

Regressionsanalysen

Y = Zielvariable/abhängige Variable

X = Prädiktor/unabhängige Variable

Ziele Regressionsanalyse

- I. Zusammenhänge zwischen Prädiktoren und Zielvariablen (X & Y)
 - a. Verstehen
 - b. Quantifizieren
- II. Vorhersagen machen

Lineares Modell

- metrische Zielgröße
- umfasst verschiedene Spezialfälle:
 - Lineare Regression
 - t-Test
 - Varianzanalyse (ANOVA)

Logistische Regression

- binäre Zielgröße
- „verallgemeinertes lineares Modell“

- Einflussgrößen können sowohl metrisch, als auch kategorial sein
- Einfache Regression (eine Einflussgröße) & multiple Regression (mehrere Einflussgrößen)

Lineare Modelle

Lineare Modelle

- einfache lineare Regression: ein metrischer Prädiktor
- multiple lineare Regression: mehrere metrische Prädiktoren
- Varianzanalyse: kategoriale Prädiktoren
- Kovarianzanalyse: metrische und kategoriale Prädiktoren

Umsetzung in R

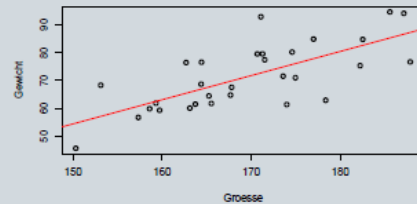
Alle Formen des linearen Modells können einheitlich mithilfe der `lm`-Funktion umgesetzt werden.

Einfache lineare Regression

Bestandteile lineare Regression

I. Systematischen Teil (intercept und slope)

Regressionsgerade



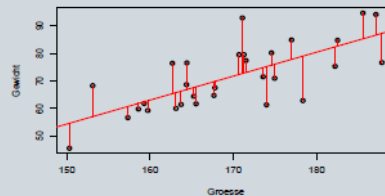
$$\text{Gewicht} = \beta_0 + \beta_1 \cdot \text{Grösse}$$

β_0 Schnittpunkt mit y-Achse (*intercept*)

β_1 Anstieg (wenn Größe um 1 wächst) (*slope*)

II. Zufälligen Teil (Residuen)

Lineares Modell



$$\text{Gewicht}_i = \beta_0 + \beta_1 \cdot \text{Grösse}_i + \epsilon_i$$

Residuum ϵ_i beschreibt Abweichung von der Geraden mit $E[\epsilon_i] = 0$

Parameter gehen linear in **systematischen Teil** ein!

Beispiele für lineare Modelle

- $y = \beta_0 + \beta_1 \cdot x + \epsilon$
- $y = \beta_0 + \beta_1 \cdot x^2 + \epsilon$
- $y = \beta_0 + \beta_1 \cdot \log(x) + \epsilon$

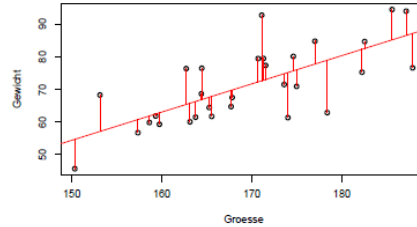
da die Parameter β_0 und β_1 linear eingehen

Beispiele für **nicht**-lineare Modelle

- $y = \frac{\beta_1 \cdot x}{\beta_2 + x} + \epsilon$ (Michaelis-Menten Modell)
- $y = \beta_0 + \beta_1 \cdot x^{\beta_2} + \epsilon$

Methode der kleinsten Quadrate

- Die Anstiege werden so gewählt, dass die Summe der quadratischen Abweichungen (Residuen) so gering wie möglich ist und am wenigsten von der Gerade abweichen



Methode der kleinsten Quadrate

Wähle $\hat{\beta}_0$ und $\hat{\beta}_1$ so, dass die Summe der quadratischen Abweichungen minimal ist:

$$\min_{\beta_0, \beta_1} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 \cdot x_i))^2$$

Voraussetzungen linearer Regression

- Normalverteilung
- Konstante Varianz
- Unabhängigkeit

Hat X einen linearen Einfluss auf Y?

- $H_0 : \beta_1 = 0$ vs $H_A : \beta_1 \neq 0$
- Test-Statistik $\frac{\hat{\beta}_1}{se_{\hat{\beta}_1}} \stackrel{H_0}{\sim} t_{n-p}$
- Wald Signifikanz-Test

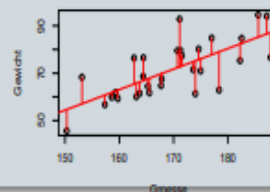
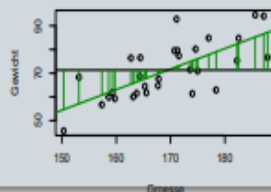
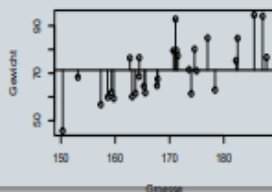
Extrapolation - Gefahr

- Vorhersagen sollte nur für Werte die innerhalb der Beobachtungen liegen gemacht werden

Varianzzerlegung

Teile der Varianz von Y kann durch die Einflussgröße X erklärt werden:

$$\text{Gesamtvarianz} = \text{erklärte Varianz} + \text{Residualvarianz}$$



Bestimmtheitsmaß B (R²)

Anpassungsgüte des linearen Modells

Anteil der erklärten Varianz an der Gesamtvarianz:

$$B = R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\text{erklärte Varianz}}{\text{Gesamtvarianz}}$$

Bei der einfachen linearen Regression (d.h. eine Einflussgröße) entspricht das Bestimmtheitsmaß B dem Quadrat des Pearsonschen Korrelationskoeffizienten r.

Modelldiagnostik

- Notwendig, um zu überprüfen, ob das statistische Modell adäquat ist

Die drei Säulen

- Annahmen über die Residuen
- Ungewöhnliche Beobachtungen
- Modellstruktur

1. Annahme über die Residuen

Annahmen über die Residuen:

- Normalverteilung
- konstante Varianz
- Unabhängigkeit

Ungewöhnliche Beobachtungen

Gibt es ungewöhnliche Beobachtungen, so sollten diese besonders auf Plausibilität überprüft werden.

Modellstruktur

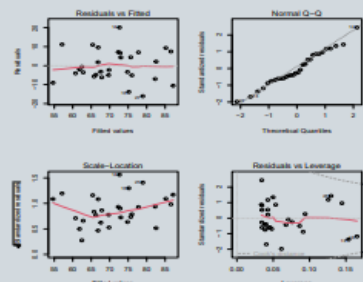
Ist die Annahme der Linearität gerechtfertigt?

Modelldiagnostik:

- Diagnostische Residuenplots

Diagnostische Residuen-Plots

```
par(mfrow = c(2,2))  
plot(model1)
```



Diagnostische Residuenplots

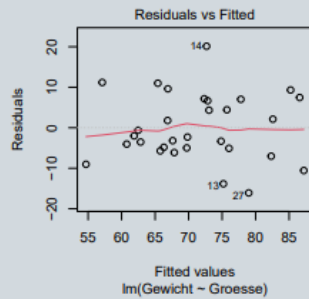
```
plot(model1)
```

```
par(mfrow = c(2,2))
```

```
par(mfrow = c(1,1))
```

Linearitätsannahme

Residuen vs Angepasste Werte

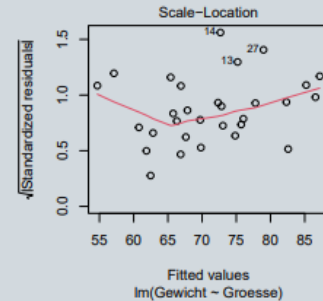


Residuen sollten um die Nulllinie streuen, anderenfalls könnte die Linearitätsannahme verletzt sein.

Struktur des Modells korrekt?
Varianz konstant?

Konstante Residuenvarianz

transformierte Residuen vs Angepasste Werte

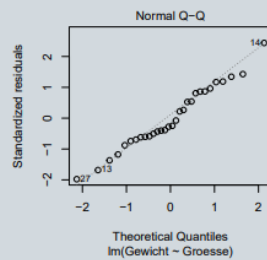


Überprüfung, ob Residuenvarianz konstant ist: Rote Linie sollte näherungsweise bei 0.8 verlaufen.

Steigt die Linie an, ist dies ein Zeichen zur Verletzung der Varianzgleichheit

Normalverteilungsannahme

Q-Q-Plot: Normalverteilungsannahme

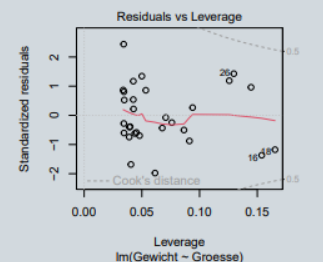


Bei Normalverteilung der Residuen bilden Punkte eine Gerade.

Gegenüberstellung der bei einer Normalverteilung erwarteten mit beobachteten Residuen

Ausreißer

Ausreißer und einflussreiche Beobachtungen



Leverage = Einfluss eines Datenpunkts auf den Fit
Cook's distance = Maß für Ausreißer und Leverage

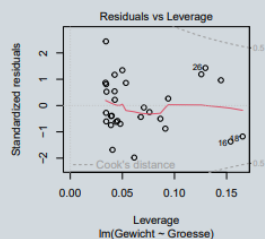
Untersuchung auf einflussreichen Beobachtungen → Ausreißern

x-Achse = Leverage jedes Datenpunktes (Maß, wie stark der Fall die Parameterschätzung beeinflusst (z.B. in Varianzanalyse deutlich weniger Fälle in einer Gruppe)

y-Achse = standardisierte Residuen

Cook's Distance = Punkte darin sollten nochmal kritisch überprüft werden

Ausreißer und einflussreiche Beobachtungen



Ausreißer (d.h. große Residuen) und Punkte mit großer Leverage oder Cook-Distanz sollten besonders auf Plausibilität überprüft werden.

Überprüfung angepasster Modelle auf Adäquatheit“

3. Modell erstellen mit lm-Funktion

lm-Funktion

```
(modell1 <- lm(Gewicht ~ Groesse, data = lmExample))  
  
##  
## Call:  
## lm(formula = Gewicht ~ Groesse, data = lmExample)  
##  
## Coefficients:  
## (Intercept)      Groesse  
##      -75.494       0.866
```

2. Ergebnisse: summary Funktion

```
summary(modell1)  
  
##  
## Call:  
## lm(formula = HF ~ Belastung, data = ergo)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -25.611  -7.311  -3.957   6.825  22.171   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) 76.95671    5.34763   14.39 1.14e-11 ***  
## Belastung    0.46091    0.04574   10.08 4.66e-09 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 12.69 on 19 degrees of freedom  
## Multiple R-squared:  0.8424, Adjusted R-squared:  0.8341  
## F-statistic: 101.5 on 1 and 19 DF,  p-value: 4.659e-09
```

1. Ergebnissinterpretationen

- Intercept = Schnittpunkt mit der Y-Achse, gibt den Wert der Zielgröße (Herzfrequenz) bei $X = 0$ (keine Belastung) an, sprich eine mittlere HF von 77 bei keiner Belastung
- Belastung = gibt den Anstieg der Zielvariablen (Herzfrequenz) pro Anstieg des Prädiktors (Belastung) um eine Einheit, sprich pro 1 W Belastung steigt die HF um 0,46, der Anstieg ist statistisch signifikant
- R^2 = Maß zur Güte der Anpassung eines Modells an den beobachteten Daten
 - = Anteil der durch das Modell erklärten Varianz an der Gesamtvarianz
 - Wert zwischen 1 und 0
 - Je höher der Wert, desto größer ist der Anteil der Variation in der abhängigen Variable zu erklären

Konfidenzintervalle Modell-Parameter: confint(model)

95%-Konfidenzintervall zu den Modell-Parametern mit `confint`. Das KI zum Anstieg enthält nicht die 0. Dies entspricht dem signifikanten Testergebnis in der `summary`-Ausgabe.

```
confint(model1)

##              2.5 %      97.5 %
## (Intercept) 65.7639934 88.1494265
## Belastung   0.3651667  0.5566515
```

Modellvorhersagen für bestimmte Werte: predict(model) inklusive Konfidenzintervalle

```
vorhersage <- data.frame(Belastung = c(125, 1000))
vorhersage$HF <- predict(model1, newdata = vorhersage)
vorhersage
```

```
##  Belastung      HF
## 1      125 134.5703
## 2     1000 537.8658 CAVE: Extrapolation!
```

Für diese vorhergesagten mittleren Herzfrequenzen bei gegebener Belastung können auch Konfidenzintervalle angegeben werden, um die Unsicherheit der Schätzung abzubilden:

```
predict(model1, newdata = vorhersage, interval = "confidence")
```

```
##      fit      lwr      upr
## 1 134.5703 128.2982 140.8425 Mittelwert mit dazugehörigen KI
## 2 537.8658 451.5029 624.2288
```

Mithilfe der `predict`-Funktion kann zusätzlich zum **Mittelwert** bei gegebener Belastung und seinem Konfidenzintervall auch ein **Intervall vorhergesagt** werden, in welchem 95% der **Einzelwerte** liegen:

```
predict(model1, newdata = vorhersage, interval = "prediction")
```

```
##      fit      lwr      upr
## 1 134.5703 107.2726 161.8681 Mittelwert mit dazugehörigen KI für
## 2 537.8658 447.5088 628.2228 Einzelwerte
```

Dieses Intervall berücksichtigt nun die Varianz zwischen den Probanden. _____

Modellanpassung: Ausreißer ausschließen subset()

```
model3 <- lm(y ~ x, data = modelldiagnostik, subset = -12)
summary(model3)

##
## Call:
## lm(formula = y ~ x, data = modelldiagnostik, subset = -12)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.6560 -1.0432  0.2932  1.3056  2.8831
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  21.41904    1.33695   16.02 1.08e-11 ***
## x            -0.07717    0.05890   -1.31  0.208
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.108 on 17 degrees of freedom
## Multiple R-squared:  0.09171, Adjusted R-squared:  0.03828
## F-statistic: 1.716 on 1 and 17 DF,  p-value: 0.2076
```

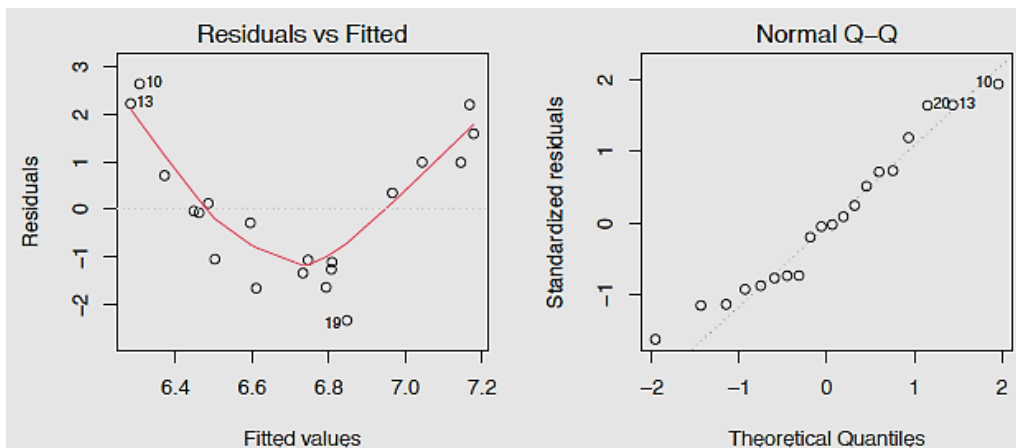
Steigung des
Prädiktor x zwar
negativ, jedoch
nicht mehr
statistisch
signifikant (<.05)

Modelldiagnostik: quadratischer Zusammenhang

```
modelldiagnostik2 <- read.table(file = "data/modelldiagnostik2.dat")
plot(y ~ x, data = modelldiagnostik2)

model4 <- lm(y ~ x, data = modelldiagnostik2)
abline(model4, col = "red", lty = "dotted")
```

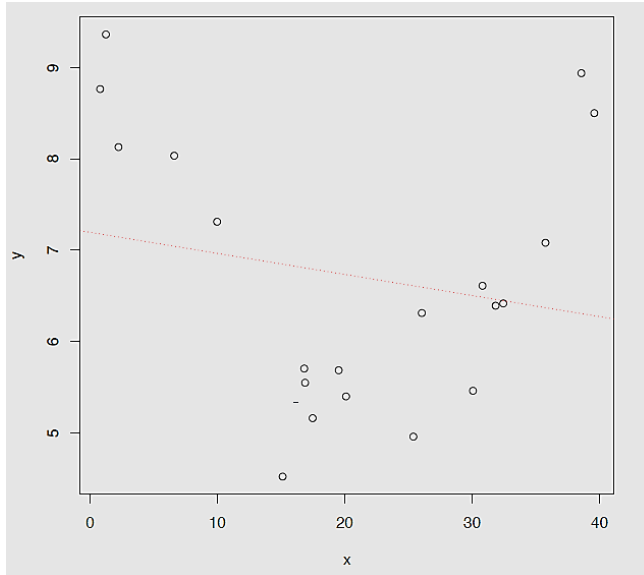
Quadratischer Zusammenhang in Modelldiagnostik erkennbar



CAVE: im einfach Streudiagramm wäre erst einmal kein linearer Zusammenhang erkennbar!

```
modelldiagnostik2 <- read.table(file = "data/modelldiagnostik2.dat")
plot(y ~ x, data = modelldiagnostik2)

model4 <- lm(y ~ x, data = modelldiagnostik2)
abline(model4, col = "red", lty = "dotted")
```



Quadratischer
Zusammenhang!

Erkennbar in Residuals
vs. fitted plot

Quadratischen Einfluss im lm-Befehl fitten

```
lm(y ~ x + I(x^2), data = <dataframe>)
```

Modell bleibt linear, da die Koeffizienten (Beta1, Beta2, etc.) linear in das Modell eingehen

!!!!!!!!!!!!!!

Nochmal
ansehen

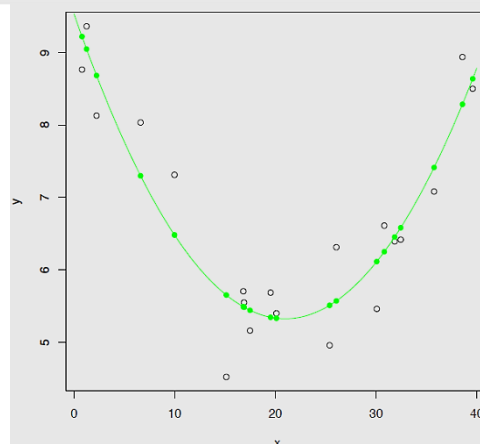
+

Wofür
steht das
I?

```
model5 <- lm(y ~ x + I(x^2), data = modelldiagnostik2)
```

```
plot(y ~ x, data = modelldiagnostik2)
points(fitted(model5) ~ x, data = modelldiagnostik2,
       col = "green", pch = 16)
```

```
# komplette Vorhersage
predData <- data.frame(x = seq(0, 40, by = .5))
predData$y_pred <- predict(model5, newdata = predData)
lines(y_pred ~ x, data = predData, col = "green")
```



Ergebnisse quadratisches Modell

```
summary(model5)
```

```
##
## Call:
## lm(formula = y ~ x + I(x^2), data = modelldiagnostik2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.13030 -0.36290  0.00118  0.34557  0.82989
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.5318085   0.3371351   28.27 9.82e-16 ***
## x            -0.4010146   0.0378602  -10.59 6.62e-09 ***
## I(x^2)        0.0095576   0.0009204   10.38 8.89e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5618 on 17 degrees of freedom
## Multiple R-squared:  0.869, Adjusted R-squared:  0.8535
## F-statistic: 56.36 on 2 and 17 DF,  p-value: 3.149e-08

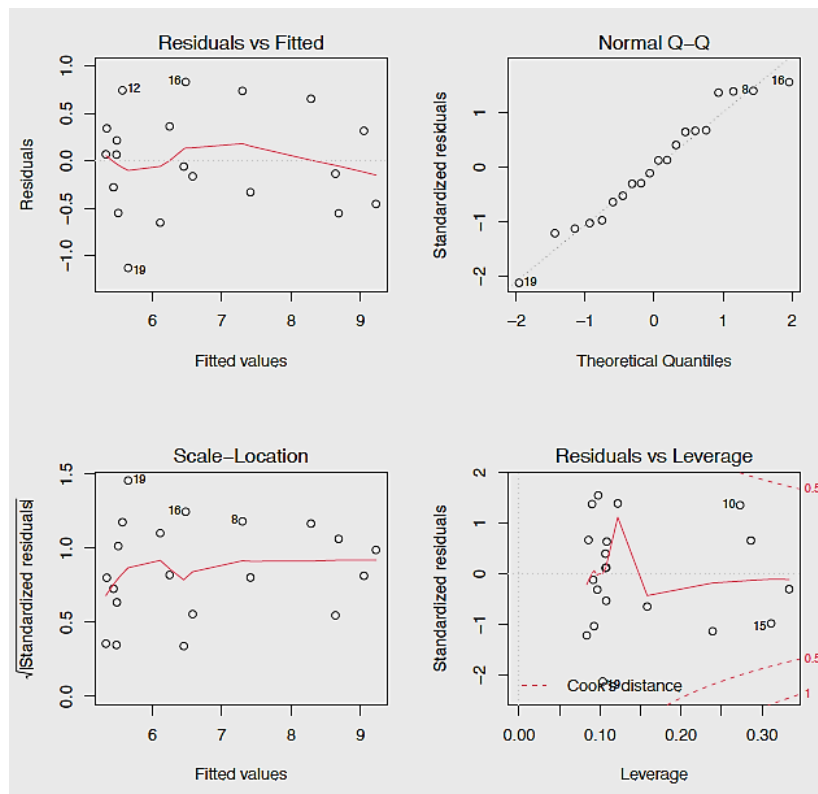
par(mfrow = c(2,2))
plot(model5)
```

Modellveränderung durch quadratischen Term:

-Koeffizient zum quadratischen Term statistisch hoch signifikant!

-Residuen Plots jetzt i.O.

-Im Residuals vs. Fitted Plot kein Trend mehr erkennbar



Weitere Möglichkeiten inadäquaten Modellen:

- Log-Transformation
- Wurzeltransformation
- Arcus-Sinus-Transformation

Multiple lineare Regression

Modell

$$\text{Geburtsgewicht} = \beta_0 + \beta_1 \cdot \text{Alter} + \beta_2 \cdot \text{Gewicht} + \beta_3 \cdot \text{Groesse} + \epsilon$$

Modellanpassung

Analog zur einfachen Regression:

Schätzung der Parameter über Methode der kleinsten Quadrate

Umsetzung in R

Prädiktoren werden mit + in der Formel verknüpft:

```
lm(Geburtsgewicht ~ Alter + Gewicht + Groesse, data=babies)
```

Einfluss mehrerer Prädiktoren

Multiples lineares Modell	=	Prädiktoren ausschließlich metrisch
ANOVA	=	Prädiktoren ausschließlich kategorial
ANCOVA	=	Prädiktoren metrisch & kategorial

Deskription/Visualisierung der verschiedenen Prädiktoren

Kategorial und kategorial:

Kreuztabellen

```
xtabs(~type + activity, data = fruitfly)
```

```
## activity  
## type no yes  
## A 30 30  
## B 30 30
```

Kategorial und metrisch:

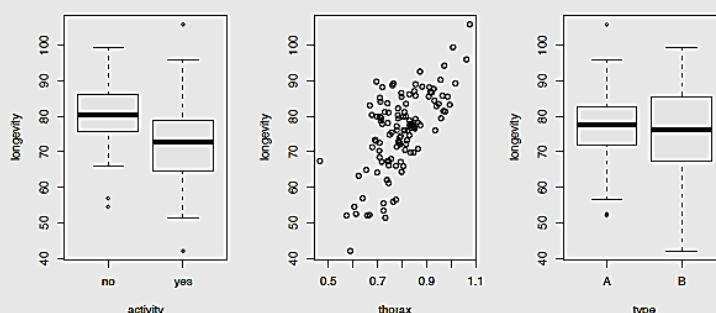
Boxplot

Metrisch und metrisch:

Streudiagramm

Deskriptive Plots: `par(mfrow=c(1,3))`

```
boxplot(longevity ~ activity, fruitfly)  
plot(longevity ~ thorax, fruitfly)  
boxplot(longevity ~ type, fruitfly)
```



-Überlebensdauer sinkt bei Aktivität

-Je größer Thorax, desto größer Überlebensdauer

-Subtyp ohne Einfluss auf Überlebensdauer

Ausführung der multiplen linearen Regression

```
mod_fruitfly1 <- lm(longevity ~ activity + thorax + type, fruitfly)
summary(mod_fruitfly1)

##
## Call:
## lm(formula = longevity ~ activity + thorax + type, data = fruitfly)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.1110  -5.1165   0.6272   4.5039  16.6682
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  23.3681     5.1018   4.580 1.18e-05 ***
## activityyes  -9.3216     1.3008  -7.166 7.69e-11 ***
## thorax       71.8871     6.2154  11.566 < 2e-16 ***
## typeB       -0.2519     1.3036  -0.193  0.847
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.105 on 116 degrees of freedom
## Multiple R-squared:  0.6019, Adjusted R-squared:  0.5916
## F-statistic: 58.46 on 3 and 116 DF,  p-value: < 2.2e-16
```

Interpretation:

-Referenzgruppe in R die Kategorie, die alphabetisch zuerst kommt (no/yes = no, Typ A/Typ B = Typ A)

-Intercept: 23 = bei einer Thoraxgröße von 0

-Aktivität: Aktive Fliegen, bei gleichbleibender Thoraxgröße und vom selben Typ leben -9 Tage kürzer, als nicht aktive

-Typ: Fliegen von Typ B leben 0,25 Tage kürzer, als Fliegen von Typ A (nicht statistisch signifikant)

Lm-Funktion erkennt automatisch, dass es sich um eine ANCOVA handelt

-Modell beschreibt Zusammenhang zwischen Thorax und Lebensdauer mit einem Anstieg von 72

-Die Verankerung der Gerade hängt jedoch von den beiden kategorialen Variablen ab (Aktivität & Typ) → diese sorgen für eine Parallelverschiebung der Gerade

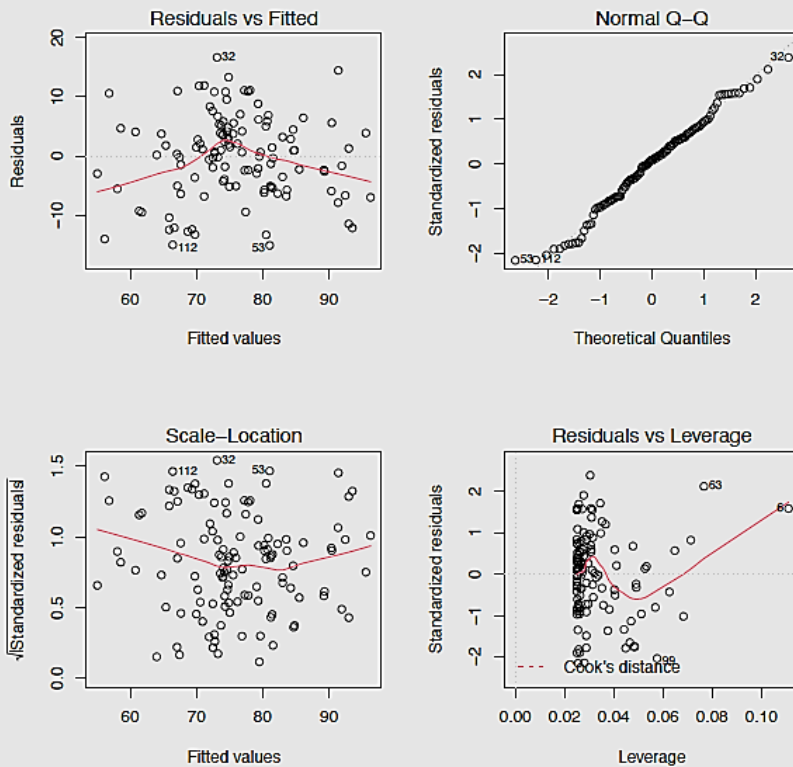
-Modell beschreibt eigentlich 4 Geraden

Modelldiagnostik – analog zur einfachen Regression

- Klassischer plot()-Befehl
- Außerdem: Partial Residual Plots prplot(), package: faraway
 - Überprüfung Linearität für jeden einzelnen metrischen Prädiktor (der geschätzte Einfluss aller anderen Prädiktoren wird herausgerechnet)
 - Testet, ob der übrig gebliebenen Zusammenhang zwischen Prädiktor und abhängiger Variable mit angenommener Linearität vereinbar ist
 - Linearer Zusammenhang = gleichmäßige Streuung der Datenpunkte über und unter der Geraden über den gesamten Wertebereich des Prädiktors

Modelldiagnostik plot()

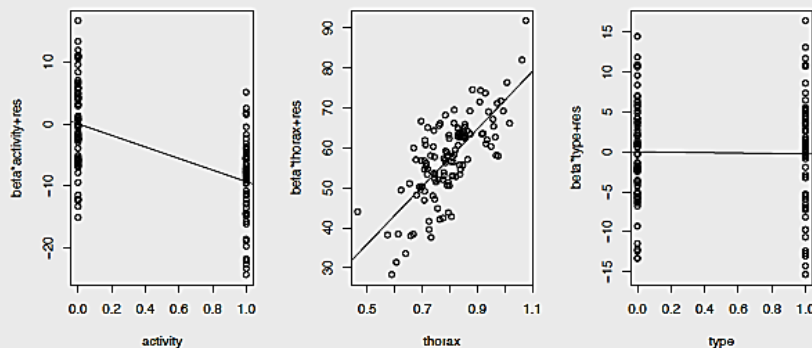
```
par(mfrow=c(2,2)) # Graphikfenster mit 2x2 Diagrammen
plot(mod_fruitfly1) # Ausgabe der vier Diagnostikplots
```



- keine gravierende Verletzung der Modellannahmen
- Residuals vs. fitted in Ordnung (jedoch an Rändern negative Residuen = neigt zur Überschätzung der Lebensdauer)
- Residuenverteilung widerspricht nicht der Normalverteilungsannahme im QQ-Plot
- Ebenfalls erkennbar im Scale-Plot durch erhöhte Varianz
- Datenpunkt Nr. 6 hat einen vergleichsweise starken Einfluss auf den Modell-Fit

Partial residuals plot: prplot()

```
library(faraway) # Laden des Pakets
par(mfrow=c(1,3))
for(i in 1:3) prplot(mod_fruitfly1,i)
```

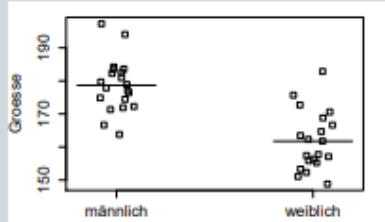


- gleichmäßige Verteilung der Residuen auf beiden Seiten
- Streuung der Beobachtungen erscheinen auf den verschiedenen Ausprägungen gleich

Kategoriale Prädiktoren: Varianzanalyse

Kategoriaker Prädiktor mit zwei Ausprägungen

Beispiel 1: Prädiktor mit zwei Kategorien



Zielgröße: Körpergröße
Prädiktor: Geschlecht

Modell

$$\text{Groesse} = \beta_0 + \beta_1 \cdot \text{Geschlecht} + \epsilon$$

Kodierung der kategorialen Variable:

$$\text{Geschlecht} = \begin{cases} 0, & \text{wenn männlich,} \\ 1, & \text{wenn weiblich} \end{cases}$$

Interpretation der Parameter

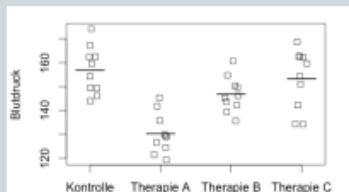
β_0 : Schätzer für Mittelwert bei Männern

β_1 : Schätzer für Mittelwertunterschied zwischen Geschlechtern

$\rightarrow \beta_0 + \beta_1$: Schätzer für Mittelwert der Frauen

Kategoriaker Prädiktor mit mehreren Ausprägungen

Beispiel 2: Prädiktor mit vier Kategorien



Zielgröße: Blutdruck
Prädiktor: Behandlung

Modell

$$\text{Blutdruck} = \beta_0 + \beta_1 \cdot \text{Therapie}_A + \beta_2 \cdot \text{Therapie}_B + \beta_3 \cdot \text{Therapie}_C + \epsilon$$

Kodierung der kategorialen Variable:

$$\text{Therapie}_A = \begin{cases} 1, & \text{wenn Therapie A,} \\ 0, & \text{sonst} \end{cases} \quad \text{andere Variablen analog}$$

Interpretation der Parameter

β_0 : Schätzer für Mittelwert in Kontrollgruppe

β_1 : Schätzer für Unterschied zwischen Therapie A und Kontrolle

β_2 : Schätzer für Unterschied zwischen Therapie B und Kontrolle

β_3 : Schätzer für Unterschied zwischen Therapie C und Kontrolle

Modellbildung

Kodierung von kategorialen Variablen durch Einführung von zusätzlichen Variablen, sogenannter Dummy-Variablen

bei n Kategorien müssen n-1 Parameter geschätzt werden

Art der Kodierung ist abhängig von Fragestellung, z. B.:

- Vergleich zu Kontrollgruppe (Standard in R)
- Vergleich jeweils benachbarter Kategorien (bei ordinalen Variablen)

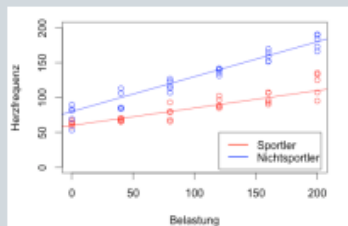
Umsetzung in R

```
summary(lm(Blutdruck ~ Behandlung))

##
## Call:
## lm(formula = Blutdruck ~ Behandlung)
##
## Residuals:
##   Min     1Q   Median     3Q      Max
## -11.47  -4.60  -1.34   3.80  20.77
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    159.82     2.27   70.41 < 2e-16 ***
## BehandlungTherapie A    -27.48     3.21  -8.56 3.3e-10 ***
## BehandlungTherapie B    -16.11     3.21  -5.02 1.4e-05 ***
## BehandlungTherapie C     -1.68     3.21  -0.52  0.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.18 on 36 degrees of freedom
## Multiple R-squared:  0.731, Adjusted R-squared:  0.709
## F-statistic: 32.7 on 3 and 36 DF,  p-value: 2.22e-10
```

Metrische & kategoriale Prädiktoren: Kovarianzanalyse

Beispiel 2: Ergometrie mit Interaktion



Zielgröße: Herzfrequenz
Prädiktoren: Belastung,
Trainingszustand

Modell

$$\text{Herzfrequenz} = \beta_0 + \beta_1 \cdot \text{Bel.} + \beta_2 \cdot \text{Training} + \beta_3 \cdot \text{Bel.} \cdot \text{Training} + \epsilon$$

Kodierung der kategorialen Variable:

$$\text{Training} = \begin{cases} 1, & \text{wenn Sportler} \\ 0, & \text{wenn Nichtsportler} \end{cases}$$

Interpretation der Parameter

β_0 : Schnittpunkt mit y-Achse bei Nichtsportlern

β_1 : Anstieg der Geraden bei steigender Belastung bei Nichtsp.

