

## 12 Logistische Regression

### 12.1 Einleitung

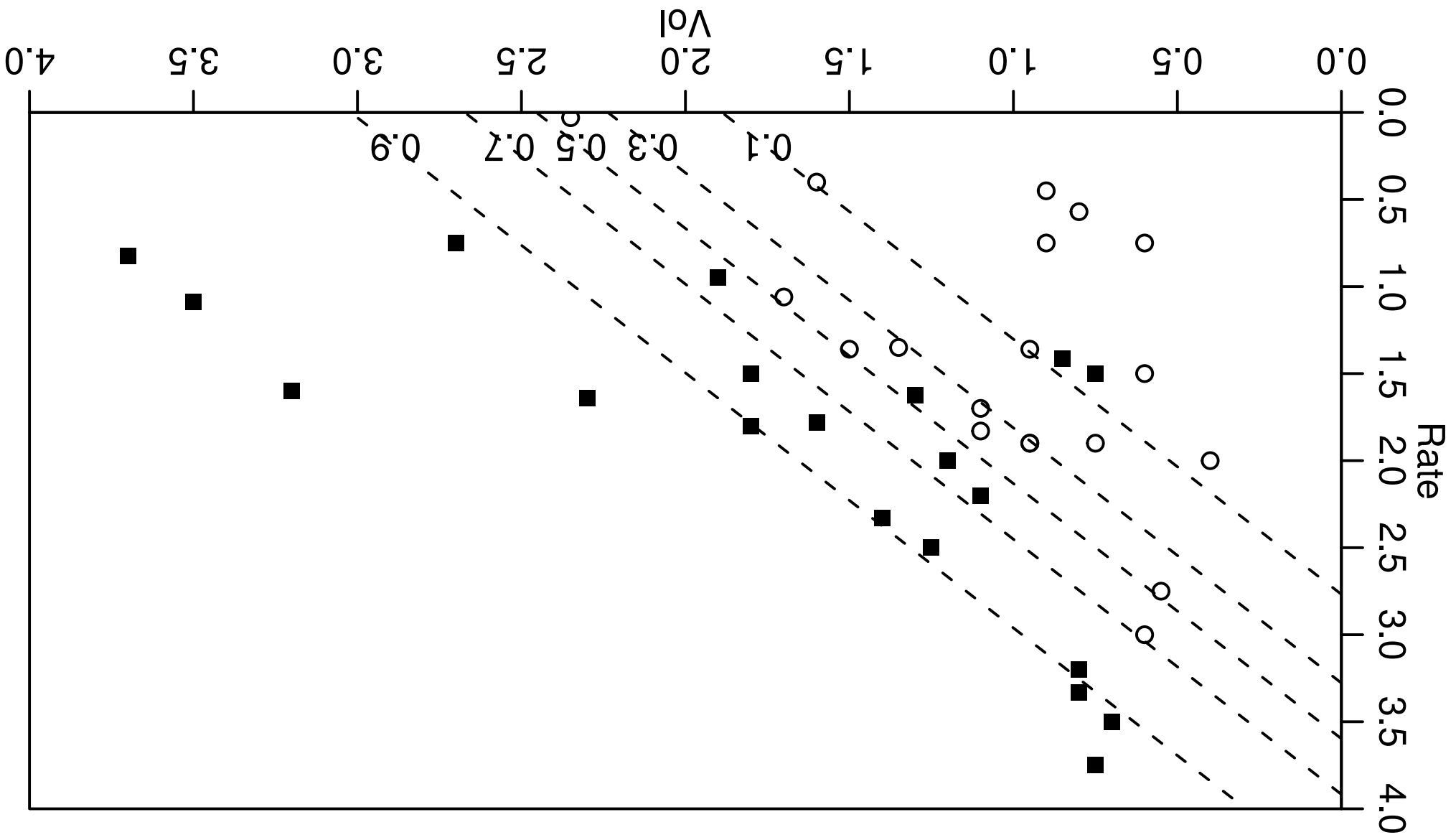
b **Beispiel Ader-Verengung**

$Y$ : Ader-Verengung ja (1) / nein (0)

Eing.: Atem-Volumen ( $V_{O_2}$ ) und Atem-Frequenz (Rate)

**Ziel:**  $P\langle Y = 1 \mid V_{O_2}, \text{Rate} \rangle$  modellieren!

c  $P\langle Y_i = 1 \rangle = h\langle x_{i(1)}^{(1)}, x_{i(2)}^{(2)}, \dots, x_{i(m)}^{(m)} \rangle$



$$P\langle Y_i = 1 \rangle = h\langle x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(m)} \rangle$$

d Weshalb keine gewöhnliche lineare Regression?

$$Y_i = \beta_0 + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \dots + \beta_m x_i^{(m)} + E_i$$

• Zufallsabweichung  $E_i$  ?

$$\mathcal{E}\langle Y_i \rangle = \beta_0 + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \dots + \beta_m x_i^{(m)}$$

Es gilt  $P\langle Y_i = 1 \rangle = \mathcal{E}\langle Y_i \rangle$ . Also wäre gleiche Form möglich!

• Geschätzte Werte können  $> 0$  und  $> 1$  werden.

→ Transformation von  $Y_i$ ? 2 Werte bleiben 2 Werte!

→ Transformation von  $\mathcal{E}\langle Y_i \rangle = P\langle Y_i = 1 \rangle$ !

