SPM-Kurs 2018

Allgemeines Lineares Modell
General Linear Model

Tobias Sommer-Blöchl,
NeuroImage Nord,
Institut für Systemische Neurowissenschaften

Universitätsklinikum
Hamburg-Eppendorf

General linear model (GLM)

-vereinigende Alternative zur klassischen Statistik

-in den späten 60ern eingeführt basierend auf der multiplen Regressionsanalyse

-T-tests, ANOVAs, ANCOVAs, und multiple Regressionen "übersetzt" in multiple Regressionsprobleme

-gelöst mit den selben Formeln und basierend auf derselben Logik

-leicht erweiterbar für nicht-lineare Fragen

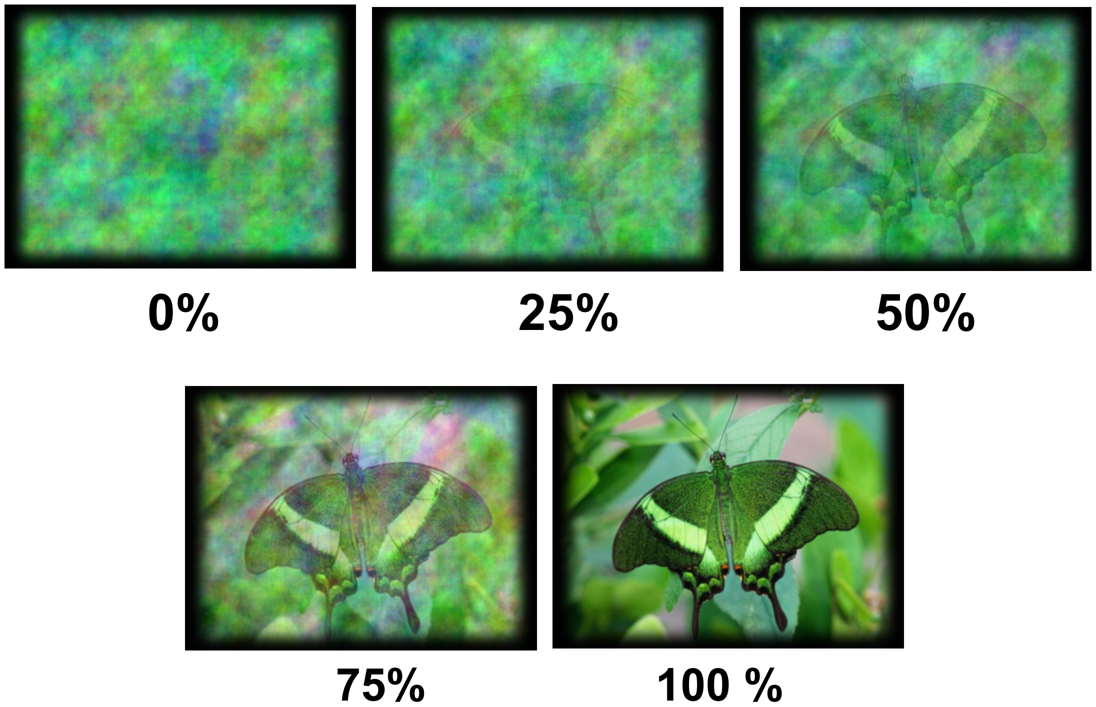
-SPM, andere fMRI Analyse Softwares, allgemeine statistische Software Packages

Übersicht über den Vortrag

- sehr kurze allgemeine Einführung in das GLM
- Mini Einführung in relevante Aspekte der Matrix Algebra
- detaillierte Einführung in das GLM
  - -mittels eines behaviouralen (multiplen) Regressionsbeispiels
- Erweiterung auf ANOVAs im GLM
- das GLM in SPM (Vorschau auf morgigen Vortrag ‚Das ALM in der fMRT')
  - -1st Level Analysen: multiple Regressionen
  - -meisten 2nd Level Models: t-tests oder ANOVAs

Beispielstudie (Rose et al., Cereb Cortex, 2005)

- Aufgabe im Scanner:
    - Bilder (1 Sekunde) in **5 unterschiedlichen Sichtbarkeitsstufen**
    - jeweils Blöcke von je 10 Bildern einer Sichtbarkeitsstufe (20 Sekunden)



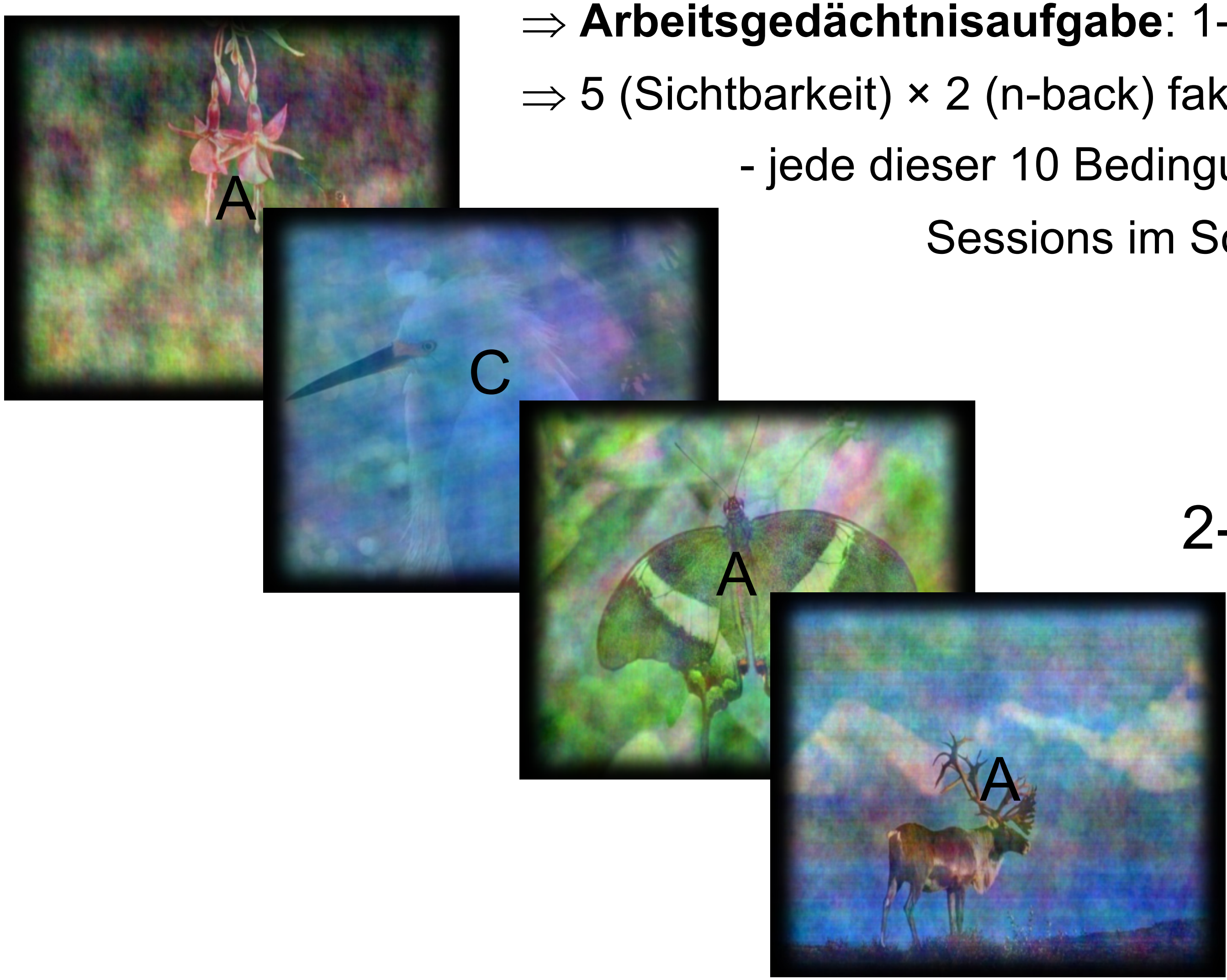
Beispielstudie (Rose et al., Cereb Cortex, 2005)

- Aufgabe im Scanner:
  - Bilder (1 Sekunde) in **5 unterschiedlichen Sichtbarkeitsstufen**
  - jeweils Blöcke von je 10 Bildern einer Sichtbarkeitsstufe (20 Sekunden)
  - jeweils einer der Buchstaben A, B, C, D, E, F auf einem Bild

⇒ **Arbeitsgedächtnisaufgabe**: 1-back vs. 2-back

⇒ 5 (Sichtbarkeit) × 2 (n-back) faktorielles Design

- jede dieser 10 Bedingungen 2 mal in jeder der 3 Sessions im Scanner



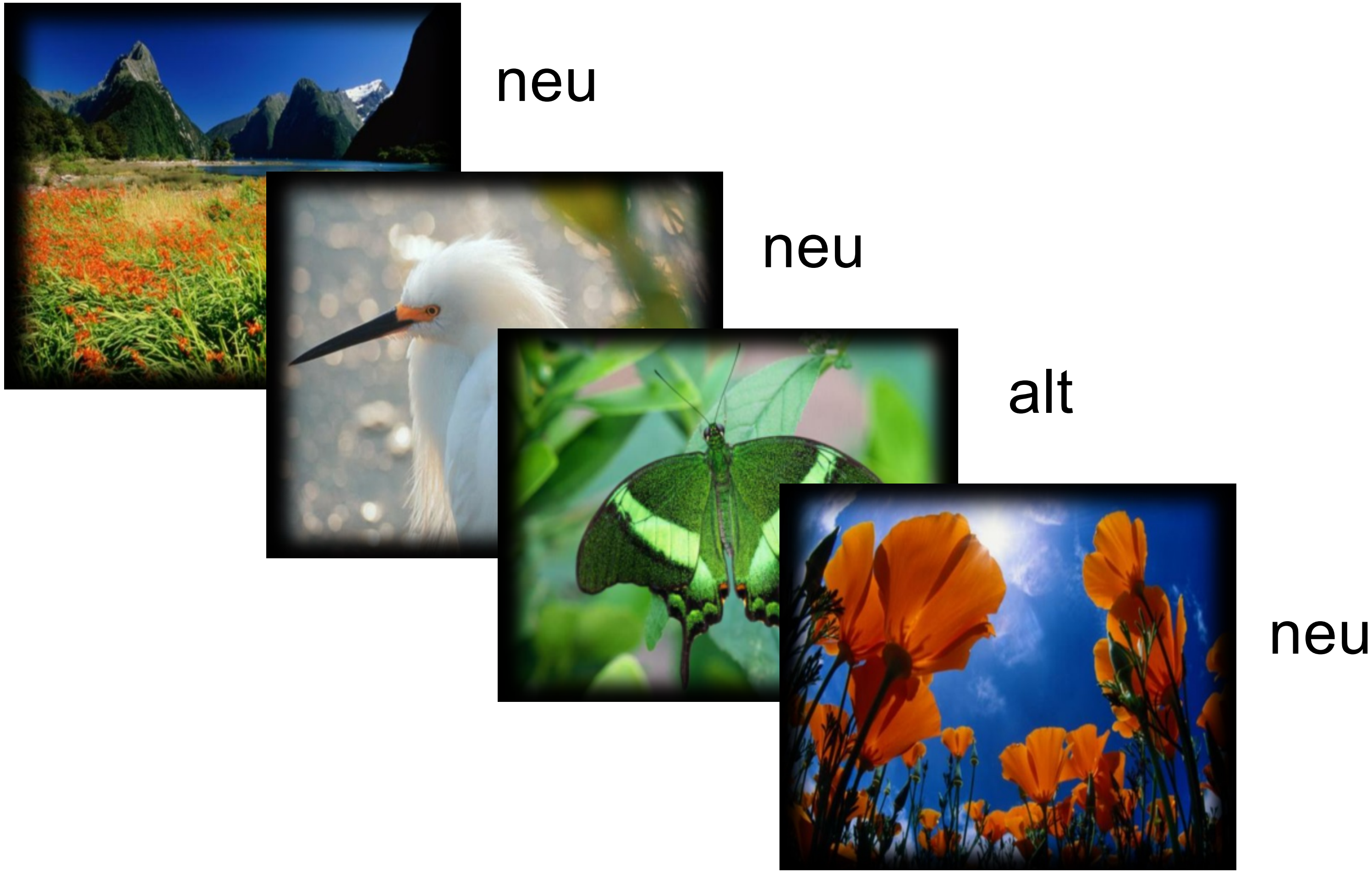
2-back target

1-back target

Beispielstudie

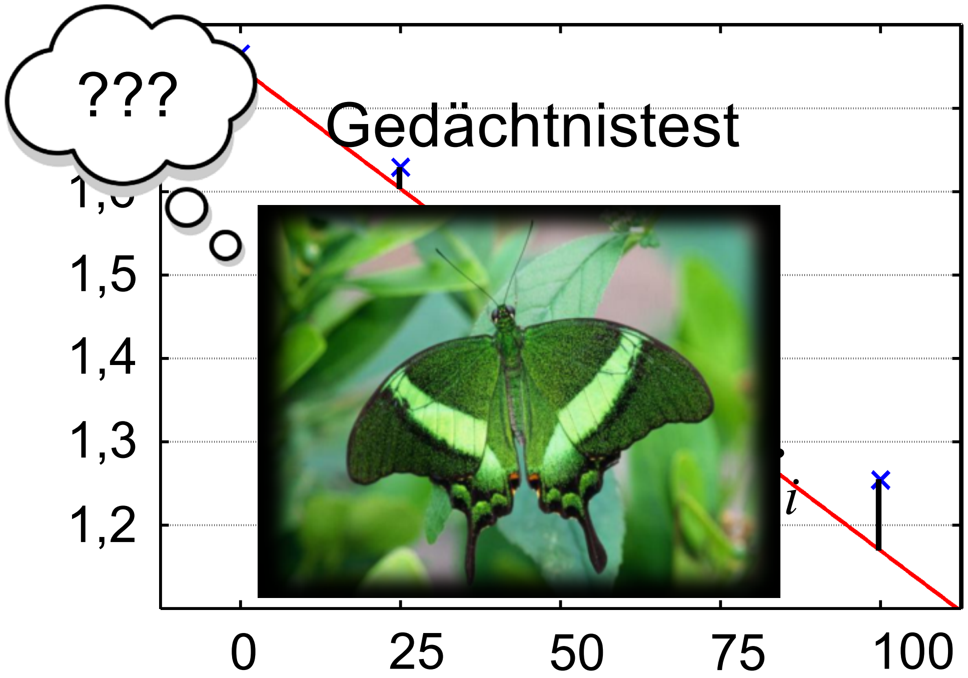
Beispielstudie (Rose et al., Cereb Cortex, 2005)