β_2 : Abstand der Geraden: Unterschied Sportler/Nichtsp. ohne Belastung

β_3 : Unterschied des Anstiegs zwischen Sp./N.sp. mit steigender Belastung

Umgang mit Interaktionen

Interaktionen zwischen Prädiktoren

- Werden im linearen Modell als Produkt modelliert
- Interaktion ist nicht Korrelation!
- Sind **nicht-additiv** (Wirkung von Prädiktor a hängt von Prädiktor b ab)
= **gegenseitige Beeinflussung der Prädiktoren (Interaktionseffekte)**

Umsetzung in R

- ohne Interaktion der Prädiktoren

```
lm(Herzfrequenz ~ Belastung + Training)
```

- mit Interaktion der Prädiktoren

```
lm(Herzfrequenz ~ Belastung + Training + Belastung:Training)
```

oder äquivalent

```
lm(Herzfrequenz ~ Belastung*Training)
```

Umsetzung in R

```
mod_fruitfly3 <- lm(longevity ~ thorax + activity + thorax:activity,  
  data = fruitfly)
```

```
## oder einfacher:
```

```
mod_fruitfly3 <- lm(longevity ~ thorax * activity, data = fruitfly)
```

```
## lm(formula = longevity ~ thorax * activity, data = fruitfly)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max  
## -17.0874  -4.4162   0.5808   4.0780  14.0723  
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept)      43.841     6.339   6.916 2.70e-10 ***  
## thorax           45.821     7.951   5.763 6.93e-08 ***  
## activityyes     -51.855     9.132  -5.679 1.02e-07 ***  
## thorax:activityyes  53.277    11.341   4.698 7.28e-06 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 6.513 on 116 degrees of freedom  
## Multiple R-squared:  0.6654, Adjusted R-squared:  0.6568  
## F-statistic: 76.9 on 3 and 116 DF,  p-value: < 2.2e-16
```

Ergebnisinterpretation in R

```
## lm(formula = longevity ~ thorax * activity, data = fruitfly)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.0874  -4.4162   0.5808   4.0780  14.0723
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      43.841      6.339   6.916 2.70e-10 ***
## thorax           45.821      7.951   5.763 6.93e-08 ***
## activityyes     -51.855      9.132  -5.679 1.02e-07 ***
## thorax:activityyes  53.277     11.341   4.698 7.28e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.513 on 116 degrees of freedom
## Multiple R-squared:  0.6654, Adjusted R-squared:  0.6568
## F-statistic: 76.9 on 3 and 116 DF, p-value: < 2.2e-16
```

Interpretation

1. „Intercept“:
-mittlere Lebenszeit in der Referenzgruppe (Nichtaktive Fliegen) bei theoretischer Thoraxgröße 0 = 44 Tage
2. „activities“:
-Unterschied zwischen aktiven zu nicht-aktiven Fliegen, bei der Thoraxgröße 0
= - 52, sprich aktive Fliegen leben -52 Tage kürzer
3. „thorax“:
-nimmt die Thoraxgröße um eine Einheit zu, nimmt die das Überleben um 44 Tage zu bei der Referenzgruppe
4. Interaktionsterm „thorax:activities“
-Interaktion bedeutet, dass die Geraden für Lebensdauer vs. Thorax nicht parallel verlaufen in beiden Aktivitätsgruppen (d.h. der Zuwachs der Lebenszeit bei zunehmender Thoraxgröße verläuft in beiden Gruppen unterschiedlich)
-der Unterschied zwischen den Anstiegen beider Gruppen beträgt 53 Tage
-Anstieg der Geraden für aktive Fliegen (thorax + thorax:activities) ist 99,1, daher profitieren aktive Fliegen mehr von ihrer Thoraxgröße, als nicht-aktive
-kleine aktive Fliegen sterben nach dem Modell im Mittel früher, als kleine nicht-aktive
→ dieser Unterschied verringert sich jedoch und ändert sich mit zunehmender Thoraxgröße (bei großem Thorax leben aktive Fliegen länger)

Visualisierung des angepassten Modells

1. Plot erstellen, kategorisiert dargestellt

- Prädiktor Thorax und Zielvariable Lebensdauer sollen in einem Streudiagramm dargestellt werden
- Der kategoriale Prädiktor Aktivität farblich dargestellt werden

```
plot(longevity ~ thorax, data = fruitfly, main = "Fruitflies",  
     col = ifelse(fruitfly$activity == 'yes', 'red', 'blue'),  
     pch = ifelse(fruitfly$activity == 'yes', 17, 15)  
     )
```



Dreick /Quadrat

- Mit diesem Behl wird zwar das Streudiagramm mit den einzelnen Datenpunkten, jedoch nicht die jeweilige Gerade angezeigt

Visualisierung der linearen Geraden

1. Geschätzten Koeffizienten (Anstiege und Intercept aus dem Modell extrahieren) aus dem Modell extrahieren: `coef(model)`
2. Koeffizienten für die beiden Geraden übernehmen und in den `abline()` Befehl integrieren
3. Legenden hinzufügen

Visualisierung des angepassten Modells

2. Visualisierung der linearen Geraden

Mithilfe der `coef()`-Funktion werden die geschätzten Koeffizienten (Anstiege und Schnittpunkte mit der y-Achse) aus dem Modell extrahiert und unter dem Namen `betas` gespeichert:

```
betas <- coef(mod_fruitfly3)
betas
```

```
##      (Intercept)      thorax      activityyes thorax:activity:
##      43.84103      45.82093      -51.85536      53.27
```

Abhängig davon, in welcher Reihenfolge die Prädiktoren in das Modell aufgenommen wurden, wählen wir nun die entsprechenden Koeffizienten für die beiden Geraden und zeichnen diese mithilfe der Funktion `abline()`. Für die nichtaktiven Fliegen ergibt sich:

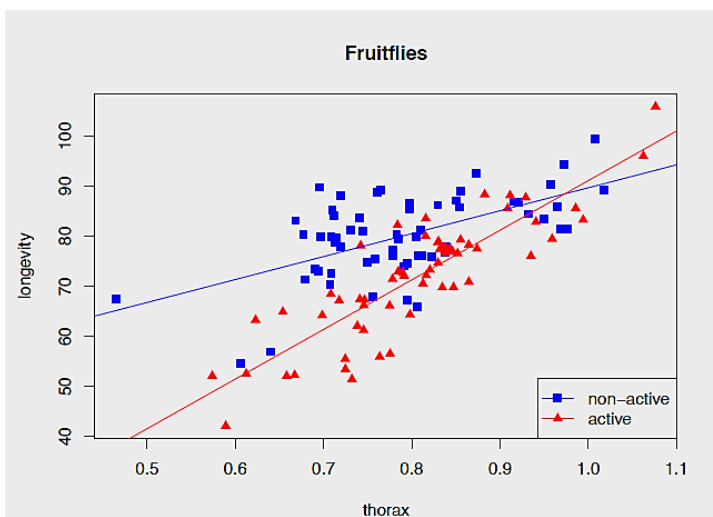
```
abline(a = betas[1] , b = betas[2] , col = 'blue' )
```

Für die aktiven Fliegen ergeben sich Anstieg und Schnittpunkt mit der y-Achse aus der Summe der jeweiligen Koeffizienten:

```
abline(a = betas[1] + betas[3] , b = betas[2] + betas[4] ,col='red')
```

Zur Verdeutlichung fügen wir noch eine Legende hinzu:

```
legend("bottomright", legend = c("non-active","active"),
      col=c('blue','red'), pch = c(15, 17),
      lwd=1)
```



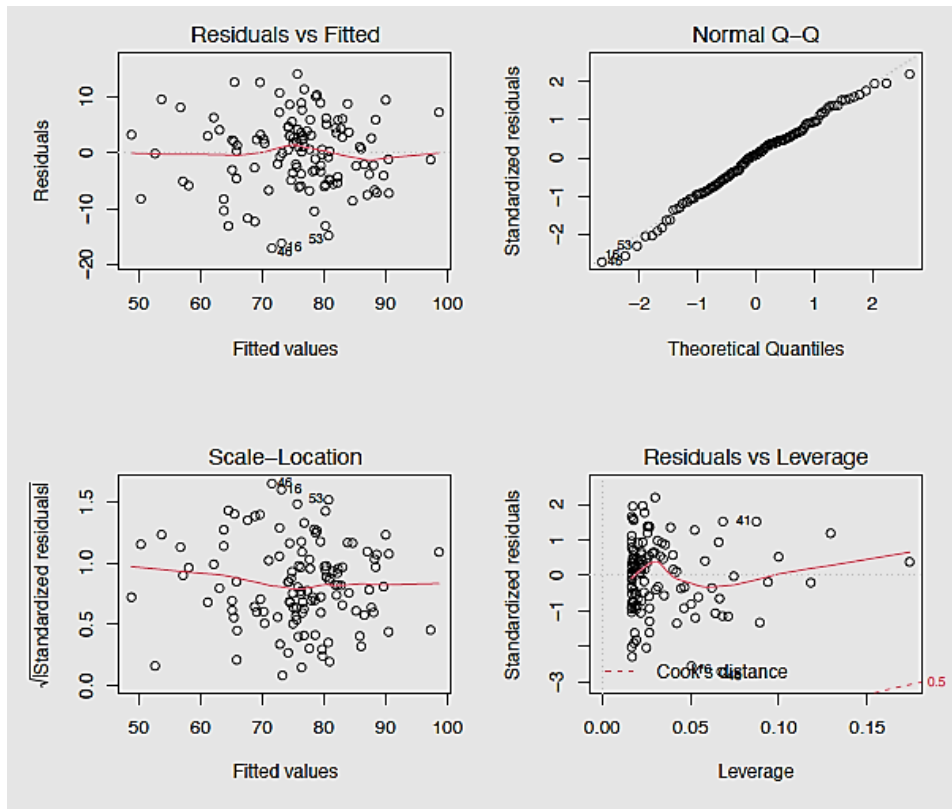
Je größer der Thorax der männlichen Fruchtfliegen ist, desto länger leben diese im Mittel (positiver Koeffizient für `thorax` und auch positiver Interaktions-Koeffizient). Dieser Zusammenhang ist ausgeprägter bei sexuell aktiven Fliegen (positiver Interaktions-Koeffizient). Sexuell aktive Männchen mit kleinem Thorax (≤ 1) haben im Mittel eine geringere Lebenszeit im Vergleich mit enthaltsamen Männchen (negativer Koeffizient für `activity`).

Modelldiagnostik bei Interaktionen

- Analog zur weiteren Modelldiagnostik

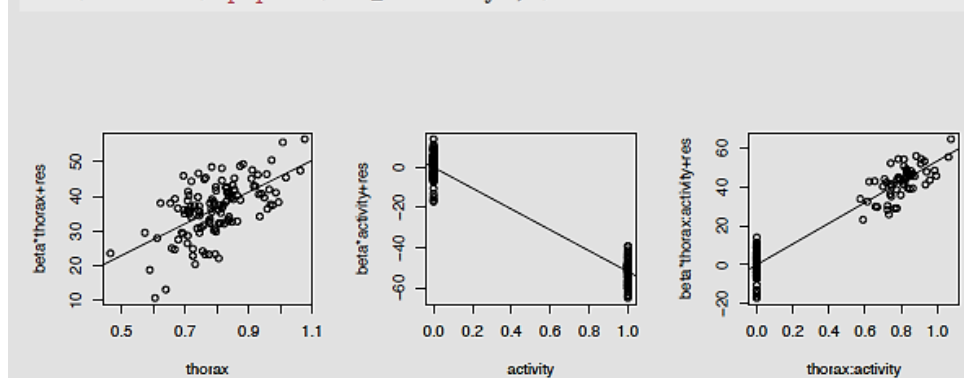
Modelldiagnostik plot()

```
par(mfrow = c(2,2))  
plot(mod_fruitfly3)
```



Partial residuals plot: prplot()

```
par(mfrow = c(1,3))  
for(i in 1:3) prplot(mod_fruitfly3,i)
```



Auf Grundlage der Diagnostikplots spricht nichts gegen das angepasste Modell

Es sollte jedoch auch der Anteil der vom Modell erklärten Varianz (Bestimmtheitsmaß R^2) berücksichtigt werden!

Modellauswahl

Ziel

- Relevante Einflüsse aus einer Vielzahl Prädiktoren herauszufiltern
- Modell auf die wesentlichen Prädiktoren beschränken

Die Zahl denkbarer Modelle ist praktisch unbeschränkt

Neue Prädiktoren können generiert werden

- Prädiktoren können miteinander verknüpft werden
z.B. Produkte → Interaktionen von Prädiktoren
- Prädiktoren können transformiert werden, z.B. Logarithmus, Wurzeln oder Polynome
→ bleibt trotzdem ein lineares Modell in Bezug auf die Parameter!

Modellauswahl

- Subjektiver Prozess
 - Analyse eines angepassten Modells: eher Handwerk
 - Auswahl eines Modells: eher Kunst
- Vorwissen über Zusammenhänge kann einfließen

mögliche Verfahren

- Backward selection
- Auswahl des Modells nach Akaikes Informationskriterium
- *Purposeful Selection*
- ...

Backward selection

- Beginn mit allen Prädiktoren (Maximales Modell)
- Entfernt den Prädiktor mit größten p-Wert = **Entfernen nicht signifikanter Prädiktoren**
- Anpassung des kleineren Modells, solange bis alle p-Werte kleiner als der kritische Alpha-Wert ist (i.d.R. 0,05)

R: drop1-Funktion

Forward selection

= Schrittweise hinzunehmen von Prädiktoren

= p-Wert basierte Verfahren

Akaikes Informationskriterium (AIC)

Berechnung

nicht über p-Werte, sondern über Restvarianz (bzw. Likelihood):

$$AIC = 2k + n \cdot \log(\sigma^2)$$

k : Zahl der Parameter des Modells

Bedeutung

Das AIC für ein Modell ist ein relatives Maß dafür, wie sehr die Daten dieses Modell stützen. Je kleiner der Wert, umso besser.

Umsetzung in R

```
selectedModel <- step(maximalModel)
```

R: step() Funktion

„Purpesful selection“

- Bei vielen Prädiktoren: Vorauswahl von Prädiktorkandidaten
 - nach univariater Signifikanz (etwa $P < 0.25$)
 - oder nach inhaltlicher Relevanz
- Start mit vollem additiven Modell nach Vorauswahl
- Schrittweises Entfernen von Prädiktoren
 - nach Signifikanz (etwa $P > 0.05$ oder > 0.1)
 - und falls kein Confounder (keine große Änderung der β s)
- Überprüfe Hinzunahme nicht vorausgewählter Prädiktoren
- Überprüfe Linearität und Kategorien der Prädiktoren
- Interaktion zwischen den Prädiktoren im Modell
 - falls Interaktion plausibel
 - einzeln nach Signifikanz (etwa $P < 0.05$ oder < 0.01)
 - volles Interaktionsmodell vereinfachen (nach P-Wert)
- Überprüfe Modell-Fit: Passt das Modell zu den Daten?

Quelle: Hosmer, Lemeshow, Applied Logistic Regression, 2000

= Kriterienbasierte Verfahren

R: Ausführung backward selection – drop1()

Procedere

1. Alle Prädiktoren – Maximales Modell

```
drop1(<model>, test="F")
```

2. Entfernung Prädiktor mit größten p-Wert und Erstellung neuen Modells

3. Weitere Anpassungen, bis alle p-Werte signifikant sind

Beispiel

```
drop1(mod_fruitfly1, test="F")

## Single term deletions
##
## Model:
## longevity ~ activity + thorax + type
##           Df Sum of Sq    RSS    AIC F value    Pr(>F)
## <none>                5855.1 474.51
## activity  1     2591.8  8446.9 516.49  51.3484 7.689e-11 ***
## thorax   1     6752.2 12607.2 564.54 133.7735 < 2.2e-16 ***
## type     1         1.9   5856.9 472.55   0.0373   0.8471
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

mod_fruitfly2 <- update(mod_fruitfly1, . ~ . -type)
# verbleibende Praediktoren erneut beurteilen!
drop1(mod_fruitfly2, test="F")

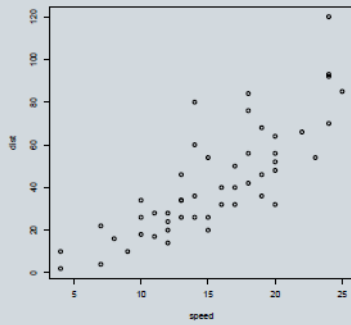
## Single term deletions
##
## Model:
## longevity ~ activity + thorax
##           Df Sum of Sq    RSS    AIC F value    Pr(>F)
## <none>                5856.9 472.55
## activity  1     2593.0  8449.9 514.53  51.798 6.359e-11 ***
## thorax   1     6842.4 12699.3 563.42 136.685 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

R-Befehle gebündelt

Plots

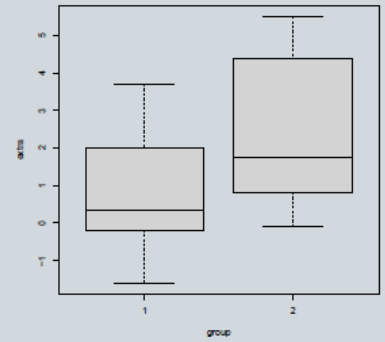
metrisch vs metrisch: Streudiagramm / Scatterplot

```
plot(dist ~ speed, data=cars)
```



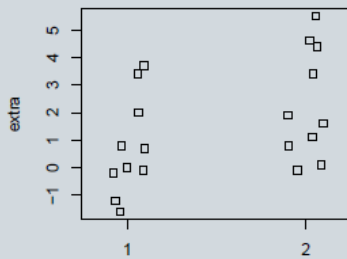
metrisch vs kategorial: Boxplot

```
boxplot(extra ~ group, data=sleep)
```



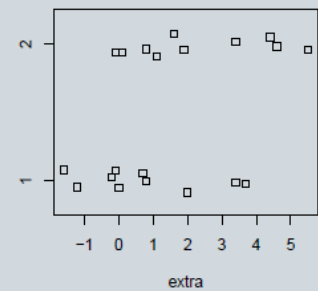
metrisch vs kategorial: Dotplot

```
stripchart(extra ~ group, data=sleep, vertical=TRUE, method="jitter")
```



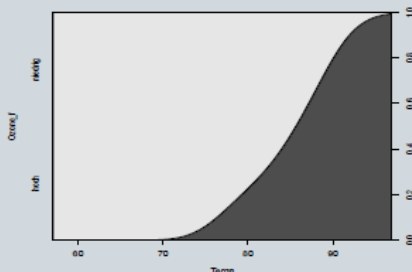
kategorial vs metrisch: Dotplot

```
stripchart(extra ~ group, data=sleep, vertical=FALSE, method="jitter")
```



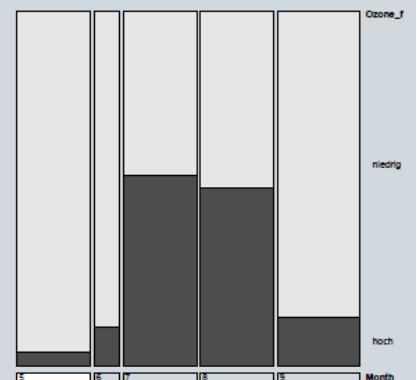
kategorial vs metrisch: konditionaler Dichteplot

```
cdplot(Ozone_f ~ Temp, data=airquality)
```



kategorial vs kategorial: Mosaik-Plot

```
library("vcd")  
doubledecker(Ozone_f ~ Month, data=airquality)
```



Einfache lineare Regression

1.

```
lm-Funktion
(modell1 <- lm(Gewicht ~ Groesse, data = lmExample))

##
## Call:
## lm(formula = Gewicht ~ Groesse, data = lmExample)
##
## Coefficients:
## (Intercept)      Groesse
##      -75.494       0.866
```

2.

```
Summary-Befehl
summary(modell1)

##
## Call:
## lm(formula = Gewicht ~ Groesse, data = lmExample)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.09  -5.06  -2.15   6.93  20.12
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -75.494     26.977   -2.80  0.0092 **
## Groesse       0.866       0.159    5.45  8.1e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.39 on 28 degrees of freedom
## Multiple R-squared:  0.515, Adjusted R-squared:  0.497
## F-statistic: 29.7 on 1 and 28 DF, p-value: 8.1e-06
```

3.

```
Koeffizienten
coef(modell1)

## (Intercept)      Groesse
##      -75.494       0.866
```

```
angepasste Werte
fitted(modell1)

##      1      2      3      4      5      6      7      8      9     10     11     12     13
## 66.87 69.73 85.22 72.33 72.84 86.57 75.71 60.77 65.77 67.86 82.32 74.84 75.19
## 14     15     16     17     18     19     20     21     22     23     24     25     26
## 72.68 66.91 87.19 76.03 54.69 77.79 67.63 62.47 69.83 62.84 65.41 66.31 57.11
## 27     28     29     30
## 78.98 73.05 61.86 82.58
```

4.

```
Residuen (Abstände zwischen beobachteten und angepassten Werten)
residuals(modell1)

##      1      2      3      4      5      6      7      8
## 1.7932 -4.9795  9.3202  7.1556  6.6689  7.4649  4.4367 -4.0350
##  9     10     11     12     13     14     15     16
## -5.7005 -6.1223 -7.0266 -3.3171 -13.8370 20.1177  9.6144 -10.5563
## 17     18     19     20     21     22     23     24
## -5.0818 -9.0227  7.0195 -3.1876 -0.6183 -2.2979 -3.5242 10.9888
## 25     26     27     28     29     30
## -4.8157 11.1837 -16.0849  4.3315 -2.0072  2.1197
```

5.

```
Summe quadratischer Abstände (RSS, deviance)
deviance(modell1)

## [1] 1972

mit R
confint(modell1)

##              2.5 %   97.5 %
## (Intercept) -130.7528 -20.235
## Groesse      0.5405    1.191
```

Konfidenzintervalle
Die Breite des Konfidenzintervalls hängt ab von:

- der Varianz der Daten
- der Stichprobengröße

Vorhersagen für Werte in linearer Regression

Funktion predict()

```
neueGroessen <- data.frame(  
  Groesse = c(160,175,190))  
  
predict(modell1, neueGroessen)  
  
##      1      2      3  
## 63.06 76.05 89.04
```

Vorhersage KIs für neuen Fall und Mittelwert

Vorhersage für neuen Fall

```
predict(modell1, neueGroessen, interval = "prediction")  
  
##      fit   lwr   upr  
## 1 63.06 45.31 80.81  
## 2 76.05 58.48 93.62  
## 3 89.04 70.34 107.74
```

Vorhersage für Mittelwert

```
predict(modell1, neueGroessen, interval = "confidence")  
  