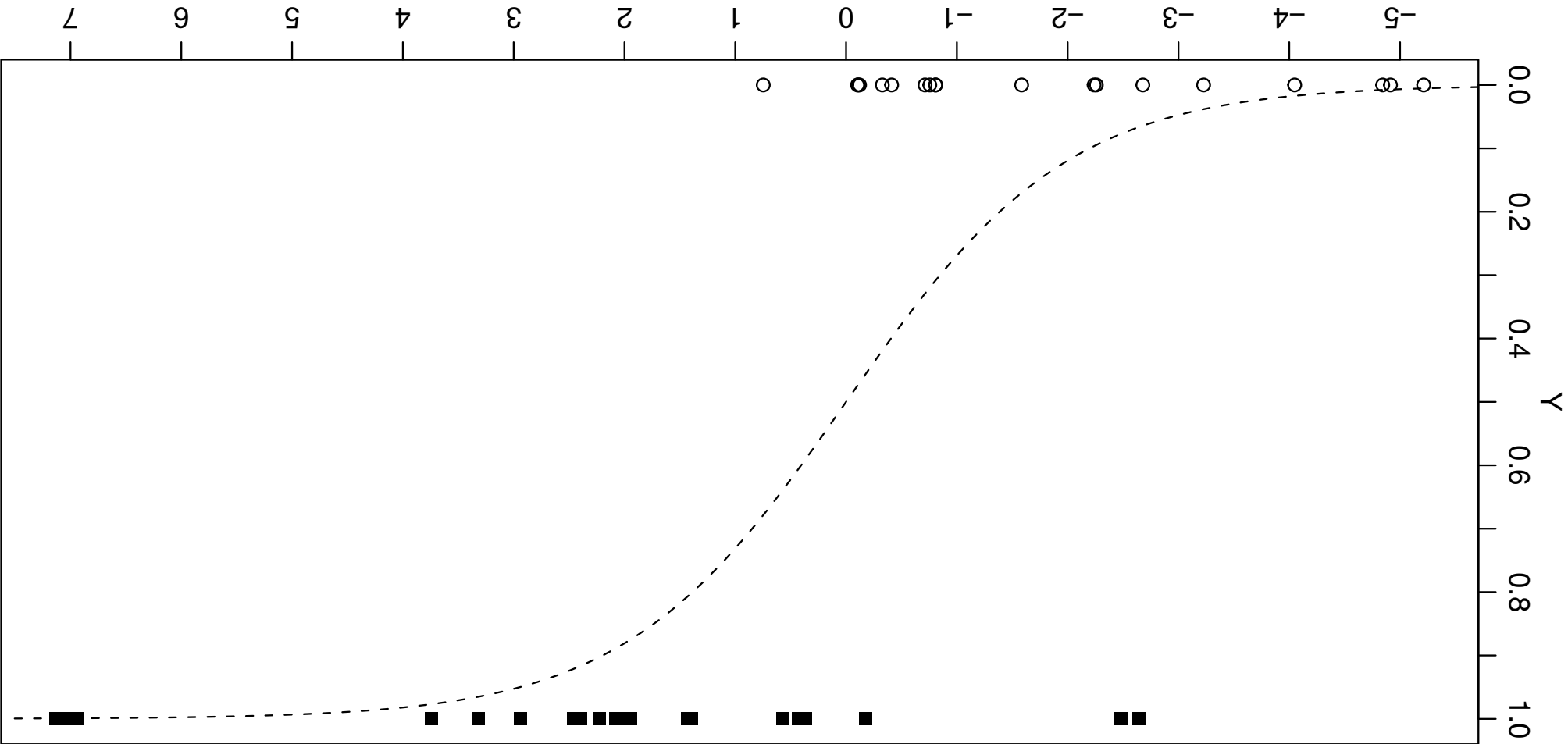
e **Modell**. Logit-Funktion  $g\langle\pi\rangle = \log\left\langle\frac{1-\pi}{\pi}\right\rangle = \log\langle\pi\rangle - \log\langle 1-\pi\rangle$

$$g\langle P\langle Y_i = 1 \rangle \rangle = \log\langle P\langle Y_i = 1 \rangle \rangle - \log\langle P\langle Y_i = 0 \rangle \rangle = \eta_i$$

$$\eta_i = \beta_0 + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \dots + \beta_m x_i^{(m)}.$$

$\eta$ : „linearer Prädiktor“.

f Beispiel:  $g\langle P\langle Y = 1 \rangle \rangle = -9.53 + 3.88 \cdot \text{Vol} + 2.65 \cdot \text{Rate}.$



**Diskriminanzanalyse:** $Y_i$  Gruppen-Zugehörigkeit $X_i^{(j)}$  multivariate Beobachtungen.

Logistische Regression:

1. Schätzen:  $\hat{\pi}_i$ 2. Zuordnen:  $Y = 1$ , wenn  $\hat{\eta}_i > 0$  ( $\hat{\pi}_i > 0.5$ )

h Weitere **Anwendungen:**

- Toxikologie: Gift tödlich für Mäuse?
- Medizin: Behandlung erfolgreich?
- Ausfall von Geräten,
- Auftreten eines Fehlers an einem Produkt,
- Vorhandensein eines bestimmten Merkmals bei Lebewesen,
- Kundenscoring

Allgemein: Zugehörigkeit zu einer von zwei Gruppen.

## 12.2 Betrachtungen zum Modell

- a Gleiche Flexibilität wie lineare Regression.  
 Häufig: **Nominale Eingangs-Variablen** (→ „dummy“-Variable = Indikator-V)

- b **Beispiel Beurteilung des Arbeitsplatzes.**

$Y_i$  „zufrieden“ (1), „unzufrieden“ (0).  
 $X_i^{(j)}$  Region, Alter, Geschlecht, Hautfarbe.

Nur ein Faktor →  $2 \times k$ -**Kreuztabelle**.

	NE	Mid-Atl.	S	Midwest	NW	SW	Pacific	total
unzutr.	738	166	514	749	711	482	209	3569
zufrieden	1161	406	916	1240	1221	971	465	6380
total	1989	572	1430	1899	1932	1453	674	9949

