## (c) Einfache lin. Korrelation und Regression:

$$S = \begin{bmatrix} \overbrace{S_{xx}}^{p} & S_{xy} \\ S_{xy}' & \underbrace{S_{yy}}_{q} \end{bmatrix}$$

$p = 1$ ; $q = 1$

① Bestimmtheitsmaß:

$$r^2 = \frac{S_{xy}}{S_{yy}} \cdot \frac{S_{xy}}{S_{xx}} = S_{yy}^{-1} \cdot S_{xy} \cdot S_{xx}^{-1} \cdot S_{xy}$$

② Regressionskoeffizient:

$$b_1 = \frac{S_{xy}}{S_{xx}} = S_{xy} \cdot S_{xx}^{-1}$$

in Matrix- (und) Vektorschreibweise geeignet auch für

## (D) (Multiple) Multiple lineare Korrelation und Regression:

$$S = \begin{bmatrix} \underbrace{S_{xx}}_{p} & S_{xy} \\ S_{xy}' & \underbrace{S_{yy}}_{q} \end{bmatrix}$$

$p > 1$

$q = 1$

transponierter Vektor $S_{xy}$

(1) multiples Bestimmtheitsmaß

$$\boxed{R^2 = S_{yy}^{-1} \cdot S_{xy}' \cdot S_{xx}^{-1} \cdot S_{xy} = PM_S}$$

Produktmatrix

~~Vektor~~ Matrix — Matrix — Vektoren

(2) Vektor d. Regressionskoeffizienten:

$$\boxed{b = S_{xy}' \cdot S_{xx}^{-1}}$$

ein Zeilenvektor — Vektor ($p \times q$ dimensional) — Matrix ($q$-dimensional)

(3) ähnliche Darstellung d. multipel lin. Regressionsgleichung:

GLEICHUNG DER REGRESSIONS-EBENE

$$\hat{Y}_i = b_\phi + b_{1i} X_1 + b_{2i} X_2 \longrightarrow Y_i = \hat{Y}_i + \epsilon_i$$

Kriterium ('criterion') Ordinatenabstand ('constant')

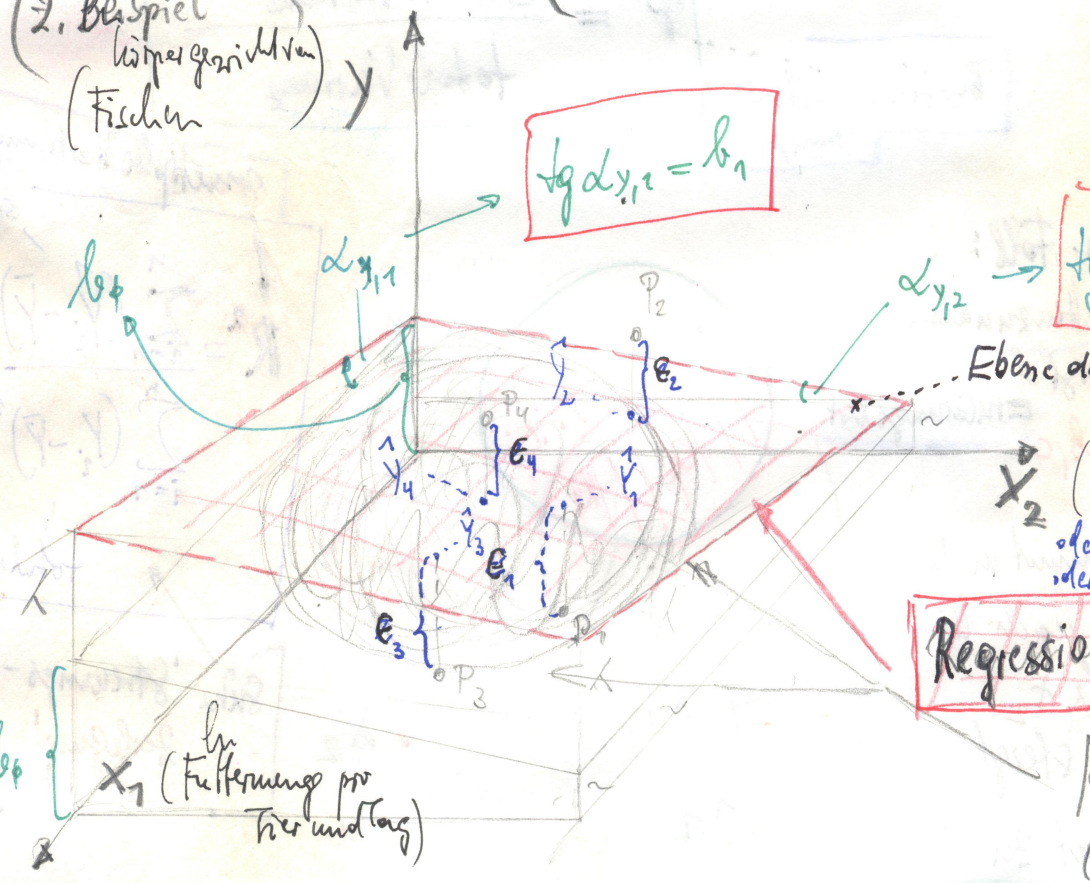
Prädiktoren ('predictors')

(semi)partielle Regressionskoeffizienten ('semipartial coefficient of regression')

Residualabweichung ('residuals')

Y... Körpergewicht Fische, g.
$X_1$... Futtermenge pro Tier und Tag, g
$X_2$... Nettogewicht Parasiten pro Wirt g