##      fit   lwr   upr  
## 1 63.06 58.65 67.47  
## 2 76.05 72.44 79.66  
## 3 89.04 81.68 96.40
```

Logistische Regression

Hintergrund

- „verallgemeinertes lineares Modell
- Umfasst **binäre keine metrischen Zielgrößen**

Problem:

- Wahrscheinlichkeiten zwischen 0 und 1 (binäre Ausprägung)
→ jedoch bei linearer Regression zwischen $\pm \infty$
- **Normalverteilungsannahme und Varianzgleichheit bei binären Ausprägungen nicht erfüllt**
- Transformation der Wahrscheinlichkeiten notwendig mittels \ln zur Basis e

Transformation der Wahrscheinlichkeiten

- Odds und Wahrscheinlichkeiten sind 1:1 umwandelbar

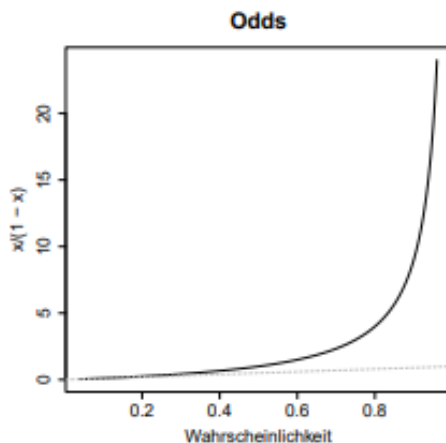
$$\text{W'keit zu Odds: } p \mapsto o = \frac{p}{1-p}$$

$$\text{Odds zu W'keit: } o \mapsto p = \frac{o}{1+o}$$

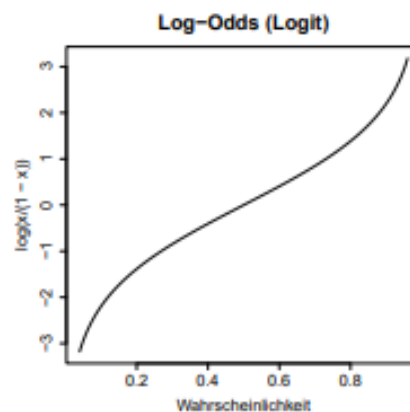
- Da Odds jedoch nur für positive reelle Zahlen gelten, wird die Transformation mittels Logarithmus in Log-Odds (Logit) durchgeführt

$$\text{logit} = \ln(o) = \ln\left(\frac{p}{1-p}\right) \in \mathbb{R}$$

Odds



Log-Odds (Logit)



→ Modellierung von Logit statt Wahrscheinlichkeiten

→ Logit in Abhängigkeit von Prädiktor X, lineare Abhängigkeit

$$\text{logit}_x = \beta_0 + \beta_1 \cdot x$$

Interpretation der logistischen Regression

- Analog zur linearen Regression
 - Steigt x um 1, änders sich das Logit um β_1
 - $\text{logit}_x = \ln(o_x) = \beta_0 + \beta_1 \cdot x$
 - Odds: $o_x = e^{\beta_0 + \beta_1 \cdot x}$
 - Odds o_x in Wahrscheinlichkeit p_x umwandeln via $p_x = \frac{o_x}{1+o_x}$

$$p_x = \frac{o_x}{1 + o_x} = \frac{e^{\beta_0 + \beta_1 \cdot x}}{1 + e^{\beta_0 + \beta_1 \cdot x}}$$

Interpretation des Koeffizienten (kategorialer Prädiktor)

- Zweistufig codiert (x = 0, und x=1 für jeweilige Ausprägungen)

$$\text{logit}_x = \ln(\text{odds}_x) = \beta_0 + \beta_1 \cdot x$$

$$x = 0 : \quad \text{logit}_{x=0} = \beta_0$$

$$x = 1 : \quad \text{logit}_{x=1} = \beta_0 + \beta_1$$

$$x = 0 : \quad \text{odds}_{x=0} = e^{\beta_0}$$

$$x = 1 : \quad \text{odds}_{x=1} = e^{\beta_0 + \beta_1} = e^{\beta_0} \cdot e^{\beta_1}$$

$$\text{OR} = \frac{\text{odds}_{x=1}}{\text{odds}_{x=0}} = \frac{e^{\beta_0} \cdot e^{\beta_1}}{e^{\beta_0}} = e^{\beta_1} = \text{OR zwischen den beiden Ausprägungen}$$

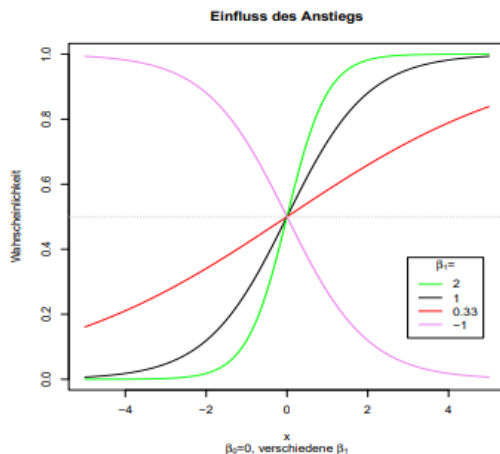
Interpretation des Koeffizienten (metrischer Prädiktor)

- Ändert sich x um eine Einheit, dann ändert sich:

Logit	additiv	$+\beta_1$
Odds	multiplikativ	$\cdot e^{\beta_1}$

- Erhöhung des Logit mit Zunahme des Prädiktors um eine Einheit
- e^{β_1} ist daher das OR, wenn sich der Prädiktor um eine Einheit erhöht

Auswirkungen verschiedener Anstiege



Modellanpassung an Daten

- Methode der kleinsten Quadrate nicht optimal (Varianz von Y nicht homogen)

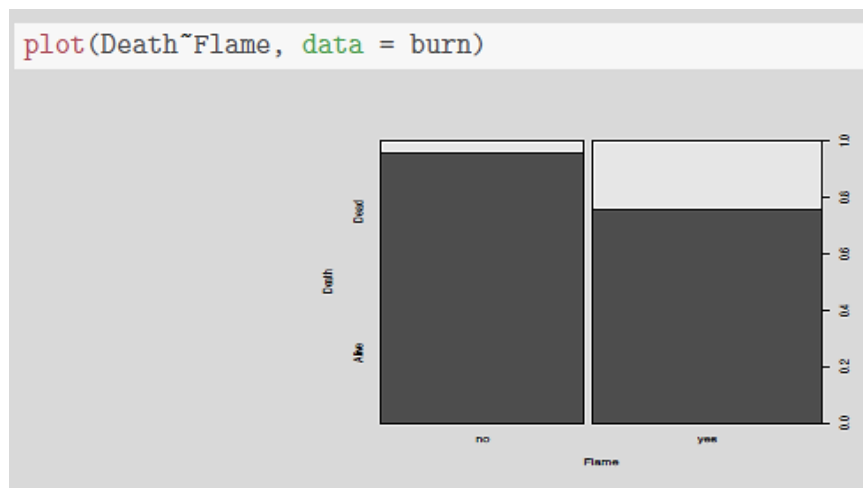
Maximum Likelihood (ML) Methode

- Überprüfung, welche Parameter die beobachteten Daten am wahrscheinlichsten
- ML liefert Punkt-Schätzung und Intervall-Schätzung über Standardfehler
- Weitere Erläuterungen

Logistische Regression – Umsetzung in R

1. Visualisierungen in R

Kategoriale Prädiktoren: `plot(Zielvariable~Prädiktor, data = xxx)`



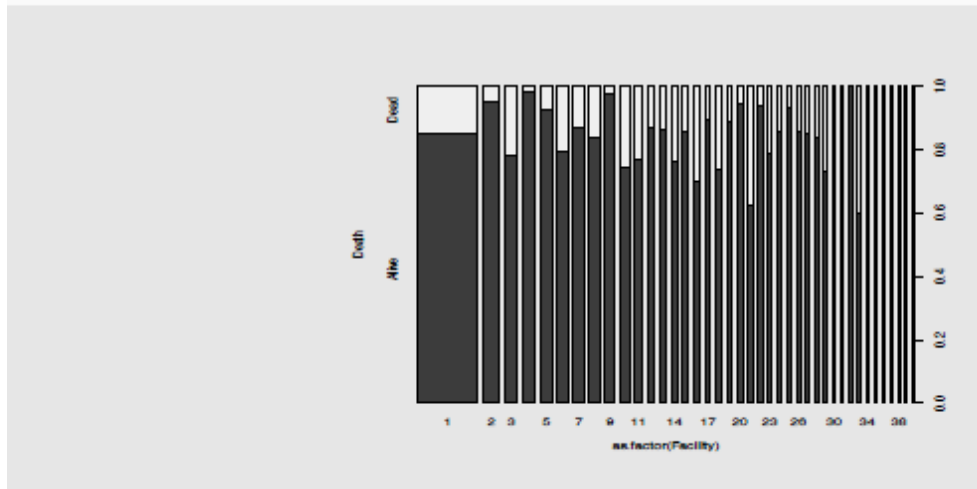
Metrische Prädiktoren: `cdplot()`



1. Visualisierungen in R

Wenn mehrere kategoriale Prädiktoren: mit `as.factor` arbeiten

```
# mit as.factor() kann sie als  
# kategoriale Groesse verwendet werden:  
plot(Death~as.factor(Facility), data = burn)
```



R: Einfache logistische Regression mit kategorialen Prädiktoren

- Logistische Regression gehört zur Familie der generalisierten linearen Modelle
 - R-Befehl: `glm()` und Zusatz: `family = binomial`

```
logmodel1 <- glm(Death~Flame, data = burn, family = binomial)  
summary(logmodel1)
```

```
##  
## Call:  
## glm(formula = Death ~ Flame, family = binomial, data = burn)  
##  
## Deviance Residuals:  
##      Min       1Q   Median       3Q      Max   
## -0.7510  -0.7510  -0.2946  -0.2946   2.5136   
##  
## Coefficients:  
##              Estimate Std. Error z value Pr(>|z|)      
## (Intercept)  -3.1157     0.2285 -13.637 < 2e-16 ***  
## Flameyes      1.9943     0.2498   7.984 1.42e-15 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for binomial family taken to be 1)  
##  
##      Null deviance: 845.42  on 999  degrees of freedom  
## Residual deviance: 755.46  on 998  degrees of freedom  
## AIC: 759.46
```

**Signifikanter
Zusammenhang zwischen
Flammen und Risiko zu
versterben**

R: Einfache logistische Regression mit kategorialen Prädiktoren

CAVE: Logit-Ebene

```
coef(logmodel1)

## (Intercept)    Flameyes
##   -3.115735     1.994308
```

Koeffizienten mittels coef() Befehl:

-Intercept = β^0 für die Referenzgruppe (Flames no)

-Flameyes = β^1 (Unterschied zwischen Flame yes und Referenzgruppe)

Odds-Ratio Ebene (exponieren)

```
exp(coef(logmodel1))

## (Intercept)    Flameyes
##   0.0443459     7.3471173
```

Koeffizienten mittels exp(coef()) Befehl:

-Intercept = Odds zu versterben für Personen in der Referenzkategorie

-Flameyes = Odds-Ratio, Risiko zu

versterben für Flameyes 7,3-fach erhöht

-Odds der zweiten Gruppe = Intercept (Odds Referenz) * Flameyes (Odds-Ratio)

R: Einfache logistische Regression mit metrischen Prädiktoren

```
logmodel2 <- glm(Death~TBSA, data = burn, family = binomial)
summary(logmodel2)

##
## Call:
## glm(formula = Death ~ TBSA, family = binomial, data = burn)
##
## Deviance Residuals:
##   Min       1Q   Median       3Q      Max
## -3.1709  -0.3976  -0.3112  -0.2745   2.5680
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.345107   0.175648  -19.04  <2e-16 ***
## TBSA         0.085367   0.006956   12.27  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

- Es besteht ein signifikanter Zusammenhang zwischen den Anteil der verbrannten Hautfläche und dem Risiko zu versterben
- Intercept(β^0) beschreibt Logit bei einem TBSA-Wert von 0
- TBSA (β^1) beschreibt den Anstieg der Geraden auf Logit-Ebene, erhöht sich der TBSA um eine Einheit, steigt das Logit um β^1 ein

R: Einfache logistische Regression mit metrischen Prädiktoren

- Erneut exponentieren, um Odds und Odds Ratios zu erhalten

```
exp(coef(logmodel2))  
  
## (Intercept)      TBSA  
## 0.03525643  1.08911645
```

Intercept = Odds für TBSA-Wert = 0

TBSA = Odds-Ratio, zwischen zwei Personen, die sich im TBSA-Wert um einer Einheit unterscheiden

Wichtig dafür: Gilt die Linearitätsannahme (Modelldiagnostik prüfen), dann ist das OR über alle Werte konstant

Überprüfung: OR, wenn Personen mehr als eine Einheit auseinanderliegen

- Dann muss der expnierte Koeffizient noch mit x-Einheiten potenziert werden
→ $e^{\text{Beta}1}$
- Beispiel: OR ermitteln für 10 TBSA-Einheiten Unterschied

Um das Odds Ratio zwischen zwei Patienten, die sich um 10 TBSA-Einheiten unterscheiden, zu ermitteln, wird der zweite Koeffizient zehnmal angewendet, d. h. $(e^{\beta_1})^{10}$ berechnet:

```
exp(coef(logmodel2)[2])^10  
  
##      TBSA  
## 2.348244
```

Die gleiche Zahl ergibt sich, wenn der Koeffizient zunächst mit 10 multipliziert und dann exponenziert wird ($e^{10 \cdot \beta_1}$):

```
exp(10*coef(logmodel2)[2])  
  
##      TBSA  
## 2.348244
```

Logits in Wahrscheinlichkeiten umrechnen

- Verwendung der logistischen Funktion

Grafische Darstellung Zusammenhang in Wahrscheinlichkeiten

1. Neue Spalte erstellen, damit Werte als 1 und 0 (TRUE/FALSE) codiert werden

```
burn$Death2 <- burn$Death=="Dead"
```

2. Streudiagramm mit neuer Spalte/Variable erstellen

```
plot(Death2 ~ TBSA, data = burn)
```

3. Koeffizienten (Beta0 und Beta1) aus dem angepassten Modell (mit Death2) entnehmen

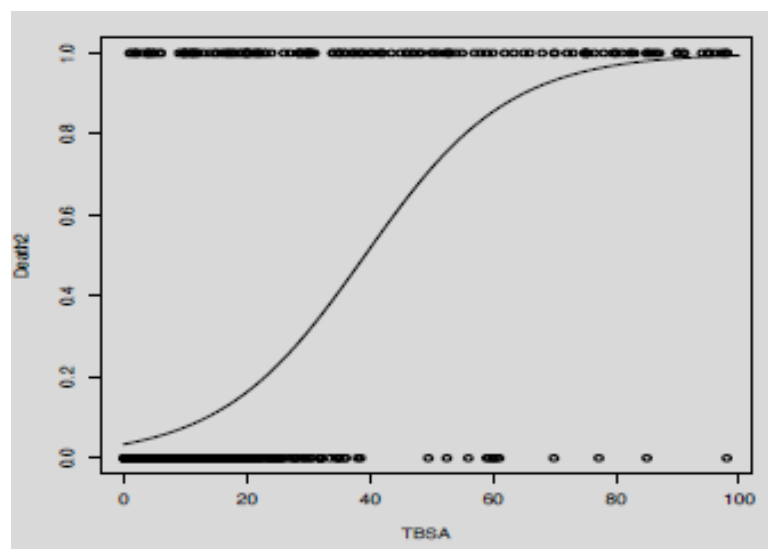
```
betas <- coef(logmodel2)
```

4. Kurve in das Streudiagramm einzeichnen: curve()
Logits als Wahrscheinlichkeiten umwandeln

```
curve( exp(betas[1] + betas[2]*x ) /  
       ( 1 + exp(betas[1] + betas[2]*x ) ),  
       from = 0, to = 100, add=TRUE )
```

Achseneinteilung:
Wahrscheinlichkeiten
von 0 bis 100

Damit die Kurve in den
obigen Plot geplottet
wird eingezeichnet wird



Logistische Regression mit mehreren Prädiktoren – additives Modell

- Analog dem glm() Befehl in der einfach logistischen Regression

```
logmodel3 <- glm(Death ~ Age + Gender + TBSA + Inhalation + Flame,
                family=binomial, data=burn)
summary(logmodel3)

##
## Call:
## glm(formula = Death ~ Age + Gender + TBSA + Inhalation + Flame,
##      family = binomial, data = burn)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.04961  -0.25696  -0.09212  -0.03670   2.52857
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -7.719152   0.677906 -11.387 < 2e-16 ***
## Age           0.078780   0.008174   9.638 < 2e-16 ***
## GenderMale   -0.275483   0.300888  -0.916    0.36
## TBSA         0.087889   0.009041   9.722 < 2e-16 ***
## Inhalationyes 1.380090   0.354293   3.895 9.81e-05 ***
## Flameyes     0.435599   0.347289   1.254    0.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

- Einige Prädiktoren nicht statistisch signifikant
- Jedoch: Likelihood-Ratio-Test (im drop1 Befehl) besser geeignet, als der Wald-Test auf Signifikanzen (im summary Befehl)
 - Wird berechnet über den Befehl test="Chisq"

```
drop1(logmodel3, test="Chisq")

## Single term deletions
##
## Model:
## Death ~ Age + Gender + TBSA + Inhalation + Flame
##           Df Deviance   AIC    LRT Pr(>Chi)
## <none>          341.66 353.66
## Age           1   499.19 509.19 157.526 < 2.2e-16 ***
## Gender        1   342.50 352.50   0.833   0.3615
## TBSA          1   521.32 531.32 179.660 < 2.2e-16 ***
## Inhalation    1   356.87 366.87  15.211 9.616e-05 ***
## Flame         1   343.28 353.28   1.614   0.2040
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Gender und Flame auch hier nicht statistisch signifikant
- ➔ Backward selection sinnvoll zur Modellanpassung

Backward selection für logistische Regression

```
logmodel4 <- update(fm.burn.1, . ~ . - Gender)
logmodel4

##
## Call: glm(formula = Death ~ Age + TBSA + Inhalation + Flame, family = binomial,
## data = burn)
##
## Coefficients:
## (Intercept)      Age      TBSA Inhalationyes  Flameyes
## -7.95897      0.07989      0.08740      1.38280      0.44376
##
## Degrees of Freedom: 999 Total (i.e. Null); 995 Residual
## Null Deviance:      845.4
## Residual Deviance: 342.5 AIC: 352.5
```

- Gender entfernen und anschließend erneute Prädiktorenbetrachtung

```
drop1(logmodel4, test="Chisq")

## Single term deletions
##
## Model:
## Death ~ Age + TBSA + Inhalation + Flame
##          Df Deviance   AIC    LRT Pr(>Chi)
## <none>          342.50 352.50
## Age           1   506.01 514.01 163.514 < 2e-16 ***
## TBSA          1   521.33 529.33 178.831 < 2e-16 ***
## Inhalation    1   357.79 365.79 15.294 9.2e-05 ***
## Flame         1   344.18 352.18 1.683 0.1946
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Flame weiterhin statistisch nicht signifikant (obwohl diese im einfachen Modell signifikant war)
→ trägt nicht zur Erklärung der Zielgröße bei gemeinsamer Betrachtung mit anderen Einflussgrößen bei → ebenfalls entfernen

```
logmodel5 <- update(logmodel4, . ~ . - Flame)
logmodel5

##
## Call: glm(formula = Death ~ Age + TBSA + Inhalation, family = binomial,
## data = burn)
##
## Coefficients:
## (Intercept)      Age      TBSA Inhalationyes
## -7.74279      0.08089      0.08885      1.49971
##
## Degrees of Freedom: 999 Total (i.e. Null); 996 Residual
## Null Deviance:      845.4
## Residual Deviance: 344.2 AIC: 352.2
```

Backward selection für logistische Regression

```
drop1(logmodel5, test="Chisq")

## Single term deletions
##
## Model:
## Death ~ Age + TBSA + Inhalation
##           Df Deviance   AIC    LRT Pr(>Chi)
## <none>           344.18 352.18
## Age           1   520.73 526.73 176.556 < 2.2e-16 ***
## TBSA          1   533.54 539.54 189.361 < 2.2e-16 ***
## Inhalation    1   362.98 368.98  18.798 1.453e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

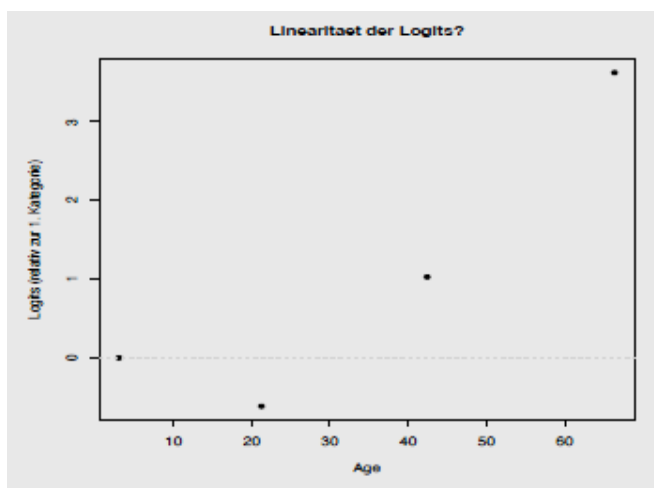
- Keine nicht-signifikanten Koeffizienten mehr vorhanden
- Vorhersagekraft und Adäquatheit des Modells sollte untersucht werden
- Modell könnte noch um mögliche Interaktionen zwischen Einflussgrößen ergänzt werden

Modelldiagnostik logistische Regression – metrische Prädiktoren

Linearitätsüberprüfung

- Mittels R-Befehl `testLinearity()` → eine Grafik eröffnet sich

```
testLinearity(data = burn, variable = "Age", model = Modellname )
```



- Alter wird anhand Quantile in 4 gleich große Gruppen eingeteilt (kategorisiert anhand der Quantile)
- Erste Kategorie (jüngsten) ist Referenzkategorie mit Logit = 0
- Anderen Punkte stellen die Mittelwerte jeder Alterskategorie dar
- Liegen die Punkt etwa auf einer Gerade, ist die Linearitätsannahme gerechtfertigt
 - Die metrische Variable kann verwendet werden, jedoch:
 - Eine zusätzliche Variable sollte eingeführt werden, die kleinere Kinder vom Rest unterscheiden

Einführung eines neuen Faktors in die Regression

1. Erstellung eines neuen Faktors: factor()

Hintergrund: Junge Patienten zeigen einen untypischen Alterseffekt

→ Faktor: Kinder kleiner gleich 10 Jahre

Mithilfe der Funktion `factor` legen Sie eine neue Variable an, die Kinder mit einem Alter von kleiner oder gleich 10 Jahren vom Rest unterscheidet:

```
burn$Child <- factor(burn$Age <= 10, levels=c(FALSE,TRUE),
                    labels=c("no", "yes"))
```

2. Erweiterung des Modells um den zusätzlichen Faktor

```
logmodel6 <- update(logmodel5, .~. + Child)
```

3. Erneut Überprüfung mittel `drop1()` Befehl

```
drop1(logmodel6, test="Chisq")

## Single term deletions
##
## Model:
## Death ~ Age + TBSA + Inhalation + Child
##           Df Deviance   AIC    LRT Pr(>Chi)
## <none>          330.03 340.03
## Age           1   497.27 505.27 167.240 < 2.2e-16 ***
## TBSA          1   525.28 533.28 195.256 < 2.2e-16 ***
## Inhalation    1   347.09 355.09  17.062 3.618e-05 ***
## Child         1   344.18 352.18  14.152 0.0001686 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Indikatoren stat.
signifikant

4. Umwandlung Logits in Odds Ratios `exp(coef)`

```
exp(coef(logmodel6))

##      (Intercept)           Age          TBSA Inhalationyes      Childyes
## 1.263135e-04 1.104047e+00 1.096294e+00 4.277487e+00 1.816696e+01
```

- Patienten unter 10 Jahre haben ein höheres Risiko zu versterben
- Jedoch: sinnvolle Interpretation erschwert, da das Alter auch noch als metrische Größe eingeschlossen (besser: nur mit 4 Alterskategorien auswerten)

Faktorielle statt metrische Auswertung in der Regression

5. Alter in vier Altersgruppen darstellen

a) Quartile berechnen:
quantile()

```
ageQuantiles <- quantile(burn$Age)
ageQuantiles

##      0%      25%      50%      75%     100%
##  0.100 10.850 31.950 51.225 89.700
```

b) Daten anhand Quartile in vier Gruppen einteilen cut()

```
burn$Age_cat <- factor(
  cut(x=burn$Age, breaks=ageQuantiles, include.lowest=T),
  labels= c(1,2,3,4))
```

Namen der Intervalle

Grenze der
Intervalle

Damit wird der kleinste
Wert de, ersten Intervall
zugeordnet

6. Modell mit kategoriale statt metrischer Variable

```
logModel_Age_cat <- update(logmodel5, .~. - Age + Age_cat)
summary(logModel_Age_cat)

##
## Call:
## glm(formula = Death ~ TBSA + Inhalation + Age_cat, family = binomial,
##      data = burn)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.00584  -0.21265  -0.11048  -0.07107   2.62207
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.706188   0.588121  -9.702  < 2e-16 ***
## TBSA         0.085384   0.008499  10.046  < 2e-16 ***
## Inhalationyes 1.465006   0.331692   4.417 1.00e-05 ***
## Age_cat2     -0.615312   0.762303  -0.807  0.4196
## Age_cat3      1.028316   0.601499   1.710  0.0873 .
## Age_cat4      3.624702   0.560463   6.467 9.97e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

→ Nur die oberste Alterskategorie weist einen signifikanten Unterschied zur Basiskategorie (Kinder) auf

Modellgüte: Vorhersagekraft – Hosme-Lemeshow Anpassungstest

„Wie gut passt die Vorhersage zu den Daten?“

Hosmer-Lemeshow-Anpassungstest

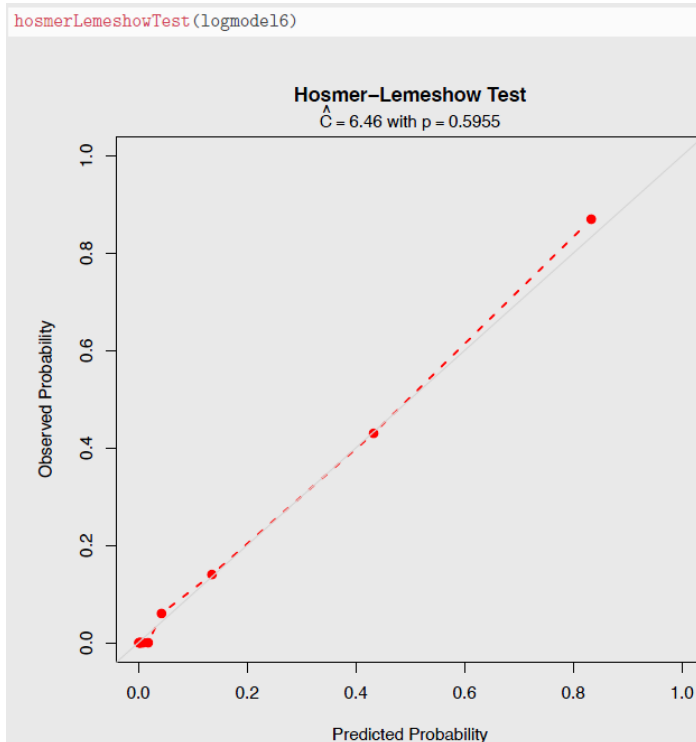
- Gruppieren die Fälle nach der *vorhergesagten* Wahrscheinlichkeit
- Meistens $g = 10$ Gruppen (mindestens $g \geq 6$)
- H_0 : beobachtete Wahrscheinlichkeiten \sim vorhergesagte Wahrscheinlichkeiten in den Gruppen
- Kleiner p-Wert heißt schlechte Anpassung

Grundidee:

- Übereinstimmung der vorhergesagten mit tatsächlich beobachteten Zielvariable
 - Einteilung der vorhergesagten Zielvariable in mindestens 6 gleichgroße Gruppen
 - Überprüfung ob die beobachtete mit der mittleren vorhergesagten Wahrscheinlichkeit übereinstimmen
- Nullhypothese
 - Die vom Modell berechneten Wahrscheinlichkeiten passen zu den beobachteten Wahrscheinlichkeiten in den Daten

Anwendung in R

- Befehl: `hosmerLemeshowTest(model)`



Grafische Betrachtung:

-vorhergesagte und beobachtete Wahrscheinlichkeiten in den Gruppen stimmen gut überein (weitestgehend auf der Winkelhalbierenden)

p-Wert:

-bei 0,6 d.h. die Nullhypothese müsse abgelehnt werden (keine signifikante Abweichung zwischen vorhergesagten und beobachteten Wahrscheinlichkeiten)
→ Modell passt zu Daten

Klassifizierung von logistischen Regressionsmodellen

- Logistische Modell werden of zur Wahrscheinlichkeitsvorhersage gemacht
 - Mithilfe eines Schwellenwert („Cutoff“) erhält man eine Klassifizierung der Patienten
 - Zur Bestimmung der Sensitivität und Spezifität des Modells
1. Nutzung der predict-Funktion zur Vorhersage der Wahrscheinlichkeiten für Patienten → Erstellung neuer Spalte

```
burns$pred <- predict(logmodel6, type="response")
```

⌵
Damit Wahrscheinlichkeiten und
keine Logits ausgerechnet werden

2. Wahrscheinlichkeiten in zwei Gruppen einteilen (>0.5 = Tod)

```
burns$pred_cat <- factor(burns$pred > .5, levels = c(FALSE,TRUE),  
                          labels = c("Alive", "Dead"))
```

3. Vierfeldertafel erstellen (stellt Zusammenhang zwischen vorhergesagten Status und Beobachtungen dar)

```
vierfeldertafel <- xtabs(~Death + pred_cat, data=burn)  
vierfeldertafel  
  
##      pred_cat  
## Death  Alive Dead  
##  Alive  823   27  
##  Dead   46  104
```

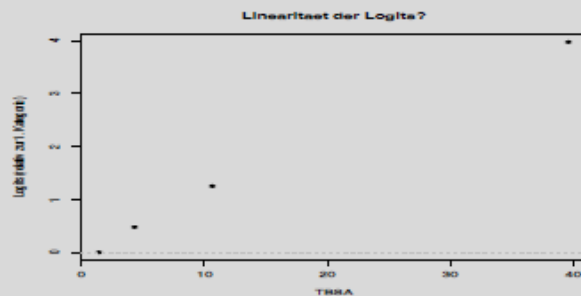
Interpretation:

- Von 150 Toten wurden 104 richtig klassifiziert (Sensitivität 70%)
- Von 850 Überlebenden wurden 823 richtig klassifiziert (Spezifität: 97%)
 - Insgesamt 927/1000 richtig klassifiziert
- Um ein Vorhersagemodell realistisch zu bewerten, sollte dies mit neuen Testdaten ohne Modellanpassung überprüft werden

Warum?

Analog zum Vorgehen für das Alter wird die zweite metrische Größe TBSA überprüft:

```
testLinearity(data = burn, variable = "TBSA", model = logmodel6 )
```



Ob die Linearitätsannahme hier verworfen werden sollte, ist eine subjektive Entscheidung. Die Punkte bilden keine optimale Gerade: Bei einem großen Anteil verbrannter Haut ist das geschätzte Logit kleiner als durch die Linearitätsannahme erwartet. Es könnte versucht werden, die Daten zu transformieren, bspw. mit der Wurzel oder dem Logarithmus. Alternativ könnte statt der metrischen Größe auch die kategorisierte Größe verwendet werden. Wir wollen es hier bei der ursprünglichen metrischen Variable belassen.