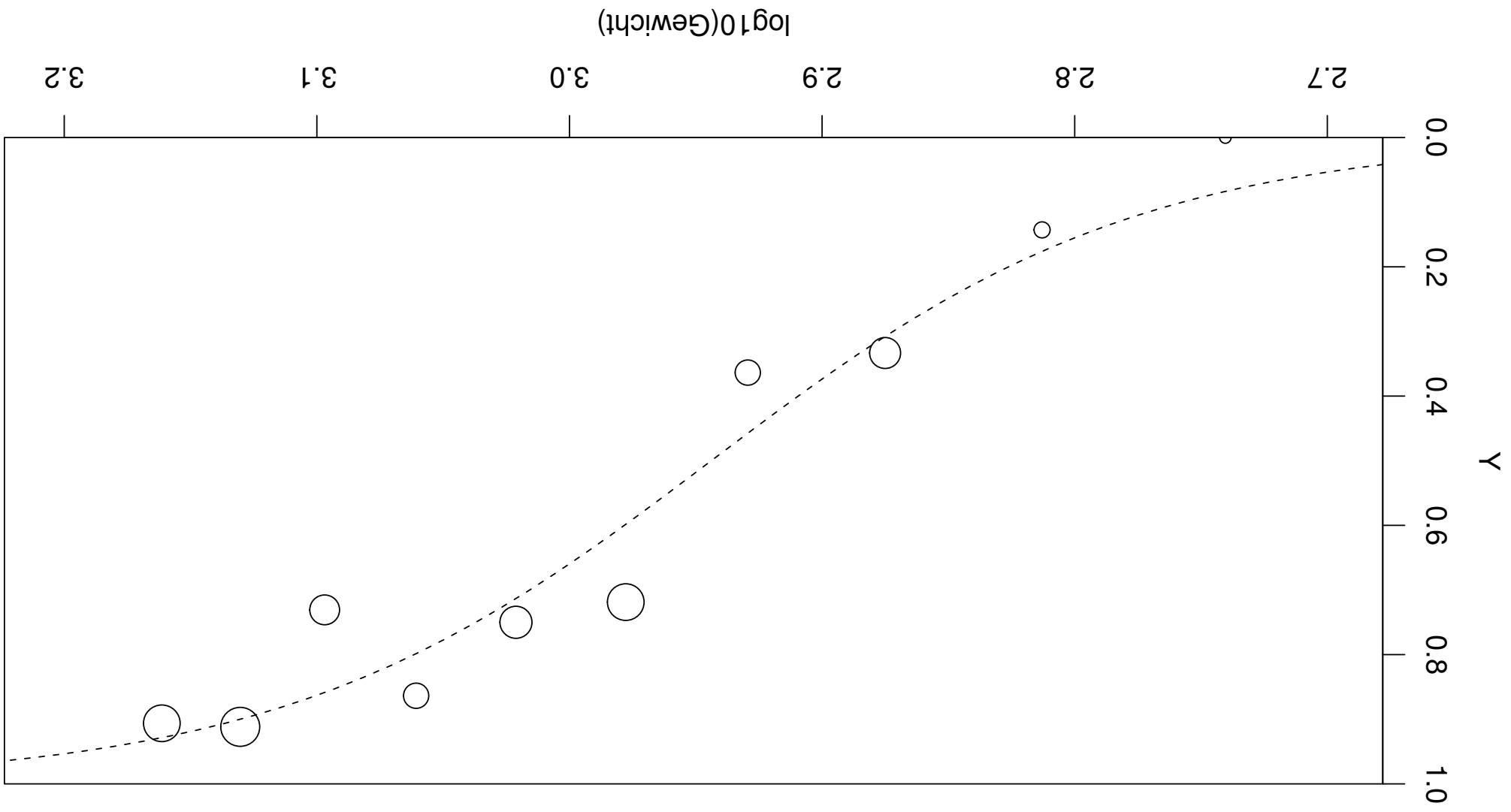
Höher-dimensionale Kontingenztafeln → log-lineare Modelle.

- c **Gruppierte Daten:**  $m_\ell$  Beob.  $Y_i$  zu gleichen  $\bar{x}_i = \tilde{x}_i$ :  
 $\tilde{Y}_\ell = \sum_i: \bar{x}_i = \tilde{x}_i \quad Y_i \sim B(m_k, \pi_k) \quad \mathcal{E}\langle \tilde{Y}_\ell / m_\ell \rangle = \pi_\ell$   
 → Logistische Regression:  $g\langle \pi_\ell \rangle = \eta_\ell$

- d **Beispiel Überleben von Frühgeburten.** 247 Säuglinge.

Erklärende Variable: Geburtsgewicht. Klassen von je 100 g Gewicht.

	n	Surv.no	Surv.yes	Weight
1	10	10	0	550
2	14	12	2	650
3	27	18	9	750
4	22	14	8	850
5	32	9	23	950
6	28	7	21	1050
7	22	3	19	1150
8	26	7	19	1250
9	34	3	31	1350
10	32	3	29	1450



e **Transformierte Beobachtungen.**

$$\mathcal{E}\langle \tilde{Y}_\ell / m_\ell \rangle = \pi_\ell, \quad g\langle \pi_\ell \rangle = \text{linearer Prädiktor.}$$

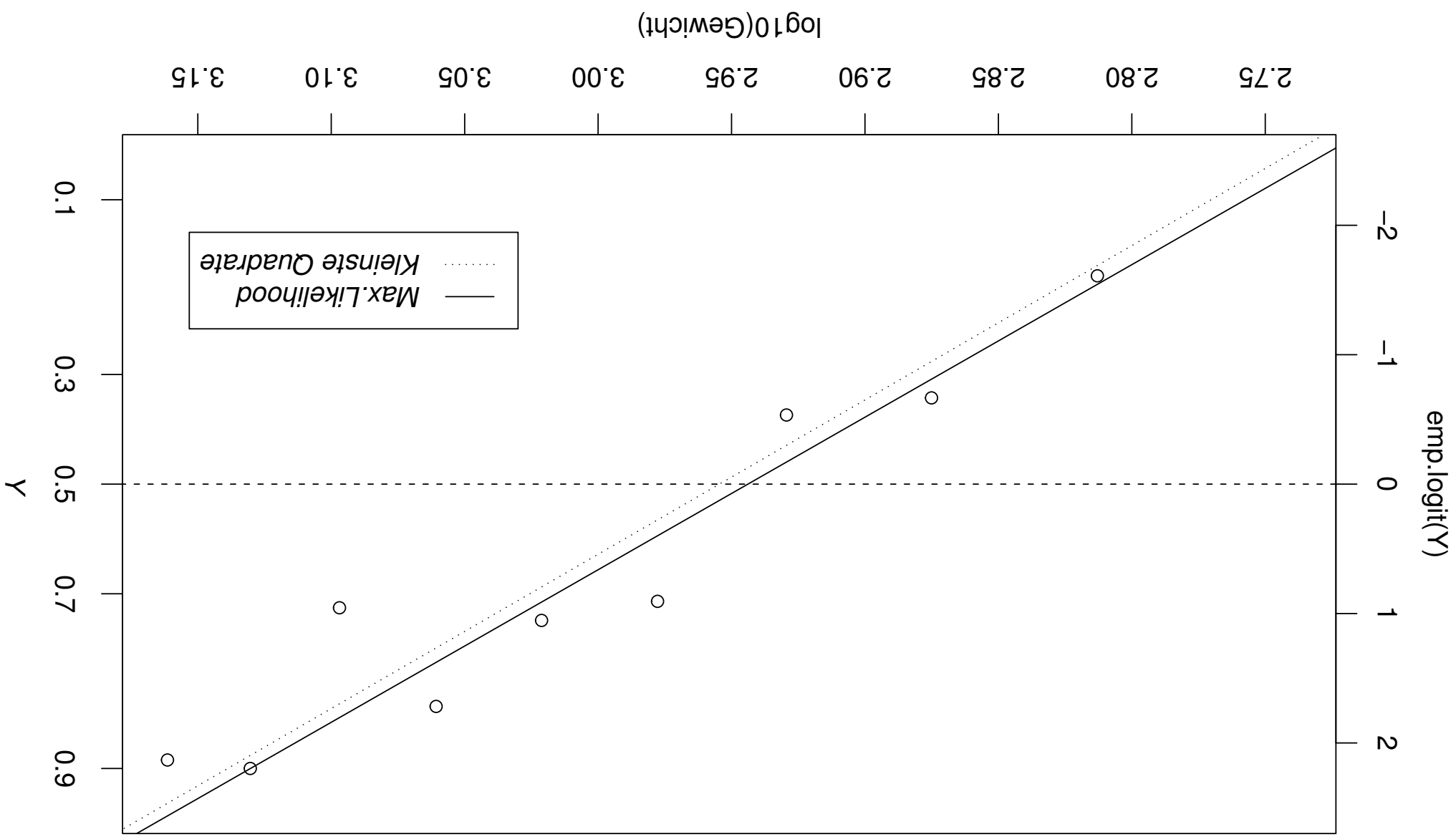
$$g\langle \tilde{Y}_\ell / m_\ell \rangle \approx \text{linearer Prädiktor.}$$