- Überraschungs-Gedächtnistest nach dem Scan:
  - alte und neue Bilder
  - alle Bilder 100% sichtbar
  - Aufgabe: möglichst schell alt/neu-Entscheidung



- GLM: (multiple) Regression – ähnlich einer Gradengleichung $y = ax + b + Fehler$
- abhängige Variable (y) und unabhängige Variable(n)/**Regressoren** (x's)
- a = Steigungsmaß/Regressionskoeffizient/**Parameter** (im GLM: β - betas)
- b = Achsenabschnitt/**Konstante**
- Fehler = **Residuen** (im GLM: ε)

Sichtbarkeit x	Reaktionszeit y
0	1,77
25	1,63
50	1,44
75	1,20
100	1,25



Ziel: ⇒ Beschreibung des Zusammenhangs zwischen x und y

– bei mehreren Regressoren x_j: $y = a_1x_1 + a_2x_2 + \ldots + a_mx_m + 1b + \varepsilon$

⇒ Ziel des GLM: y als gewichtete Linearkombination von x_j

Matrizen

- rechteckiges Feld/Tabelle von Elementen angeordnet in Zeilen und Spalten

$$\begin{array}{cc} & \text{Spalten} \rightarrow \\ \text{Zeilen} \downarrow & \begin{bmatrix} 2.4 & 3.7 & 0.9 & 2.1 & 6.3 \\ 1.9 & 3.4 & 2.1 & 0.8 & 4.1 \\ 5.6 & 2.3 & 4.7 & 0.4 & 2.9 \end{bmatrix} \end{array}$$

$$\begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} & x_{1,5} \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} & x_{2,5} \\ x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} & x_{3,5} \end{bmatrix}$$

- Elementе von Matrizen werden
  adressiert durch 1. Zeile und 2. Spalte

$$\begin{array}{cc} & \text{n Spalten, j ändert sich} \rightarrow \\ \text{m Zeilen, i ändert} \downarrow & \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,n} \\ \vdots & \ddots & & & \vdots \\ x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,n} \\ \vdots & & & \ddots & \vdots \\ x_{m,1} & \cdots & x_{m,j} & \cdots & x_{m,n} \end{bmatrix} \end{array}$$

Spezialfälle von Matrizen

- *Vektor*: nur eine Zeile oder Spalte

Zeilen-Vektor $x = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} & x_{1,5} \end{bmatrix}$

Spalten-Vektor $x = \begin{bmatrix} x_{1,1} \\ x_{2,1} \\ x_{3,1} \\ x_{4,1} \\ x_{5,1} \end{bmatrix}$

- *Skalar*: nur ein Element $\begin{bmatrix} x_{1,1} \end{bmatrix}$
- *Identitäts-Matrix* (immer eine *quadratische M*atrix)
    - analoge zur Zahl 1

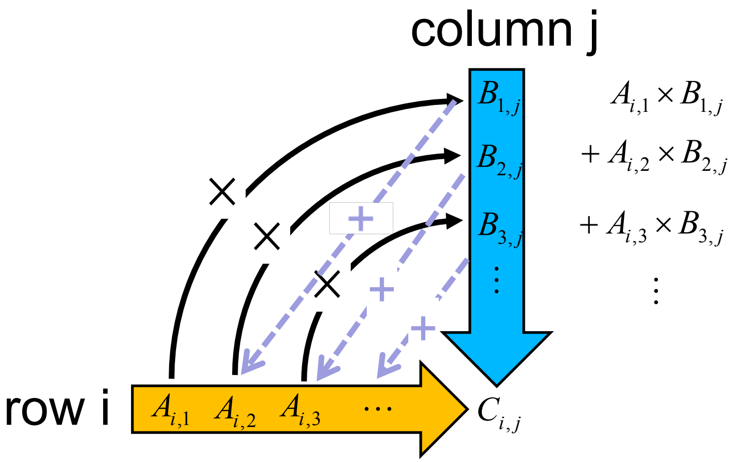
$$I_5 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

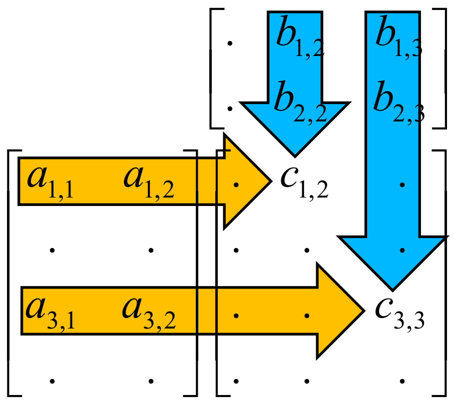
- Matrizen: Großbuchstaben; Vektoren und Skalare: Kleinbuchstaben

$$X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,n} \\ \vdots & \ddots & & & \vdots \\ x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,n} \\ \vdots & & & \ddots & \vdots \\ x_{m,1} & \cdots & x_{m,j} & \cdots & x_{m,n} \end{bmatrix}$$

Matrizen-Rechnung

- Addition, Subtraktion und Punkt Multiplikation von zwei Matrizen:
    - beide Matrizen müssen gleiche Dimensionen haben
    - Element-weise addiert, subtrahiert oder multipliziert
- Multiplikation mit Skalar: jedes Element wird mit dem Skalar multipliziert
- Multiplikation mit einer anderen Matrix oder einem Vektor
    - Anzahl der Spalten der linken Matrix muss genauso groß sein wie Anzahl der Zeilen der rechten Matrix
    - Summe der Punkt-Produkte der Elemente einer gegebenen Zeile in **A** und einer gegebenen Spalte in **B**



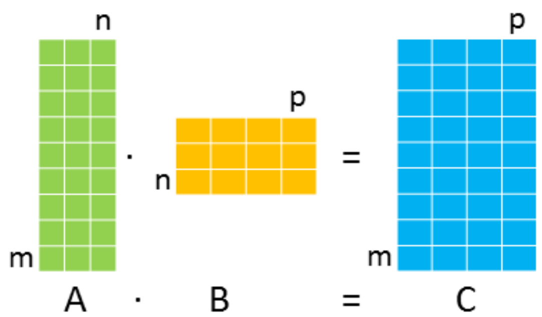


$$a_{1,1} \times b_{1,2} + a_{1,2} \times b_{2,2} = c_{1,2}$$

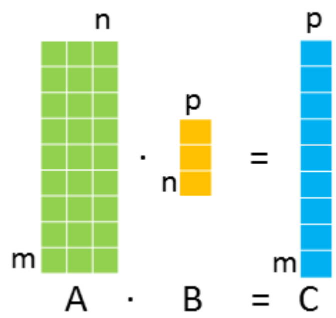
$$a_{3,1} \times b_{1,3} + a_{3,2} \times b_{2,3} = c_{3,3}$$

Matrizen-Rechnung

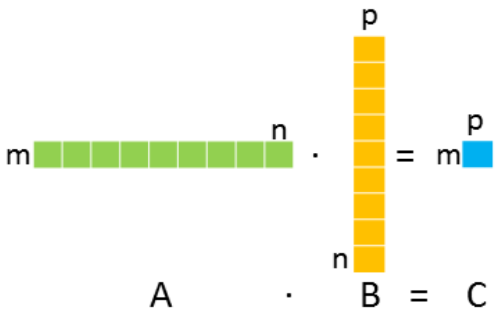
- Multiplikation mit einer anderen Matrix oder einem Vektor
  - Anzahl der Spalten der linken Matrix muss genauso groß sein wie Anzahl der Zeilen der rechten Matrix
  - Summe der Punkt-Produkte der Elemente einer gegebenen Spalte in A und einer gegebenen Spalte in B
- m-mal-*n Matrix* **A** · n-mal-p Matrix **B** = m-mal-p Matrix *Produkt* **AB**
  - **Beispiele**: Matrix · Matrix = Matrix



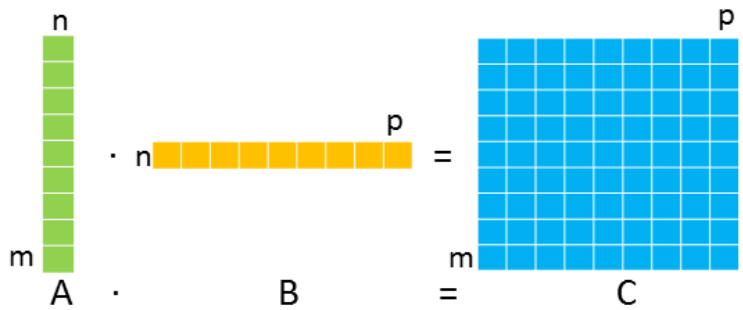
Matrix · Spalten-Vektor = Spalten-Vektor



Zeilen-Vektor · Spalten-Vektor = Skalar



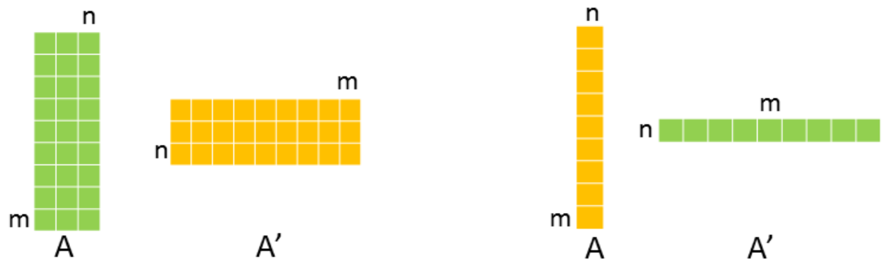
Spalten Vektor · Zeilen-Vektor = Matrix



Matrizen-Rechnung

- transponierte Matrix

- *transponierte* einer *m*-mal-*n* Matrix **A** ist die *n*-mal-*m* Matrix **A’** (auch als $\mathbf{A}^T$ notiert), die man erhält durch das Drehen von Zeilen in Spalten und umgekehrt



- die Inverse einer quadratischen Matrix

$$I_5 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

- **AB** = **BA** = $\mathbf{I}_n$ (Identitäts-Matrix, analog zu $a \times \frac{1}{a} = 1$)
- wenn **B** existiert, ist sie eindeutig und wird $\mathbf{A}^{-1}$ genannt, die inverse Matrix von **A**
- **A** ist dann invertierbar, regulär, nicht-singulär oder nicht degeneriert

- (Moore–Penrose) pseudoinverse Matrizen

- $\mathbf{A}^+$ ist eine Verallgemeinerung der inversen Matrix
- wenn die Inverse nicht berechnet werden kann, z.B. bei nicht quadratischen oder Rang defizient Matrizen (später mehr weil SPM relevant)

– GLM ist in Matrizen und Matrix-Algebra formuliert

$$y = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n + 1b + \varepsilon$$

gewichtete Regressoren

Konstante

abhängige Variable

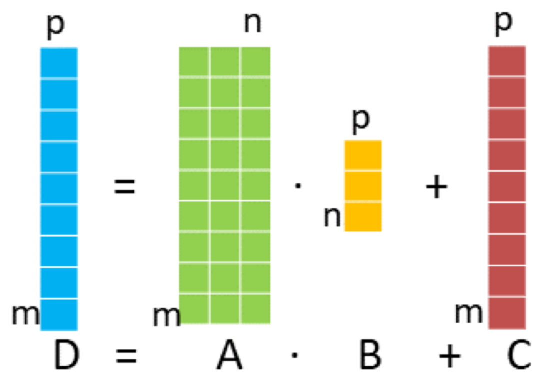
Fehler

$$\begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_m \end{bmatrix} = \beta_1 \begin{bmatrix} x_{1,1} \\ \vdots \\ x_{i,1} \\ \vdots \\ x_{m,1} \end{bmatrix} + \beta_2 \begin{bmatrix} x_{1,2} \\ \vdots \\ x_{i,2} \\ \vdots \\ x_{m,2} \end{bmatrix} + \ldots + \beta_m \begin{bmatrix} x_{1,n} \\ \vdots \\ x_{i,n} \\ \vdots \\ x_{m,n} \end{bmatrix} + \beta_0 \begin{bmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_m \end{bmatrix} = \underbrace{\begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1n} & x_{10} \\ \vdots & \vdots & & \vdots & \vdots \\ x_{i,1} & x_{i,2} & & x_{i,n} & x_{i0} \\ \vdots & \vdots & & \vdots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,n} & x_{m,0} \end{bmatrix}}_{\text{Designmatrix}} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \\ \beta_0 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

GLM in Matrizenschreibweise

$$\begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1n} & x_{10} \\ \vdots & \vdots & & \vdots & \vdots \\ x_{i,1} & x_{i,2} & \cdots & x_{i,n} & x_{i0} \\ \vdots & \vdots & & \vdots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,n} & x_{m,0} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \\ \beta_0 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_m \end{bmatrix}$$



$$y = X\beta + \varepsilon$$

Regressionsanalyse

-Regressionsanalysen im GLM intuitiver

-SPM 1st Level Analysen sind multiple Regressionen

1. **Modellauswahl/-spezifikation**
2. **Parameterschätzung**
3. **Beurteilung der Modellgüte**
4. **Inferenz/Signifikanztestung**

Modellauswahl/-spezifikation

- Wie viele und welche Regressoren werden in das Model aufgenommen?