Meta-Analyse

Hintergrund

- Anzahl wissenschaftlicher Publikationen riesig und unübersichtlich
 - Bedarf Informationsverdichtung und -bewertung von Forschungsergebnissen
- Metaanalysen = quantitative Ergebniszusammenfassungen
 - An Kriterien empirischer Forschung orientierte Methode zur Integration der Ergebnisse empirischer Untersuchungen sowie Analyse der Variabilität der Ergebnisse
 - Grundlage von Pearson
 - Methodische Weiterentwicklung bis heute
 - Benötigen immer quantitativ empirische Ergebnisse

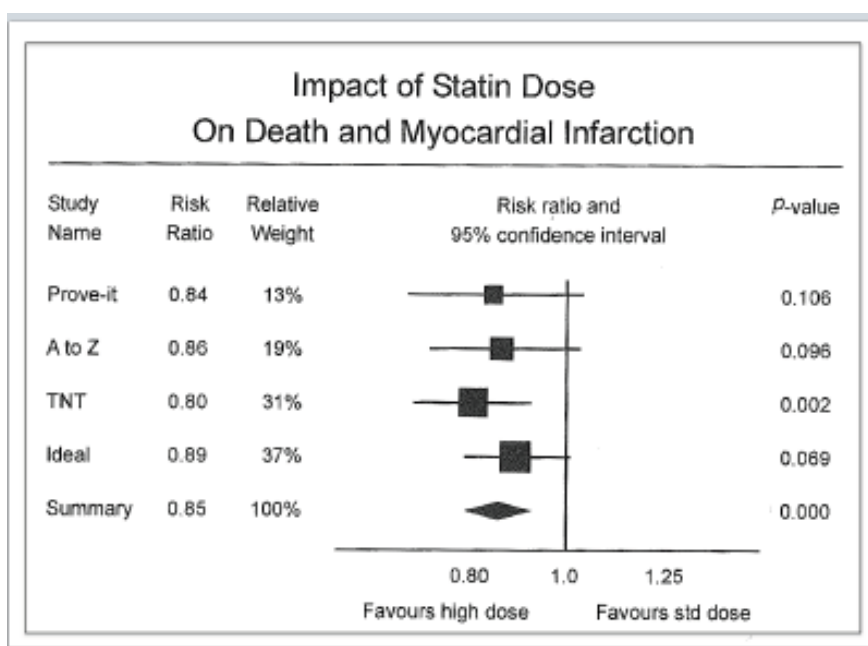
Aufbau Metaanalyse

Prototypischer Aufbau einer Metaanalyse

Vorgehensweise und Ablauf einer Metaanalyse sind mit dem entsprechenden Vorgehen in Einzeluntersuchungen vergleichbar

- Konkretisierung des Forschungsvorhabens
- Sammlung relevanter Untersuchungen
- Codierung und Bewertung der Untersuchungen
- Datenanalyse
- Darstellung und Interpretation der Ergebnisse

Elemente Metaanalyse – Forest Plot



Effektstärken

Geeignete Zielgröße abhängig von:

- Vergleichbarkeit
- Berechenbarkeit
- Interpretierbarkeit der Effektstärken

Effektstärkenbeschreibung durch Mittelwerte – unstandardisierte Differenz

1. Differenzen/standardisierte **Differenzen der Mittelwerte** als Effektgröße
 - Geeignet bei Messwerte auf üblichen Skalen (z.B. Blutdruck, Benötigung Abitur)

aus den Stichprobenmittelwerte \bar{X}_1 und \bar{X}_2 zweier unabhängiger Stichproben (bspw. Treatment and Control) lässt sich der Schätzer für die Differenz D berechnen:

$$D = \bar{X}_1 - \bar{X}_2$$

2. Gepoolte Standardabweichungen

$$S_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

3. Varianz

$$V_D = \frac{n_1 + n_2}{n_1 n_2} S_{\text{pooled}}^2$$

Berechnung Varianz & Standardabweichung: Einzel- und Mittelwerte

Zufallsvariable	Einzelmessung x	Mittelwert \bar{x}
Varianz	$\begin{aligned} \text{Var}[x] &= s_x^2 \\ &= \frac{1}{n-1} \sum_i^n (x_i - \bar{x})^2 \end{aligned}$	$\begin{aligned} \text{Var}[\bar{x}] &= s_{\bar{x}}^2 \\ &= \frac{s_x^2}{n} \end{aligned}$
Standardabweichung	$s_x = \sqrt{\text{Var}[x]}$	$\begin{aligned} s_{\bar{x}} &= \sqrt{\text{Var}[\bar{x}]} \\ &= \frac{s_x}{\sqrt{n}} \end{aligned}$

Beispiel: unstandardisierte Differenzen

Beispiel: unstandardisierten Differenz von Mittelwerten

- Studie mit zwei Gruppen: Mittelwerte $\bar{X}_1 = 103.0$ und $\bar{X}_2 = 100.0$ mit Standardabweichung $S_1 = 5.5$ und $S_2 = 4.5$ und Stichprobenumfang $n_1 = n_2 = 50$
- Mittelwertdifferenz

$$D = 103.0 - 100.0 = 3.0$$

- gepoolte Standardabweichung

$$S_{\text{pooled}} = \sqrt{\frac{(50 - 1)5.5^2 + (50 - 1)4.5^2}{50 + 50 - 2}} = 5.0249$$

- Varianz

$$V_D = \frac{50 + 50}{50 \cdot 50} \cdot 5.0249^2 = 1.01$$

Standardisierte Differenz von Mittelwerten

- Bei Studien mit unterschiedlichen Skalen geeignet
- **Normierung der Differenz der Mittelwerte mit der Standardabweichung**
 - Annahme: Standardabweichung für beide Gruppen gleich
 - Daraus ergibt sich folgender **die standardisierte Mittelwertdifferenz**:

standardisierte Differenz der Mittelwerte in einer Studie:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S_{\text{within}}}$$

mit

$$S_{\text{within}} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

- **Sowie Varianz der Mittelwertdifferenz:**

$$V_d = \frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2)}$$

- der erste Term beschreibt die Unsicherheit in der Schätzung der mittleren Differenz, der zweite Term die Unsicherheit von S_{within}
- der Standardfehler von d ist definiert als

$$SE_d = \sqrt{V_d}$$

Biasbereinigung Mittelwertdifferenz: Hedges' g

- für kleine Fallzahlen überschätzt d den wahren Wert $\delta \rightarrow$ bias-korrigierter Schätzer: Hedges' g

$$g = J \cdot d$$

und

$$V_g = J^2 \cdot V_d$$

- Korrekturfaktor

$$J = 1 - \frac{3}{4df - 1}$$

mit Freiheitsgraden $df = n_1 + n_2 - 2$

Varianz & 95% KIs Studien

Varianz als Maß der Genauigkeit

- Varianz V_Y wird für jede Effektstärke Y unterschiedlich ermittelt
- Berechnung des Standardfehler der Effektstärke

$$SE_Y = \sqrt{V_Y}$$

- damit können die Grenzen des 95% Konfidenzintervalls berechnet werden (Annahme: Normalverteilung)

$$LL_Y = Y - 1.96 \cdot SE_Y$$

$$UL_Y = Y + 1.96 \cdot SE_Y$$

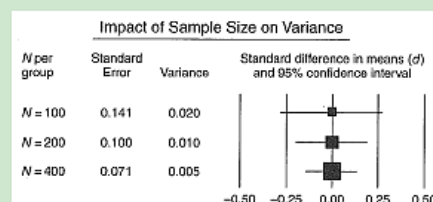
Beeinflussung der Genauigkeit eines Schätzer

- I. Stichprobenumfang = zentraler Faktor
- II. Studiendesign (matched groups vs. independent groups)

Geht beides in die Gewichtung der Meta-Analyse ein

\rightarrow Studien mit höherer Genauigkeit werden stärker berücksichtigt

- 3 Studien ($n = 100, 200, 400$) mit Mittelwertvergleich in je zwei unabhängigen Gruppen
- standardisierte Mittelwertdifferenz $d = 0.0$



- Fläche der Box ist invers proportional zur Varianz
- CI ist proportional zum Standardfehler

Parameter	Beschreibung
Q	Summe der quadratischen Abweichungen der Effektstärke vom Summary-Effect
df	Anzahl der Freiheitsgrade
$Q - df$ (excess variation)	entspricht dem Anteil der Variabilität, der durch die Heterogenität der wahren Effektstärken (nicht der Stichprobenfehler) verursacht wird
p -Wert für (Q, df)	der p -Wert entspricht der Wahrscheinlichkeit, dass die beobachtete Heterogenität unter der Annahme einer gleichen, wahren Effektstärke zufällig zustande gekommen ist.
I^2	relatives Maß für den Anteil der excess variation an der gesamten Dispersion
τ^2	entspricht T^2 , Schätzer für die Varianz der wahren Effektstärke
H^2	Verhältnis der beobachteten Heterogenität zur erwarteten Heterogenität unter der Annahme einer gleichen, wahren Effektstärke für alle Studien ($H^2 = Q/df$)

Vergleich beider Modelle

	Fixed effect model	Random effect model
Summary Effekt	Identische „wahre“ Effektstärke für alle Studien	Schätzung des Mittelwertes einer Verteilung von Effekten
	Stichprobenfehler führen zur Variabilität der Resultate	Jede Studie Informationsgehalt bzgl. der Verteilung
	Kleinere Studien = geringere Gewichtung	Kleine Studien höhere Gewichtung
	Große Studien = großer Einfluss auf summary Effekt	Gewichtungen balancierter, als im fixed-effect Modell
Konfidenzintervalle	-	Besitzt eine zusätzliche Quelle der Unsicherheit = between-study variance
	Kleinerer Standardfehler und KI	Größerer Standardfehler und KI
Modellauswahl	Annahme einer gemeinsamer Effektstärke gegeben	Für unabhängige Studien verschiedener Wissenschaftler
		Generalisierbar für andere Populationen

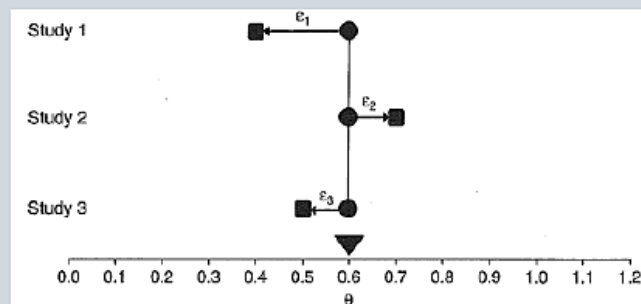
Fixed-effect & random-effect Modell

1. Fixed effect Modell

- Annahme: Es gibt eine **wahre Effektstärke**, alle **Unterschiede der beobachteten Effektstärken sind Stichprobenfehler**
- **Alle Studien eine gemeinsame wahre Effektstärke**

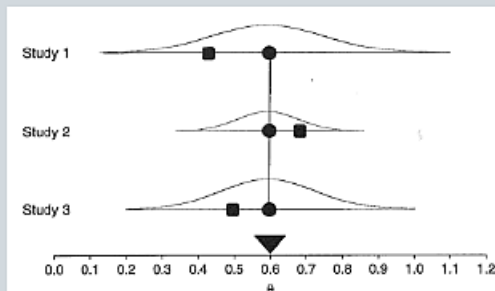
beobachtete Effektstärken

Variationen zwischen den Studien beruhen nur auf zufälligen Fehlern ϵ_j durch den endlichen Stichprobenumfang



beobachtete Effektstärken

- beobachtete Effektstärke $Y_j = \theta + \epsilon_j$
- Fehler ϵ_j ist zufällig für jede Studie, allerdings kann die Verteilung der Fehler durch eine Normalverteilung approximiert werden
- Breite der Normalverteilung skaliert mit dem Standardfehler



Berechnungsgrundlagen fixed-effect Modell

- **Studiengewichtung - Genauere Studien mit höherem Gewicht:**

$$W_i = \frac{1}{V_{Y_i}}$$

- das gewichtete Mittel M berechnet sich nach

$$M = \frac{\sum_{i=1}^k W_i Y_i}{\sum_{i=1}^k W_i}$$

1. Gemittelte Effektstärke – Summary effect: Varianz ermitteln

$$V_M = \frac{1}{\sum_{i=1}^k W_i}$$

2. Standardfehler der Effektstärke

$$SE_M = \sqrt{V_M}$$

3. 95% Konfidenzintervalle

$$LL_M = M - 1.96 \cdot SE_M$$

$$UL_M = M + 1.96 \cdot SE_M$$

4. Test gegen die Nullhypothese (z-Wert)

Test gegen die Nullhypothese, dass der wahre Effekt $\theta = 0$ ist
z-Wert berechnen

$$z = \frac{M}{SE_M}$$

für einen zweiseitigen Test ergibt sich

$$p = 2(1 - \Phi(|z|))$$

wobei Φ der kumulativen Standardnormalverteilung entspricht

1. Allgemeine Vorbereitung

1. Metafor-Package laden

```
install.packages("metafor")
```

```
library(metafor)
```

2. Daten/Datentabellen einlesen

a. Arbeitsverzeichnis anlegen

b. Datentabellen einlesen

```
# Setzen des aktuellen Arbeitsverzeichnisses
setwd("[Pfad_einfügen]")

# read.table = Einlesen von Daten
# header = Tabellenköpfe (nicht) übernehmen
# sep = Angabe des Daten-Separators, hier = Tabulator

smd_data <- read.table("SMD_Data.txt",header=TRUE, sep="\t")
b_data <- read.table("B_Data.txt",header=TRUE, sep="\t")
```

3. Berechnung der Effektstärken und zugehörigen Varianzen: escalc()

```
calc_smd <- escalc(measure="SMD"
  ,m1i=smd_data$Treated_Mean
  ,sd1i=smd_data$Treated_SD
  ,n1i=smd_data$Treated_N
  ,m2i=smd_data$Control_Mean
  ,sd2i=smd_data$Control_SD
  ,n2i=smd_data$Control_N
  ,data=smd_data)
```

CAVE: escalc() führt keine Biaskorrektur nach Hedge's g durch (manuell)

```
# Berechnung des Korrekturfaktors
J <- (1-3/(4*(calc_smd$Treated_N+calc_smd$Control_N-2)-1))
# Überschreiben der alten Varianz-Werte
calc_smd$vi <- J^2*calc_smd$vi
# Anzeigen der benannten Spalten
calc_smd[,c("Study", "yi", "vi")]

##      Study    yi    vi
## 1 Carroll 0.0945 0.0329
## 2   Grant 0.2774 0.0307
## 3    Peck 0.3665 0.0499
## 4   Donat 0.6644 0.0105
## 5 Stewart 0.4618 0.0426
## 6   Young 0.1852 0.0234
```

2. Modelldurchführung

- Analyse mit dem: `rma`-Befehl()
 - Annahme: Stichprobenfehler durch Normalverteilung approximiert
 - Method="FE" für fixed-effect model

```
# FE-Modell für kontinuierliche Daten
res_smd.FE <- rma(yi, vi, data=calc_smd, method="FE", slab=Study)

# Anzeige FE-Modell
res_smd.FE

##
## Fixed-Effects Model (k = 6)
##
## I^2 (total heterogeneity / total variability): 58.35%
## H^2 (total variability / sampling variability): 2.40
##
## Test for Heterogeneity:
## Q(df = 5) = 12.0047, p-val = 0.0347
##
## Model Results:
##
## estimate      se      zval      pval      ci.lb      ci.ub
## 0.4143  0.0640  6.4747  <.0001  0.2889  0.5397  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Estimate = gewichtetes Mittel Effektstärken
- se = Standardabweichung der gemittelten Effektstärken
- zval/pval = z-Wert/p-Wert

Darstellung der relativen Gewichte `weights()` und `summary effects predict()`

```
weights(res_smd.FE)

## Carroll Grant Peck Donat Stewart Young
## 12.426289 13.335105 8.210527 38.946581 9.600002 17.481496

predict(res_smd.FE)

##
## pred      se ci.lb ci.ub
## 0.4143 0.0640 0.2889 0.5397
```

3. Erstellen eines Forest Plots

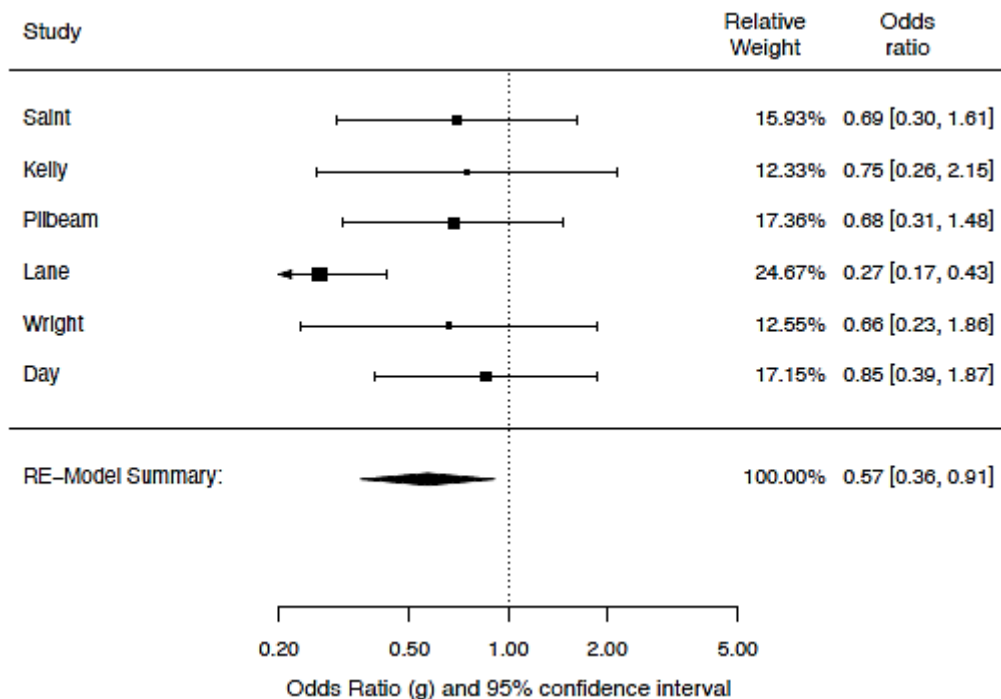
- Forest()

```
# Grundeinstellung Grafikparameter
par(mar=c(6,4,1,2),cex=.9, font=1)

# Random-Effects-Model mit binären Daten
forest(res_b.RE,showweights = TRUE,digits=2
       ,xlim=c(-3.5,3.5)
       ,ylim=c(-3,9)
       ,refline=log(1)
       ,at=log(c(0.2,0.5,1,2,5)) # Einteilung x-Achse
       ,xlab="Odds Ratio (g) and 95% confidence interval"
       ,mlab="RE-Model Summary:"
       ,atransf=exp) # logarithm. Transformierung x-Achse

# Spaltenbeschriftungen
par(font=1)
text(c(-3.2,1.8,2.8), 7.7, c("Study","Relative\nWeight"
                             ,"Odds\nratio"))
```

→ Weitere Informationen wie T^2 oder I^2 müssen zusätzlich ergänzt werden



Rechenbeispiel Fixed effect model

1. Bestehende Studien inkl. Mittleren Effektschätzer und SD

Zusammenfassung der Studien mittels fixed-effect Modell

Study	Treated			Control		
	Mean	SD	n	Mean	SD	n
Carroll	94	22	60	92	20	60
Grant	98	21	65	92	22	65
Peck	98	28	40	88	26	40
Donat	94	19	200	82	17	200
Stewart	98	21	50	88	22	45
Young	96	21	85	92	22	85

2. Berechnung standardisierte Differenz Mittelwerte

- Berechnung der standardisierten Differenz der Mittelwerte

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S_{\text{within}}}, \quad S_{\text{within}} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

- für die *Carroll* Studie ergibt sich

$$S_{\text{within}} = \sqrt{\frac{(60 - 1)22^2 + (60 - 1)20^2}{60 + 60 - 2}} = 21.0238$$

und

$$d = \frac{94 - 92}{21.0238} = 0.0951$$

3. Berechnung Varianz und Korrekturfaktor J

- Varianz

$$V_d = \frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2)} = \frac{60 + 60}{60 \cdot 60} + \frac{0.0951^2}{2(60 + 60)} = 0.0334$$

- der Korrekturfaktor *J* ergibt sich zu

$$J = 1 - \frac{3}{4df - 1} = 1 - \frac{3}{4 \cdot 118 - 1} = 0.9936$$

4. Hedges' g (biaskorrigierte Mittelwertdifferenz) und deren Varianz

damit berechnet sich die bias-korrigierte, standardisierte Mittelwertdifferenz (Hedges' g) und deren Varianz

$$g = J \cdot d = 0.9936 \cdot 0.0951 = 0.0945$$

und

$$V_g = J^2 \cdot V_d = 0.9936^2 \cdot 0.0334 = 0.0329$$

das Gewicht der Studien ergibt sich aus der inversen Varianz

$$W = \frac{1}{V_Y} = \frac{1}{0.0329} = 30.3515$$

Study	Effect size γ	Variance Within V_Y	Weight W	Calculated quantities		
				$W\gamma$	$W\gamma^2$	W^2
Carroll	0.095	0.033	30.352	2.869	0.271	921.214
Grant	0.277	0.031	32.568	9.033	2.505	1060.682
Peck	0.367	0.050	20.048	7.349	2.694	401.931
Donat	0.664	0.011	95.111	63.190	41.983	9046.013
Stewart	0.462	0.043	23.439	10.824	4.999	549.370
Young	0.185	0.023	42.698	7.906	1.464	1823.115
Sum			244.215	101.171	53.915	13802.325

Study	Effect size γ	Variance Within V_Y	Weight W	Calculated quantities		
				$W\gamma$	$W\gamma^2$	W^2
Carroll	0.095	0.033	30.352	2.869	0.271	921.214
Grant	0.277	0.031	32.568	9.033	2.505	1060.682
Peck	0.367	0.050	20.048	7.349	2.694	401.931

5. Gewichtetes Mittel und die Varianz

- das gewichtete Mittel ergibt sich aus

$$M = \frac{\sum_{i=1}^k W_i Y_i}{\sum_{i=1}^k W_i}$$

und die Varianz der gemittelten Effektstärke

$$V_M = \frac{1}{\sum_{i=1}^k W_i}$$

- im konkreten Fall:

$$M = \frac{101.171}{244.215} = 0.4143$$

und

$$V_M = \frac{1}{244.215} = 0.0041$$

6. Standardfehler und 95%KI

Rechenbeispiel

- aus der Varianz errechnet sich der Standardfehler

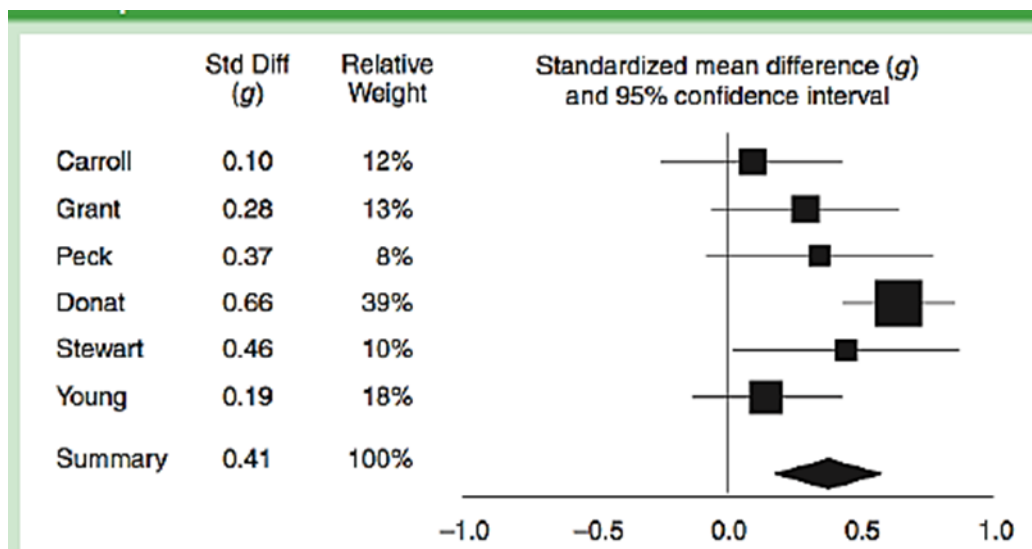
$$SE_M = \sqrt{V_M} = \sqrt{0.0041} = 0.0640$$

- für die Grenzen des 95%-Konfidenzintervalles ergibt sich

$$LL_M = M - 1.96 \cdot SE_M = 0.4143 - 1.96 \cdot 0.0640 = 0.2889$$

$$UL_M = M + 1.96 \cdot SE_M = 0.4143 + 1.96 \cdot 0.0640 = 0.5397$$

7. Ergebnisse



Random effects model

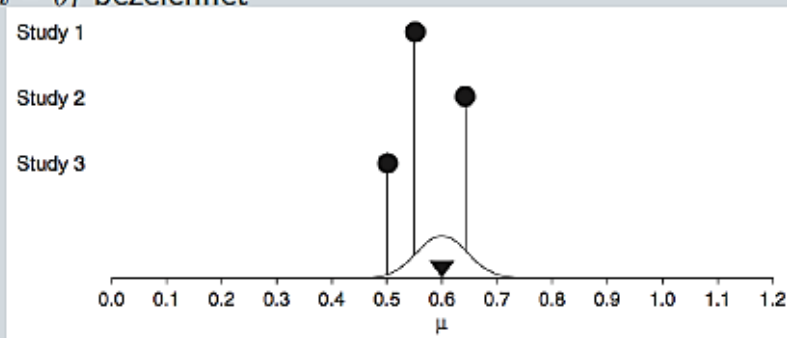
Allgemeine Annahme:

- **Effektstärken sind ähnlich, jedoch nicht identisch**
 - Zusammensetzungen der Stichproben führen zu unterschiedlichen Effektstärken
 - Individuelle Effektstärken sind um einen Mittelwert normalverteilt

Effektstärken

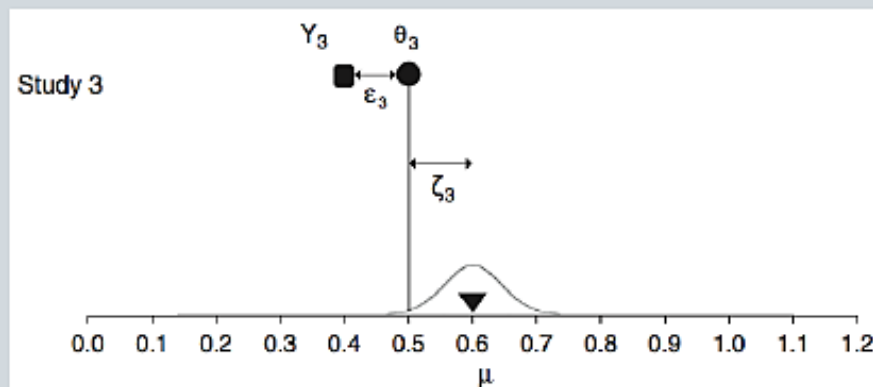
wahre Effektstärken

- für Studien mit unendlicher Stichprobenanzahl stellen die Effektstärken (hier: θ_1, θ_2 und θ_3) Realisierung aus der Verteilung der wahren Effekte θ dar
- die Abweichungen vom Mittelwert der wahren Effekte μ wird mit $\zeta_i = \mu - \theta_i$ bezeichnet



beobachtete Effektstärken

durch den endlichen Stichprobenumfang in jeder Studie gibt es einen zusätzlichen Stichprobenfehler ϵ_i

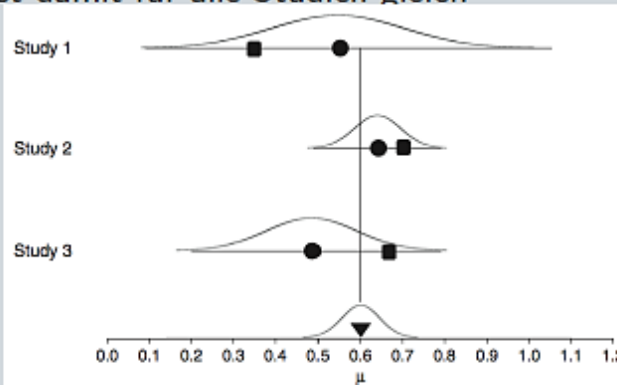


beobachtete Effektstärken

- beobachtete Effektstärke Y_i ergibt sich als Summe der Abweichungen ζ_i und ϵ_i :

$$Y_i = \mu + \zeta_i + \epsilon_i$$

- die Verteilung der ζ_i wird durch die Varianz τ^2 beschrieben; die Varianz τ^2 ist damit für alle Studien gleich



Varianz

Varianz der Studien

- für bekannte, wahre Effektstärken θ_i und eine unendliche Anzahl von Studien könnte man die Varianz τ^2 bestimmen
- Abschätzung von τ^2 mit der Methode von DerSimonian und Laird

$$T^2 = \frac{Q - df}{C}$$

wobei

$$Q = \sum_{i=1}^k W_i (Y_i - M)^2 = \sum_{i=1}^k W_i Y_i^2 - \frac{(\sum_{i=1}^k W_i Y_i)^2}{\sum_{i=1}^k W_i}$$

$$df = k - 1$$

und

$$C = \sum_{i=1}^k W_i - \frac{\sum_{i=1}^k W_i^2}{\sum_{i=1}^k W_i}$$

Berechnungsvorschrift

- für die Schätzung der Gewichte sind jetzt beide Varianzquellen zu berücksichtigen

$$W_i^* = \frac{1}{V_{Y_i}^*}$$

- $V_{Y_i}^*$ entspricht der summierten Varianz

$$V_{Y_i}^* = V_{Y_i} + T^2$$

- das gewichtete Mittel M^* berechnet sich nach

$$M^* = \frac{\sum_{i=1}^k W_i^* Y_i}{\sum_{i=1}^k W_i^*}$$

Berechnungsvorschrift

- Varianz der gemittelten Effektstärke (summary effect) wird abgeschätzt durch

$$V_{M^*} = \frac{1}{\sum_{i=1}^k W_i^*}$$

- Standardfehler der gemittelten Effektstärke

$$SE_{M^*} = \sqrt{V_{M^*}}$$

- Grenzen des 95% Konfidenzintervalls

$$LL_{M^*} = M^* - 1.96 \cdot SE_{M^*}$$

$$UL_{M^*} = M^* + 1.96 \cdot SE_{M^*}$$

Rechenbeispiel Random effects model (F. 27)

- identische Analyse mittels eines random-effects Modells
- Schätzung der Varianz τ^2 der wahren standardisierten Mittelwerte mittels der *DerSimonian und Laird* Methode:

$$T^2 = \frac{Q - df}{C}$$

1. Berechnung von Q, df und C

- Berechnung von Q, df und C:

$$Q = \sum_{i=1}^k W_i Y_i^2 - \frac{(\sum_{i=1}^k W_i Y_i)^2}{\sum_{i=1}^k W_i} = 53.915 - \frac{101.171^2}{244.215} = 12.0033$$

$$df = k - 1 = 6 - 1 = 5$$

$$C = \sum_{i=1}^k W_i - \frac{\sum_{i=1}^k W_i^2}{\sum_{i=1}^k W_i} = 244.215 - \frac{13802.325}{244.215} = 187.698$$

- daraus folgt

$$T^2 = \frac{Q - df}{C} = \frac{12.0033 - 5}{187.698} = 0.0373$$

Study	Effect size Y	Variance Within V_Y	Variance Between T^2	Variance Total $V_Y + T^2$	Weight W^*	Calculated quantities W^*Y
Carroll	0.095	0.033	0.037	0.070	14.233	1.345
Grant	0.277	0.031	0.037	0.068	14.702	4.078
Peck	0.367	0.050	0.037	0.087	11.469	4.204
Donat	0.664	0.011	0.037	0.048	20.909	13.892
Stewart	0.462	0.043	0.037	0.080	12.504	5.774
Young	0.185	0.023	0.037	0.061	16.466	3.049
Sum					90.284	32.342

2. Gewichtete Mittel, Varianz, SEM der gemittelten Effektstärke & 95% KI

- damit berechnet sich das gewichtete Mittel und die Varianz:

$$M^* = \frac{\sum_{i=1}^k W_i^* Y_i}{\sum_{i=1}^k W_i^*} = \frac{32.342}{90.284} = 0.3582$$

$$V_{M^*} = \frac{1}{\sum_{i=1}^k W_i^*} = \frac{1}{90.284} = 0.0111$$

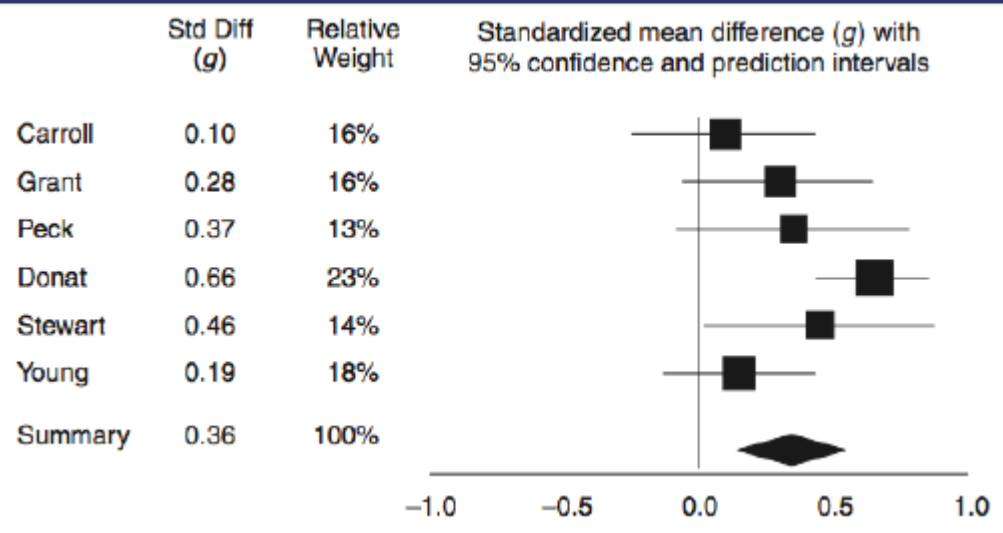
- Standardfehler der gemittelten Effektstärke

$$SE_{M^*} = \sqrt{V_{M^*}} = \sqrt{0.0111} = 0.1052$$

- Grenzen des 95% Konfidenzintervalls

$$LL_{M^*} = M^* - 1.96 \cdot SE_{M^*} = 0.3582 - 1.96 \cdot 0.1052 = 0.1520$$

$$UL_{M^*} = M^* + 1.96 \cdot SE_{M^*} = 0.3582 + 1.96 \cdot 0.1052 = 0.5645$$



R-Durchführung: Random effects model

- Die Analyse eines RE-Modells in R ist grundlegend identisch mit dem FE-Modell
 - Unterschied: method="DL" (DerSimonian-Laird-Schätzer)

Mögliche Ergebnisse des RE-Modells (hier mit Binärdaten)

```
# RE-Modell für binäre Daten mit DerSimonian-Laird-Schätzer
res_b.RE <- rma(yi, vi, data=calc_b_OR, method="DL", slab=Study)

# Anzeige RE-Modell
res_b.RE

##
## Random-Effects Model (k = 6; tau^2 estimator: DL)
##
## tau^2 (estimated amount of total heterogeneity): 0.1729 (SE = 0.2148)
## tau (square root of estimated tau^2 value):      0.4158
## I^2 (total heterogeneity / total variability):    52.61%
## H^2 (total variability / sampling variability):   2.11
##
## Test for Heterogeneity:
## Q(df = 5) = 10.5512, p-val = 0.0610

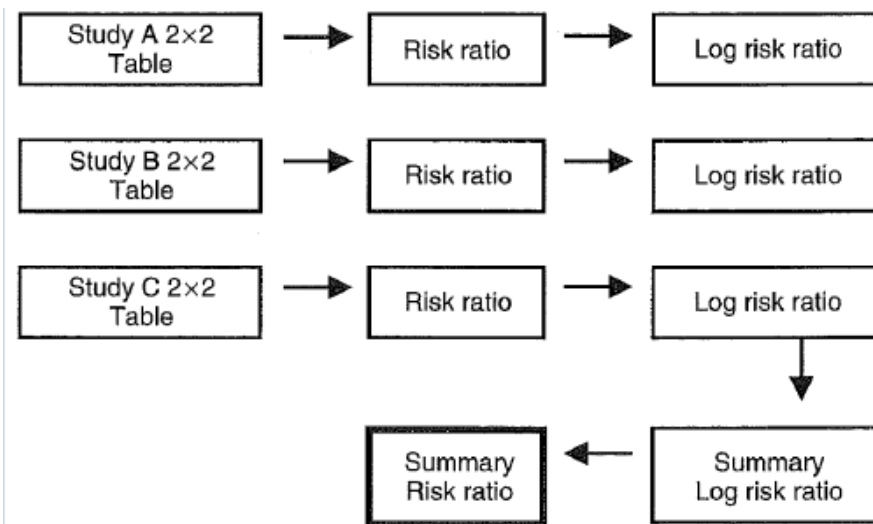
## Model Results:
##
## estimate      se      zval      pval      ci.lb      ci.ub
## -0.5663  0.2388  -2.3711  0.0177  -1.0344  -0.0982 *
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

→ Für logarithmische Relative Risiken gilt:

- $\text{LogRR} = \ln(\text{RR})$
- $\text{RR} = \exp(\text{LogRR})$
- Selbes gilt für Odds-Ratios

Warum die logarithmische Transformation?