Was tun mit  $Y_\ell / m_\ell = 0$  oder  $= 1$ ?  $g\langle 0 \rangle = -\infty$ ,  $g\langle 1 \rangle = \infty$ .

Abhilfe: **Empirische Logits**

$$\tilde{Z}_\ell = \log \left\langle \frac{\tilde{Y}_\ell + 0.5}{m_\ell - \tilde{Y}_\ell + 0.5} \right\rangle.$$

← Gewöhnliche multiple Regression mit  $Z_\ell$ ? → Näherung.



‡ Interpretation von Koeffizienten? Braucht die folgenden Begriffe:

Wettverhältnisse (odds)

$$\text{odds} = \frac{P\langle Y_i = 1 \rangle}{1 - P\langle Y_i = 1 \rangle}$$

$\pi = 1/4$  : Wettverhältnis 1:3 ( 3 × häufiger Misserfolg )

$\log(\text{odds}) = g\langle Y_i = 1 \rangle$ ,  $g$ : Logit-Funktion.

$\log(\text{odds}) = \eta \rightarrow$  Wahrscheinlichkeit  $\pi = g^{-1}\langle \eta \rangle = \frac{\exp\langle \eta \rangle}{1 + \exp\langle \eta \rangle}$ .

$g^{-1}$ : "logistische Funktion".

Logistische Regression:  $\log(\text{odds}) = \text{linearer Prädiktor}$   $\sum_j \beta_j x_i^{(j)}$ .  
 $\pi_i = \text{logistische Funktion} \left\langle \sum_j \beta_j x_i^{(j)} \right\rangle$ .

g **Odds ratio** (Doppelverhältnis): Vergleich zweier Beobachtungen

$$\log \left\langle \frac{\text{odds}\langle \bar{x}_1 \rangle}{\text{odds}\langle \bar{x}_2 \rangle} \right\rangle = \log \langle \text{odds}\langle \bar{x}_1 \rangle \rangle - \log \langle \text{odds}\langle \bar{x}_2 \rangle \rangle$$

$$= \eta_1 - \eta_2 = (\bar{x}_1 - \bar{x}_2) \bar{\beta}$$

Koeff.  $\beta_j$ : Vergrößerung von  $x^{(j)}$  um 1 erhöht odds ratio um Faktor  $e^{\beta_j}$ .

h Beispiel Ader-Verengung:

Wert für Vol = **0.5**, Rate = 1.75

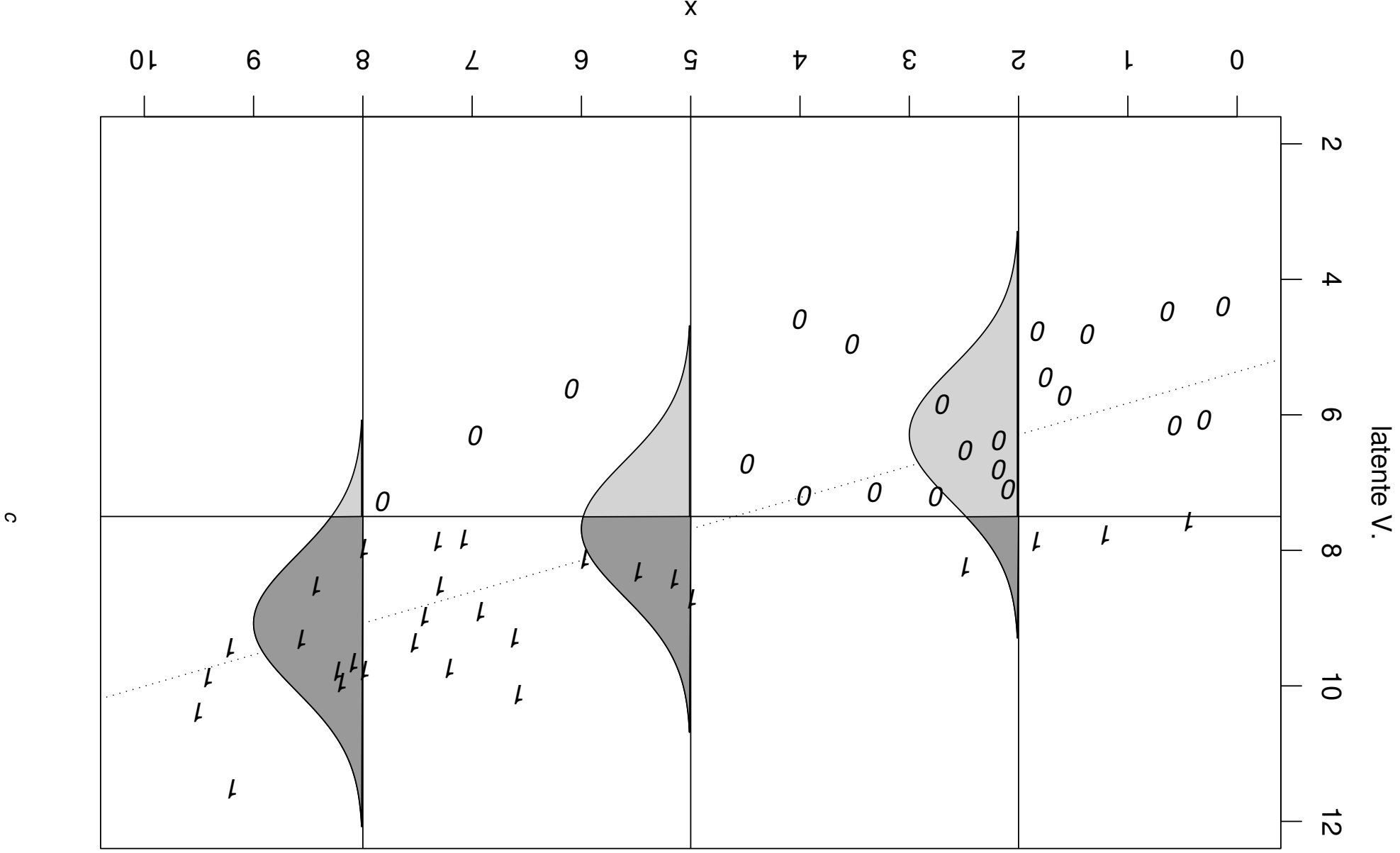
$$\log(\text{odds}) = -9.56 + 3.88 \cdot 0.5 + 2.65 \cdot 1.75 = -2.85$$