- Welche Art von Zusammenhang zwischen y und x's ist erwartet?
  - erweiterbar zu nichtlinearen Zusammenhängen, z.B. ein logarithmischer Einfluss der unabhängigen Variablen $y_i = \beta_1 \log(z_i) + \beta_0 + \varepsilon_i$
  - mit $x_i = \log(z_i)$ ein lineares Model $y_i = \beta_1 x_i + \beta_0 + \varepsilon_i$

Beispielstudie:

⇒ abhängige Variable:

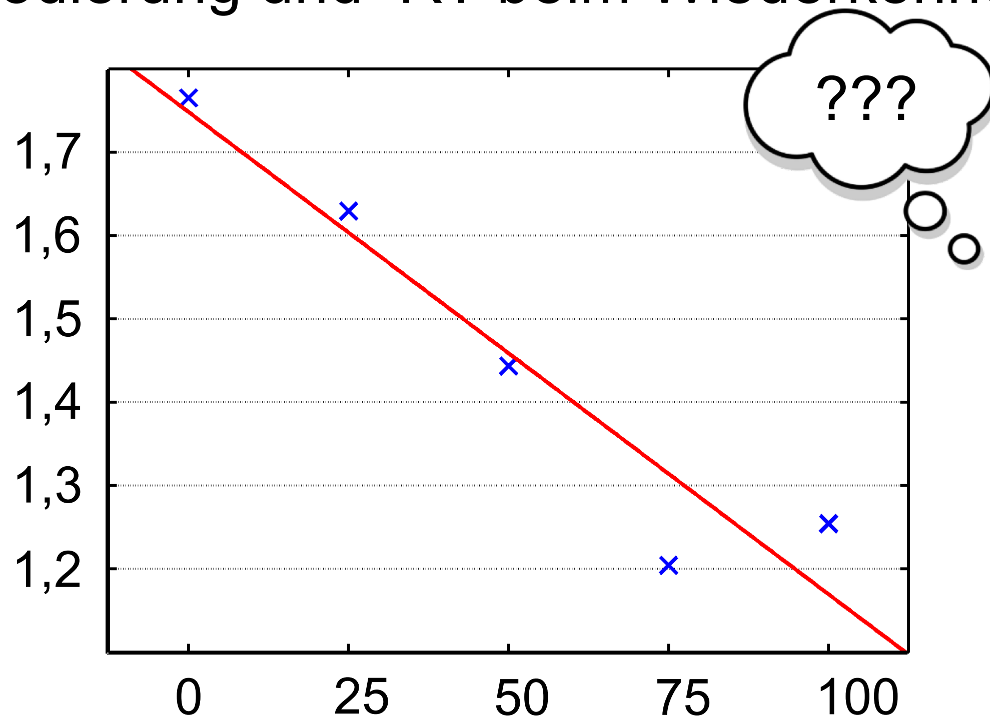
- mittlere Reaktionszeit (RT) im Gedächtnistest
- RT ≈ Gedächtnisstärke

⇒ 1 unabhängige Variable/Regressor (+ 1 Konstante)

- Sichtbarkeit der Bilder während der Enkodierung in %

⇒ Hypothese: linearer Zusammenhang zwischen Sichtbarkeit während der Enkodierung und RT beim Wiederkennen

Reakt. y	Sichtb. x
1.77	0
1.63	25
1.44	50
1.20	75
1.25	100



$$y = X\beta + \varepsilon$$

y	x
1,77	0
1,63	25
1,44	50
1,20	75
1,25	100

x_1 :Sichtbarkeit

$$\underset{\text{Reaktionszeiten}}{\begin{bmatrix} 1,77 \\ 1,63 \\ 1,44 \\ 1,20 \\ 1,25 \end{bmatrix}} = \underset{\text{Designmatrix}}{\begin{bmatrix} \overbrace{0}^{x_1} & 1 \\ 25 & 1 \\ 50 & 1 \\ 75 & 1 \\ 100 & 1 \end{bmatrix}} \begin{bmatrix} \beta_1 \\ \beta_0 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

β_0 : y-Achsenabschnitt/Konstante

β_1 : Steigung

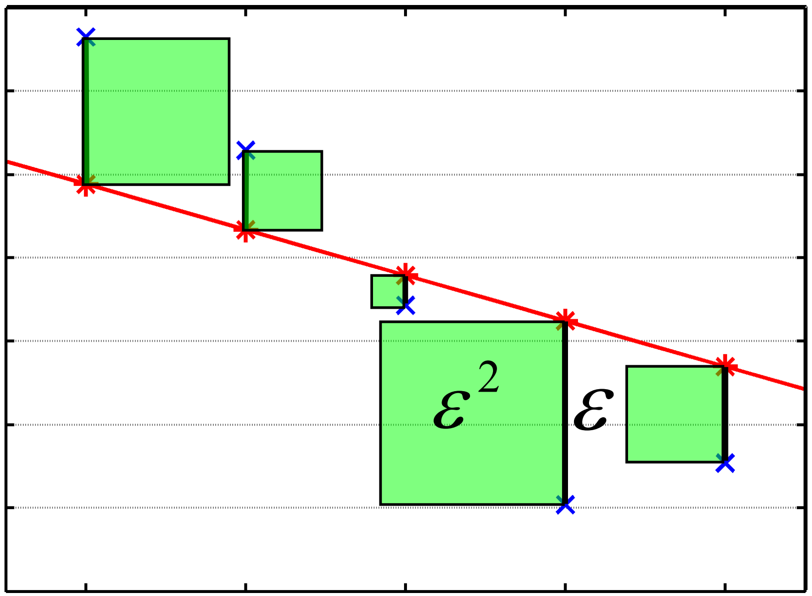
Parameterschätzung

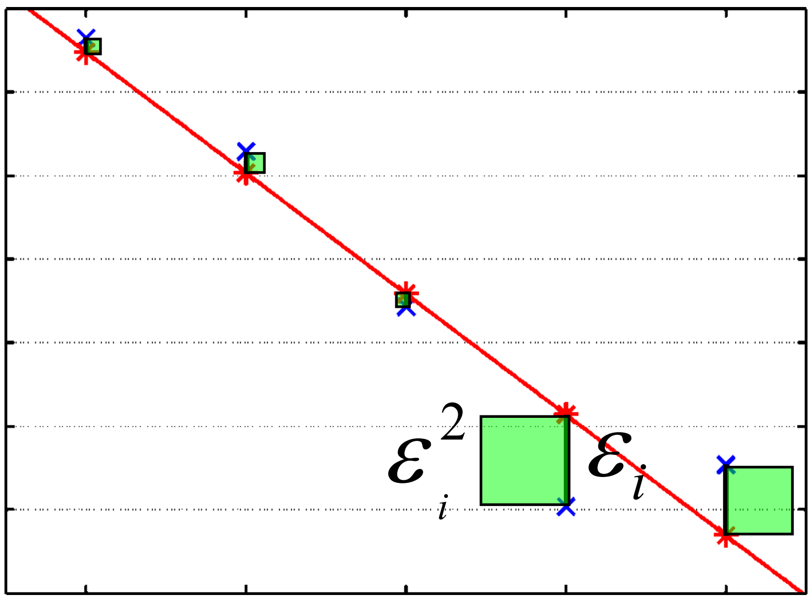
Ziel:

- Regressionsgewichte bestimmen, so dass die abhängige Variable möglichst genau vorhergesagt wird
- Fehler soll minimiert werden

⇒ **Kriterium der kleinsten Abweichungsquadrate:**

- **SSE** (**S**um of **S**quared **E**rrors) soll minimiert werden





Beispielstudie:

- 1 Regressor + 1 Konstante

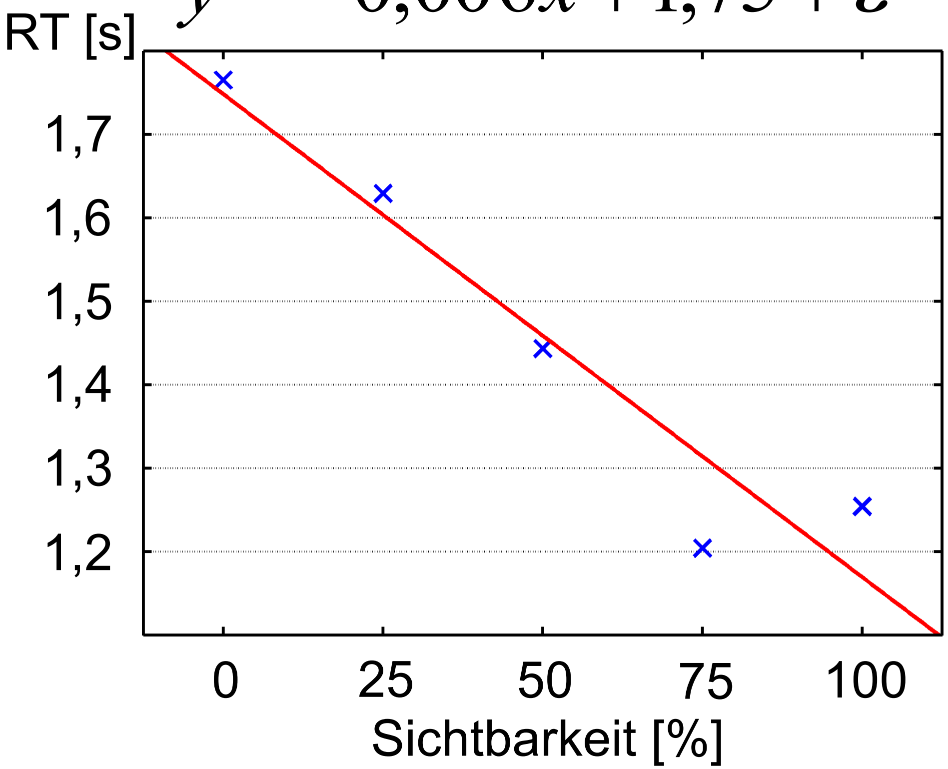
$\Rightarrow$ 2 Regressionsgewichte/Parameter zu schätzen

$\Rightarrow$ im GLM eine Formel für das *ordinary least square* Verfahren für alle statistischen Anwendungen

$$\beta = (X'X)^{-1} X'y$$

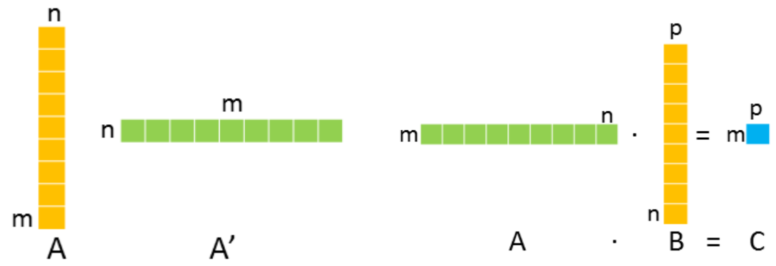
$$\beta = (X'X)^{-1} X'y = \begin{bmatrix} -0{,}006 \\ 1{,}752 \end{bmatrix}$$

$$y = -0{,}006x + 1{,}75 + \varepsilon$$



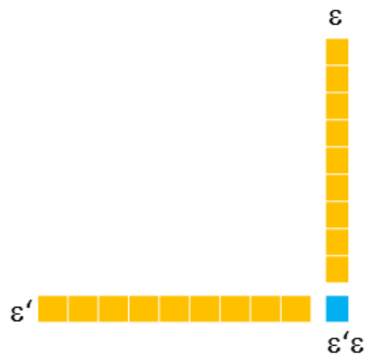
$\Rightarrow$ SSE soll minimiert werden

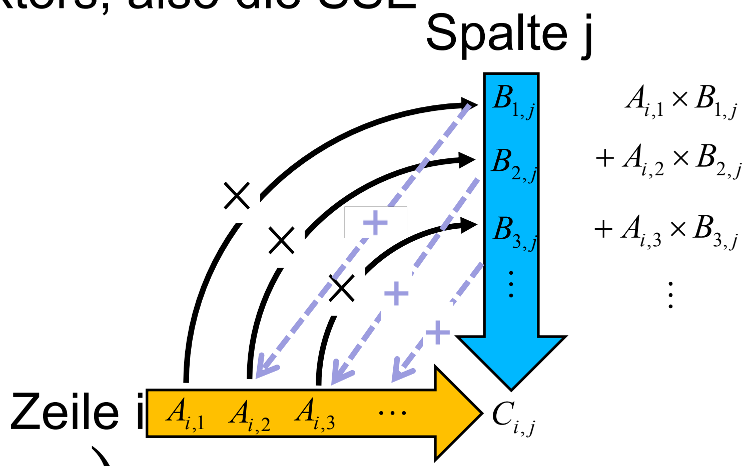
$$SSE = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon$$



- Fehler ε: Spalten-Vektor
- ε' : Zeilen-Vektor
- Multiplikation eines Zeilen- mit einem Spalten-Vektor ergibt ein Skalar
- insbesondere, Multiplikation eines transponierten Spalten-Vektors (ε') mit sich selbst (ε) ergibt die Summe der Quadrate der Elemente des Vektors, also die SSE

$$\boxed{\begin{aligned} y &= X\beta + \varepsilon \\ \varepsilon &= y - X\beta \end{aligned}}$$





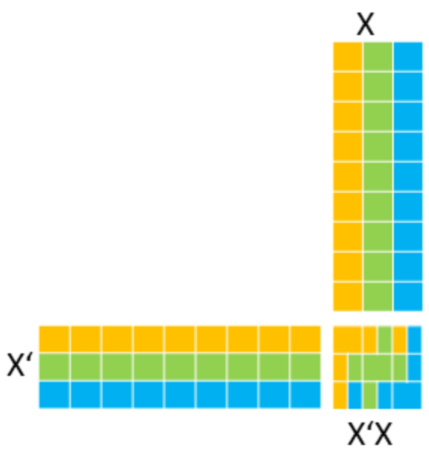
$$SSE = \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta)$$

$$SSE = y'y - 2\beta'X'y + \beta'X'X\beta$$

$$SSE = y'y - 2\beta'X'y + \beta'X'X\beta$$

$\Rightarrow$ Multiplikation eines transponierten **Spalten-Vektors** (ε’) mit sich selbst (ε) ergibt die **Summe der Quadrate der Elemente** des Vektors, also die SSE

$\Rightarrow$ Multiplikation einer transponierten **Matrix** mit sich selbst ergibt die “Produkt-Summen-Matrix”



- auf der Hauptdiagonalen die **Summe der Quadrate der Elemente** der Spalten
- auf den anderen Diagonalen zeilenweise die Produkt-Summen der Spalte und der anderen Spalten