- Sicherung der Symmetrie der Analyse
- Berechnung des Summary Schätzer sollte auf Log-Ebene geschehen
 - Rücktransformation erst auf Ebene des Summary Schätzers



- Varianzabschätzung und Standardfehler erfolgt ebenfalls auf Log-Ebene

Varianzabschätzung Relatives Risiko

- zur Berechnung eines Summary Effect werden Varianzmaße nicht in der originalen Metrik des Relativen Risikos berechnet, sondern in der logarithmischen Transformation
- Varianzschätzung

$$\text{Var}_{\text{LogRR}} = \frac{1}{A} - \frac{1}{n_1} + \frac{1}{C} - \frac{1}{n_2}$$

- approximierter Standardfehler

$$\text{SE}_{\text{LogRR}} = \sqrt{\text{Var}_{\text{LogRR}}}$$

Varianzabschätzung Odds Ratio

- analog zum Relativen Risiko werden Varianzmaße in der logarithmischen Metrik bevorzugt
- Varianzschätzung

$$Var_{\text{LogOR}} = \frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}$$

- approximierter Standardfehler

$$SE_{\text{LogOR}} = \sqrt{Var_{\text{LogOR}}}$$

R: logarithmische Transformierung – Umgang Binärdaten

1. Berechnung Effektstärken und Varianz mit escalc()

Typ	Argument	Beschreibung
OR	ai	Vektor 4-Feldertafel oben links
	bi	Vektor 4-Feldertafel oben rechts
	ci	Vektor 4-Feldertafel unten links
	di	Vektor 4-Feldertafel unten rechts

Und dazu die zu verwendende R-Syntax für die Berechnung:

```
calc_b_OR <- escalc(measure="OR"  
  ,ai=b_data$Treated_Events  
  ,bi=b_data$Treated_Non_Events  
  ,ci=b_data$Control_Events  
  ,di=b_data$Control_Non_Events  
  ,data=b_data)
```

2. Spezifizierung des Modells

```
# RE-Modell für binäre Daten mit DerSimonian-Laird-Schätzer  
res_b.RE <- rma(yi, vi, data=calc_b_OR, method="DL", slab=Study)  
  
# Anzeige RE-Modell  
res_b.RE  
  
##  
## Random-Effects Model (k = 6; tau^2 estimator: DL)  
##  
## tau^2 (estimated amount of total heterogeneity): 0.1729 (SE = 0.2148)  
## tau (square root of estimated tau^2 value): 0.4158  
## I^2 (total heterogeneity / total variability): 52.61%  
## H^2 (total variability / sampling variability): 2.11  
##  
## Test for Heterogeneity:  
## Q(df = 5) = 10.5512, p-val = 0.0610  
##
```

3. Ergebnisse

```
## Model Results:
##
## estimate      se      zval      pval      ci.lb      ci.ub
## -0.5663  0.2388  -2.3711  0.0177  -1.0344  -0.0982 *
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

CAVE: Die Odds-Ratios liegen logarithmiert vor!

Dies gilt es auch bei der Plotterstellung zu berücksichtigen (x-Achse muss ebenfalls transformiert werden)

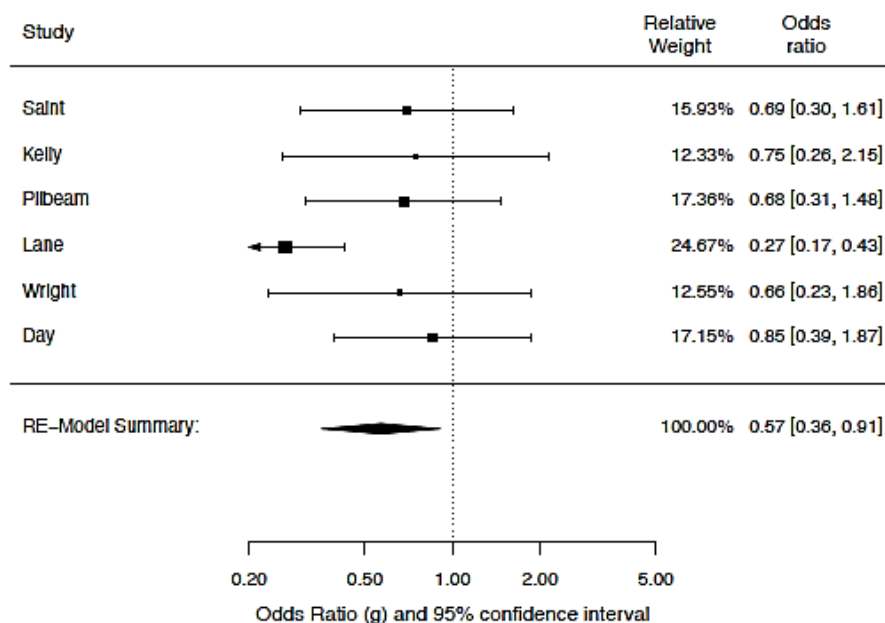
4. Plotterstellung

- Mit „atransf“ wird der Wertebereich logarithmiert
- Mit „at“ die X-Achse skaliert
- Mit „refline“ kann die Position der Y-Achse auf der x-Achse verändert werden

```
# Grundeinstellung Grafikparameter
par(mar=c(6,4,1,2),cex=.9, font=1)

# Random-Effects-Model mit binären Daten
forest(res_b.RE,showweights = TRUE,digits=2
       ,xlim=c(-3.5,3.5)
       ,ylim=c(-3,9)
       ,refline=log(1)
       ,at=log(c(0.2,0.5,1,2,5)) # Einteilung x-Achse
       ,xlab="Odds Ratio (g) and 95% confidence interval"
       ,mlab="RE-Model Summary:"
       ,atransf=exp) # logarithm. Transformierung x-Achse

# Spaltenbeschriftungen
par(font=1)
text(c(-3.2,1.8,2.8), 7.7, c("Study","Relative\nWeight"
                             ,"Odds\nratio"))
```



R: Optimierung Forest Plots

```
# Fixed-Effect-Modell mit SMD
forest(res_smd.FE #rma-Objekt
       ,showweights = TRUE # Anzeige der relativen Gewichte
       ,digits=2 # Anzahl Nachkommastellen
       ,xlim=c(-2,3) # horizontale Begrenzung der Plot-Region
       ,alim=c(-1,1) # Minimum und Maximum der x-Achse
       ,ilab=cbind(round(calc_smd$vi,4)) # zusätzliche Spalte
       ,ilab.xpos=c(-1)) # Position der zusätzlichen Spalte
```

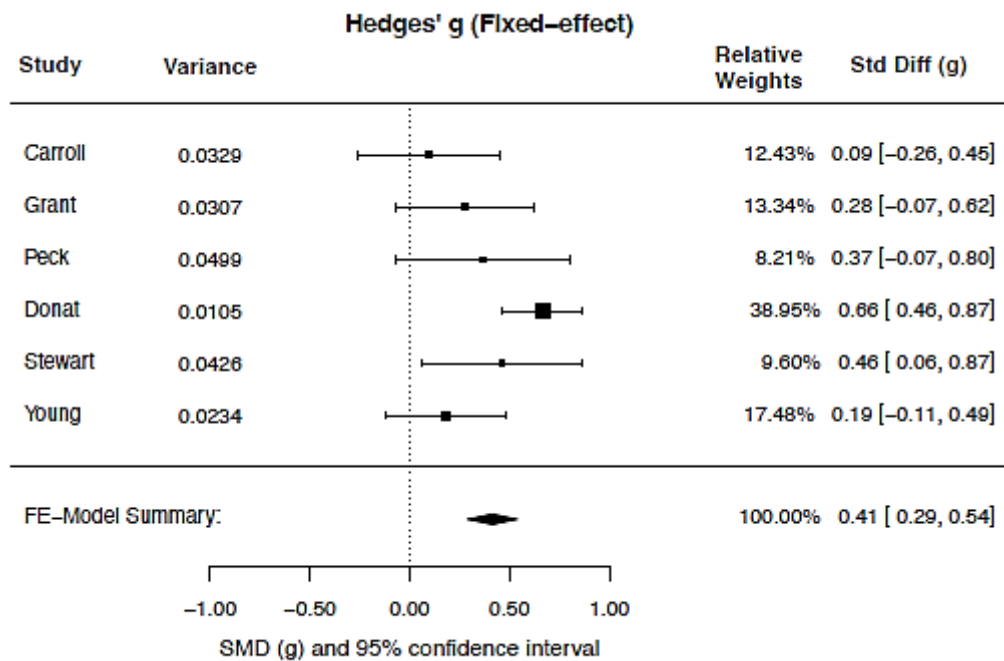
- Über die Funktion `par()` können weitere Grundeinstellungen für Plots gesetzt werden
 - Z.B. Schrifttyp/-größe, Positionierungen, etc.
 - Mit `par(„usr“)` [1:4] erhält man die aktuellen Einstellungen der Plotregion nach `xmin/max` sowie `ymin/max`

```
# par: für Layout-Parameter
# mar: zum Setzen von Rand (margin)-Größen
#       c(unten, links, oben, rechts)
# cex: Schriftgröße
# font: definiert Schriftsatz
par(mar=c(6,4,1,2),cex=.9, font=1)

# Fixed-Effect Modell mit SMD
forest(res_smd.FE #rma-Objekt
       ,showweights = TRUE # Anzeige der Gewichte
       ,digits=2 # Anzahl Nachkommastellen
       ,xlim=c(-2,3) # horizontale Begrenzung der Plot-Region
       ,alim=c(-1,1) # Minimum und Maximum der x-Achse
       ,ilab=cbind(round(calc_smd$vi,4)) # zusätzliche Spalte
       ,ilab.xpos=c(-1) # Position der zusätzlichen Spalte
       ,xlab="SMD (g) and 95% confidence interval" #Plot-Bezeichng.
       ,mlab="FE-Model Summary:") # Summary-Bezeichnung

# Änderung der Grafikparameter für zusätzliche Bezeichnungen
# Spaltenbeschriftungen
par(font=2)
text(c(-1.80,-1,1.75,2.5), 7.7
     ,c("Study", "Variance", "Relative\nWeights"
        , "Std Diff (g)"))

# Diagrammüberschrift
par(cex=1, font=2)
text(0.4,8.5,"Hedges' g (Fixed-effect)")
```



Zusätzliche Angaben wie T² oder I² mit text()

```
# Grundeinstellung Grafikparameter
par(mar=c(6,4,1,2),cex=.9, font=1)

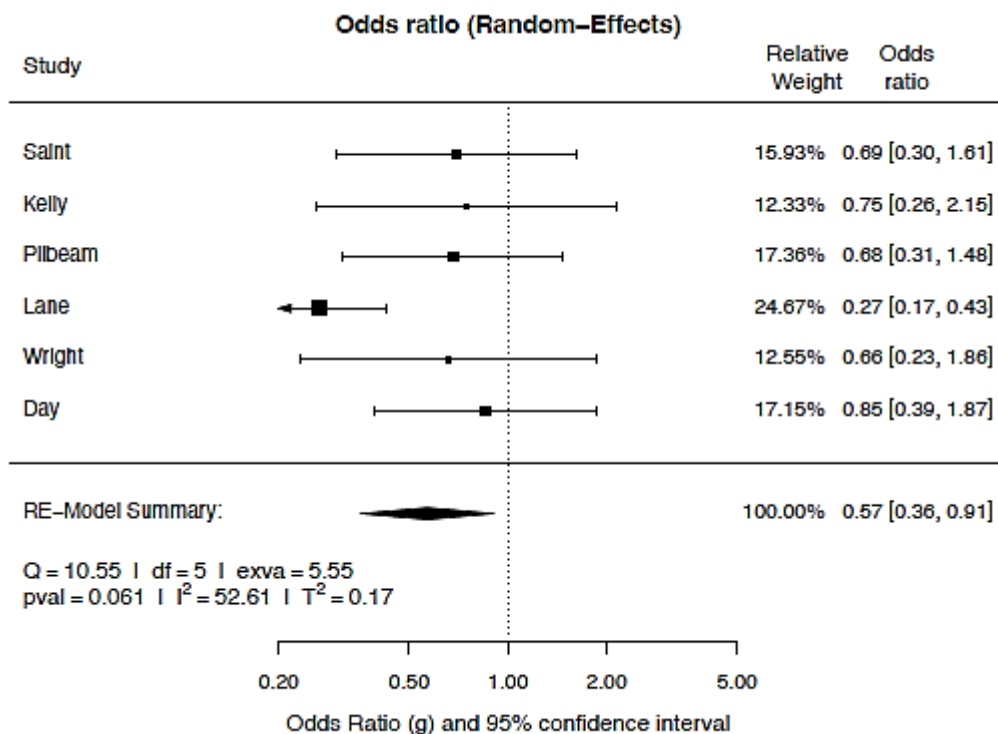
# Random-Effects Model mit binären Daten

forest(res_b.RE,showweights = TRUE,digits=2
, xlim=c(-3.5,3.5)
, ylim=c(-3,9)
, refline=log(1)
, at=log(c(0.2,0.5,1,2,5)) # Einteilung x-Achse
, xlab="Odds Ratio (g) and 95% confidence interval"
, mlab="RE-Model Summary:"
, atransf=exp) # logarithm. Transformierung x-Achse

# Spaltenbeschriftungen
par(font=1)
text(c(-3.2,2.1,2.8), 7.7
, c("Study","Relative\nWeight","Odds\nratio"))

# Extra-Informationen
text(-3.5,-2.2,bquote(Q == .(round(res_b.RE$QE,2)) ~ " | "
~ df == .(res_b.RE$k-1) ~ " | "
~ exva == .(round(res_b.RE$QE-(res_b.RE$k-1),2))), pos=4)
text(-3.5,-2.7,bquote(pval == .(round(res_b.RE$QEp,4)) ~ " | "
~ I^{2} == .(round(res_b.RE$I2,2)) ~ " | "
~ T^{2} == .(round(res_b.RE$tau2,2))), pos=4)

# Diagrammüberschrift
par(cex=1, font=2)
text(0,8.5,"Odds ratio (Random-Effects)")
```



Heterogenität

Maßzahlen

- I. Q-Statistik
- II. T² (between-study Varianz)
- III. I² (Verhältnis der wahren Heterogenität zur gesamt gemessenen Variabilität)

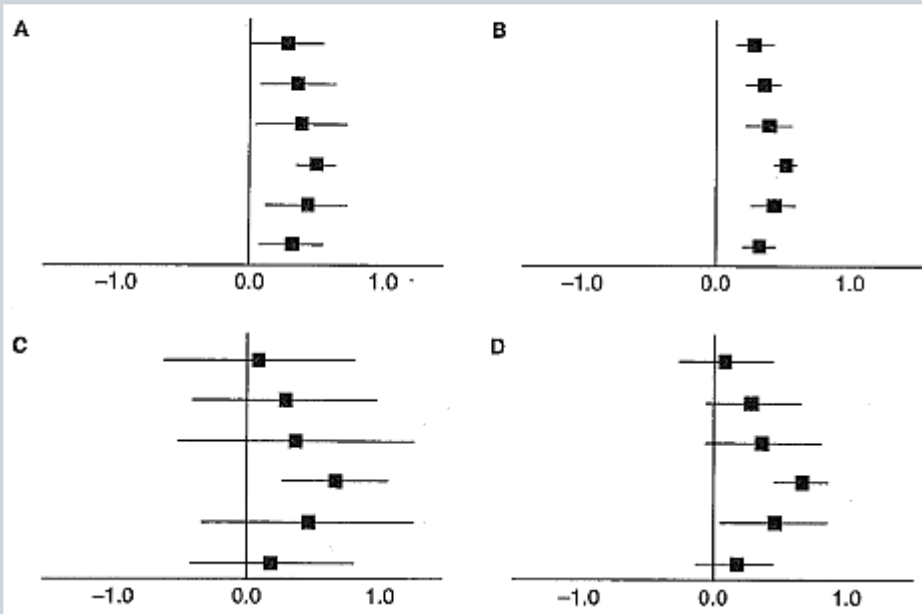
Nomenklatur

- Heterogenität = Heterogenität der wahren (unterliegenden Population, unendlicher Stichprobe, kein Stichprobenfehler)
- Variabilität und Dispersion werden verwendet um Unterschiede in Effektstärken zu beschreiben (wahre und gemessene)

Aufteilung gemessener Variabilität in: a) Heterogenität der Effektstärken & b) Stichprobenfehler (within-study variance)

- I. **Berechnung absoluter Variabilität zwischen allen Studien**
- II. **Abschätzung der Variabilität, bei Annahme gleicher wahrer Effekte für alle Studien**
- III. **Excess variation (Differenz zwischen beiden Variabilitäten) = Schätzer für Heterogenität**

Dispersion vs. Stichprobenfehler



Stichprobenfehler = Abweichung zwischen Ergebnissen einer Stichprobe und den tatsächlichen Werten in der Population (verursacht durch Zufallsvariation)

→ kann durch eine größere Stichprobengröße verringert werden

Dispersion = Streuung von Datenpunkten um einen Mittelwert herum

→ zeigt die Heterogenität in einer Stichprobe/Population

Q-Statistik

- Messung der **Heterogenität/Variation zwischen den Studien** in der Metaanalyse
- Häufig im Zusammenhang mit dem I²-Wert (Gibt Prozentsatz der Variation zwischen den Studien an, der aufgrund der Heterogenität und nicht Zufall zu erwarten ist)

Berechnung von Q

- Berechnung von Q als gewichtete Summe der quadratischen Abweichungen der Effektstärken Y_i vom Summary Effect M :

$$Q = \sum_{i=1}^k W_i (Y_i - M)^2$$

- Q ist ein standardisiertes Maß; es ist daher unabhängig von der Metrik der Effektstärke
- äquivalente Formulierung

$$Q = \sum_{i=1}^k W_i Y_i^2 - \frac{(\sum_{i=1}^k W_i Y_i)^2}{\sum_{i=1}^k W_i}$$

Erwartungswert von Q

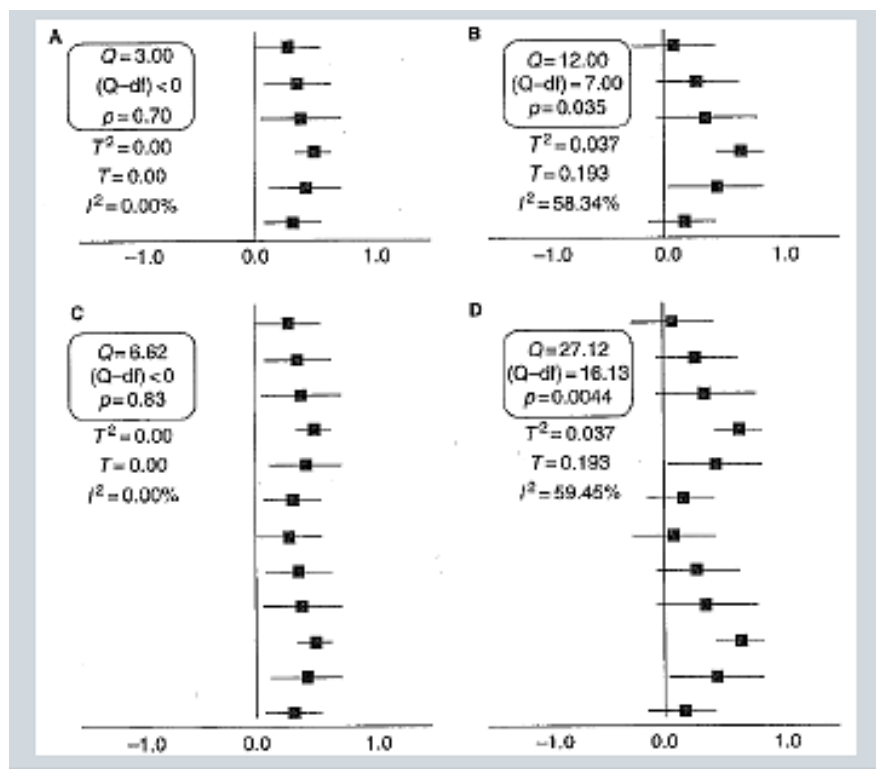
- Abschätzung des Erwartungswertes für Q unter der Annahme, dass alle Studien eine identische Effektstärke aufweisen (Variation der Effektstärken wird nur durch die Stichprobenfehler bestimmt)
- da Q ein standardisiertes Maß ist, ist der Erwartungswert unabhängig von der Metrik der Effektstärke und entspricht der Anzahl der Freiheitsgrade df

$$df = k - 1$$

Abschätzung der *excess variation*

- die Differenz $Q - df$ beschreibt die *excess variation*
- entspricht dem Anteil der Variabilität, der durch die Heterogenität (d.h. die zugrunde liegende Verteilung der wahren Effektstärken) verursacht wird

- Interpretation im Zusammenhang mit dem p-Wert
 - Überprüfung, ob die Variation zwischen den Studien größer ist, als erwartet
 - **Signifikanter Q-Wert deutet auf höhere Heterogenität zwischen den Studien hin, als erwartet → Ergebnisse der einzelnen Studien nicht konsistent**



- P-Werte adressieren die Signifikanz der Resultate nicht die Stärke der Effekte

T² - Varianz der wahren Effekte

- Beruht im Gegensatz zur Q-Statistik (Chi²-Verteilung) auf die multivariate Normalverteilung
- Misst die Streuung der Effektgrößen über verschiedene Studien hinweg
- Zur Bewertung der globalen Heterogenität zwischen den Studien
- Häufig im Zusammenhang mit der Q-Statistik, insbesondere bei multivariaten Meta-Analysen mit mehreren Effektgrößen nützlich

- τ^2 ist bestimmbar für eine unendliche Anzahl von Studien mit jeweils unendlicher Stichprobengröße (d.h. es wird jeweils der wahre Effekt θ_i in jeder Studie bestimmt)

- Abschätzung durch T^2

$$T^2 = \frac{Q - df}{C}$$

Schätzung der Varianz der wahren Effekte

- die Konstante C entspricht einer Normierung, die die *excess variation* (a) in die ursprüngliche Metrik ("quadratisch") zurück transformiert und (b) einen Mittelwert bildet

$$C = \sum_{i=1}^k W_i - \frac{\sum_{i=1}^k W_i^2}{\sum_{i=1}^k W_i}$$

- für $Q < df$ wird T^2 negativ und wird auf $T^2 = 0$ gesetzt

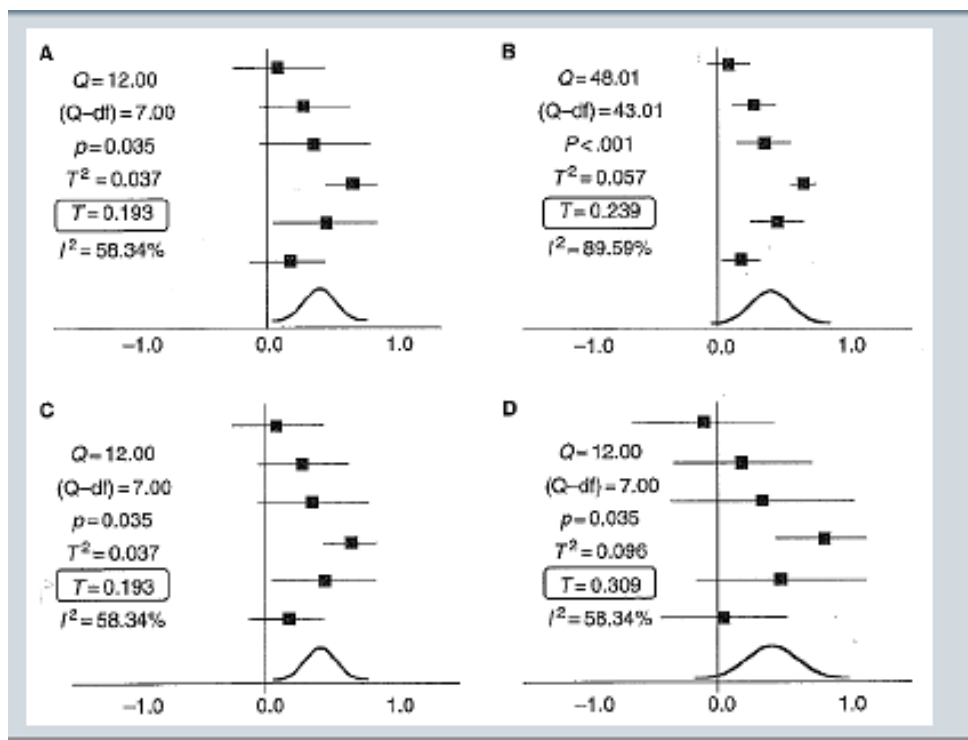
Interpretation:

- Hoher T²-Wert deutet auf große Heterogenität hin
- Signifikanz: signifikanter T²-Wert deutet auf große Heterogenität hin
- Nicht-Signifikanz deutet auf nicht größere Variation als erwartet hin, wenn nur Zufallsvariation vorhanden ist

T – Standardabweichung der wahren Effektstärke

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

- T ist in der gleichen Metrik wie die Effektstärke
- ein hinreichend genauer Schätzer für T erlaubt es, die Verteilung der wahren Effektstärken anzugeben (bzw. die entsprechenden Konfidenzintervalle)



- ➔ T & T^2 sind absolute Maße, die die Abweichung in der Metrik der Effektstärke quantifizieren
- ➔ Jedoch: Beschreibung der Heterogenität auf einer unabhängigen Skala wünschenswert!
- ➔ I^2 = relatives Maß, beschreibt Anteil der excess variation an der gesamtgemessenen Dispersion

I² Statistik

- Misst Anteil der excess variation an der gesamtgemessenen Dispersion

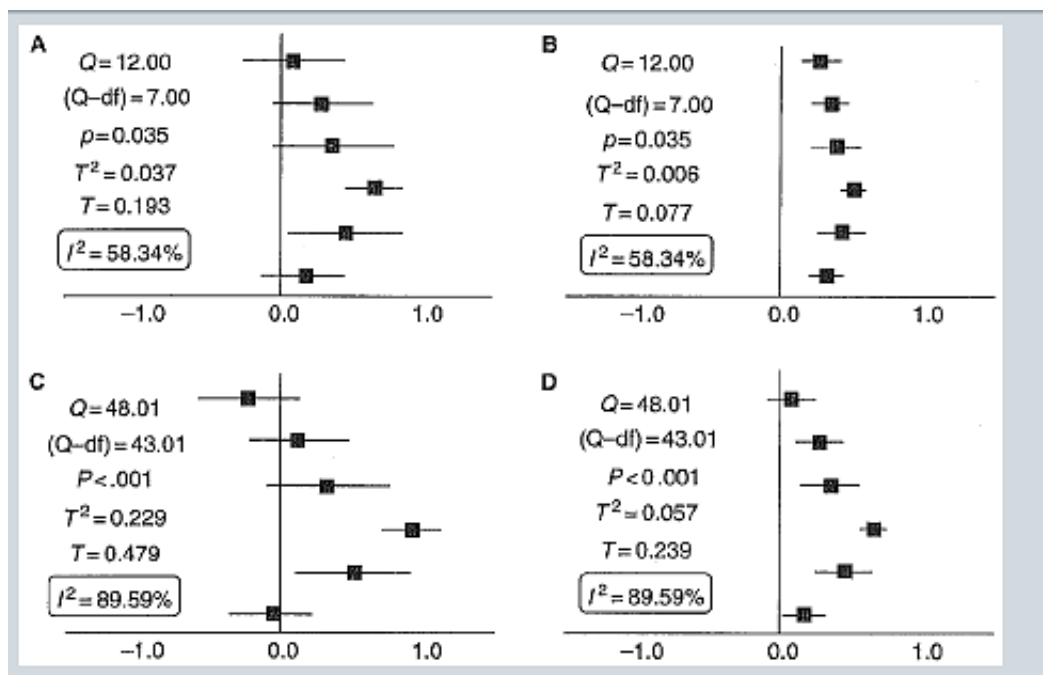
$$I^2 = \frac{Q - df}{Q} \cdot 100\%$$

I² lässt sich konzeptionell beschreiben als

$$I^2 = \frac{Var_{between}}{Var_{total}} \cdot 100\% = \frac{\tau^2}{\tau^2 + V_y} \cdot 100\%$$

Interpretation:

- Je niedriger der Wert, desto besser ist die Variabilität durch den auftretenden Stichprobenfehler zu beschreiben (I² → 0)
- Bei hohen Werten gilt es die Ursache der Heterogenität herauszufinden (Subgruppenanalyse, Metaregression)
- I² 25% = niedrig, 50% = moderat, 75% = hoch
- I² = 100% sagt nur aus, dass die Variabilität der Daten auf eine Heterogenität der wahren Effekte zurückzuführen ist, nicht dass die Effekte weit streuen!



Weitere Maßzahlen zur Heterogenität

weitere Maßzahlen zur Heterogenität in Metaanalysen

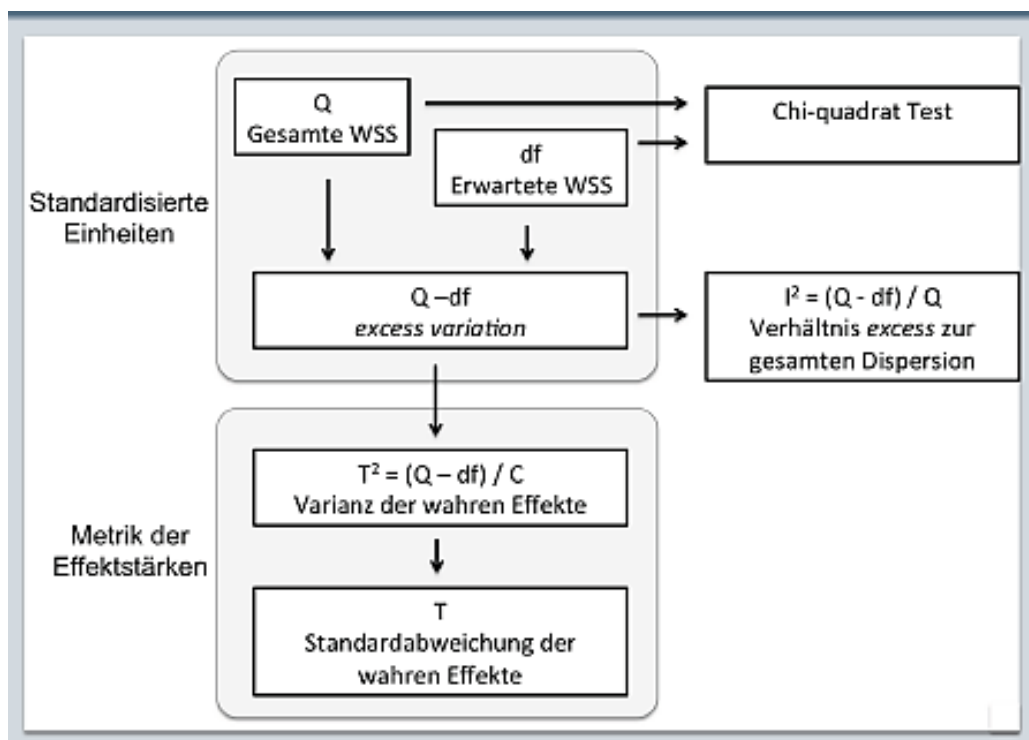
- H^2 beschreibt das Verhältnis der beobachteten Heterogenität zur erwarteten Heterogenität unter der Annahme einer gleichen, wahren Effektstärke für alle Studien

$$H^2 = \frac{Q}{df}$$

- R^2 beschreibt das Verhältnis des Konfidenzintervalls unter dem random effects Models mit dem Konfidenzintervall des fixed effect Modells

$$R^2 = \frac{CI_{RE}}{CI_{FE}}$$

Zusammenfassend - Heterogenitätsmaße



- Q-Statistik und p-Wert als Signifikanzmaß, abhängig von der Studienzahl, jedoch unabhängig von der Metrik der Effektstärke
- T^2 und T schätzen die Varianz und Standardabweichung der Verteilung der wahren Effektstärken in den untersuchten Studien, unabhängig von der Studienanzahl
- I^2 stellt das Verhältnis der wahren Heterogenität zur gesamten Variabilität in den einzelnen Studien, unabhängig von Metrik und Anzahl der Studien

p-Wert

- der p -Wert der Q -Statistik ist kein Maß der Effektstärke

Standardabweichung T

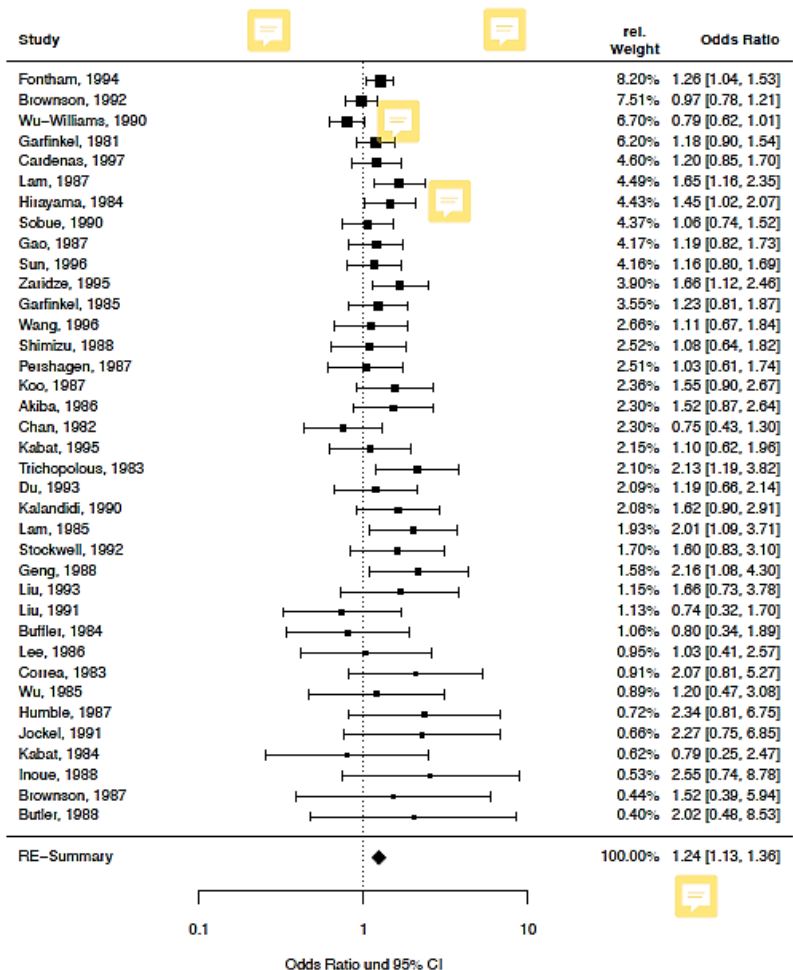
- T ist ein geeignetes Maß, um die Variabilität der wahren Effekte abzuschätzen

Genauigkeit

- die Unsicherheit bei der Schätzung von T^2 und I^2 ist häufig sehr groß
- für Studien mit geringer Genauigkeit (große Konfidenzintervalle) kann das zu einer Verdeckung der wahren Heterogenität führen ($T^2 \rightarrow 0$ und $I^2 \rightarrow 0$) \rightarrow die Schlussfolgerung, dass die Effektstärken gleich sind, wäre dann nicht gerechtfertigt

Publicationbias

- = Tendenz, dass positive Ergebnisse häufiger publiziert werden, als negative Ergebnisse
- Besonders bei kleineren Studien der Fall
- Auffällig: Großteil der kleineren Studien weisen größere, jedoch nicht-signifikante Ergebnisse auf



Bias konzeptuell verstehen

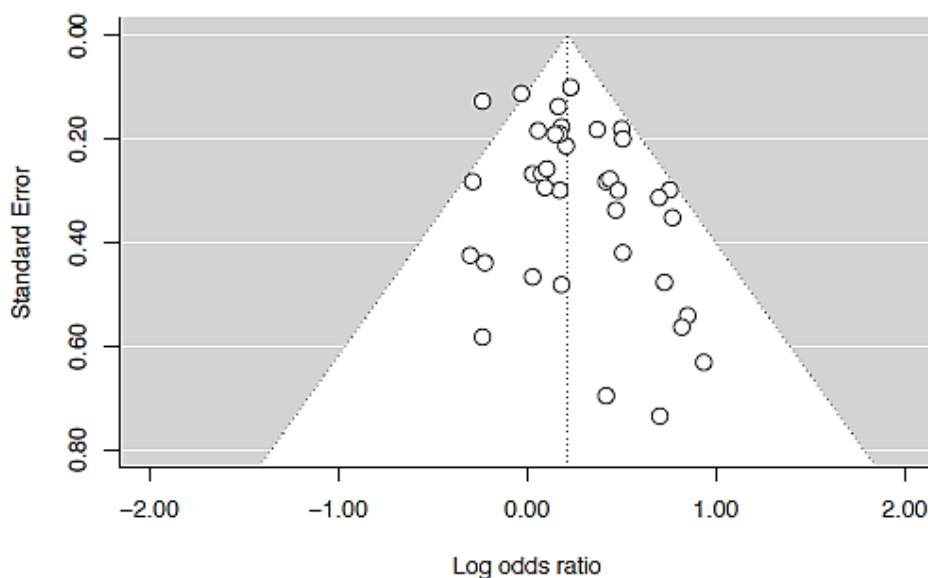
- Verständnis, welche Studien wahrscheinlich fehlen werden:
 - große Studien werden wahrscheinlich publiziert, unabhängig von ihrer statistischen Signifikanz
 - mittelgroße Studien werden vorrangig unpubliziert bleiben, wenn sie keine signifikanten Effekte aufweisen
 - kleine Studien haben das größte Risiko, nicht publiziert zu werden; nur sehr große Effektstärken führen (mit größerer Wahrscheinlichkeit) zu signifikanten Ergebnissen
- Publikationsbias wird größer, je kleiner die Studien werden

1. Visualisierung Funnel-Plot: `funnel()`

- Darstellung der Effektstärke (x-Achse) gegen ein Varianzmaß, meist Standardfehler (y-Achse) der einzelnen Studien
 - Mit wachsender Stichprobengröße nimmt der Standardfehler ab
 - Bedeutet: mit abnehmenden Werten auf der y-Achse liegen die Effektstärken näher am wahren Wert der Effektstärke
 - Größeren Studien „oben“, im „unteren“ Bereich des Diagramms werden kleinere Studien dargestellt
 - Ein mögliches Ungleichgewicht hervorgerufen durch Veröffentlichungen mit stärkeren Effektstärken, kann so leichter identifiziert werden