$$\rightarrow \text{odds} = 0.0578, \quad g^{-1}(-2.85) = 0.0546$$

Vergleich Vol = **1.5**, Rate = 1.75: odds ratio:  $e^{3.88} = 48.4$

$$\rightarrow \text{odds} = 0.0578 \cdot 48.4 = 2.80, \quad 2.80/3.80 = 0.73$$

! Modell der latenten Variablen = Schwellenwert-Modell.



$$\begin{aligned}
 Z_i &= \tilde{x}_i^T \tilde{\beta} + E_i \\
 \pi_i = P\langle Y_i = 1 \rangle = P\langle Z_i \geq c \rangle &= P\langle E_i \geq c - \tilde{x}_i^T \tilde{\beta} \rangle \\
 &= 1 - F\langle c - \beta_0 + \sum_j \beta_j x_i^{(j)} \rangle \\
 \bar{\beta} = [\tilde{\beta}_0 - c, \tilde{\beta}_1, \dots, \tilde{\beta}_m] &\Rightarrow P\langle Y_i = 1 \rangle = g^{-1}\langle \tilde{x}_i^T \bar{\beta} \rangle \text{ mit } g^{-1}\langle \eta \rangle = 1 - F\langle -\eta \rangle
 \end{aligned}$$

$F$ : kumulative Verteilungsfunktion des Zufallsfehlers  $E_i$

$E_i \sim$	logistische Vt.:	logistische Regression
$E_i \sim$	Normal-Vt.:	Probitmodell
$E_i \sim$	Extremwertv.:	komplementäres log-log Modell

## 12.3 Schätzungen und Tests

a) Methodik der Likelihood. Programme existieren!

b) Log-Likelihood:

$$\begin{aligned} \ell(\tilde{y}; \bar{\beta}) &= \log \prod_{\ell} P_{\tilde{Y}_{\ell}}(y_{\ell}) = \left\langle \sum_{\ell} \log \binom{m_{\ell}}{y_{\ell}} \pi_{y_{\ell}}^{\ell} (1 - \pi_{\ell})^{m_{\ell} - y_{\ell}} \right\rangle \\ &= \left\langle \sum_{\ell} \log \binom{m_{\ell}}{y_{\ell}} \right\rangle + \left\langle \sum_{\ell} (y_{\ell} \log \pi_{\ell} + (m_{\ell} - y_{\ell}) \log(1 - \pi_{\ell})) \right\rangle \end{aligned}$$

mit  $\text{logit}(\pi_{\ell}) = \bar{x}_{\ell}^T \bar{\beta}$

Ungruppierte Daten:  $m_{\ell} = 1$ .

$$\ell(\tilde{y}; \bar{\beta}) = \sum_{y_i=1} \log \langle \pi_i \rangle + \sum_{y_i=0} \log \langle 1 - \pi_i \rangle.$$

## f Beispiel Ader-Verengung.

```
Call: glm(formula = y ~ Vol + Rate, family = binomial,
data = d.adern)
Deviance Residuals:  . . .
Coefficients:
```

	Value	Std. Error	z <sub>-appr.</sub>	Pr(> z )	Signif.
(Intercept)	-9.529	3.2140	-2.96	0.003	**
Vol	3.882	1.4202	2.73	0.006	**
Rate	2.649	0.9095	2.91	0.004	**

(Dispersion Parameter for Binomial family taken to be 1)  
Null Deviance: 54.04 on 38 degrees of freedom  
Residual Deviance: 29.77 on 36 degrees of freedom  
Number of Fisher Scoring Iterations: 5

Correlation of Coefficients:  
(Intercept) Vol  
Vol -0.9358  
Rate -0.9228  
0.7631

$$D\langle \hat{y}; \hat{\pi} \rangle = 2 \left( \ell_{(M)} - \ell_{\langle \hat{y}; \hat{\beta} \rangle} \right) \cdot$$

Maximale erreichbare Log-Likelihood ( $\hat{\pi}_\ell = y_\ell / m_\ell$ ):

$$\ell_{(M)} = \sum_{\ell} \binom{m_\ell}{y_\ell} \log \langle y_\ell \rangle + y_\ell \log \langle y_\ell \rangle$$

$$+ (m_\ell - y_\ell) \log \langle m_\ell - y_\ell \rangle - m_\ell \log \langle m_\ell \rangle \cdot$$

h Modelle vergleichen: **Likelihood-Ratio-Tests**. Test-Statistik:

$$\tilde{D}\langle \hat{y}; \hat{\pi}_{(K)} \rangle, \hat{\pi}_{(G)} \rangle = D\langle \hat{y}; \hat{\pi}_{(K)} \rangle - D\langle \hat{y}; \hat{\pi}_{(G)} \rangle = 2(\ell_{(G)} - \ell_{(K)})$$

asymptotisch chiquadrat-verteilt, wenn das kleine Modell stimmt.

! **Residuen-Devianz** vergleicht geschätztes Modell mit maximalem Modell

→ „Anpassungstest“

Achtung: Geht nur bei nicht zu kleinen  $m_\ell$ , also nur bei gruppiereten Daten.

! Kleinstes Modell:  $\pi_i$  für alle Beobachtungen gleich.

$$\ell(0) = \sum_{\ell} \log \binom{y_{\ell}}{m_{\ell}} + \log \left\langle \frac{1 - \pi}{\pi} \right\rangle \sum_{\ell} y_{\ell} + n \log \langle 1 - \pi \rangle$$

mit  $\pi \approx \sum_{\ell} y_{\ell} / n$ .