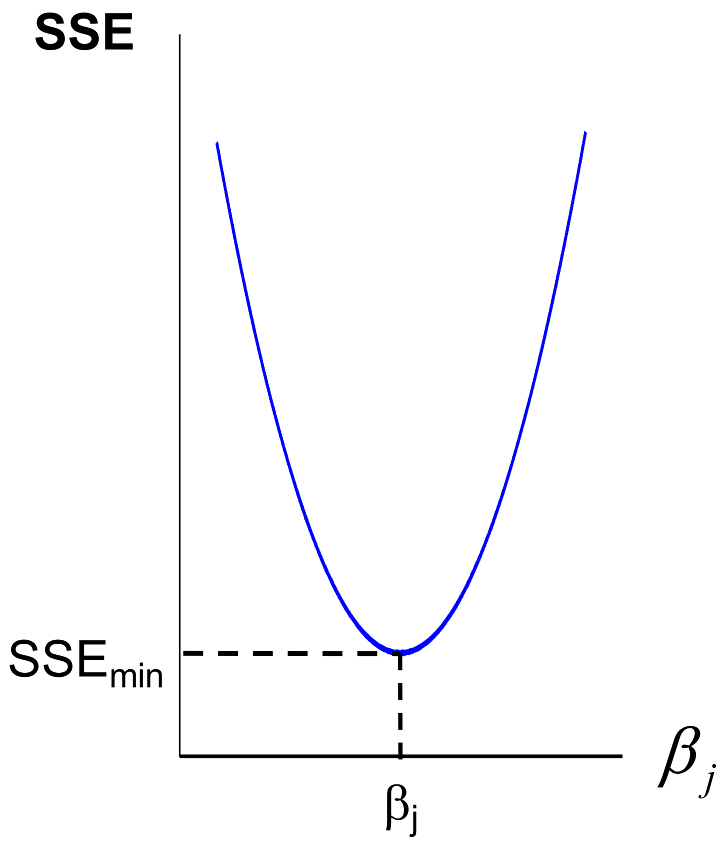
$\Rightarrow$ $SSE = \cdots + \beta'\boxed{X'X}\beta$ ist eine quadratische Gleichung wie $y = x^2$

$$SSE = y'y - 2\beta'X'y + \beta'X'X\beta$$

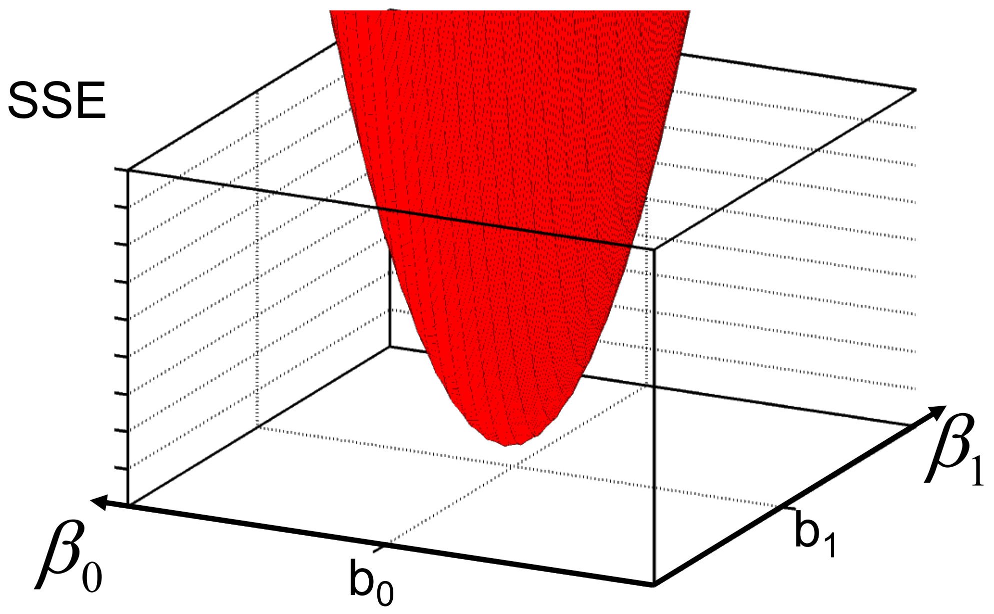
- bei einem Parameter formen die SSE eine Parabel als Funktion von β
- die partielle Ableitung der SSE nach 0 ergibt das β, für das die SSE minimal sind

$$\frac{\partial SSE}{\partial \beta'} = -2X'y + 2X'X\beta = 0$$

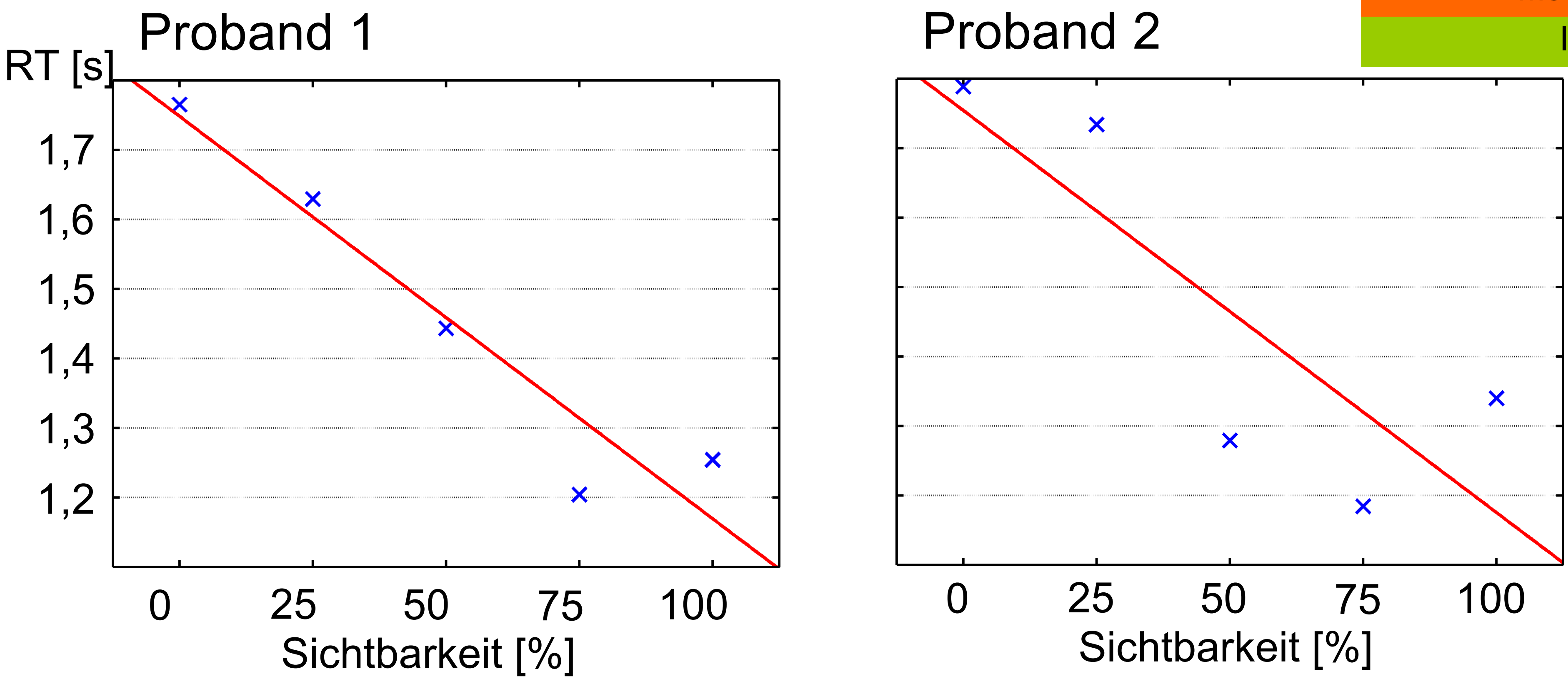
$$\hat{\beta} = \left(X'X\right)^{-1} X'y$$



- bei zwei Parameters (z.B. 1 Regressor und Konstante), formen die SSEs als Funktion der βs einen 3-dimensionalan Kegel mit genau einem Minimum
- die partielle Ableitung nach 0 ergibt die Kombination von zwei βs bei denen die SSE minimal sind
- **Implikationen**
  - eine eindeutige Lösung
  - Schätzung aller Parameter beim OLS Verfahren simultan in einem Schritt



Beurteilung der Modellgüte



$$y = -0{,}006x + 1{,}75 + \varepsilon$$

Welchen Anteil der beobachteten Varianz erklärt das Modell?

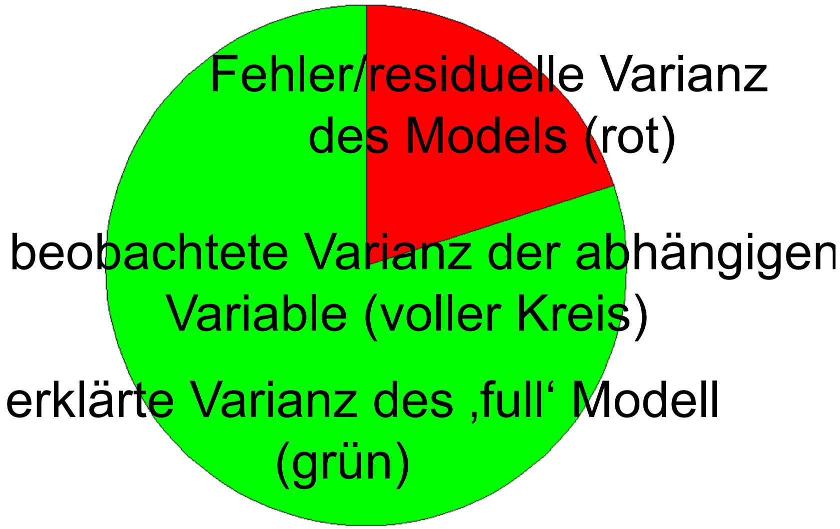
Definitionen:

i) uneingeschränktes oder ‚**full**' Modell:

- Model mit allen Regressoren
- Im Beispiel 1 regressors und die Konstante

$$y = \beta_1 x_1 + \beta_0 1 + \varepsilon$$

$$\begin{bmatrix} 1.77 \\ 1.63 \\ 1.44 \\ 1.20 \\ 1.25 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 25 & 1 \\ 50 & 1 \\ 75 & 1 \\ 100 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_0 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}$$

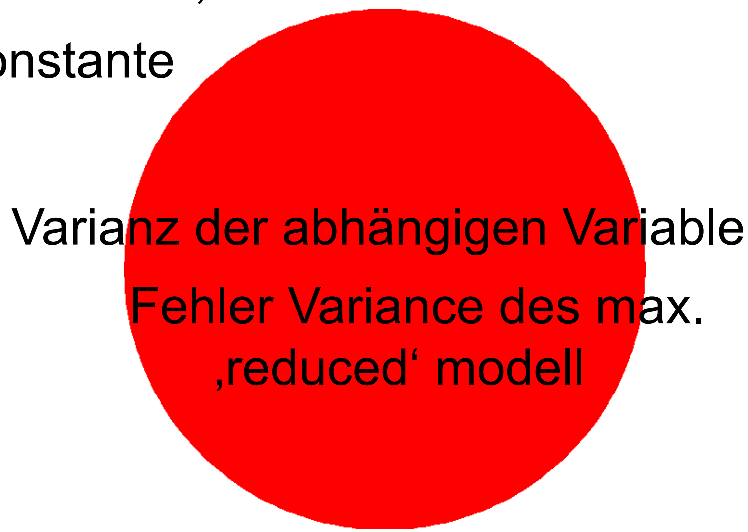


ii) eingeschränktes oder ‚**reduced**' Modell

- reduziert um mindestens einen Regressor verglichen zum ‚full' Modell
- max. reduziert um alle Regressoren bis auf die Konstante

$$y = \beta_0 1 + \varepsilon$$

$$\begin{bmatrix} 1.77 \\ 1.63 \\ 1.44 \\ 1.20 \\ 1.25 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \beta_0 + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}$$

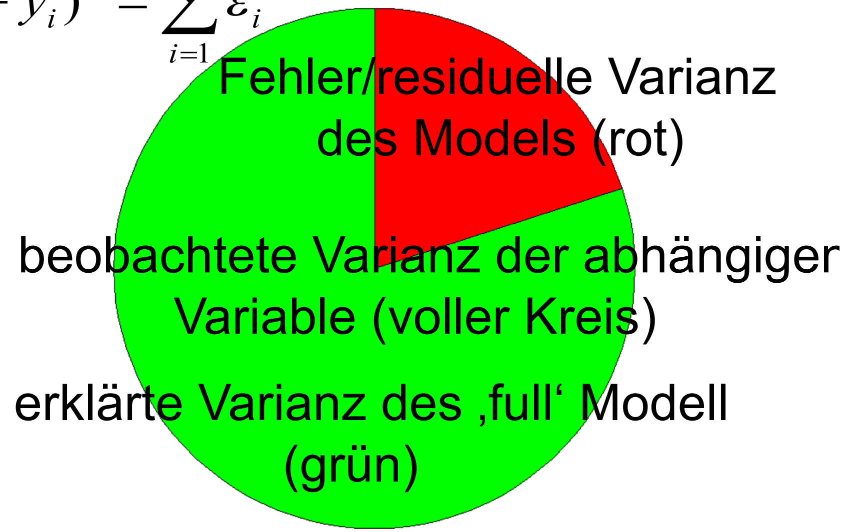


Berechnung der relevanten Varianzen

- beobachtete Varianz der abhängigen Variable $$s_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$
- **S**um of **S**quares **T**otal der abhängigen Variable $$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$$
- geschätzte Parameter $$\hat{\beta} = \left(X'X\right)^{-1} X'y$$
- durch das Modell vorhergesagte Werte $$\hat{y} = X\hat{\beta}$$
- **S**um of **S**quares **M**odel, die durch das Modell erkläre Varianz $$SSM = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$
- **S**um of **S**quares **E**rrors, residuale oder Fehler Varianz des Modells $$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\varepsilon_i^2$$

$$SST = SSM + SSE$$

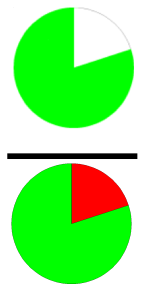
➢ Berechnung der relevanten Varianzen des ‘full’ und ‘reduced’ Modells durch entsprechende Modifikation der Designmatrix X.