```
funnel(res_hh.RE #rma-Object
,main="Funnel Plot of Standard Error by Log Odds ratio"
,pch=21 #Punkte-Typ
,xlab="Log odds ratio"
,xlim=c(-2,2) #Skalierung x-Achse
,ylab="Standard Error"
,ylim=c(0,0.8) #Skalierung y-Achse
,digits=2, cex=1.5)
```

Funnel Plot of Standard Error by Log Odds ratio



Darstellung des Standardfehlers anstatt Varianz streckt die Darstellung auf der y-Achse für kleinere Studien

2. Quantifizierung des Publicationbias

Quantifizierung des Publikationsbias

- *Rosenthal's (und Orwin's) Fail-safe N:*
 - Hinzufügen von N kleinen Studien mit Effektstärke $Y_i = 0$ ($Y_i = Y_{min}$) bis der Summary Effekt nicht mehr signifikant ist
 - große N vermitteln mehr Sicherheit, dass der Summary Effekt nicht zu stark durch Publikationsbias von kleinen Studien verzerrt wird
- *Trim and Fill:*
 - iteratives Entfernen kleiner Studien mit extremen Effektstärken → *unbiased* Schätzung der Effektstärke (aber zu stark reduzierte Varianz)
 - auffüllen aller originalen Studien und deren "Spiegelbilder" (Imputation) → kein Einfluss auf den Mittelwertschätzer der Effektstärke, aber Korrektur der Varianz

a. Rosenthals Fail-safe N: fsm(type="Rosenthal")

- Grundlegend: signifikanter p-Wert bzgl. des Summary Effektes
- Stellt die Frage, wieviele Studien man übersehen haben müsse, um die Signifikanz in Frage zu stellen
- Versucht so lange Studien ohne nachweisbaren Effekt ($Y=0$ bzw. bei RR/OR $Y=1$) hinzuzufügen, bis der Summary Effekt nicht mehr signifikant ist
 - Anzahl künstlicher Studien ist Indikator (große n vermitteln Sicherheit)

Problem:

- Methode nicht unumstritten und heutzutage kaum noch erforscht

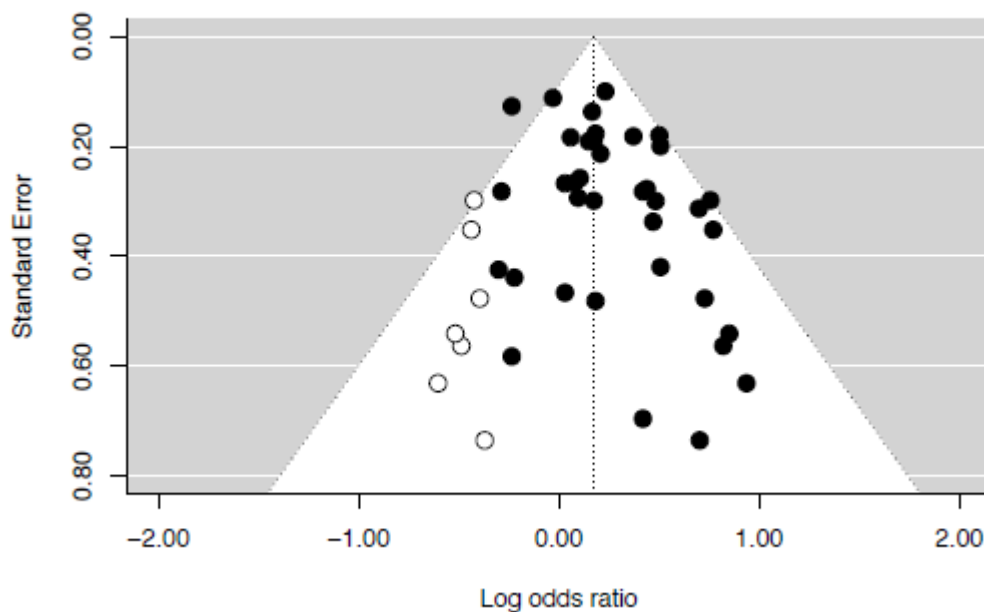
```
fsm(yi,vi, data=dat.hackshaw1998, type="Rosenthal")  
  
##  
## Fail-safe N Calculation Using the Rosenthal Approach  
##  
## Observed Significance Level: <.0001  
## Target Significance Level: 0.05  
##  
## Fail-safe N: 393
```

b. Trim and Fill Methode trimfill()

- Zentrale Annahme: symmetrische Verteilung der Effektstärke um den wahren Wert
- Zur Herstellung der Symmetrie werden zuerst iterativ kleine Studien aus der Metaanalyse entfernt, bis die Studien symmetrisch um die neue unverzerrte Effektstärke verteilt sind
 - Reduziert jedoch auch die Varianz und resultiert in kleineren Kis
- Um diese Varianzschätzung wieder zu korrigieren, werden alle entfernten Studien mit ihren Spiegelbildern wieder in die Metaanalyse aufgenommen
 - Dadurch wird der Summary Effekt nicht verändert, die Varianz jedoch wieder verbessert

```
taf <- trimfill(res_hh.RE)
funnel(taf #rma-Object
,main="Funnel Plot of Standard Error by Log odds ratio"
,xlab="Log odds ratio"
,xlim=c(-2,2) #Skalierung x-Achse
,ylab="Standard Error"
,ylim=c(0,0.8) #Skalierung y-Achse
,digits=2, cex=1.5)
```

Funnel Plot of Standard Error by Log odds ratio



Schwarzen Punkte stellen die vorhandenen und weiß die hinzugefügten Studien dar

Ergebnisinterpretation – weitere Verfahren

Subgruppenanalyse

- Zur Erklärung der Muster von Effekten, z.B. bestimmte Patientengruppen, Randomisationsschemata, etc.

3 Verfahren für Subgruppenanalysen:

- I. Fixed effect → unter Annahme gemeinsamer Effektstärken
- II. Random effects mit individuellem Schätzer für T^2
- III. Random effects mit einem gemittelten Schätzer für T^2

Dabei wird ähnlich vorgegangen, wie in einzelnen Studien:

- a) Mittlere Effektstärke und Varianzschätzung für jede Subgruppe
- b) Vergleich der mittleren Effekte zwischen den Gruppen

Fixed effect Beispiel: Vergleich zweier Studien mit verschiedenen Interventionen (A & B)

Study	Effect size Y	Variance Within V_Y	Variance Between T^2	Variance Total V	Weight W	Calculated quantities		
						WY	WY^2	W^2
Thornhill	0.110	0.0100	0.0000	0.0100	100.000	11.000	1.210	10000.000
Kendall	0.224	0.0300	0.0000	0.0300	33.333	7.467	1.673	1111.111
Vandamm	0.338	0.0200	0.0000	0.0200	50.000	16.900	5.712	2500.000
Leonard	0.451	0.0150	0.0000	0.0150	66.667	30.067	13.560	4444.444
Professor	0.480	0.0100	0.0000	0.0100	100.000	48.000	23.040	10000.000
Sum A					350.000	113.433	45.195	28055.556
Jefferies	0.440	0.0150	0.0000	0.0150	66.667	29.333	12.907	4444.444
Fremont	0.492	0.0200	0.0000	0.0200	50.000	24.600	12.103	2500.000
B Doyle	0.651	0.0150	0.0000	0.0150	66.667	43.400	28.253	4444.444
Stella	0.710	0.0250	0.0000	0.0250	40.000	28.400	20.164	1600.000
Thorwald	0.740	0.0120	0.0000	0.0120	83.333	61.667	45.633	6944.444
Sum B					306.667	187.400	119.061	19933.333
Sum					656.667	300.833	164.255	47988.889

1. Effektstärke und Varianz für Gruppe A

Effektstärke für Subgruppe A

Vorgehen analog zur fixed effect Metaanalyse aus VL 2

- mittlere Effektstärke M_A berechnet sich nach

$$M_A = \frac{\sum_{i=1}^k W_{A_i} Y_{A_i}}{\sum_{i=1}^k W_{A_i}} = \frac{113.433}{350.0} = 0.3241$$

- Varianz der gemittelten Effektstärke

$$V_{M_A} = \frac{1}{\sum_{i=1}^k W_{A_i}} = \frac{1}{350.0} = 0.0029$$

2. Standardfehler und 95%KI ermitteln für Gruppe A

Effektstärke für Subgruppe A

daraus ergibt sich

- Standardfehler der gemittelten Effektstärke

$$SE_{M_A} = \sqrt{V_{M_A}} = \sqrt{0.0029} = 0.0535$$

- Grenzen des 95% Konfidenzintervalls

$$LL_{M_A} = M_A - 1.96 \cdot SE_{M_A} = 0.3241 - 1.96 \cdot 0.0535 = 0.2193$$

$$UL_{M_A} = M_A + 1.96 \cdot SE_{M_A} = 0.3241 + 1.96 \cdot 0.0535 = 0.4289$$

3. Analog für Gruppe B

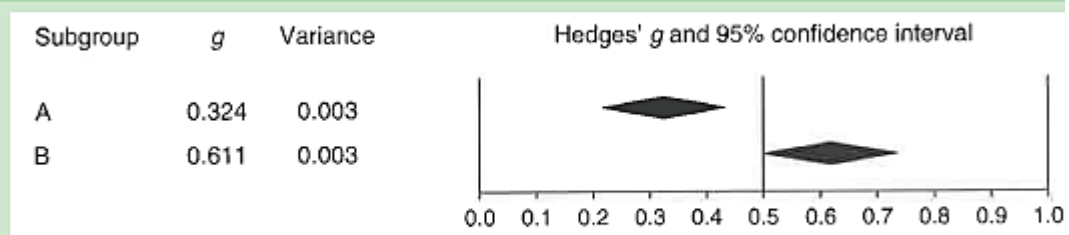
Effektstärke für Subgruppenanalyse

Analoge Berechnung für Subgruppe B und für beide Gruppen gemeinsam

	A	B	Gesamt
M	0.3241	0.6111	0.4581
V	0.0029	0.0033	0.0015
SE_M	0.0535	0.0571	0.0390
LL_M	0.2193	0.4992	0.3816
UL_M	0.4289	0.7230	0.5346

4. Gruppenvergleiche

Vergleich der Subgruppen



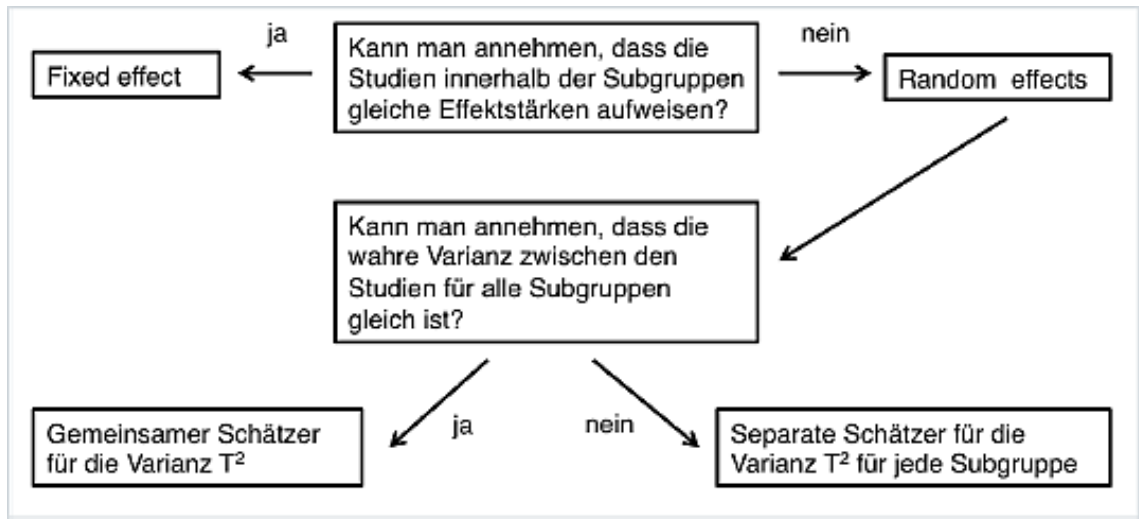
→ mehrere Möglichkeiten, die Gruppen miteinander zu vergleichen

5. Testverfahren

- T-Test anhand der Differenz der Mittelwerte zwischen den Gruppen
- Mehr als zwei Gruppen: ANOVA → Aufteilung der Varianzkomponenten zwischen den Gruppen und innerhalb der Gruppen
- Außerdem: KIs zu der Differenz der Mittelwerte

Random effects model

- Realistischeres Szenario, bei denen die wahren Effektstärken einer Verteilung unterliegen



Bietet sich an bei gleich Varianz innerhalb der Gruppen und kleiner Anzahl verfügbarer Studien pro Gruppe

a. Random effects model mit einem gemeinsamen Schätzer für T^2 within

- Unter der Annahme: die wahre Varianz der Studien innerhalb der Gruppen ist ähnlich (ähnliche Verteilung um die jeweiligen mittleren Effektstärken in den Gruppen)

- Abschätzung durch $T^2 = \frac{Q-df}{C}$, wobei $C = \sum_{i=1}^k W_i - \frac{\sum_{i=1}^k W_i^2}{\sum_{i=1}^k W_i}$
- Ein gemittelter Schätzer für τ_{within}^2 ergibt sich aus der Summe der Q , df und C (für jede Subgruppe j) entsprechend:

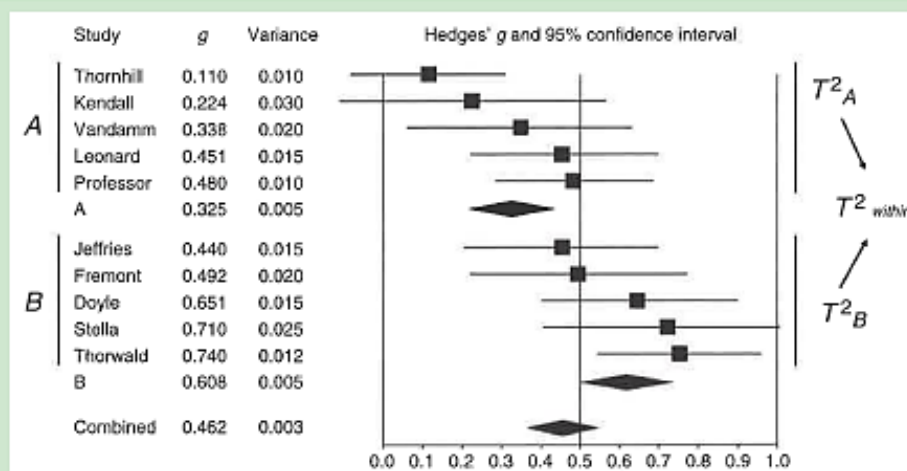
$$T_{within}^2 = \frac{\sum_{j=1}^m Q_j - \sum_{j=1}^m df_j}{\sum_{j=1}^m C_j}$$

random-effects model mit einem gemeinsamen Schätzer für τ^2_{within}

Für den konkreten Fall ist die Spalte T^2 gleich für alle Studien.

Study	Effect size γ	Variance Within V_γ	Variance Between T^2	Variance Total V	Weight W	Calculated quantities		
						$W\gamma$	$W\gamma^2$	W^2
Thornhill	0.110	0.0100	0.0097	0.0197	50.697	5.577	0.613	2570.150
Kendall	0.224	0.0300	0.0097	0.0397	25.173	5.639	1.263	633.678
Vandamm	0.338	0.0200	0.0097	0.0297	33.642	11.371	3.843	1131.752
Leonard	0.451	0.0150	0.0097	0.0247	40.445	18.241	8.226	1635.767
Professor	0.480	0.0100	0.0097	0.0197	50.697	24.334	11.681	2570.150
Sum A					200.652	65.161	25.627	8541.498
Jefferies	0.440	0.0150	0.0097	0.0247	40.445	17.796	7.830	1635.767
Fremont	0.492	0.0200	0.0097	0.0297	33.642	16.552	8.143	1131.752
Doyle	0.651	0.0150	0.0097	0.0247	40.445	26.329	17.140	1635.767
Stella	0.710	0.0250	0.0097	0.0347	28.798	20.446	14.517	829.299
Thorwald	0.740	0.0120	0.0097	0.0217	46.030	34.062	25.206	2118.721
Sum B					189.358	115.185	72.837	7351.306
Sum					390.010	180.346	98.463	15892.804

random-effects model mit einem gemeinsamen Schätzer für τ^2_{within}



Achtung: Angabe eines summary effects sollte gut begründet sein

R: Subgruppenanalyse mit Random effects model, T^2 für alle Subgruppen geschätzt

1. Normale Schritte des random effect models durchführen mit `escalc()` und `rma()`
2. Über das Argument „mods“ kann der Moderator festgelegt werden, der die Subgruppen festlegt → diese werden als Kovariablen definiert und ausgewertet

```
# Berechnung des rma-Objekts
res_bcg.RE_alloc <- rma(yi, vi
                        ,mods = ~alloc # Angabe des Moderators
                        ,data = bcg_RR # Datenquelle
                        ,method="DL")

# Anzeigen des rma-Objekts
res_bcg.RE_alloc

##
## Mixed-Effects Model (k = 13; tau^2 estimator: DL)
##
## tau^2 (estimated amount of residual heterogeneity):      0.5596 (SE =
## tau (square root of estimated tau^2 value):            0.7480
## I^2 (residual heterogeneity / unaccounted variability): 92.45%
## H^2 (unaccounted variability / sampling variability):   13.24
## R^2 (amount of heterogeneity accounted for):            0.00%
##
## Test for Residual Heterogeneity:
## QE(df = 10) = 132.3676, p-val < .0001
##
## Test of Moderators (coefficients 2:3):
## QM(df = 2) = 1.4349, p-val = 0.4880
##
## Model Results:
##
##              estimate      se      zval      pval      ci.lb      ci.ub
## intrcpt          -0.5125  0.5421  -0.9454  0.3444  -1.5751  0.5500
## allocrandom       -0.4780  0.6286  -0.7605  0.4470  -1.7099  0.7540
## allocsystematic   0.1042  0.6822   0.1528  0.8786  -1.2329  1.4414
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Das Maß R^2

- Wird genutzt, um den Anteil durch eine Kovariable erklärte Varianz an der Gesamtvarianz zu beschreiben

- Im Kontext einer Metaanalyse wird R^2 über die wahre Varianz τ^2 definiert als:

$$R^2 = 1 - \left(\frac{T_{within}^2}{T_{total}^2} \right)$$

- Voraussetzung: gleiches τ^2 für alle Subgruppen

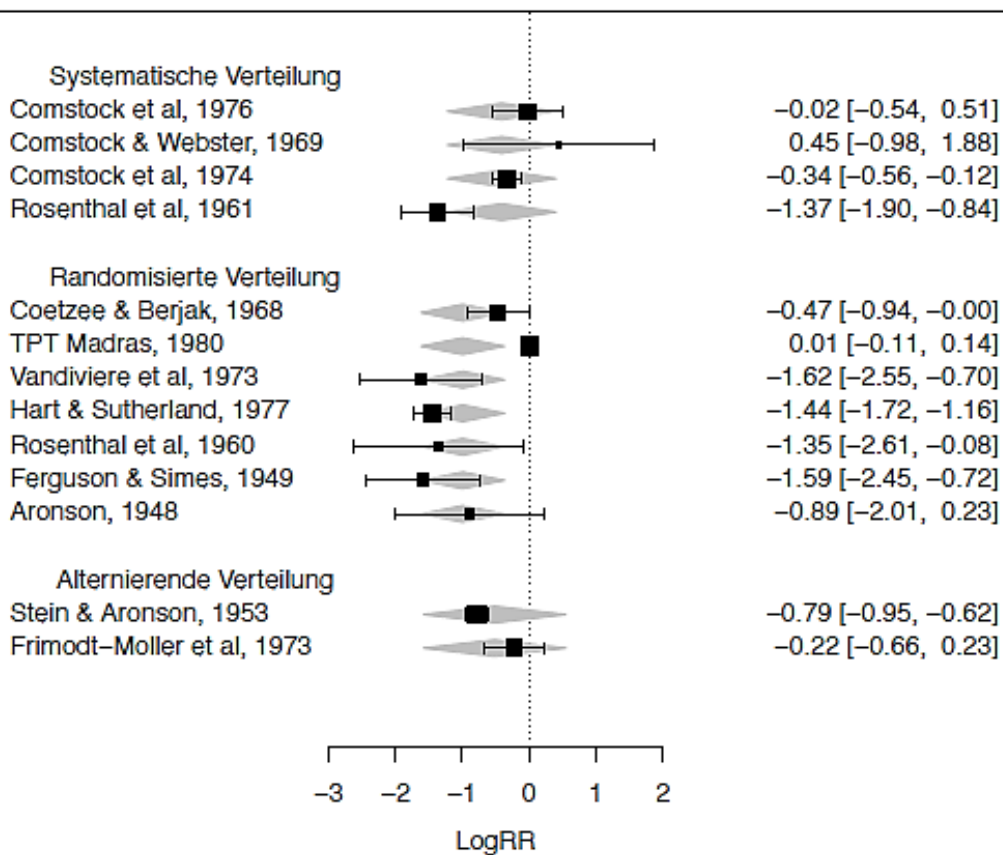
$$R^2 = 1 - \left(\frac{0.0097}{0.0299} \right) = 0.6745$$

67 % der gesamten Varianz können durch die Gruppenzugehörigkeit erklärt werden.