**Null-Devianz:**  $D \langle \bar{y}; \pi \rangle = 2 \left( \ell(M) - \ell(0) \right)$

→ Gesamt-Test für das Modell. ( $H_0$ : alle  $\beta_s = 0$ )

## 12.4 Residuen-Analyse

a **Rohe Residuen** (response residuals)

$$R_\ell = \tilde{Y}_\ell / m_\ell - \hat{\pi}_\ell, \quad \hat{\pi}_\ell = g^{-1} \langle \tilde{x}_\ell^T \hat{\beta} \rangle$$

**Pearson residuals:**  $R_\ell^{(P)} = R_\ell / \sqrt{\hat{\pi}_\ell(1 - \hat{\pi}_\ell)} / m_\ell$

**Deviance residuals:** Beitrag der  $i$ -ten Beobachtung zur Devianz

**Working residuals:**

Berechnung der logist. Regr.: iterativ gewichtete Kl.Qu. (vgl. nichtlin. Regr.)  
 → lineare Näherung → Residuen: „working residuals“.

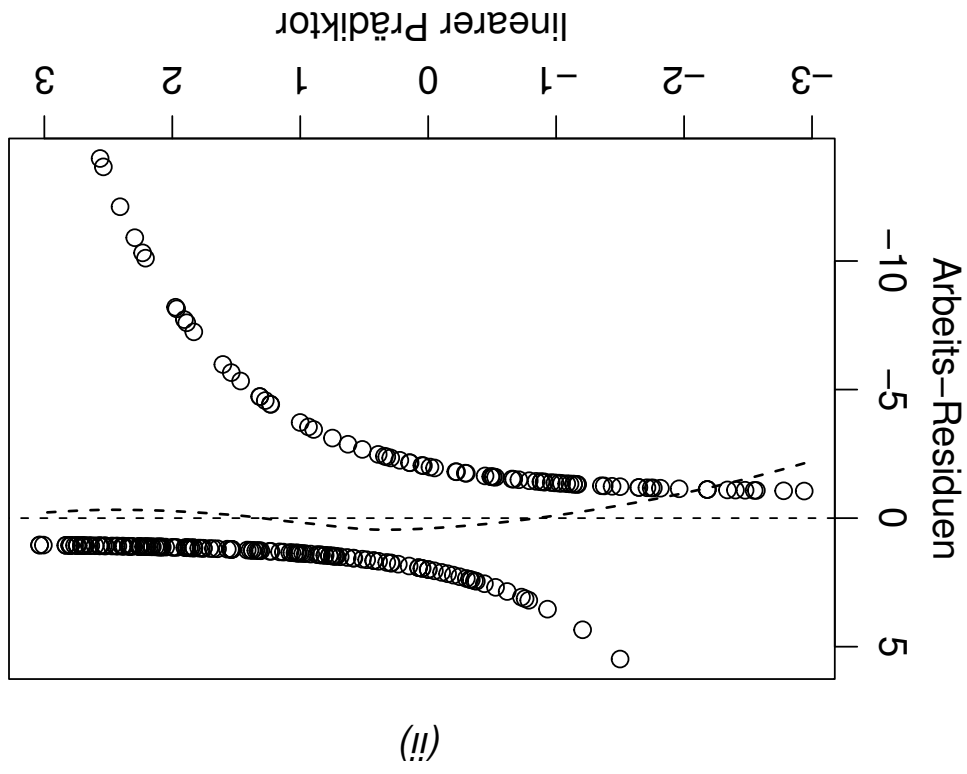
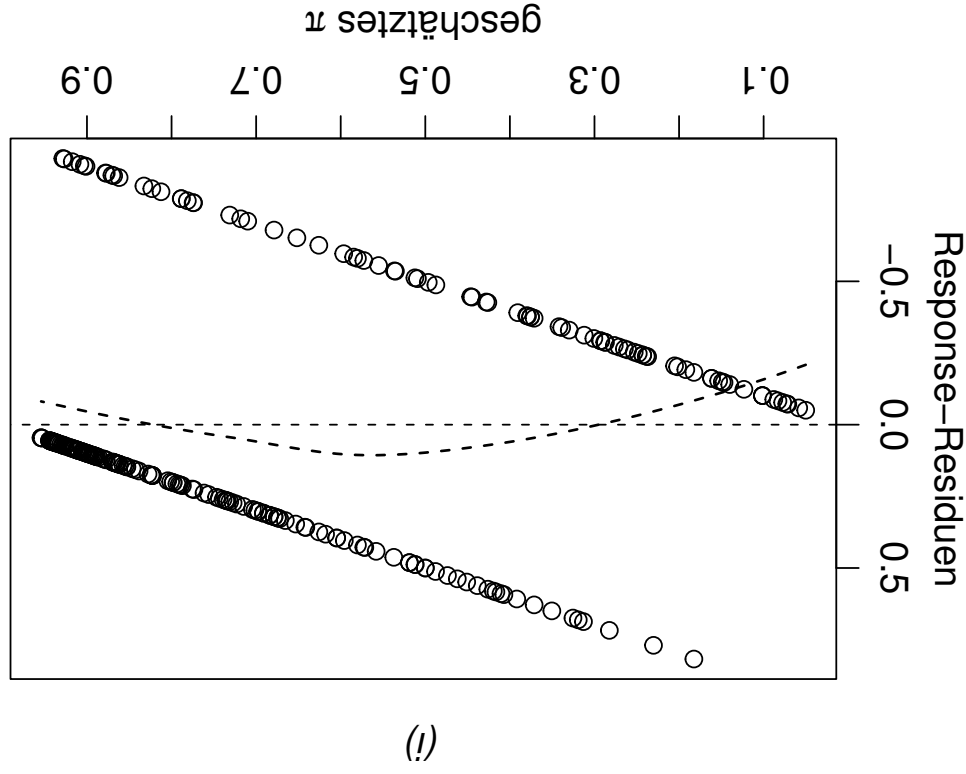
b Grafische Darstellungen:

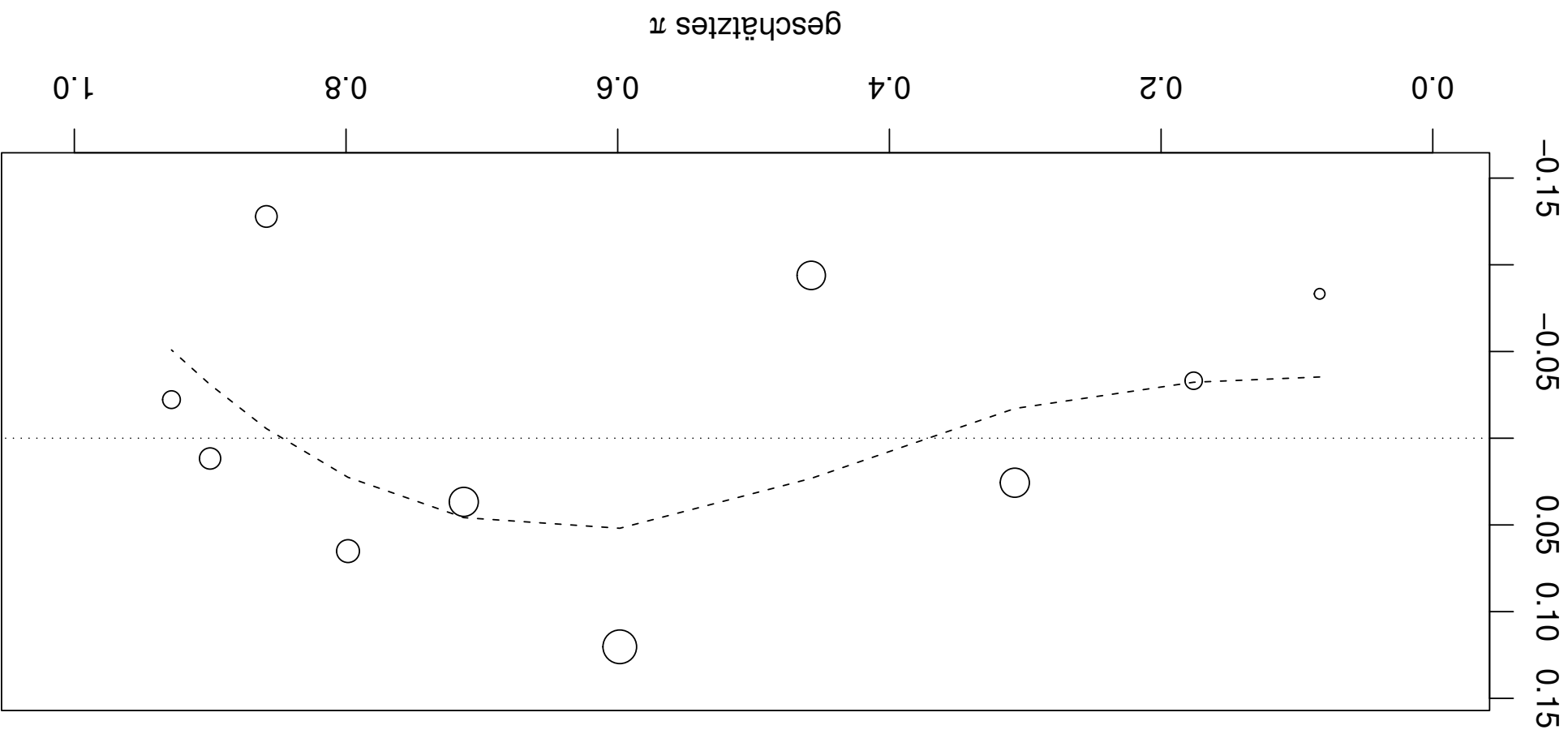
Q-Q- (normal) plot meist unnützlich!

## Tukey-Anscombe-Diagramm:

Rohe Res. / geschätzte  $\pi_i$  oder Arbeitsres. / lin. Prädiktor

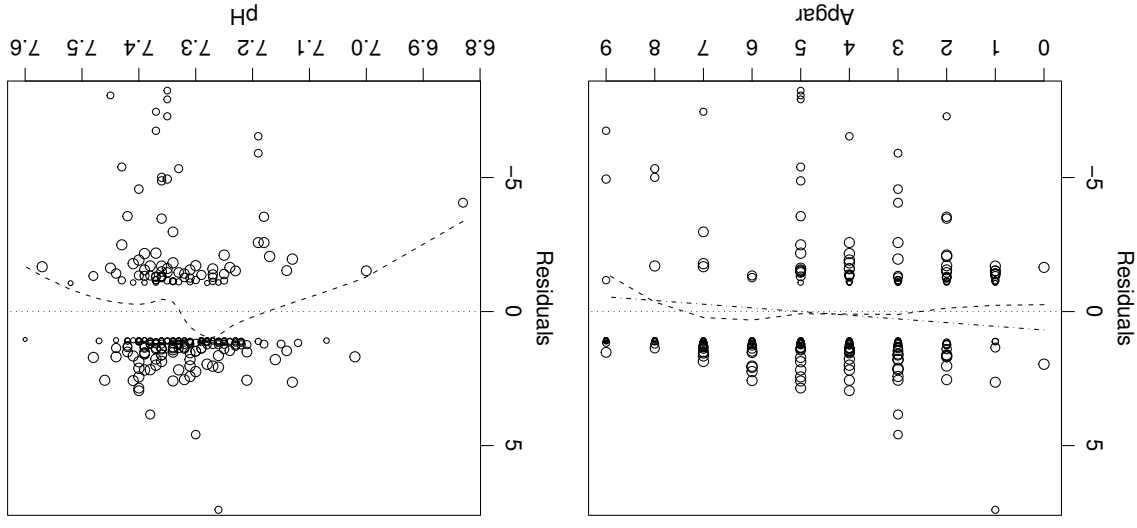
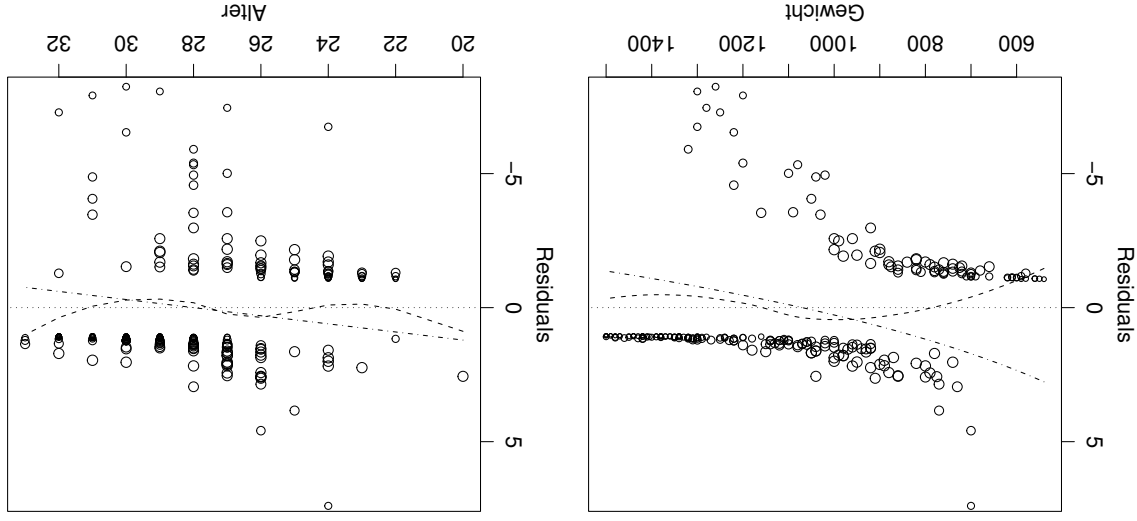
braucht Glättung.





