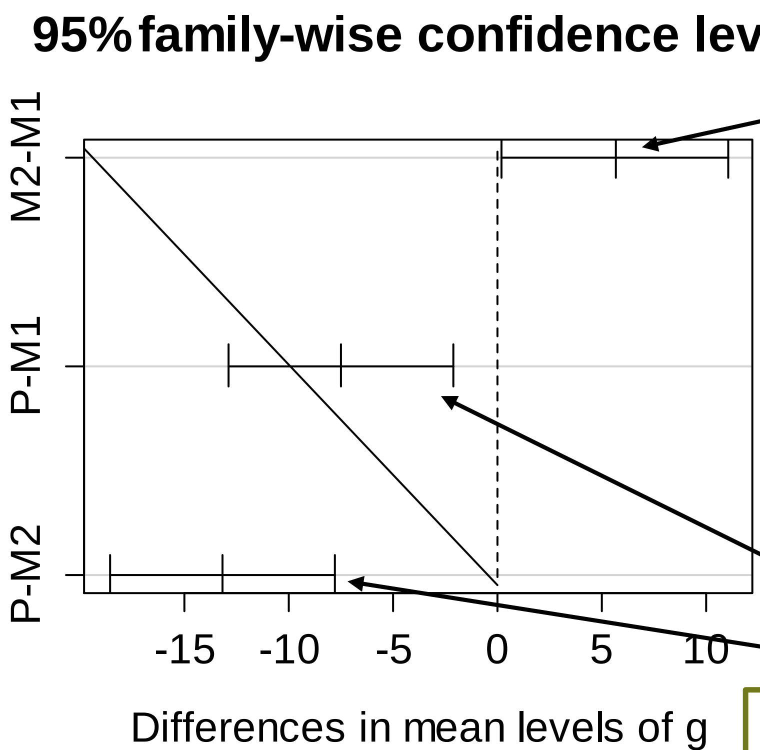
- Vorteil:
  - Vertrauensintervalle für Differenzen der Gruppenmittelwerte
  - Wa., dass **alle** wahren Differenzen in den Vertrauensintervallen liegen: 95%
- Alternative zum paarweisen t-Test

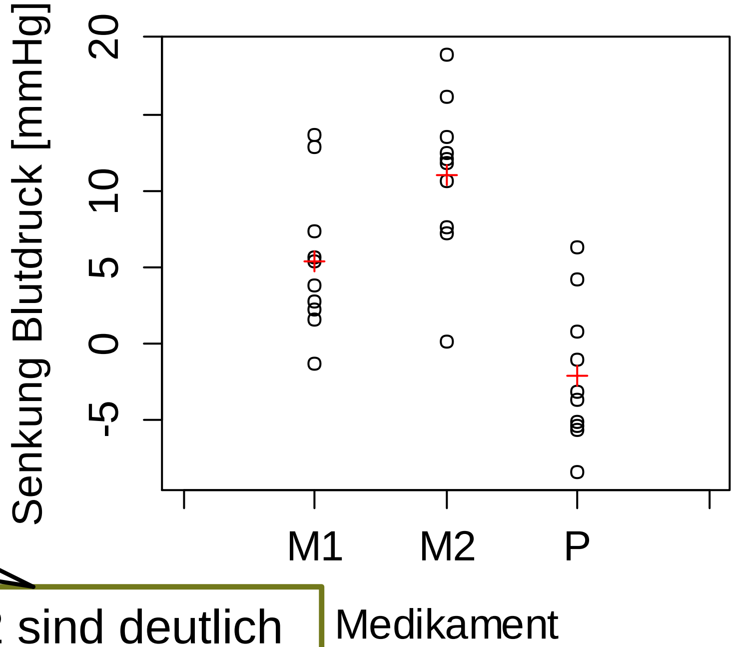
- Empfehlung: Tukey HSD verwenden

Beispiel in R: TukeyHSD

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

M2 ist deutlich wirksamer als M1



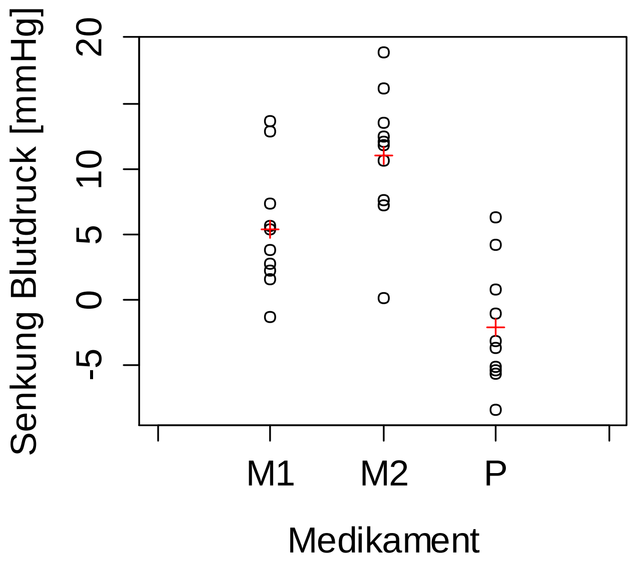


M1 und M2 sind deutlich wirksamer als Placebo

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

Konstraste – Bsp 1: Paarweise Vergleiche

(Alternative zu TukeyHSD)

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

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

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

Kontraste – Bsp 2: Gruppe der Medikamente vs. Placebo

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

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

Die Medikamente sind deutlich wirksamer als Placebo

M2 ist deutlich wirksamer als M1

Kontraste

Angenommen, es gibt zwei Medikamente (M1, M2) und auch zwei mögliche Formen von Placebo (P1, P2). Folgende Matrix ist dann eine mögliche Kontrastmatrix für die Vergleiche:

- (M1, M2) vs. (P1, P2)
- M1 vs. M2
- P1 vs. P2

$$\begin{pmatrix} 0.5 & 0.5 & -0.5 & -0.5 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix} * \begin{pmatrix} \mu_{M1} \\ \mu_{M2} \\ \mu_{P1} \\ \mu_{P2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

- Ja
- Nein

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

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

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