Forest Plot dazu darstellen

```
forest(res_bcg.RE_alloc
  # Ordnen nach Behandlungsschema
  ,order=order(dat.bcg$alloc)
  # vertikale Begrenzung der Plot Region
  ,ylim=c(1, 24)
  # vertikale Angabe der Position für die Subgruppen
  ,rows=c(3:4,7:13,16:19)
  # x-Achsen Beschriftung
  ,xlab="LogRR")

# Beschriftungen für die Subgruppen
text(-5 # horizontale Positionierung (linker Rand)
  ,c(20,14,5) # vertikale Positionierung
  ,c("Systematische Verteilung",
     "Randomisierte Verteilung",
     "Alternierende Verteilung"))
```



Graue Raute = Mittelwertschätzung

Metaregression

- Einzelne Studien treten an der Stelle von individuellen Messwerten
- Untersuchung des Einfluss von Kovariablen auf die Effektstärken einzelner Studien
 - Möglichkeit beobachtete Heterogenität zu erklären
 - Handelt sich im Gegensatz zur Subgruppenanalyse um metrische Kovariablen
- Der Operator „mods“ in der rma-Funktion kann genutzt werden, um metrische Kovariablen zu berücksichtigen

```
# Berechnung des rma-Objekts
res_bcg.RE_abl <- rma(yi, vi
                    ,mods = ~ablat # Angabe des Moderators
                    ,data = bcg_RR # Datenquelle
                    ,method="DL")
# Anzeigen des rma-Objekts
res_bcg.RE_abl

##
## Mixed-Effects Model (k = 13; tau^2 estimator: DL)
##
## tau^2 (estimated amount of residual heterogeneity):      0.0633 (SE = 0.0548)
## tau (square root of estimated tau^2 value):             0.2516
## I^2 (residual heterogeneity / unaccounted variability): 64.21%
## H^2 (unaccounted variability / sampling variability):    2.79
## R^2 (amount of heterogeneity accounted for):             79.50%
##
## Test for Residual Heterogeneity:
## QE(df = 11) = 30.7331, p-val = 0.0012
##
## Test of Moderators (coefficient 2):
## QM(df = 1) = 18.8452, p-val < .0001
##
## Model Results:
##
##           estimate      se    zval    pval    ci.lb    ci.ub
## intrcpt    0.2595  0.2323   1.1172  0.2639  -0.1958  0.7149
## ablat     -0.0292  0.0067  -4.3411 <.0001  -0.0424 -0.0160 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Individuelle Patientendaten (IPD) Meta Analyse

- Zugriff auf individuelle Patientendaten, die in allen Studien zugrundeliegen
 - Goldstandard
 - Ermöglicht Re-Analysen der Studien
 - Höhere Qualität, Kontrolle und Korrektur möglich
 - Flexiblere Auswertung,
 - Jedoch: selten durchführbar, Ressourcenintensiv, höhere Anforderung Datenaustausch und -sicherheit, methodisch anspruchsvoll

Auswertung:

- I. Re-Analyse der einzelnen Studien nach einheitlichen Protokoll
- II. Dann: fixed- oder random effects Metaanalyse

Grundlagen der Versuchsplanung

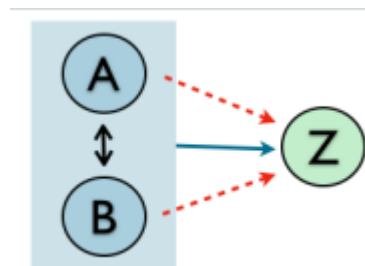
- Grundlage jeder Studie ist eine klare Fragestellung
 - Diese muss in eine wissenschaftliche Hypothese überführt werden, um sie statistisch untersuchen zu können
- Hypothese = klare Beschreibung für eine potentielle Erklärung des betrachteten Phänomens, die es zulässt überprüfbare Vorhersagen abzuleiten
 - Präzise Formulierung
 - Daraus lässt sich eine testbare Vorhersage ableiten

Einfluss sportlicher Betätigung auf die Gesundheit

- **Wissenschaftliche Hypothese:** Aktive sportliche Betätigung verbessert den allgemeinen Gesundheitszustand.
- **Vorhersage (1):** Mitglieder in Sportvereinen gehen weniger häufig zum Arzt als der Durchschnitt der Bevölkerung.
→ *"schwache" (unspez.) Vorhersage, da viele Gründe für Zutreffen der Vorhersage in Frage kommen (z.B. Alter, Einstellung zum Sport)*
- **Vorhersage (2):** Teilung der Teilnehmer eines Vorsorgeprogrammes in 2 Gruppen: a) Teilnahme an regelmäßigen Aktivitäten (z.B. Laufen, Schwimmen), b) kein spezifisches Programm. In Gruppe 1 gibt es eine positivere Veränderung individueller Kenngrößen (z.B. BMI)
→ *spezifischere Vorhersage: durch Intervention und anzunehmende Homogenität der Gruppen, liefert Test der Vorhersage mehr Informationen bzgl. der Hypothese*

Confounding

- Hintergrund: Komplexe Systeme und multiple Abhängigkeiten/Wechselwirkungen zwischen Variablen
- Zwei Einflussfaktoren sind confounded (miteinander verwoben), wenn das experimentelle Design es nicht erlaubt, die isoliert hinsichtlich ihrer Zielgröße zu untersuchen

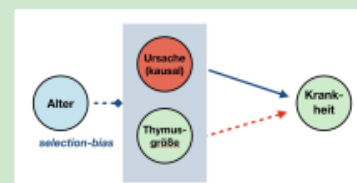


Confounding durch Selection bias:

- Miteinander verweben unabhängiger Zielgrößen durch falsche Auswahl der Proband:innen

Beispiel: Thymusentfernung bei Atemwegsproblemen

- Um 1900 gab es die Empfehlung, bei Atemwegsproblemen von Kleinkindern den Thymus zu entfernen
- Zugrunde liegende Beobachtung: Kleinkinder mit Atemwegsproblemen haben größeren Thymus als gesunde Erwachsene



Exploration vs. Konfirmation

- Exploration = Aufdeckung wichtiger unbekannter Varianzquellen und Suche nach neuen Effekten → Hypothesengenerierung
- Konfirmation = Bestätigung dieser und statistische Absicherung → Hypothesenprüfung

Beobachtungs- vs. Interventionsstudie

Beobachtungsstudie	
Vorteile <ul style="list-style-type: none">• Einfacher, weniger aufwendig• Keine externe "Beeinflussung"• <i>natürliche</i> Varianz	Nachteile <ul style="list-style-type: none">• Kausalitäten nicht nachweisbar• Unkontrollierte Effekte• Gefahr des Confounings

Interventionsstudie	
Vorteile <ul style="list-style-type: none">• gezielte Kontrolle ausgewählter Einflussgrößen• Kausalitäten u.U. nachweisbar	Nachteile <ul style="list-style-type: none">• Oft aufwendiger• Gefahr "unnatürlicher" Effekte• Intervention z.T. nicht möglich

Pilotstudien

- Spezialfall explorativer Studien
 - Interventionsstudien zur Auswahl informativer Interventionen (z.B. mit geringer Fallzahl)
 - Zur Vorbereitung konkreter konfirmatorischer Studien
 - Überprüfung Fragestellung, Design, Messtechnik, Verteilungseigenschaften etc.

Variabilität

- Kann durch geeignete Planung und Analyse von Studien so zerlegt werden:
 - Dass reale Effekte
 - Von ungewollten (zufälligen) Einflüssen anhand ihrer Varianzkomponenten unterschieden werden

3 Quellen der Variabilität

- I. Gewollt – durch die untersuchten Bedingungen
- II. Ungewollt – durch den Messprozess
- III. Ungewollt – Im Untersuchungsmaterial/-prozess

Untersuchung des Einflusses akustischer Reize auf die Aufmerksamkeit von Probanden anhand von EEG Messungen

- 1 Variabilität der EEG Werte bei verschiedenen Reizen (*gewollt*)
- 2 (Technische) Variabilität der EEG Messungen (*ungewollt*)
- 3 Variabilität des Ansprechens individueller Probanden (*ungewollt*)

3 Arten von Variabilität

1. Gewollt – Geplant, systematische Variabilität
 - i. z.B. Zuordnung mit erwarteter Ähnlichkeit

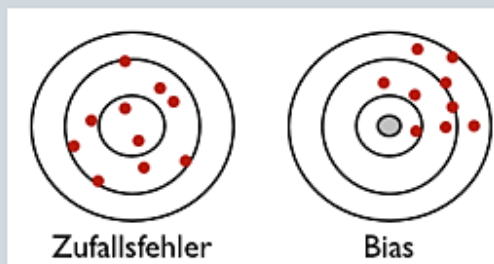
Beispiel: Überleben bei Bronchialkarzinom unter Therapie A und B

- Gezielte Zuordnung von Patienten in Therapiearm A bzw. B

2. Ungewollt jedoch unkritisch – Zufällige, nicht-systematische Variabilität, z.B. individuelle Unterschiede der Untersuchungsobjekte
 - i. Zufallsfehler
 - ii. „Within treatment“ Variabilität, „noise“, Messfehler
 - iii. Zufallscharakter, Möglichkeit Schätzung der Größe des Zufallsfehlers bei korrektem Design
3. Ungewollt, jedoch gefährlich da Verzerrungsgefahr – ungeplant systematische Variabilität
 - i. Durch unbekannte Störgrößen, systematische Verzerrung (Bias)
 - ii. Gefahr falscher Ergebnisse und Schlussfolgerungen

Unterschiede von Zufallsfehler und Bias

- Sowohl Bias als auch Zufallsfehler führen zu Abweichungen der Beobachtung vom "wahren" Wert
- Allerdings mitteln sich diese Abweichungen im Rahmen vieler Beobachtungen beim Zufallsfehler aus, wohingegen dies beim Bias nicht der Fall ist



Ziele der Versuchsplanung

- I. Vermeidung/Minimierung ungeplanter systematischer Variabilität (Bias, Confounding)
- II. Minimierung zufälliger, nicht-systematischer Variabilität
- III. Effiziente Schätzung spezifischer Effekte

Designprinzipien zur Umsetzung der Ziele

Verringerung des Zufallsfehlers

→ Replikation

Vermeidung / Minimierung ungeplanter systematischer Verzerrungen

→ Randomisierung

→ Blockbildung (blocking)

Effizienter Vergleich von mehr als zwei Versuchsbedingungen

→ Faktorielle Kombination

1. Replikation

= Ein Mittel zur Berechnung des Zufallsfehlers

→ Dafür werden jedoch echte Beobachtungen (statistisch unabhängig voneinander) benötigt

Statistisch abhängige Replikate = **Pseudoreplikate**

- Führen zu fehlerhaften Resultaten und Schätzungen der Varianz
- Wenn biologische und/oder technische Variabilität nicht abgebildet werden kann

Häufige Ursachen für Pseudoreplikaten

- Verwandtschaftsbeziehungen
- Mehrfachmessungen an identischen Objekten
- Zeitverlaufsmessungen
- Gemeinsamer Kontext von Beobachtungen (z.B. soziales Umfeld, Einzugsgebiet von Kliniken)
- Induzierte Pseudoreplikation (z.B. identischer Stimulus für verschiedene, andernfalls unabhängige Probanden innerhalb der Versuchsgruppen → mögliches Confounding des Stimulus mit unbekanntem Effekten)

Umgang mit Pseudoreplikaten

- Mittellung der "inneren" Replikate (einfach, aber nicht optimal)
- Vermeidung von Pseudoreplikaten durch geeignete Versuchsplanung (erfordert u.U. größeren Aufwand)
- Anwendung geeigneter statistischer Modelle, die die Abhängigkeitsstruktur der Messungen korrekt berücksichtigen; erfordert geeigneten Versuchsplan (siehe Blockbildung) und ggf. Unterstützung durch Biometriker

2. **Randomisierung**

= zufällige Zuordnung von Objekten zu Studiengruppen (jedes Objekt gleiche Wahrscheinlichkeit für eine bestimmte Gruppe ausgewählt zu sein)

- Konvertiert ungeplante, systematische in zufällige, nicht-systematische Variabilität → diese kann durch Replikation begegnet werden
- Gewährleistung Stukturgleichheit
- Verminderung Bias und Confounding

Fehlerquellen:

- Willkürliche (nicht zufällige) Zuweisung
- „Selbstausswahl“

3. **Blockbildung**

= Zusammenfassung ähnlicher Objekte zu Blöcken (Einführung Blockvariablen)

- Zuweisung der Objekt zu den Versuchsbedingungen separat in den Blöcken
- Konvertiert ungeplante, systematische in geplante systematische Variabilität
- Aufteilung in „within block“ und „between block“ Varianz, erlaubt effizientere Einschätzung des Gruppeneffektes innerhalb der Blöcke
- Macht Confounder/Biasquellen explizit und kann dadurch Fehlinterpretation verhindern

Beispiel: Einfluss eines Rehabilitationsprogrammes auf Hospitalisierungsdauer

- Hypothese: Teilnahme an Reha-Programm senkt Hospitalisierungsdauer
- Studienkollektiv: Alle Patienten mit bestimmter Diagnose/Behandlung
- Es ist bekannt, dass das Alter bei Diagnose einen erheblichen Einfluss auf Hospitalisierungsdauer hat (unabhängig von Reha)
- vollständige Randomisierung auf die Gruppen Reha ja / nein ist ggf. nicht optimal, da das Alter eine zusätzliche Variabilität induziert, die einen ggf. vorhanden Effekt verdecken könnte und derer Einfluss durch die Randomisierung u.U. nicht vollständig vermieden werden kann
- sinnvoll: Verwendung von Altersgruppen als Blöcke, innerhalb derer dann randomisiert wird

Nachteile Blockbildung

- Verlust statistischer Power
 - Wenn Blöcke nach Variablen ohne Effekt (kaum Varianz erklären) gebildet werden
 - Blockbildung nur bei hinreichender Evidenz für Effekt der Blockvariable
- Leerbleiben von Blöcken bei kleiner Fallzahl
 - unvollständige Versuchspläne
 - Nutzen Blockbildung steigt mit höheren Fallzahlen
- Gepaarte Designs (gepaarter t-Test) = Spezialfall eines Block-Designs
- Wenn Block-Variable kontinuierlich, macht es unter Umständen auch Sinn diese als Covariable zu betrachten

Vollständig randomisiertes 1-Faktor Design

Beispiel: BF[1] in interventioneller Studie

- Untersuchung von 2 neuen physiotherapeutischen Behandlungsstrategien (A und B), sowie einer Kontrolle (d.h. Standardbehandlung, C) in der Rehabilitation nach Oberschenkelfrakturen bezüglich der Dauer des stationären Aufenthaltes
- Alle in die Studie eingeschlossenen Patienten werden zufällig auf die 3 Gruppen verteilt:

Gruppe	Datenpunkte (Stationärer Aufenthalt in Tagen)
A	15 17 21 16 13 14 21 18 16 14
B	16 12 18 18 19 15 16 20 17
C	18 21 19 14 24 17 18 20 17 16

Unbalanciert

Beispiel: BF[1] in Beobachtungsstudie

- Erhebung und Vergleich des Hygienestatus von drei Großküchen (I, II, III) anhand der Anzahl Keim-belasteter (*positiver*) Proben; als *positiv* gilt eine Probe, wenn in ihr eine vorgegebene Keimzahl von Enterobakterien überschritten ist
- Bestimmung der Anzahl *positiver* Proben (aus jeweils 20 Proben) in je 8 zufällig ausgewählten belieferten Einrichtungen:

Gruppe	Datenpunkte (pos. Test)
I	0 1 0 2 0 2 3 0
II	0 0 1 1 0 3 0 0
III	1 1 3 0 2 0 1 0

Balanciert

Balancierung

- Balanciertes Design = wenn alle Faktorstufen die gleiche Anzahl von Beobachtungen zugewiesen wird
 - Designs sollten möglichst balanciert sein
 - Ermöglicht bessere statistische Eigenschaften
 - Weniger anfällig Verletzungen von Verteilungsvoraussetzungen
 - Höhere Power

Partitionierung

= Methode Messwerte in Gruppen einzuteilen, sodass jeder Messwert zu einer und kein Wert zu mehr als eine gehört

- Ein **(Struktur-)Faktor** ist eine sinnvolle Partionierung

Erkennen von (Struktur-)Faktoren in einem Versuchplan

- 1 Teilen die Messwerte/Beobachtungen in den Gruppen der Partitionierung Eigenschaften, die von den Messwerten/Beobachtungen in anderen Gruppen nicht geteilt werden?
- 2 Ist es sinnvoll innerhalb der Gruppen Mittelwerte zu bilden und diese zu vergleichen?
- 3 Führt eine Vertauschung der Gruppenzugehörigkeit zu einer Veränderung des Sinnes der Daten?

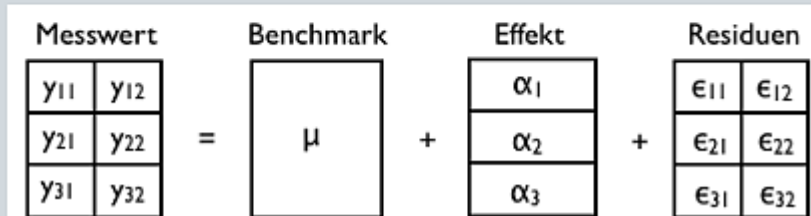
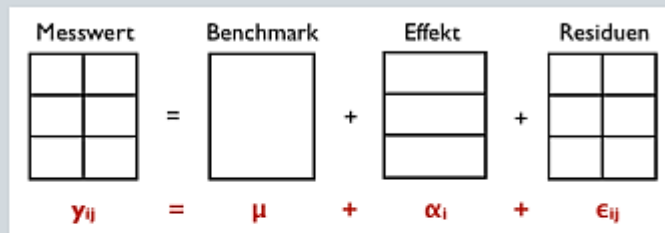
Im Falle einer positiven Antwort ("ja") auf alle drei Fragen handelt es sich bei der Partitionierung um einen (Struktur-)Faktor.

Beispiel: "Physiotherapie bei Oberschenkelfraktur"

Gruppe	Datenpunkte									
A	15	17	21	16	13	14	21	18	16	14
B	16	12	18	18	19	15	16	20	17	
C	18	21	19	14	24	17	18	20	17	16

- Die Zeilen generieren einen Faktor (gleiche Behandlung; Vergleich von Zeilenmitteln sinnvoll; Gruppenzugehörigkeit von Daten nicht sinnvoll vertauschbar)
- Die Spalten hingegen generieren keinen Faktor (Werte in gleichen Spalten haben nicht mehr gemein als Werte in verschiedenen Spalten; Vergleich von Spaltenmittel ist wenig sinnvoll; Spaltenzugehörigkeit ist prinzipiell vertauschbar)

Zuordnung des ANOVA Modells zu den (Struktur-)Faktoren



Vollständig randomisiertes 2-Faktor Design

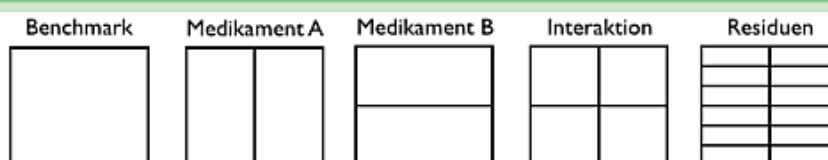
- Effiziente Betrachtung von 2 oder mehr Einflussfaktoren gleichzeitig (multivariate Analyse)
- Erlaubt Schätzung von Wechselwirkungen
 - Setzt Vorliegen von Replikaten in den Faktorkombinationen voraus
- Jedoch: Komplexes Design, schwer zu kommunizieren und interpretieren

Beispiel: BF[2] zum Studium der Wirkung von zwei Chemotherapeutika

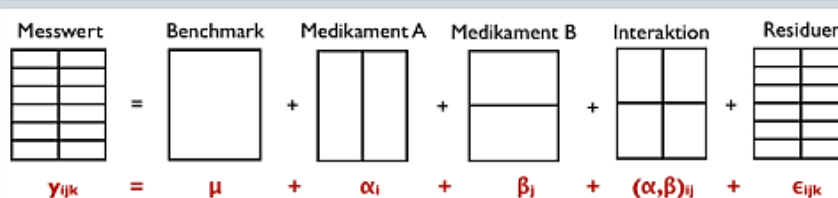
- Anwendung von zwei Medikamenten (A, B), wobei sowohl die isolierte Wirkung der jeweiligen Dosis als auch eine mögliche Interaktion hinsichtlich der Zeit bis zur Remission (in Wochen) geprüft werden soll
- Alle in die Studie eingeschlossenen Patienten werden zufällig auf die 4 Faktorkombinationen verteilt:

		Medikament A	
		10mg	40mg
Medika- ment B	100mg	22, 24, 30	25, 35, 22
	200mg	22, 28, 33	30, 35, 38

Beispiel: "Kombi-Chemotherapie"



Zuordnung des ANOVA Modells zu den (Struktur-)Faktoren



Störgrößen/Nuisance-Variablen

= alle Arten der Variabilität, die nicht durch die interessierenden Versuchsbedingungen verursacht werden

- Unkontrolliert können sie zu Verzerrungen (Bias) führen
- Blockbildung machen Nuisance-Effekt explizit (integriert als Blockvariable im Erklärungsmodell)

Vollständig geblocktes Design

= Zuordnung von „ähnlichen“ Beobachtungseinheiten zu Gruppen (Blöcke)

- Innerhalb dieser Gruppen werden die Faktorstufen erneut zufällig verteilt

Vorteile:

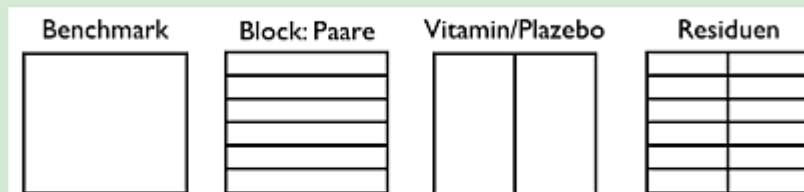
- Erlaubt bessere Schätzung der Unterschiede zwischen interessierenden Faktoren (Varianzkomponente der Blockfaktoren wird separiert)
- Effizient bei hoher inter-individueller Variabilität der Gesamtstichprobe

Nachteile:

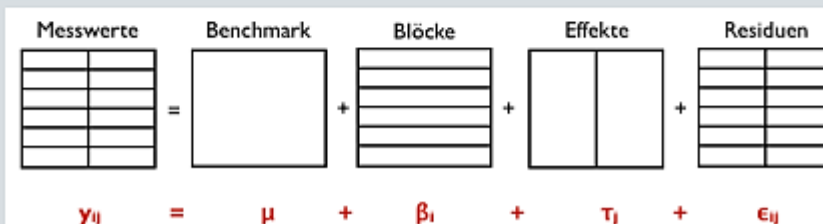
- Geblocktes Design weniger effizient als Basic faktorielle Designs (wenn Blockfaktoren keine/wenig Varianz erklären)
- Aufwändige Planung

Faktorstruktur im CB Design

Beispiel: "Vitamin B6 Effekt"



Zuordnung des ANOVA Modells zu den (Struktur-)Faktoren



Latin Square Design

- Beim Auftreten von zwei Nuisance und einem interessierenden Faktor
- Wenn Anzahl der Level für alle drei Faktoren identisch
- Alle Paare von Faktoren faktoriell kombiniert (gekreuzt) auftreten

Realisierung:

- Erster Nuisance-Faktor als Spaltenblock
- Zweiter Nuisance-Faktor als Zeilenblock
- Aufteilung der Stufen des interessierenden Faktors genau einmal pro Zeile und Spalte jeder Stufe

Beispiel: Untersuchung der Wirksamkeit verschiedener Konzentrationsübungen auf Konzentrationsfähigkeit von Patienten mit Aufmerksamkeits Defizit Hyperaktivitäts Störung (ADHS)

- Ausgangssituation:
 - Betrachte 3 Übungen zur Konzentrationssteigerung: Ü1, Ü2, Ü3
 - Es stehen 3 Patienten zur Untersuchung zur Verfügung
 - Die Untersuchungen müssen alle an einem Tag durchgeführt werden