(2. Beispiel Körpergewicht von Fischen)




Ebene der $\hat{Y}_i$-Werte (Regressionswerte)

$X_2$ (Nettogewicht Parasiten) oder Anzahl d. Parasitencercarien oder ~ ~ d. Befall Bakterien

Regressionsebene

$X_1$ (Futtermenge pro Tier und Tag)

Punktewolke ('scatter')

Die Gleichung für die obige Regressionsebene in einer alternativen Darstellung lautet:

a) $$\hat{Y}_i = b_{y.12} + b_{y1.2\,i} X_1 + b_{y2.1\,i} X_2$$

bzw. b) $$\cancel{\hat{X}_{3i} = b_{1.23} + b_{12.3\,i} X_1 + b_{13.2\,i} X_3}$$

Anpassung d. Regressionsebene (~ Regressionsgeraden) nach d. Methode d. kleinsten Fehlerquadrate mittels NORMALGLEICHUNGS-SYSTEM

(4) Begriffe „multipel" – „partiell"

einf. lin. Fall:

(a)

Erklärung mit Hilfe d. Mengenlehre ('set theory')


die Varianz (= Information) von Y, die durch $X_1$ erklärt werden soll

('total variance' – Y)

Anteil an der Y-Varianz, der von X erklärt wird.

$$r^2 = \frac{\text{erklärte Varianz}}{\text{totale Varianz}}$$

Bestimmtheitsmaß

(b) multipel lin. Fall:

multipel bedeutet: durch Hinzunahme eines weiteren Prädiktors steigt d. Erklärungswert


multiples Bestimmtheitsmaß

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$

erkl. Str.

totale Str.

[Siehe 'Streuungszerlegung'!]

(1) erklärte Varianz: a, b, c

(2) Zusätzlich zu a, b kommt noch der Anteil c hinzu nach Hinzufügen d. Variable $x_2$
> Erklärungswert steigt
> $R^2$ nimmt zu

(3) c ist der partielle Beitrag von $X_2$ zur Aufklärung von y, ebenso wie a der partielle Beitrag von $x_1$ — " — — " — ist

partiell heißt: exclusiv, ohne Einfluß der anderen, im Ansatz befindlichen Variablen.

(c) Beispiel für **Auspartialisierung**

psychische Beeinträchtigung im Falle von

AIDS-Erkrankungen (ja – nein : dummy Variable)


„Genetische, immunologische Disposition"

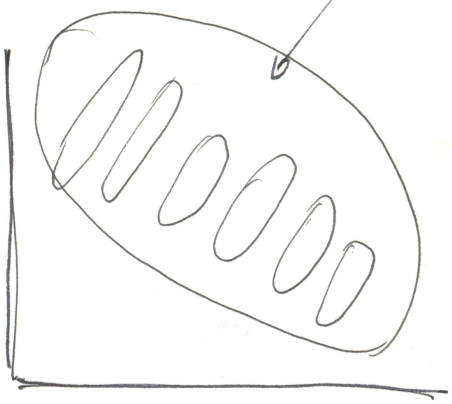
„sozialer Lebensumstände"
(z.B. Einkommen, Arbeitslosigkeit – ja od. nein)

Es wird Ev. $x_1$ aus y und aus $x_2$ auspartialisiert, um den „reinen Zusammenhang zw. Aids-Erkr. (psych. Beeinträchtigung) und Lebensumständen zu erfassen

(d) weitere Beispiele für Auspartialisieren:

(1) Geschlechtszugehörigkeit

(2) → Gruppenzugehörigkeit:

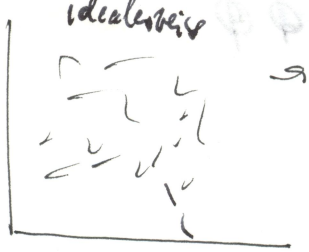


nötige Auspartialisierung d. Untergruppen

---

**Residualanalyse** wie im einf. lin. Fall durchführen

a)



$\epsilon_i$ ↑

Residuen

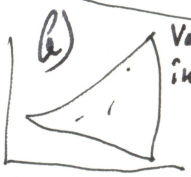
$Y_i$ ↑ Kriterium

ungewichtete Punktwolk.

→ ev. Prüfung auf Normalität bei $\epsilon_i$

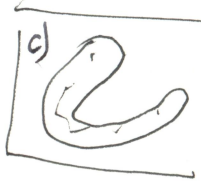
→ nicht erlaubt; die Konst. d. Vorgehens des ALM nicht erfüllen

b)



Varianz inhomogen

c)



d)

(5) Bei Regressionsanalysen mit z-standardisierten Variablen ⇒ **Pfadanalyse** ('path analysis')

(6) **Andere Regressionsansätze:**

- Fourier- (periodische / zyklische) Regressionen ('cyclic, periodic regr.')
- Polynom-Anpassungen
- Logit-Regression, Logistische Regression
- Poisson-Regression
- Probit-Regression

Modelle der 'Generalized Linear Models'

beruhen meist auf Maximum-Likelihood-Verfahren

---

Beispiel für } **Dummy-Code:**

(A) für Codierung von 2 Gruppen

$x \to \phi$, $x \to 1$

$g = 2$

(B) für Codierung von 3 Gruppen

$g = 3$

$(g-1)$ dummies

z.B.

| | $x_1$ | $x_2$ |
|---|---|---|
| $g_1$ | 1 | $\phi$ |
| $g_2$ | $\phi$ | 1 |
| $g_3$ | $\phi$ | $\phi$ |

die beiden Dummyvariablen