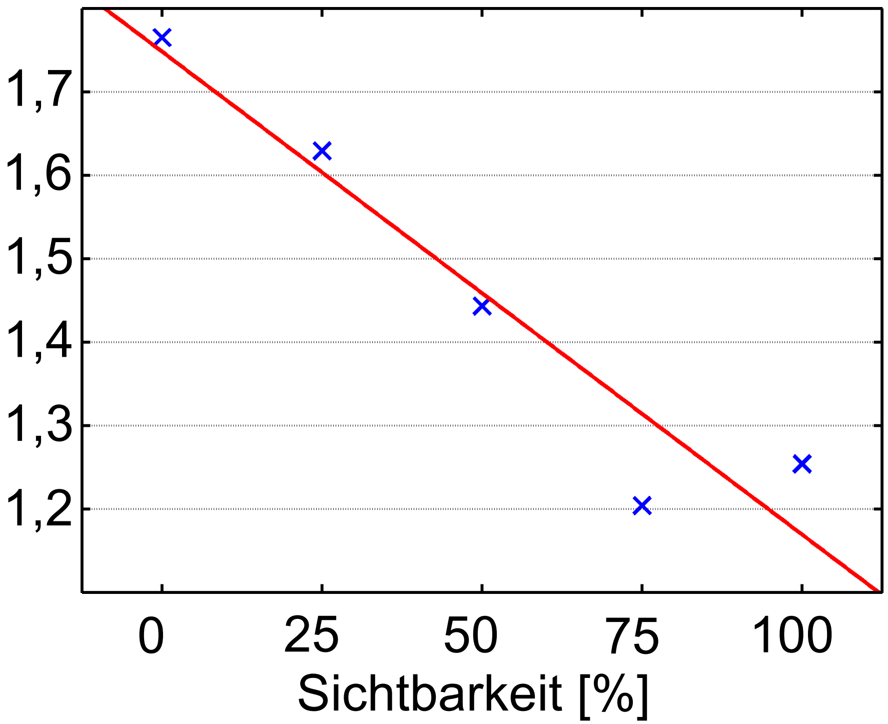
Determinationskoeffizient

Welchem Anteil der beobachteten Varianz erklärt das Modell?

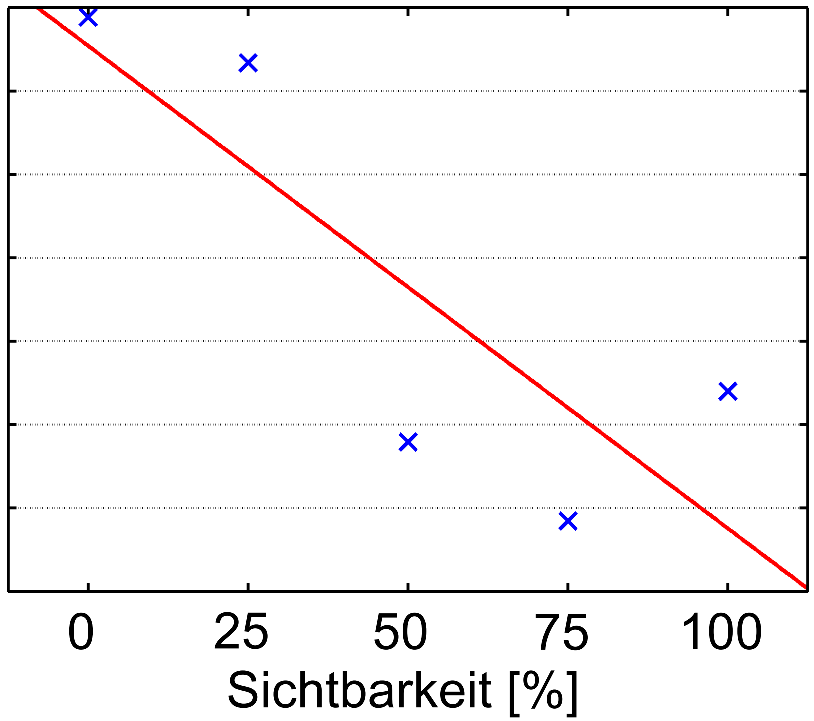
$$R^2 = \frac{\text{erklärte Varianz}}{\text{beobachtete Varianz}} =$$



Proband 1



Proband 2



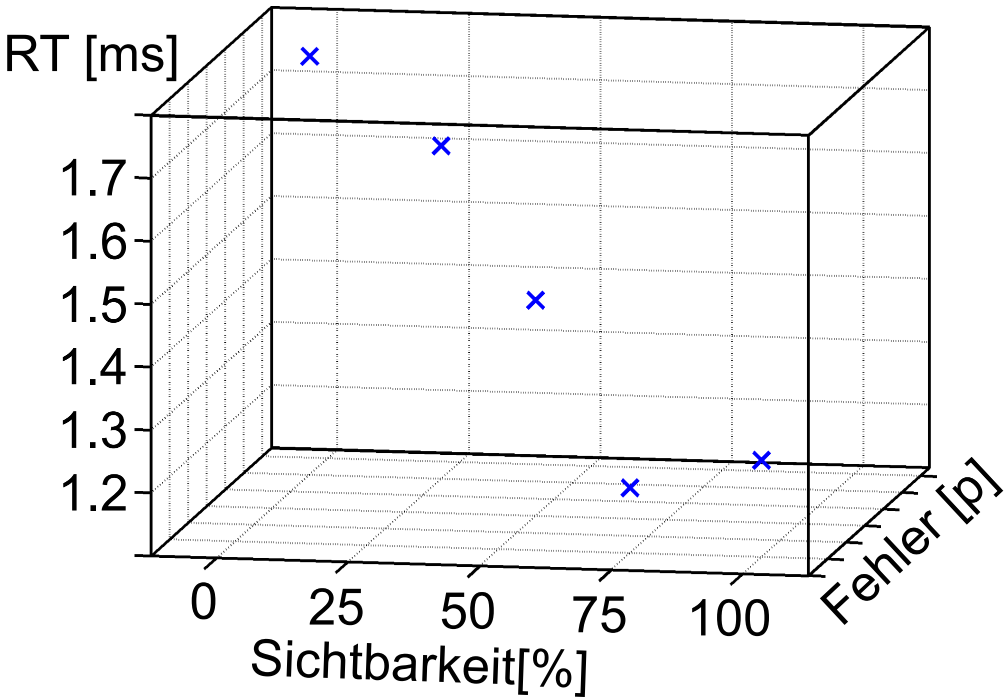
$$R^2 = 0{,}91$$

$$R^2 = 0{,}39$$

Erweiterung/Verbesserung des Modells

- zweiter Regressor:
  - Fehlerrate im Gedächtnistest um die unabhängige Variable RT zu erklären
  - RTs von falschen Antworten sind langsamer

y	x_1	x_2
1.77	0	0.5
1.63	25	0.5
1.44	50	0.3
1.20	75	0.1
1.25	100	0.1



$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_0 1 + \varepsilon$$

Sichtbarkeit Fehlerrate

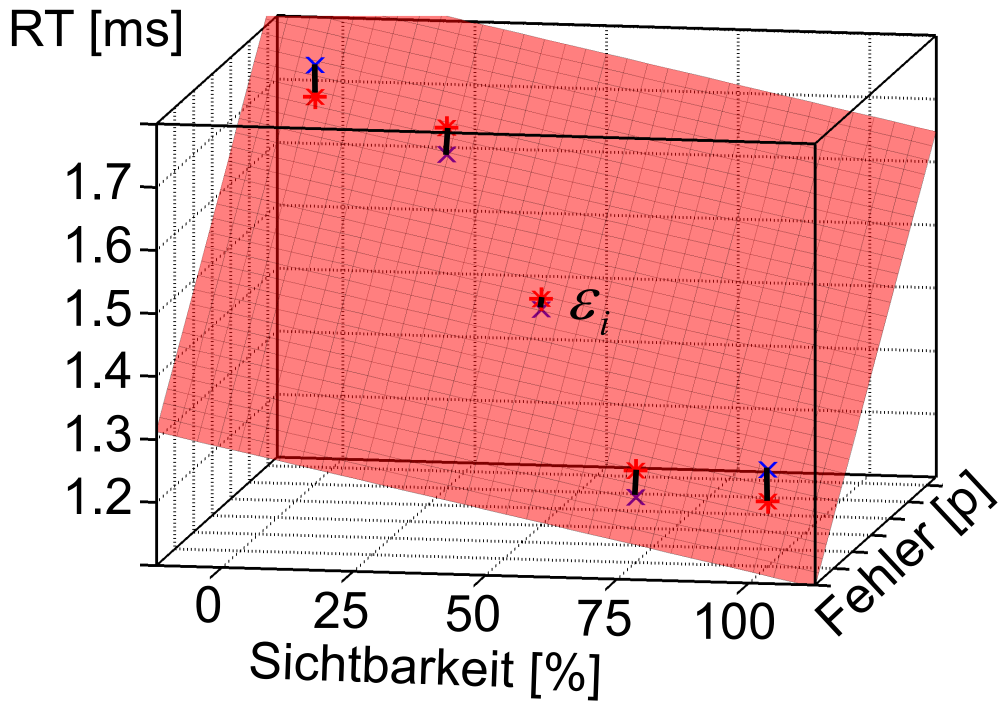
Reaktionszeit

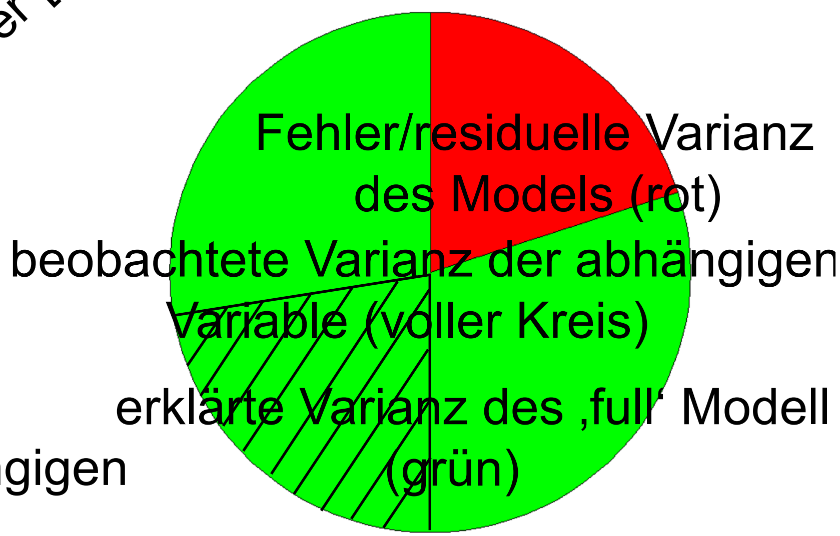
Designmatrix

$$\begin{array}{c} \\ \begin{bmatrix} 1.765 \\ 1.629 \\ 1.443 \\ 1.204 \\ 1.254 \end{bmatrix} \end{array} = \begin{array}{c} \begin{array}{ccc} x_1 & x_2 & \end{array} \\ \begin{bmatrix} 0 & 0.5 & 1 \\ 25 & 0.5 & 1 \\ 50 & [illegible] & 1 \\ 75 & 0.1 & 1 \\ 100 & 0.1 & 1 \end{bmatrix} \end{array} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_0 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}$$

$$b = (X'X)^{-1} X'y = \begin{bmatrix} -0.002 \\ 0.85 \\ 1.3 \end{bmatrix}$$

$$y = -0.002x_1 + 0.85x_2 + 1.3 + \varepsilon$$





- Determinationskoeffizient: $R^2 = 0.96$
- **Inkrement**: 5% zusätzlich erklärte Varianz
- **aber** beide Regressoren sind hoch korreliert
  - erklären teilweise dieselbe Varianz der abhängigen Variable (straffierte Fläche)

  ⇒ **werden weniger reliabel geschätzt**
  - nachteilig für die Statistik

⇒ **dieser Effekt der Korrelation von Regressoren** wichtig in fMRT Designs und Analysen (Mumford et a., 2015, PLOS ONE)

Inferenz

Ziel statistischer Inferenz

- Formulieren und Testen von Hypothesen bzgl. einer Population
  basierend auf einer Zufallsstichprobe aus dieser Population

I) Übersetzung der Forschungsfrage in statistische Hypothese

i) Null-Hypothese H_0: die Hypothese, die man zurückweisen möchte

ii) Alternativhypothese H_1: die Hypothese, die man akzeptieren möchte

II) Wahl des Signifikanzniveaus α

- Wahrscheinlichkeit, fälschlich die Alternativhypothese zu akzeptieren (Typ I Fehler)

III) F-Test oder t-Test?

– hängt von der Hypothese ab

- **F-Tests: -** Erklärt ein Regressionsmodell die Varianz in den beobachteten Daten gut?
  - Ist ein bestimmter Parameter ungleich Null? (ungerichtet)

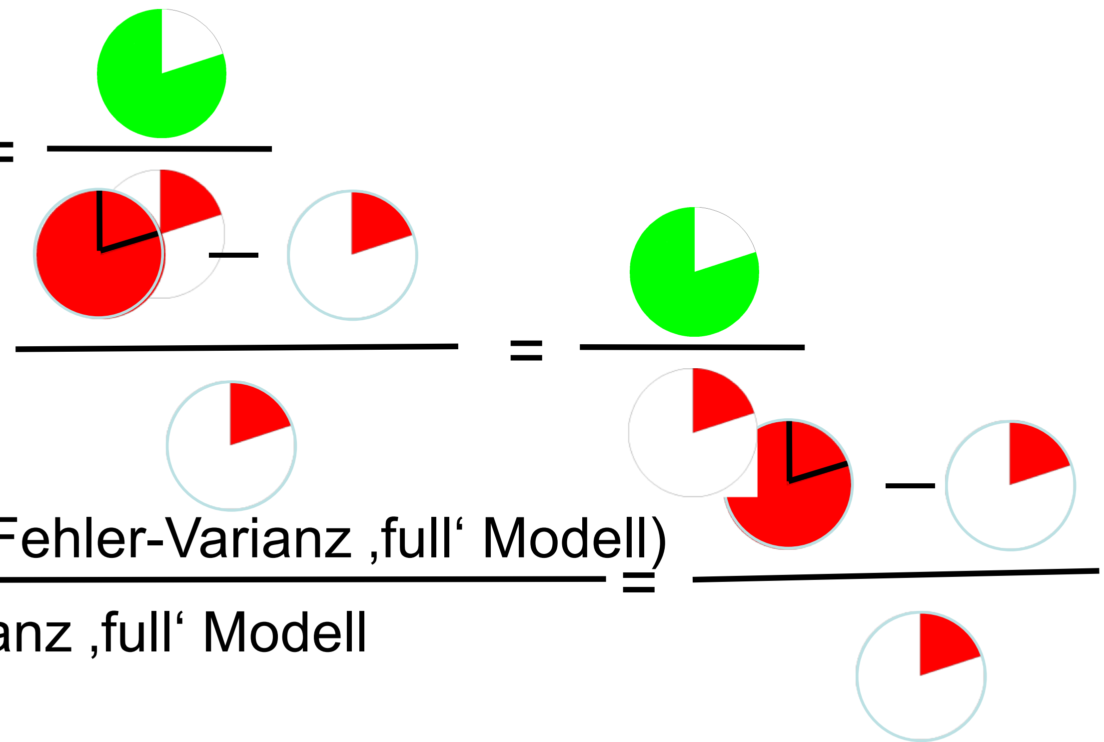
- **t-Tests: -** Ist ein bestimmter Regressionsgewicht/Parameter größer als Null?
  - Ist ein Parameter größer als ein anderer? (gerichtet)