- Zielstellung / Probleme:
 - Ziel: Vergleich der drei Übungen bzgl. der Konzentrationssteigerung
 - Probleme:
 - Hohe Variabilität der Ausgangskonzentration zwischen Patienten
 - Abnahme der Konzentrationsfähigkeit im Verlauf des Tages

Beispiel: ADHS - Fortsetzung (3)

Designvorschlag 3:

LS Design (mit Blockvariablen Patient, Tageszeit)

Patient	Tageszeit		
	vormittags	mittags	nachmittags
1	Ü1 (A)	Ü2 (B)	Ü3 (C)
2	Ü2 (B)	Ü3 (C)	Ü1 (A)
3	Ü3 (C)	Ü1 (A)	Ü2 (B)

Bemerkung:

- Nutzt Vorteile eines CB Designs (d.h. Trennung der Varianzkomponenten von interessierender und Nuisance-Variablen) und gewährleistet Balanciertheit / Vermeidung eines Confoundings

Randomisierung eines LS Designs

- 1 Zufällige Zuordnung des Subjekte/Objekte zu Zeilennummern
- 2 Zufällige Zuordnung der Zeitpunkte zu Spaltennummern
- 3 Zufällige Zuordnung der Faktorstufen zu den LS Buchstaben

Beispiel: "ADHS"

- 1 z.B.: 1 - Patient 2; 2 - Patient 1; 3 - Patient 3
- 2 z.B.: 1 - nachmittags; 2 - vormittags; 3 - mittags
- 3 z.B.: A - Übung 2; B - Übung 1; 3 - Übung 3

Patient	Tageszeit		
	nachmittags	vormittags	mittags
2	A: Ü2	B: Ü1	C: Ü3
1	B: Ü1	C: Ü3	A: Ü2
3	C: Ü3	A: Ü2	B: Ü1

Beispiel: "ADHS"

Benchmark	Block 1: Patient	Block 2: Zeit	Übungseffekt	Residuen
			A B C	
			B C A	
			C A B	

Zuordnung des ANOVA Modells zu den (Struktur-)Faktoren

Messwerte	Benchmark	Block Nuisance I	Block Nuisance II	Faktorvariable	Residuen
				A B C	
				B C A	
				C A B	

$$y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \epsilon_{ijk}$$

Vorteile Latin Square:

- Geeignet bei hoher Variabilität der Subjekte/Objekte
- Zur Balancierung der Versuchsbedingungen, wenn die Reihenfolge der Bedingungen einen systematischen Effekt induziert

Split-Plot Design

= Verzicht auf vollständige faktorielle Randomisierung, stattdessen eine vorgegebene interne Struktur, die eine vollständige Randomisation ausschließt

- Wenn ein Faktor sehr aufwendig zu variieren ist

Motivations-Beispiel: Bakterienwachstum in Abhängigkeit von Nährmedium und zugesetzten Desinfektionsmitteln (2)

→ Vollständig randomisiertes faktorielles Design:

A	B	B	A	A	B	B	A
A	A	A	B	B	A	B	B

→ SP Design:

A	B	B	A	B	A	B	A
A	B	A	B	B	A	A	B

Realisierung:

- Wahl des Blockfaktor → Zufällige Zuordnung der Stufen des Blockfaktors
- Zufällige Zuordnung der Stufen des Inner-Block Faktors

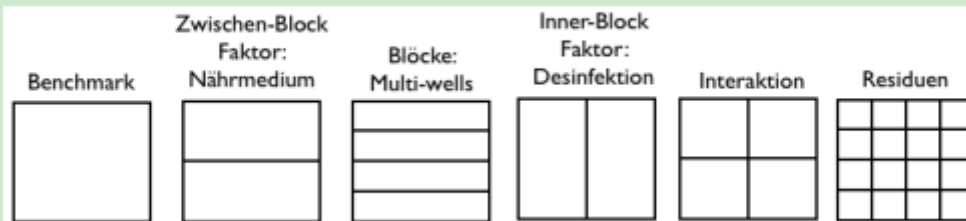
Beispiel: Realisierung eines SP Designs im "Bakterienkulturen"-Beispiel

- Wahl des Nährmediums als Blockfaktor (effiziente Verwendung von Multi-Wells innerhalb eines Kulturansatzes)
- Zufällige Zuordnung von Nährmedium I (orange) und II (blau) zu Multi-Wells

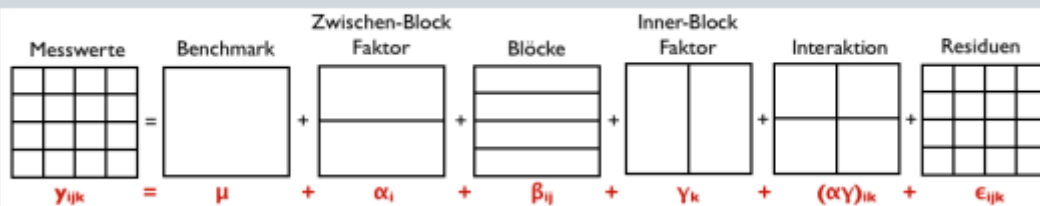
- Zufällige Zuordnung von zwei Repliaten je eines Desinfektionsmittels (A bzw. B) innerhalb der Blöcke

B	A	A	B	B	A	B	A
B	A	A	B	A	B	A	B

Beispiel: "Bakterienkolonien"



Zuordnung des ANOVA Modells zu den (Struktur-)Faktoren



Vorteile:

- Effizienter Umgang mit schwer variierbaren Blockfaktoren und hierarchischen Faktorstufen und (intra-individuellen) Mehrfachmessungen
- Erlauben Komplementierung faktorieller Designs mit Blockstrukturen
- Jedoch: Komplexe Designstruktur

Faktorielle Kombination & Nesting (Schachtelung) im Split Plot Design:

Faktorielle Kombination

- Alle Stufen von Faktor I treten gemeinsam mit allen Stufen von Faktor 2 auf
→ Möglichkeit Interaktionen zu schätzen

Faktor I	Faktor II		Kombination	
1	A	B	1A	1B
2	A	B	2A	2B

Nesting (Schachtelung)

- Alle Stufen von Faktor 2 treten genau mit einer Stufe von Faktor 1 auf
- Stufen von Faktor I nur einer Teilmenge der Stufen von Faktor II (Schachtelung)
→ Keine Schätzung der Interaktion von Faktor I & II möglich

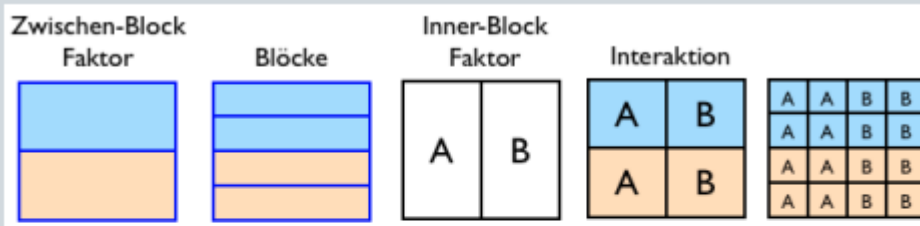
Faktor I	Faktor II	Kombination
1	A	1A
	B	1B
	C	2C
	D	2D
2		

Vollständiges Confounding

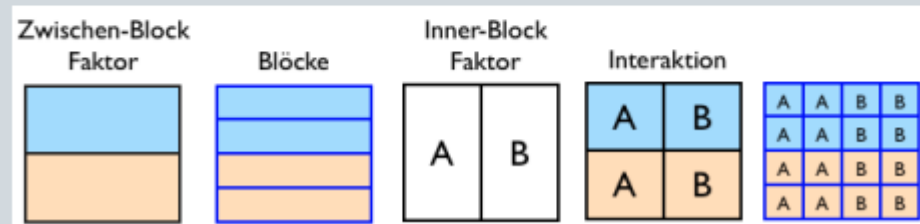
- Faktor I ist bzgl. Faktor II „nested“ und umgekehrt
→ Effekt von Faktor I & II sind nicht zu trennen

Faktor I	Faktor II	Kombination
1	A	1A
2	B	2B

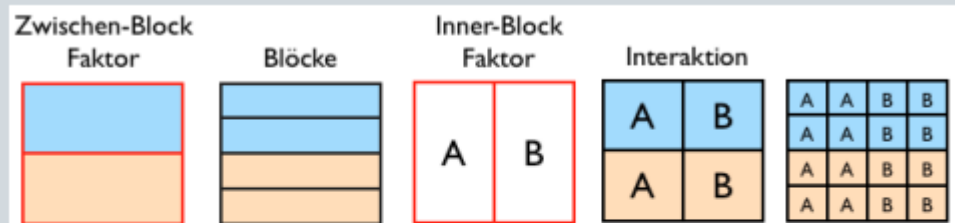
Blöcke sind "nested" innerhalb der Stufen des Zwischen-Block-Faktors



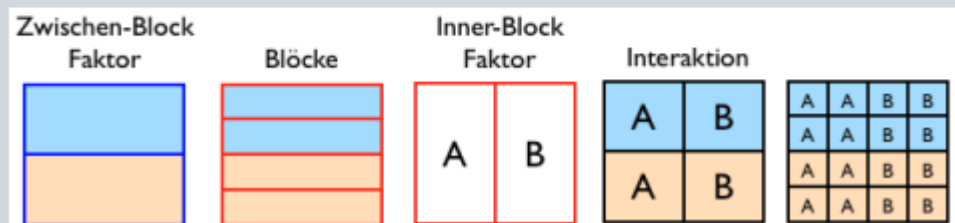
Subplot sind "nested" innerhalb der Blöcke



Zwischen-Block und Inner-Block Faktoren sind faktoriell kombiniert



Blöcke und Inner-Block Faktoren sind faktoriell kombiniert



Fehlzahlplanung

Vorgabe von Fehlerniveau

- Unabhängig von der Fallzahl, wird durch den Vergleich des P-Wertes mit einem vorgegebenen Signifikanzniveau der Fehler 1. Art kontrolliert.
- **Aber:** Fehler 2. Art (bzw. damit auch die Power) hängt unter anderem von der Stichprobengröße ab!

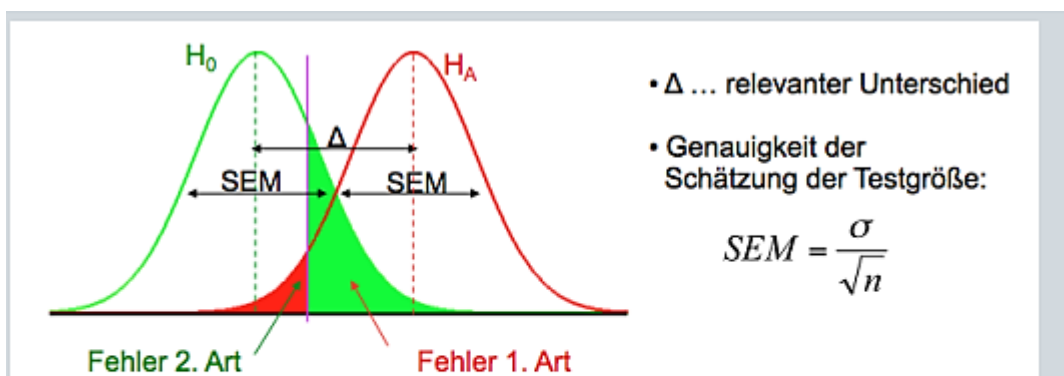
		Realität	
		H ₀ falsch	H ₀ richtig
Test-entscheidung	H ₀ ablehnen	Power: 1 - β	Fehler 1. Art: α OK
	H ₀ beibehalten	Fehler 2. Art: β ?	1 - α

Fehler 1. Art (Alpha):

- Beschreibt den Fall, dass die Nullhypothese zurückgewiesen wird, obwohl sie wahr ist (Alternativhypothese fälschlich richtig) = dem Signifikanzniveau des Tests

Fehler 2. Art (Beta)

- Nullhypothese wird abgelehnt, obwohl sie in Wahrheit richtig ist (Alternativhypothese fälschlicherweise abgelehnt)
 - ➔ Fehler II. Art und somit auch die Power hängt unter anderem von der Stichprobengröße ab!
- Beide Fehler können nicht gleichzeitig minimiert werden
- Fehlerraten sind abhängig von der Variabilität der Daten (SEM) und dem relevanten Unterschied



➔ Fehlerraten sind abhängig von Variabilität der Daten σ , von Fallzahl n (da diese die Genauigkeit der Schätzer beeinflusst) und vom zu entdeckenden Unterschied Δ (d.h. von der Wahl von H_0 und H_A)

Fallzahlplanung

- Möglichst Fehler 1. Art so, dass Signifikanzniveau einem Wert nicht übersteigt und Fehler II. Art Wert nicht übersteigt, d.h. Power eine Mindestgröße aufweist

Fallzahl steigt:

- Mit kleinerem Signifikanzniveau/höherem Power
- Kleineren relevanten Unterschied Delta
- Größerer Varianz der Endpunkts
- Höheren Drop-Out Raten

Kriterien der Fallzahlplanung:

- Welcher primärer Endpunkt
- Statistische Auswertungsmethode
- Größe des Effekts/relevanten Unterschieds
- Varianz der Daten
- Größe der Fehlerraten (1./2. Art) für die Studie

Konzepte der Fallzahlberechnung

1. Präzisions-Analyse

- Bei Schätzung von Parametern mittels Intervallschätzern
- Schätzung der Fallzahl um einen Parameter mit vorgegebener Fehlerwahrscheinlichkeit hinreichend präzise zu schätzen

Beispiel: 95%-KI für den Erwartungswert einer normalverteilten Zufallsvariablen mit bekannter Streuung σ^2

- Gesucht: Fallzahl, die es gestattet den Erwartungswert auf eine Genauigkeit (E) von $\pm 0.1 \cdot \sigma^2 = 0.2$, bei gegebenem Konfidenzniveau $1 - \alpha = 0.95$ zu schätzen. Die Streuung habe den Wert $\sigma^2 = 2$.

- Lösung:

$$E = z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{\sqrt{2}}{\sqrt{n}} = 0.2$$

und damit

$$n = z_{1-\frac{\alpha}{2}}^2 \cdot \frac{\sigma^2}{E^2} = 1.96^2 \cdot \frac{2}{0.2^2} = 192.07$$

- D.h. es sind 193 Fälle notwendig.

2. Power-Analyse

Beispiel: Einseitiger Signifikanztest zum Test der Hypothese bzgl. zweier Erwartungswerte ($H_0 : \mu_1 < \mu_2$) bei bekannter, gemeinsamer Streuung

- Gegeben:
 - Gewünschtes Signifikanzniveau $\alpha = 5\%$
 - Angestrebte Power $1 - \beta = 80\%$
 - Zu detektierender Unterschied der Erwartungswerte, $\Delta = 2$
 - Gemeinsame Streuung in beiden Populationen, $\sigma^2 = 9$
- Gesucht: notwendige Fallzahl (gleiche Gruppengröße: $n_1 = n_2 = n$)
- Lösung:
 - $z_{1-\alpha} = z_{0.95} = 1.6449$
 - $z_{1-\beta} = z_{0.80} = 0.8416$

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2 \cdot 2}{(\Delta/\sigma)^2} = \frac{(1.6449 + 0.8416)^2 \cdot 2}{(2/3)^2} = 27.8$$

- D.h. es sind 28 Beobachtungen pro Gruppe notwendig.

Beispiel: Einseitiger Signifikanztest zum Test der Hypothese bzgl. zweier Erwartungswerte ($H_0 : \mu_1 < \mu_2$) bei unbekannter, gemeinsamer Streuung

- Gegeben (wie bisher):
 - $\alpha = 5\%$, $1 - \beta = 80\%$, $\Delta = 2$
 - Gemeinsame Streuung wird geschätzt: $\hat{\sigma}^2 = s^2 = 9$
- Berechnung der notwendigen Fallzahl (gleiche Gruppengröße):

$$n = \frac{(t_{1-\alpha,df} + t_{1-\beta,df})^2 \cdot 2}{(\Delta/\sigma)^2}$$

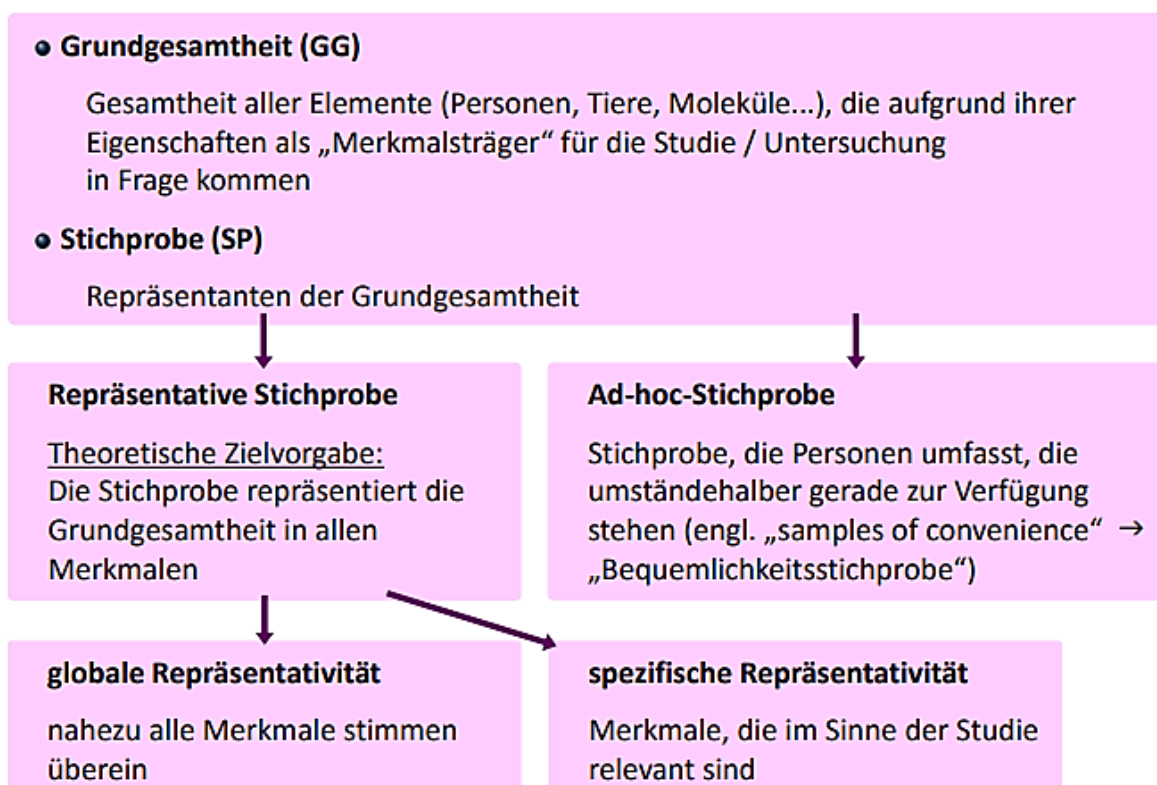
- Problem: $df = n_1 + n_2 - 2$ hängen von unbekannter Fallzahl ab!
 - Damit Bestimmung der Fallzahl nur iterativ möglich
- Anwendung eines Fallzahlprogrammes:
Hier sind 29 Beobachtungen pro Gruppe notwendig.

Sampling Strategien und Stichprobenziehung

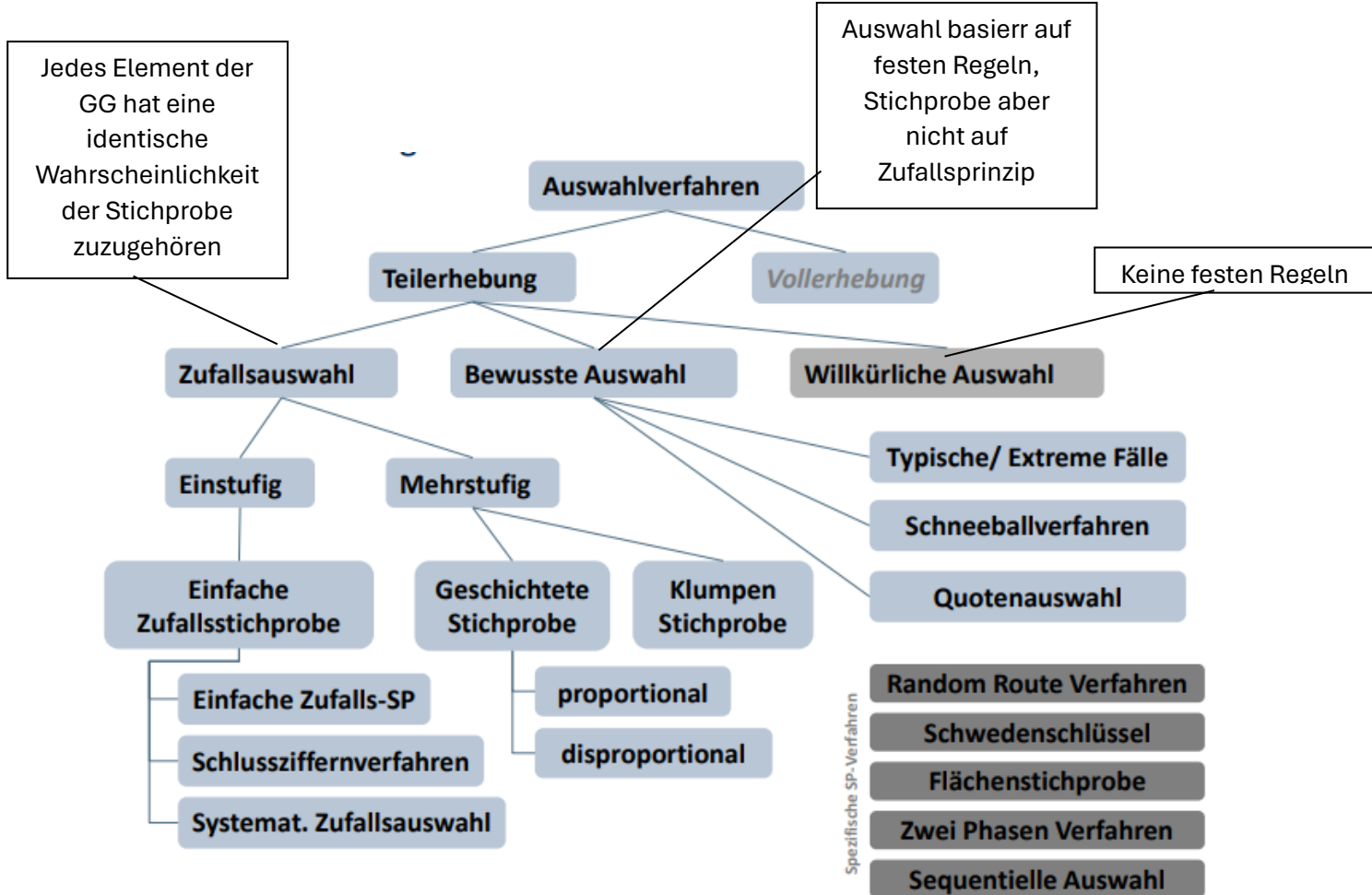
- Stichproben notwendig, um mittels konfirmatorischer Statistik Rückschlüsse auf die Grundgesamtheit abzuleiten
 - Wichtig dabei: Auf welche Grundgesamtheit bezieht sich die Fragestellung (Ein- und Ausschlusskriterien)

Kriterium: Machbarkeit

- Grundgesamtheit bekannt?
 - Ausreichende Anzahl und Chance Stichprobenziehung
- Anteil Teilnehmenden an der Zielpopulation
 - Wahrscheinlichkeit des Drop-Outs, Gefahr Unterpowered?
- Ausschlusskriterien
 - Ggf. zu viele, dass die Stichprobe nur eine Teilpopulation darstellt
- Berechnete Stichprobengröße
 - Machbar (z.B. bei seltenen Erkrankungen)
 - Finanzierbar



- Fast alle klinischen Studien basieren auf Ad-hoc Stichproben
- Randomisation schafft Strukturgleichheit zwischen Gruppen keine Repräsentativität
- Große Teilnehmerzahlen sorgen nicht für Repräsentativ (u.U. vergrößern sie den systematischen Fehler)
- Repräsentativ setzt genaueste Kenntnisse über Merkmalsverteilungen in der Grundgesamtheit zur Fragestellung voraus



Sampling Strategien für Zufallsauswahl

Vorbemerkungen

- **Voraussetzung:** Zugriff auf jedes Element der Grundgesamtheit
- **Achtung:** Bei Teilnahmeverweigerung kann eine zufällig ausgewählte Person nicht einfach durch eine andere ersetzt werden, ohne das Prinzip der zufälligen Auswahl zu verletzen!

Welche Verfahren sind möglich?

- **Einstufige Stichprobenauswahl :** Die Auswahlwahrscheinlichkeit ist für jedes Element gleich groß. Die Auswahl erfolgt in einem Schritt.
- **Mehrstufige Stichprobenauswahl aus Schichten:** Vor der Stichprobenauswahl wird die Grundgesamtheit in homogene Gruppen eingeteilt.
- **Mehrstufige Stichprobenauswahl aus Klumpen** (engl. Cluster): Vor der Stichprobenauswahl wird die Grundgesamtheit in homogene Klumpen eingeteilt, die in ihrer Zusammensetzung der Grundgesamtheit entsprechen. Als Stichprobe werden dann ein oder mehrere Klumpen zufällig ausgewählt.
- **Kombinierte Stichprobenverfahren**

1. Zufallsauswahl

Verfahren Einstufige Stichprobenauswahl

- I. **Einfache Zufallsstichprobe = simple random sampling**
(Urnen/Lotterieprinzip)
- II. **Schlussziffernverfahren**
(alle Probanden durchnummeriert und nummeriert, ausgewählt werden alle, die mit Endziffer x enden)
- III. **Systematische Zufallsauswahl (systematic sampling)**
(Zufällige Bestimmung eines ersten Probanden, dann k-Schrittweiten ($k = \frac{n}{n_{\text{Stichprobe}}}$) jeden weiteren Probanden)

Verfahren Mehrstufige Stichprobenauswahl aus Schichten

- Um Stichprobe möglichst genau zur Grundgesamtheit repräsentieren
 - Vorhandensein seltener Merkmale in ausreichender Menge
 - Sicherstellung Balanciertheit der gewählten Gruppen
 - Strata eng mit dem Untersuchungsziel zusammen
 - Z.B. Alter, Geschlecht, Wohnort, usw.
- 1. Vor der Stichprobe wird die Grundgesamtheit in homogene Strata eingeteilt
 - a) Proportionale Stichproben/Stratifizierung
 - Für jede Strata wird ein Anteil gezogen, der dem Anteil der Gruppe in der Grundgesamtheit entspricht
 - b) Disproportionale Stichproben/Stratifizierung
 - Für jede Strata wird ein anderer Anteil gezogen (z.B. um für Gruppenvergleiche gleich/genügend große Fallzahlen zu erhalten)
 - Nachteil: Repräsentativität geht ein Stück weit verloren

Mehrstufige Stichprobenauswahl aus Clustern

1. Grundgesamtheit wird vor Stichprobenauswahl in homogene Klumpen (Cluster) eingeteilt (kann auch mehrfach hintereinander erfolgen)
z.B. Stadtbezirke → Haushaltsarten → Zufällige Auswahl Personen

Kumulationsverfahren

- Zur Berücksichtigung der Einwohnerverteilung bei Bevölkerungsstichproben auf Clusterbasis
- Zerlegung in möglichst überschneidungsfreie und kumulierbare Teilstichproben (Netze) und darin enthaltener Sample Points

Spezifische Stichprobenverfahren

1. Random-Route Verfahren

- Startpunkt einer Route nach dem Zufallsprinzip und vorgegebene Laufroute mit Zufallsregeln z.B. geradeaus, 3. Haus rechts usw.
- Problem: Befragungsbias, weil besonders nicht erwerbstätige angetroffen werden

2. Schwedenschlüssel

- Zufallsauswahl von Befragungspersonen in Haushalten mit mehreren Personen
- Ziel: Jede Person im Haushalt soll die gleiche Chance erhalten
- Kriterien: Haushaltsgröße, Geschlecht, Alter der Personen
- Haushaltsmitglieder werden nach Alter und Geschlecht in eine Tabelle geordnet
- Interviewer besitzt „Schlüsselliste“ welche Person in welchem Haushalt befragt wird

3. Flächenstichprobe

- = spezifisches Clusterverfahren, besonders in Ländern ohne Einwohnerregister und Land-/Forstwirtschaft
- Aufteilung des Gebietes in Einzelflächen/Quadrate
- Per Zufall werden bestimmte Anzahl Quadrate ausgewählt in denen alle oder Stichprobe befragt wird

4. Zwei-Phasen Verfahren

- Bei Unklarer Grundgesamtheit
- 1. Phase: große Zufallsstichprobe → Merkmale der Grundgesamtheit
- 2. Phase: Merkmalsträger der 1. Stichprobe bilden Grundgesamtheit für die finale Stichprobe

5. Sequentielle Auswahl (Sequential sampling)

- Kein vorab festgelegter Stichprobenumfang
- Beginn mit kleiner Stichprobe und Festlegung nach Analyse, ob die Fragestellung ausreichen beantwortet wurde
- Wenn nicht, weitere Stichprobe & Analyse bis ausreichend Informationsstand herrscht
- Gefahr: falsch-positive Ergebnisse

Kombinierte Stichprobenverfahren

- Mehrstufige Sampling-Verfahren, bei denen verschiedene Sampling-Methoden je Stufe miteinander kombiniert werden

Beispiel

- **Multicenterstudie zur Behandlungsqualität bei Mammakarzinom**
 - 1. Stufe: Clusterverfahren zur Ermittlung der einzubeziehenden Regionen
 - 2. Stufe: Stratifizierte Auswahl der einzubeziehenden Behandlungseinrichtungen (Mammazentren vs. Sonstige)
 - 3. Stufe: Einfache Zufallsstichprobe aus den Patientinnen der letzten 12 Monate

2. Bewusste Auswahl

- Nicht alle Elemente der Grundgesamtheit haben die gleich Chance $p > 0$ in die Stichprobe aufgenommen zu werden
- Bewusste Auswahlverfahren
 - Nach Auswahlplan, Regeln angegeben und überprüfbar
- **Keine Anwendung inferenzstatistischer Techniken möglich**
 - **Wenn überhaupt: nur über Inferenzpopulation (Grundgesamtheit, über die auf Basis der vorliegenden Stichprobe tatsächlich Aussagen gemacht werden können)**

Sampling Strategien

1. Spezifische Fallauswahl

- Typische Fälle: anhand spezifischer Merkmalsausprägungen
- Extreme Fälle: anhand Art der Merkmalsausprägung
- Dominierende Fälle: anhand Konzentration/Häufigkeit der Fälle
→ Konzentrationsprinzip

•

2. Schneeballverfahren

- 1. Welle: Auswahl (ggf. Zufallsauswahl) von Individuen einer GG oder definierten Merkmalen, Probanden werden gebeten weitere Personen mit definierten Merkmalen zu nennen
- 2. Welle: Befragung aller genannten Probanden, erneutes Fragen
- 3. Welle, usw.
 - ➔ Keine Aussagen über Grundgesamtheit möglich, da diese nicht ermittelbar

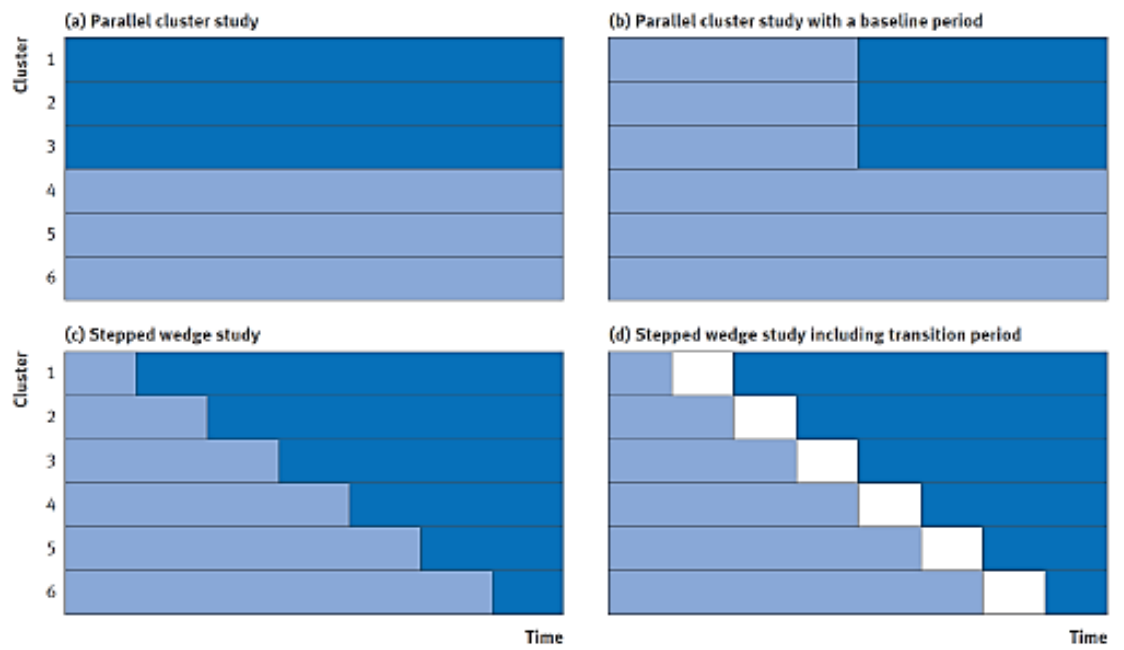
3. Quotenauswahl (Quota sampling)

- Wenn keine Probandenkartei vorhanden dann:
- Personen werden so ausgewählt, dass diese bestimmten Merkmalen entsprechen, wie in der Grundgesamtheit
- Jeder Interviewer mit festgelegter Quote zu befragender Merkmalsträger und Erhebungsgebiet
- Befragung nach Quotenplänen
 - Nicht ineinandergreifend (Quoten einzelner Merkmale unabhängig voneinander)
 - Ineinandergreifend (Quoten einer Variable müssen sich auch bei anderen Merkmalen widerspiegeln)

3. Willkürliche Auswahl

- Entscheidungen über die Aufnahme eines Elements der GG in der Stichprobe im Ermessen des Auswählenden
- I.d.R. werden diese Personen genommen, die gerade zur Verfügung stehen (Ad-hoc Stichprobe) → bei fast allen klinischen Studien
 - ➔ Abgeleitetes Design: RCT, Fall-Kontroll-Studie, etc.
 - ➔ Für Längs- und Querschnittstudien schlechter geeignet

Abgeleitetes Studiendesign – Randomisierung auf Clusterebene



Stepped Wedge Design

- Besonders bei komplexen Interventionen, die Versorgungsprozesse und -strukturen adressieren
→ Randomisierung hier schwierig aus methodischer Sicht (hohes Risiko für Kontaminationseffekt zwischen Kontrolle und Interventionen)
- Daher Randomisierung auf Ebene einer Organisationseinheit (Cluster)
→ Stepped wedge Design
- Jedes Cluster startet in Kontrollphase, Intervention startet zeitversetzt
→ Cluster werden zufällig zugeordnet

Methodische Herausforderungen:

- Planung Studienablauf (Wechselzeitpunkte, Anzahl Cluster, Zeiträume zwischen Wechsel)
- Fallzahlplanung (analog RCT, jedoch Korrektur nach Designeffekt)
- Komplexe Auswertungsmodelle (Berücksichtigung Zeiteffekte, Mehrebenenmodelle)

Einwohnermeldeämter

- Dürfen für Forschungszwecke Bevölkerungsstichproben ziehen und Adresse verfügbar machen