F-Test

- **F-Tests:** Erklärt ein Regressionsmodell die die Varianz in den beobachteten Daten gut?
  - Erklärt das Modell einen signifikanten Anteil der beobachteten Varianz?
  - Erklärt das uneingeschränkte Modell signifikant mehr Varianz als das maximal reduzierte Modell?
- **F-Wert**: Verhältnis zweier Varianzen (geteilt durch Freiheitsgrade)

$$F = \frac{\text{erklärte Varianz ‚full' Modell}}{\text{Fehler-Varianz ‚full' Modell}}$$

$$F = \frac{\text{(Fehler-Varianz ‚red.' Modell) – (Fehler-Varianz ‚full' Modell)}}{\text{Fehler-Varianz ‚full' Modell}}$$

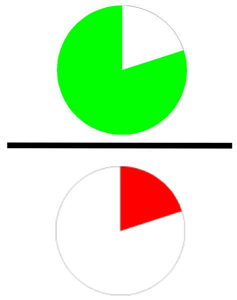


Freiheitsgrade/degrees of freedom (df)

- F-Wert : Verhältnis zweier Varianzen
- geteilt durch dfs

$$F = \frac{\left[\text{Varianz 1} \,/\, df_{\text{Zähler}}\right]}{\left[\text{Varianz 2} \,/\, df_{\text{Nenner}}\right]}$$

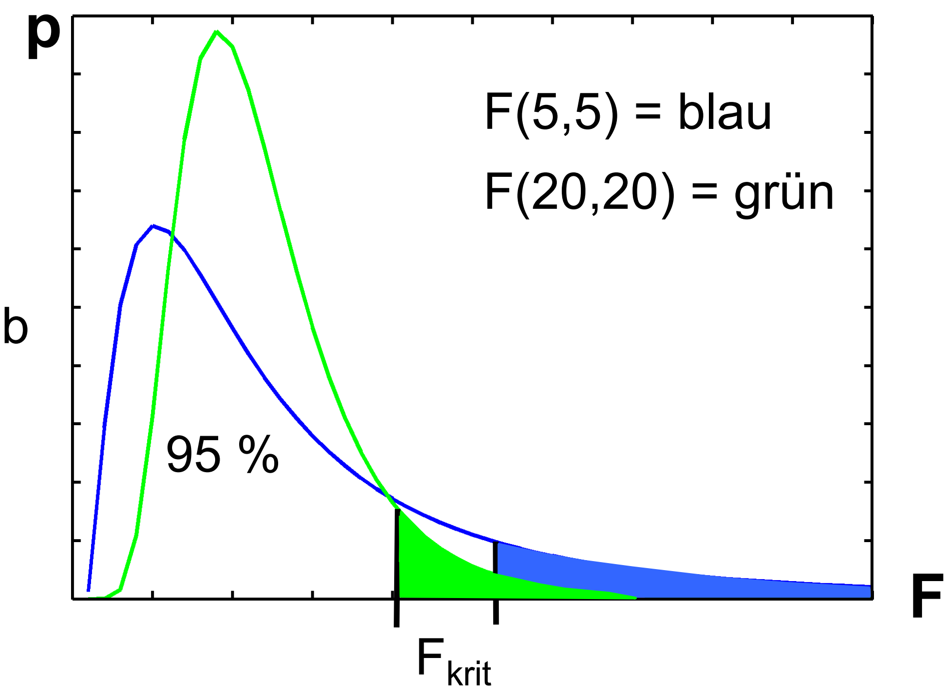
F =



- df $_{\text{Varianz des 'full' Modell (Zähler)}}$ = Anzahl der Regressoren
- df $_{\text{Fehler Varianz (Nenner)}}$ = Anzahl der Beobachtungen – Anzahl der Regressoren

F- Verteilung

- Wahrscheinlichkeitsverteilung
- Form hängt von Anzahl der Freiheitsgrade ab
- $F > F_{krit}$: H_0 kann zurückgewiesen werden mit 95% Sicherheit



Beispielstudie

- **Alternativhypothese**
  - das Modell erklärt signifikanten Anteil der beobachteten Varianz
  - Modell 1: 1 Regressor + Konstante
    - **F(1,4) = 40.4 > F_{krit} = 21.2** $\Rightarrow$ H_0 wird zurückgewiesen $R^2 = 0.91$
  - Model 2: 2 Regressoren + Konstante
    - **F(2,3) = 36.9 > F_{krit} = 30.8** $\Rightarrow$ H_0 wird zurückgewiesen $R^2 = 0.96$

- **Ergebnisse der F-Tests**
  - beide Modelle erklären einen signifikanten Anteil der Varianz in den beobachteten Daten:
    - d.h. signifikanter Zusammenhang zwischen Sichtbarkeit der Bilder sowie der Fehlerrate beim Abruf (Modell 2) und der Reaktionszeit beim Gedächtnistest
  - erklärte Varianz des Modell 2 größer
  - aber F-Wert von Modell 1 größer und F_{krit} kleiner
  - der zweite Regressor und die zusätzlich erklärte Varianz des Modell 2 kommen mit 'Kosten' (darüber hinaus, dass sie hier auch noch hochkorreliert sind!)

(nicht getestet, ob das zweite Modell signifikant mehr Varianz erklärt!)

Beispielstudie

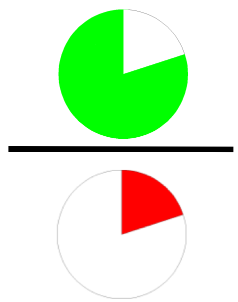
- Alternativhypothese

- das Modell erklärt signifikanten Anteil der beobachteten Varianz
- Modell 1: 1 Regressor + Konstante
  - $F(1,4) = 40.4 > F_{krit} = 21.2 \Rightarrow H_0$ wird zurückgewiesen $R^2 = 0.91$
- Model 2: 2 Regressorren + Konstante
  - $F(2,3) = 36.9 > F_{krit} = 30.8 \Rightarrow H_0$ wird zurückgewiesen $R^2 = 0.96$

- **Zusammenhang zwischen der Anzahl der Regressoren und Signifikanz**
  - df $_{\text{Varianz des 'full' Modell (Zähler)}}$ = Anzahl der Regressoren
  - df $_{\text{Fehler Varianz (Nenner)}}$ = Anzahl der Beobachtungen – Anzahl der Regressoren

$$F = \frac{\left[\text{Varianz 1} \big/ df_{\text{Zähler}}\right]}{\left[\text{Varianz 2} \big/ df_{\text{Nenner}}\right]}$$

F =



- Anzahl der Regressoren ↑ $\Rightarrow$ F-Wert ↓
- Anzahl der Beobachtungen ↑ $\Rightarrow$ F-Wert ↑

$\Rightarrow$ **mehr Regressoren erklären evtl. zusätzliche Varianz aber kosten Freiheitsgrade**

t-Test

- **t-Tests:** **-** Ist ein Parameter größer als Null?
  - Ist ein Parameter größer als ein anderer?
- **Kontrast-Vektor:** Welcher Parameter/Regressor soll getestet werden?
  - gleiche Anzahl an Elementen wie der β-Vektor
  - jedes Element des Kontrast-Vektors korrespondiert mit einem β (und Regressor)
  - Kontrastgewichte: Elemente eines Kontrast-Vektors

- Beispielstudie: Korreliert Sichtbarkeit signifikant positiv mit Reaktionszeit?

Ist das β des ersten regressors signifikant größer als 0?

Kontrast-Vektor $c' = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$

$$\begin{bmatrix} 1.765 \\ 1.629 \\ 1.443 \\ 1.204 \\ 1.254 \end{bmatrix} = \begin{bmatrix} 0 & 0.5 & 1 \\ 25 & 0.5 & 1 \\ 50 & 0.3 & 1 \\ 75 & 0.1 & 1 \\ 100 & 0.1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_0 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}$$

t-Test

- **t-Tests:** **-** Ist ein Parameter größer als Null?

  - Ist ein Parameter größer als ein anderer?

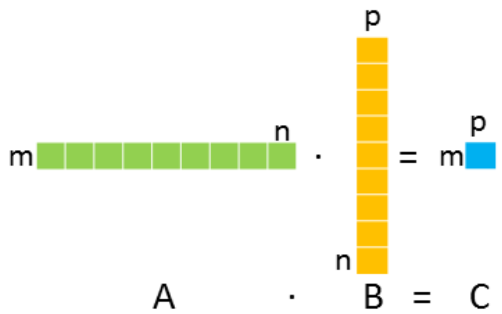
- **Kontrast-Vektor:** Welcher Parameter/Regressor soll getestet werden?
  - gleiche Anzahl an Elementen wie der β-Vektor
  - jedes Element des Kontrast-Vektors korrespondiert mit einem β (und Regressor)
  - Kontrastgewichte: Elemente eines Kontrast-Vektors

- **Beispielstudie**: Korreliert Sichtbarkeit signifikant positiv mit Reaktionszeit?

  Ist das β des ersten regressors signifikant größer als 0?

  Kontrast-Vektor $c' = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$

$$c' = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} * \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_0 \end{bmatrix}$$



- **Beispielstudie**: Korreliert Sichtbarkeit signifikant stärker mit RT als die Fehlerrate?

$$c' = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1.765 \\ 1.629 \\ 1.443 \\ 1.204 \\ 1.254 \end{bmatrix} = \begin{bmatrix} 0 & 0.5 & 1 \\ 25 & 0.5 & 1 \\ 50 & 0.3 & 1 \\ 75 & 0.1 & 1 \\ 100 & 0.1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_0 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}$$

- **Beispielstudie**: Korreliert Sichtbarkeit signifikant stärker mit Reaktionszeit als die Fehlerrate? $c' = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix}$
  - Ist das β des ersten Regressors minus das β des zweiten Regressors signifikant größer als 0?

$$c' = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} * \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_0 \end{bmatrix}$$

t-Wert: Verhältnis von (Kontrasten von) geschätzten Parametern und geschätzter Fehler-Varianz

$$t = \frac{c'\beta}{\sqrt{\frac{SSE}{df} c'(X'X)^{-1} c}}$$

$$c'\beta = 1\beta_1 + (-1\beta_2) + 0\beta_0 = \beta_1 - \beta_2$$

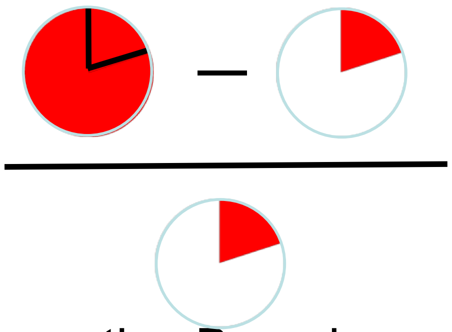
- geschätzte Fehler-Varianz der full models

- **Vergleich des empirischen mit dem kritischen t_{krit} der t-Verteilung**
  - die Nullhypothese H_o wird zurückgewiesen wenn $t > t_{krit}$

F-Test

- F-Wert: Verhältnis zweier Varianzen, kann berechnet werden z.B.

F =



- alternative Berechnung über die **Sum of Squares Hypothesis SSH**
- Kontrast-Vektor oder Matrix
    - -um die Regressors zu spezifizieren, deren erklärte Varianz getestet werden soll
    - -gleiche Anzahl von Spalten wie Designmatrix
    - -Anzahl der Zeilen hängt von der Hypothese ab
    - -Beispiel:
        - -um zu testen, ob min. einer von drei Regressoren eines Modells Varianz erklärt

$$c = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

-analog zum “effect of interest” or “Omnibus F-Test”: Erklärt das ‘full’ Modell signifikant mehr Varianz als das max. reduzierte Modell?

F-Test

- **Sum of Square of the Hypothesis SSH**

$$c = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$SSH = c\beta^{-1}[c(X'X)^{-1}c']^{-1}c\beta$$

$$F = \frac{\left(SSH / df_{Nenner} \right)}{\left(SSE / df_{Zähler} \right)}$$

- mehr (und praktischer) über Kontrast Spezifikation, t- und F-Tests in SPM im Vortrag von Philipp Taesler und Julia Jablonowski über 1st and 2nd Level fMRT Analysen

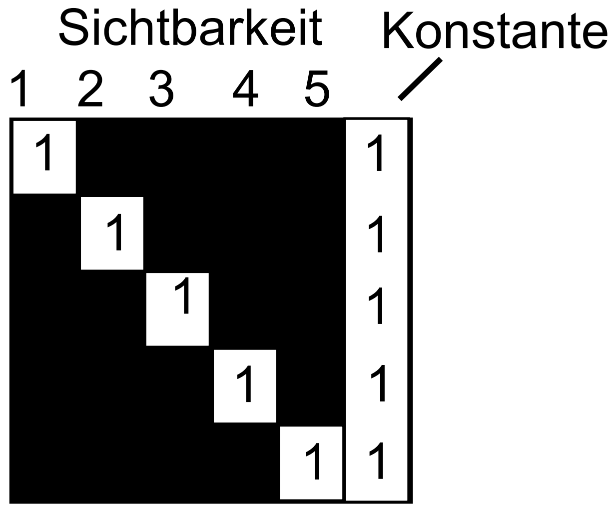
Beispielstudie:

- gleiches Model und Daten wie in der Regressionsanalyse:
  - Kriteriumsvariable: Reaktionszeit im Gedächtnistest
  - Regressoren: Sichtbarkeit der Bilder kodiert als 5 Faktorstufen

y	x
1,77	0
1,63	25
1,44	50
1,20	75
1,25	100

Kodierung →

$$\underbrace{\begin{array}{c} \text{Sichtbarkeit} \qquad\quad \text{Konstante} \\ \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & \\ \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array}}_{\text{Designmatrix X}} =$$



Visualisierung der Designmatrix X

⇒ **Grenzen des Beispielmodels**:

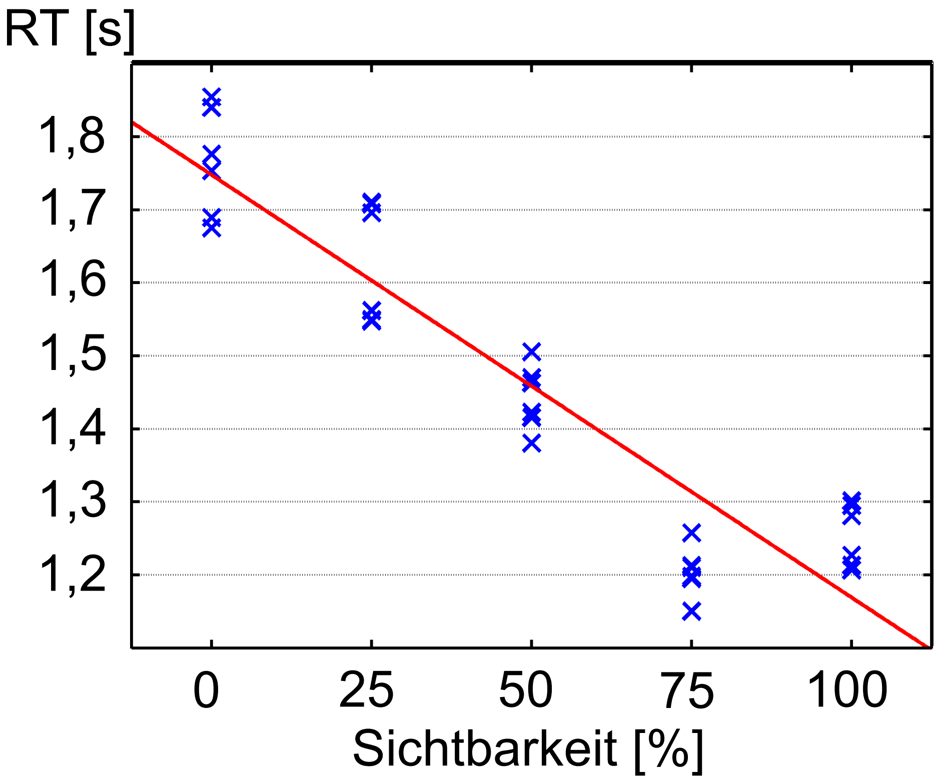
- keine Varianz in den einzelnen Faktorstufen
- Varianzanalyse nicht möglich

Modell:

- bisher: Mittelwerte der Reaktionszeit einer Person
- jetzt: einzelne Trials (n=30) im Gedächtnistest

⇒ Regressionsanalyse

y	x
1,43	50
1,58	25
1,45	25
1,85	0
1,06	100
1,59	0
1,37	75
1,46	50
1,33	75
1,21	100
..	...



$$\begin{bmatrix} 1{,}43 \\ 1{,}58 \\ 1{,}45 \\ 1{,}85 \\ 1{,}06 \\ 1{,}59 \\ 1{,}37 \\ 1{,}46 \\ 1{,}33 \\ 1.21 \\ \vdots \end{bmatrix} = \begin{bmatrix} 50 & 1 \\ 25 & 1 \\ 25 & 1 \\ 0 & 1 \\ 100 & 1 \\ 0 & 1 \\ 75 & 1 \\ 50 & 1 \\ 75 & 1 \\ 100 & 1 \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_0 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

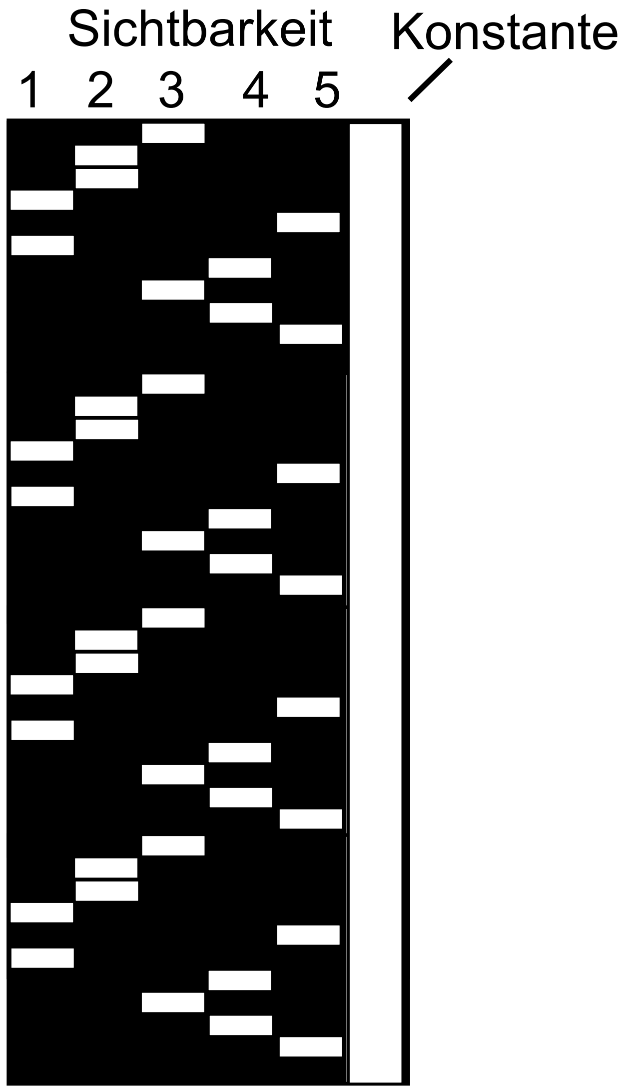
$$b = (X'X)^{-1} X'y = \begin{bmatrix} -0{,}006 \\ 1{,}75 \end{bmatrix}$$

⇒ Varianzanalyse

y	x
1,43	50
1,58	25
1,45	25
1,85	0
1,06	100
1,59	0
1,37	75
1,46	50
1,33	75
1,21	100
..	...

Kodierung

$$
\begin{array}{c}
\text{Sichtbarkeit} \qquad\qquad \text{Konstante} \\
\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & / \end{array} \\
\begin{bmatrix}
0 & 0 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{bmatrix}
\end{array}
=
$$



- Parameterschätzung

$$b = (X'X)^{-1} X'y = \left[\begin{array}{c} 0.55 \\ 0.41 \\ 0.23 \\ -0.01 \\ 0.03 \\ 1.25 \end{array}\right] \begin{array}{l} \text{100\% Sichtbark} \\ \text{75\% Sichtbarkeit} \\ \text{50\% Sichtbarkeit} \\ \text{25\% Sichtbarkeit} \\ \text{0\% Sichtbarkeit} \\ \text{Konstante} \end{array}$$

- Modellgüte

$R^2 = 0.99$ (aber nicht direkt vergleichbar zum ersten Regressionsmodel, da mehr Beobachtungen und mehr Regressoren!)

- Inferenz

- ANOVA: Vergleich einzelner Faktorstufen via t-Tests oder post hoc Tests möglich
- z.B.: H_1: RT bei 50% Sichtbarkeit größer als bei 25% Sichtbarkeit?
- Kontrast-Vektor c = [0 0 1 -1 0 0]
- Kontrast: $c' * b = 0 * \beta_1 + 0 * \beta_2 + 1 * \beta_3 + (-1 * \beta_4) + 0 * \beta_5 + 0 * \beta_6 = \beta_3 - \beta_4$

GLM in fMRT

-*mass univariate approach*:

- Aktivität wird in jedem Voxel unabhängig von allen anderen Voxeln analysiert (im Gegensatz zu multivariaten Verfahren, z.B. multivariate Mustererkennung)
- nur am Ende der Analyse werden die Ergebnisse in den übrigen Voxeln berücksichtigt ⇒ Korrektur der Statistik für multiple Vergleiche (Lena Tiedemanns Vortrag über multiple Vergleiche)

-abhängige Variable ‘y’ and unabhängige Variablen/Regressoren ‘x’:

- unterschiedlich in 1st und 2nd Level Analysen

1st level oder single subject Analysen

- first Level fMRT: **multiple Regressionsanalyse**
- abhängige Variable y:
  - Veränderungen der BOLD-Signalintensität in einem Voxel über die Zeit, also in jedem Volumen/Scan des Experimentes
  - in SPM wird Signalintensität Grau-Wert skaliert visualisiert: je dunkler, desto geringer, je heller desto höher

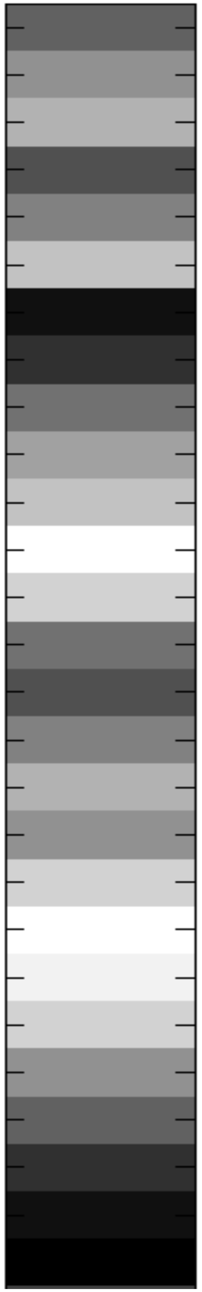
GLM in fMRI – 1st Level Regression Analyse

⇒ y = BOLD-Signalintensität in einem Voxel über die Zeit

Zeitserie

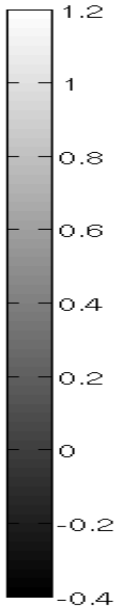
Volumen	Signalintensität
1	0.2
2	0.5
3	0.7
4	0.1
5	0.4
6	0.8
7	-0.3
8	-0.1
9	0.3
10	0.6
11	0.8
12	1.2
13	0.9
14	0.3
15	0.1
16	0.4
17	0.7
18	0.5
19	0.9
20	1.2
21	1.1
22	0.9
23	0.5
24	0.2
25	-0.1
26	-0.3
27	-0.4

grau-wert skaliert



relative Skala

Maximum: weiß

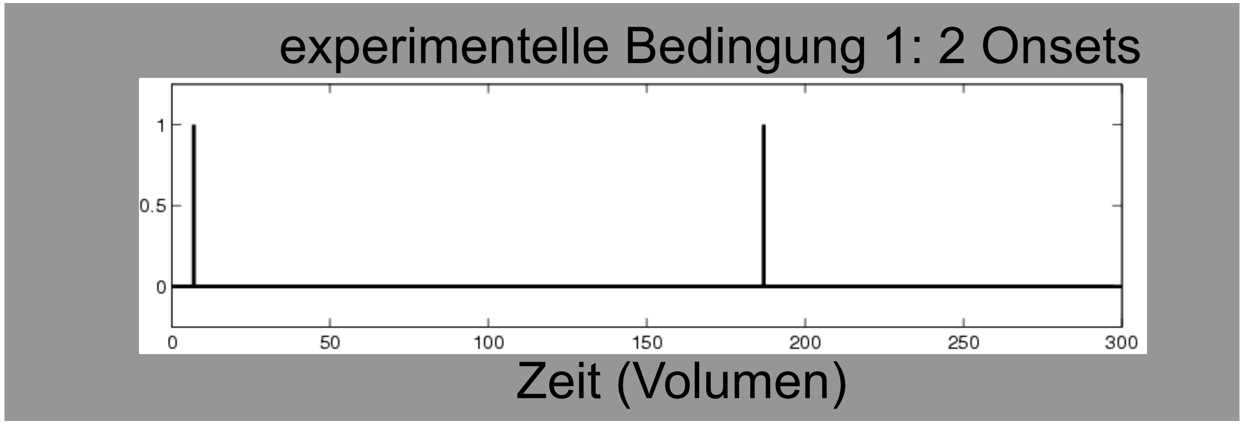


Minimum: schwarz

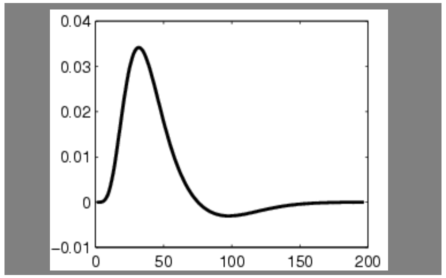
1st Level Regressoren

Regressoren = unabhängige Variablen

- unabhängig – basierend auf dem Timing der experimentellen Bedingungen
- Neurone, die in die Verarbeitung/Reaktion auf Stimuli einer experimentellen Bedingung involviert sind, zeigen zu dessen Onset erhöhte Aktivität
- elektrochemische neuronale Aktivität = sehr kurzes Ereignis
- δ-Funktion (stick-Funktion; 0 und 1) zum Onset der experimentellen Bedingung

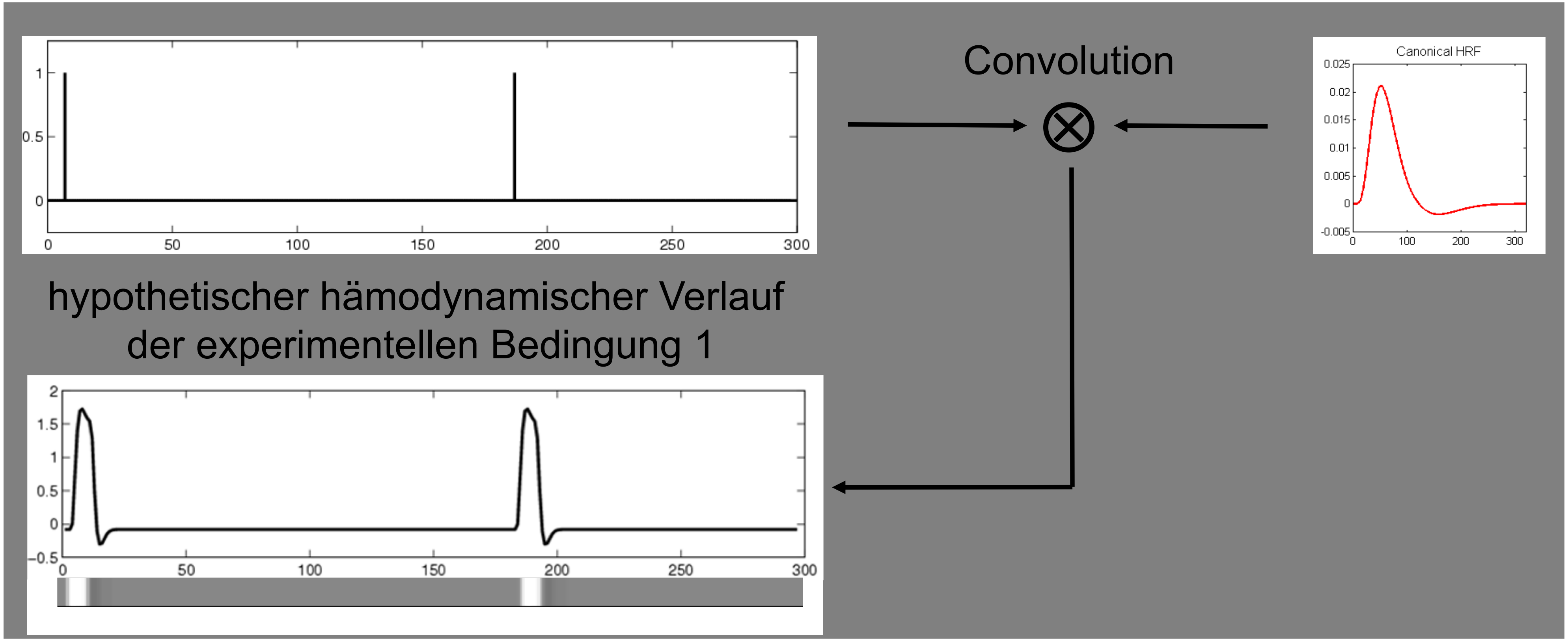


- neuronal Aktivität kann nicht direkt mittel fMRT gemessen werden
- erhöhte neuronale Aktivität triggert eine ***h**emodynamic **r**esponse **f**unction* (hrf)



1st Level Regressoren

– δ-Funktion zum Onset der experimentellen Bedingung
- neuronale Aktivität triggert hrf



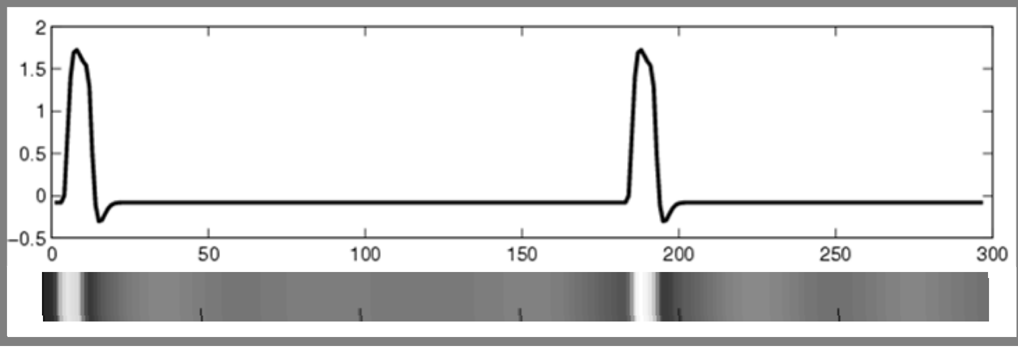
- Neurone involviert in die Prozessierung der Stimuli von Bedingung 1 sollten diese hypothetische hämodynamische Antwortfunktion zeigen ≈ BOLD Signal Veränderungen

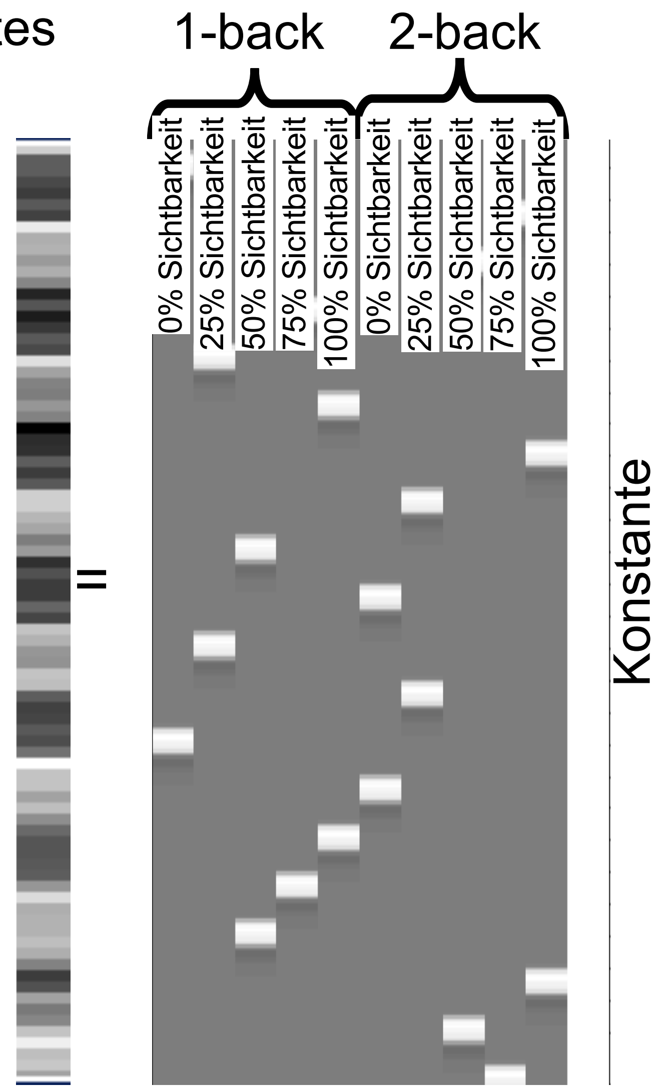
1st Level Designmatrix

- meist mehr als eine experimentelle Bedingung in Design Matrix
- Beispielstudie: 5 Sichtbarkeits- x 2 n-back Stufen = 10 Bedingungen
- Unabhängige Variable y:

  Signalintensität in jedem Volumen des Experimentes

hypothetischer hämodynamischer Verlauf
der experimentellen Bedingung 1

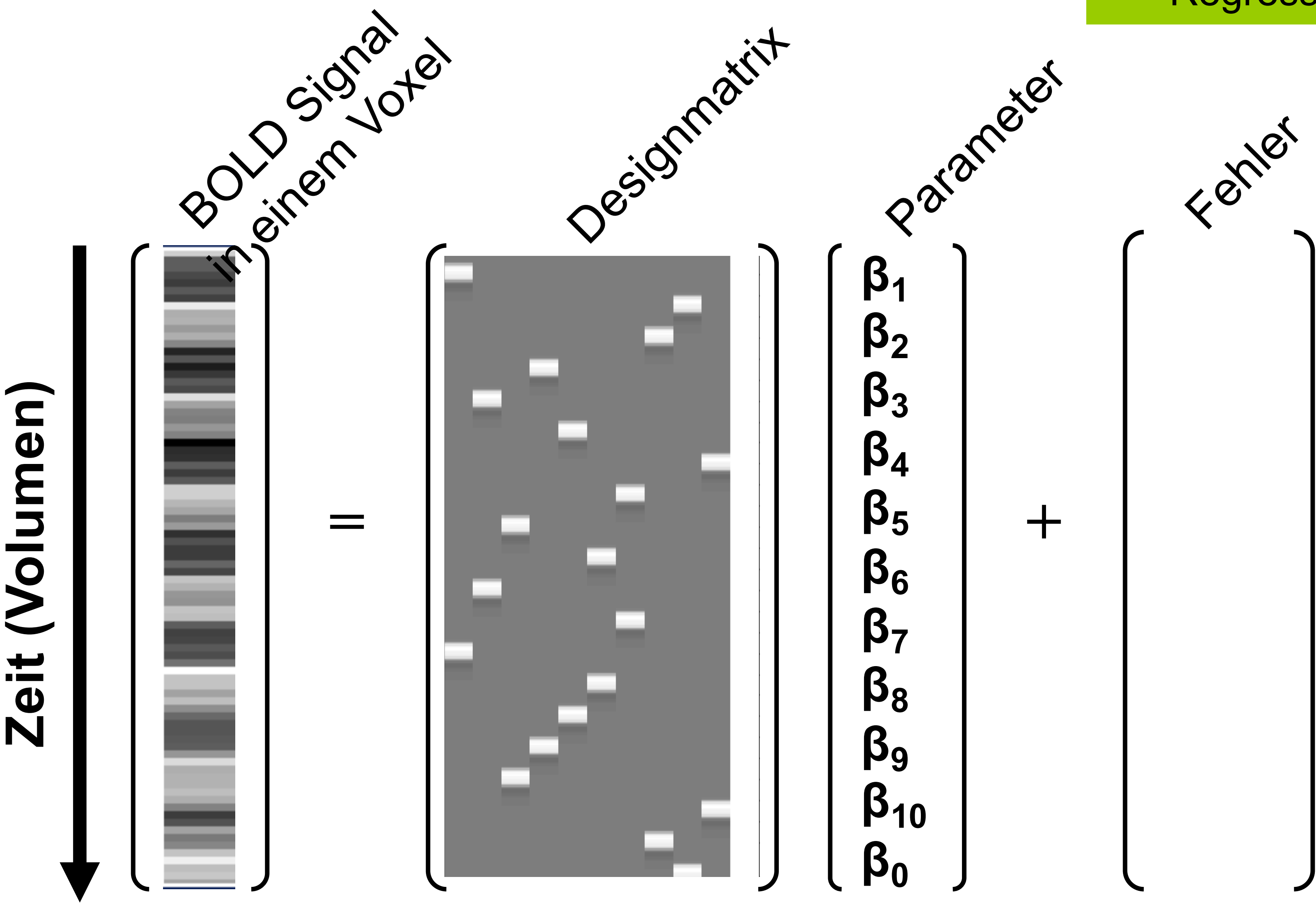




Designmatrix

mehr dazu morgen im Vortrag Model Spezifikation
und Schätzung von Jan Gläscher

GLM in 1st Level-fMRI



$$Y = X \bullet \beta + \varepsilon$$

GLM in 2nd Level-fMRT

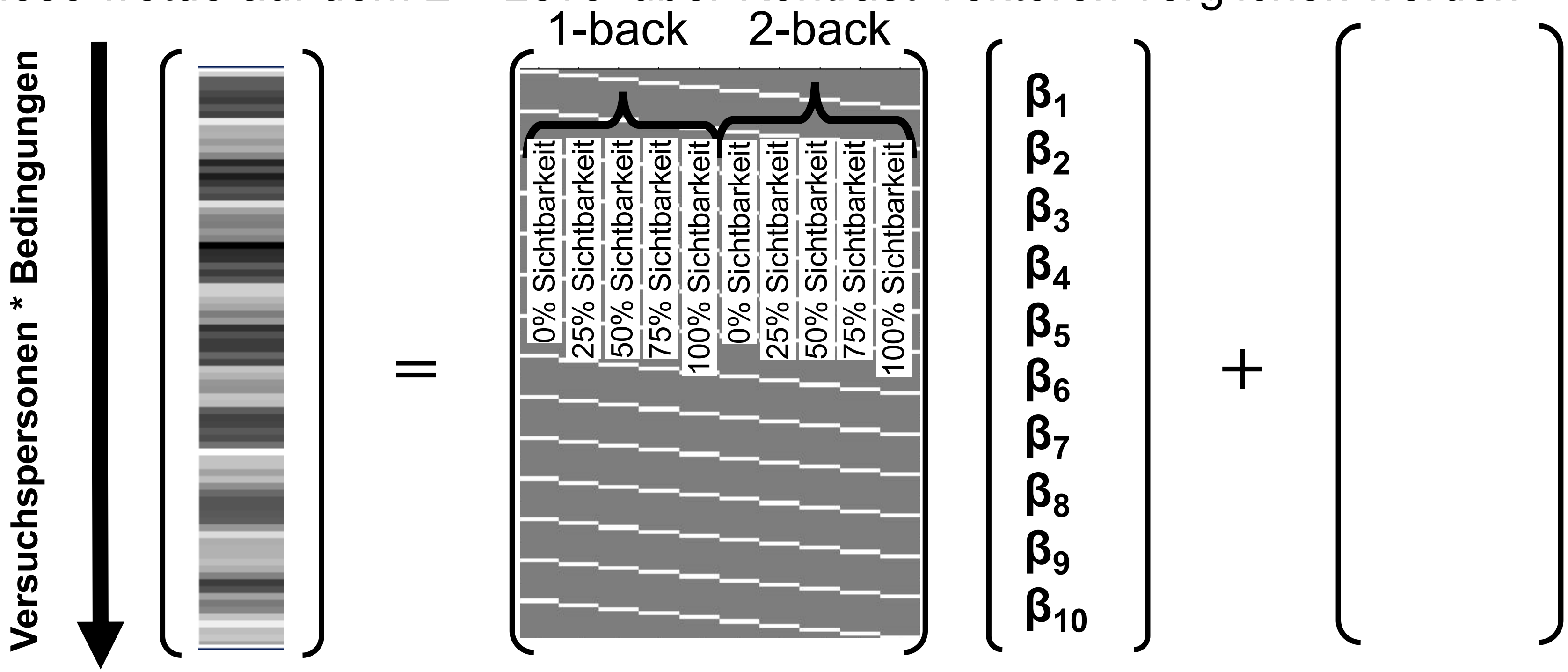
abhängige Variable

- Ergebnis der 1st-Level Analyse ein einem gegebenen Voxel über die Versuchspersonen (Werte in den con-images)

Regressoren

-0 und 1 Dummy-codierte Zuordnung dieser Ergebnisse (also der con-images) zu den Bedingungen oder Faktorstufen

- diese wetde auf dem 2nd-Level über Kontrast-Vektoren verglichen werden



$$Y = X * \beta + \varepsilon$$

- mehr über 1st und 2nd Level Modelle und Inferenz in SPM in Michael Roses Vorträgen

Unterschiede von Design-Matrizen

„full-rank“

$$\begin{bmatrix} 0 & 0{,}5 & 1 \\ 25 & 0{,}5 & 1 \\ 50 & 0{,}3 & 1 \\ 75 & 0{,}1 & 1 \\ 100 & 0{,}1 & 1 \end{bmatrix}$$

„rank-deficient“

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$\Rightarrow$ Regressor als Linearkombination der anderen Regressoren

$$1 * x_1 + 1 * x_2 + 1 * x_3 + 1 * x_4 + 1 * x_5 = x_6$$

$\Rightarrow$ nicht alle Parameter können nicht eindeutig geschätzt werden

$\Rightarrow$ nicht alle Kontraste können geschätzt werden

(Kontrastgewichte müssen in der Summe 0 ergeben)

Unterschiede von Design-Matrizen

„full-rank“

$$\begin{bmatrix} 0 & 0{,}5 & 1 \\ 25 & 0{,}5 & 1 \\ 50 & 0{,}3 & 1 \\ 75 & 0{,}1 & 1 \\ 100 & 0{,}1 & 1 \end{bmatrix}$$

$$b = \left(X'X\right)^{-1} X'y$$

Inverse von X

„rank-deficient“

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\beta = X^{+} y$$

Pseudoinverse von X
(in SPM)