e Residuen vs. Ausgangsgrößen oder „Partial residual plots“: „Effekte“ von  $x_i^{(j)}$  von  $x_i^{(j)} (= \hat{\beta}_j x_i^{(j)} - \text{Konst.})$  plus geeignete Residuen vs.  $x_i^{(j)}$ .

$$Y \sim \log_{10}(\text{Gewicht}) + \text{Alter} + \text{Apgar}$$



Jul 21, 00:25:11

**regf**

```

Call:
regf(formula = Survival ~ Weight + Age + Apgar1,
     data = t.d, family = binomial)
Terms:

```

	coef	stcoef	t.ratio	df	Chi2	p.value
(Intercept)	-8.484190	NA	NA	1	NA	NA
Weight	0.003791	1.0065	2.2780	1	22.535	0.000
Age	0.165297	0.4519	1.1254	1	4.999	0.025
Apgar1	0.142989	0.3179	0.9123	1	3.289	0.070

	deviance	df	p.value
Model	82.72	3	0
Residual	236.56	243	NA
Null	319.28	246	NA

Dispersion parameter taken to be 1. Family is binomial.  
 AIC: 244.6  
 Number of Fisher Scoring iterations: 5

```

Call:
regr(formula = cbind(Survival.1, Survival.0) ~ Weight,
      data = t.d, family = binomial)

```

Terms:

	coef	stcoef	t.ratio	df	F	p.value
(Intercept)	-4.560648	NA	NA	1	NA	NA
Weight	0.005087	1.540	3.145	1	47.98	0

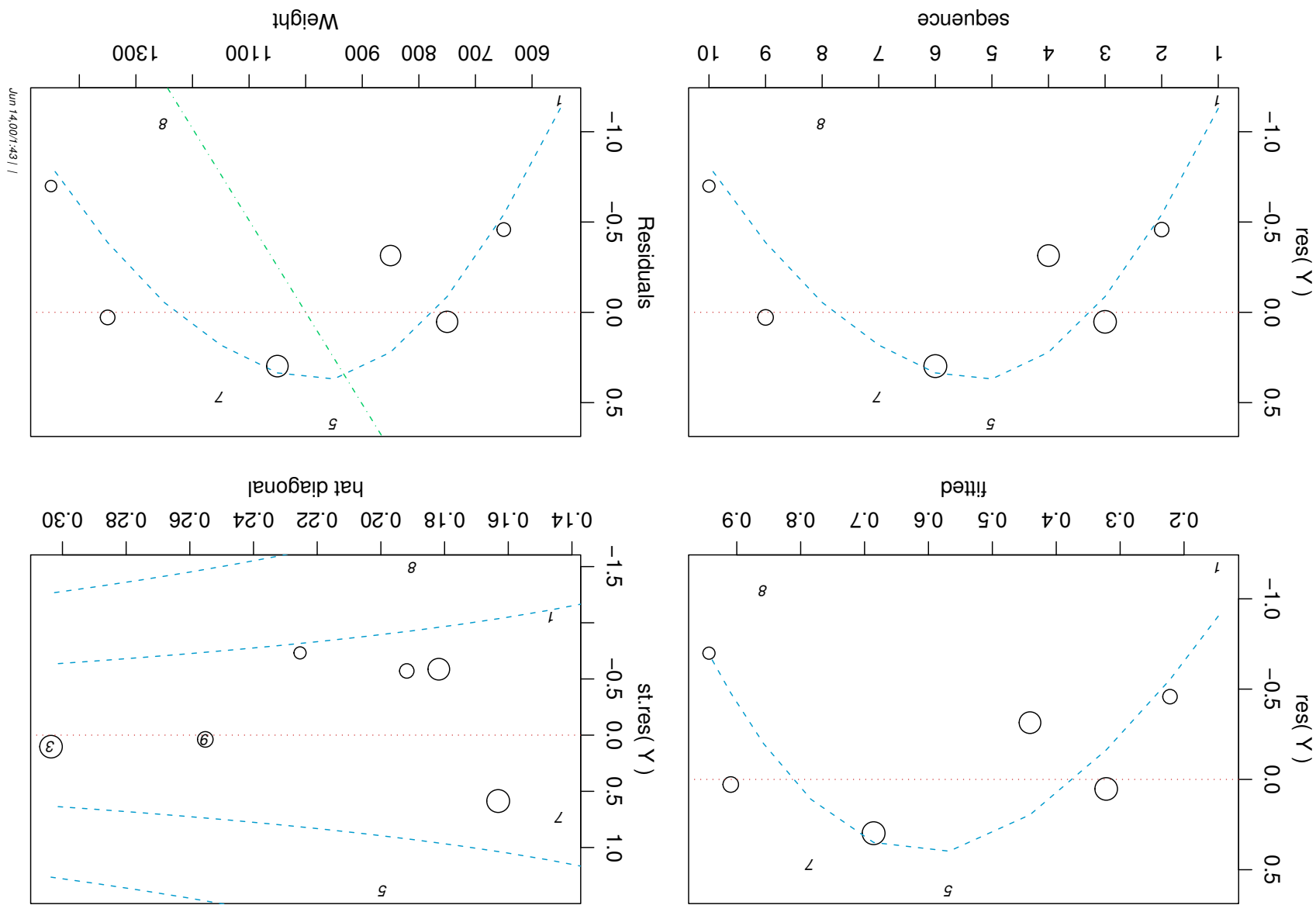
Model	deviance	df	p.value
Model	74.61	1	0.0000
Residual	12.44	8	0.1327
Null	87.05	9	NA

Dispersion parameter estimated to be 1.555.  
 Family is binomial.

AIC: 45.43

Number of Fisher Scoring iterations: 4

*cbind(Survival.1, Survival.0)~Weight*



## Logistische Regression

### Merkmale

- Logistische Regression dient für **binäre & binomial-vert.** Zielgrößen.
- Schätzungen, Tests, Vertrauensintervalle via Likelihood-Methoden und Asymptotik.
- **Residuen zeigen Artefakte**, Plots brauchen Glättungen.  
Modellwahl wie gehabt.

### 13 Verallgemeinerte Lineare Modelle

#### 13.1 Poisson-Regression

b **Beispiel Schiffs-Havarien.**

Y Anzahl Schaden-Ereignisse,  
 X Anzahl Betriebs-Monate

Schiffs-Typ

Baujahr-Periode

Betriebs-Periode

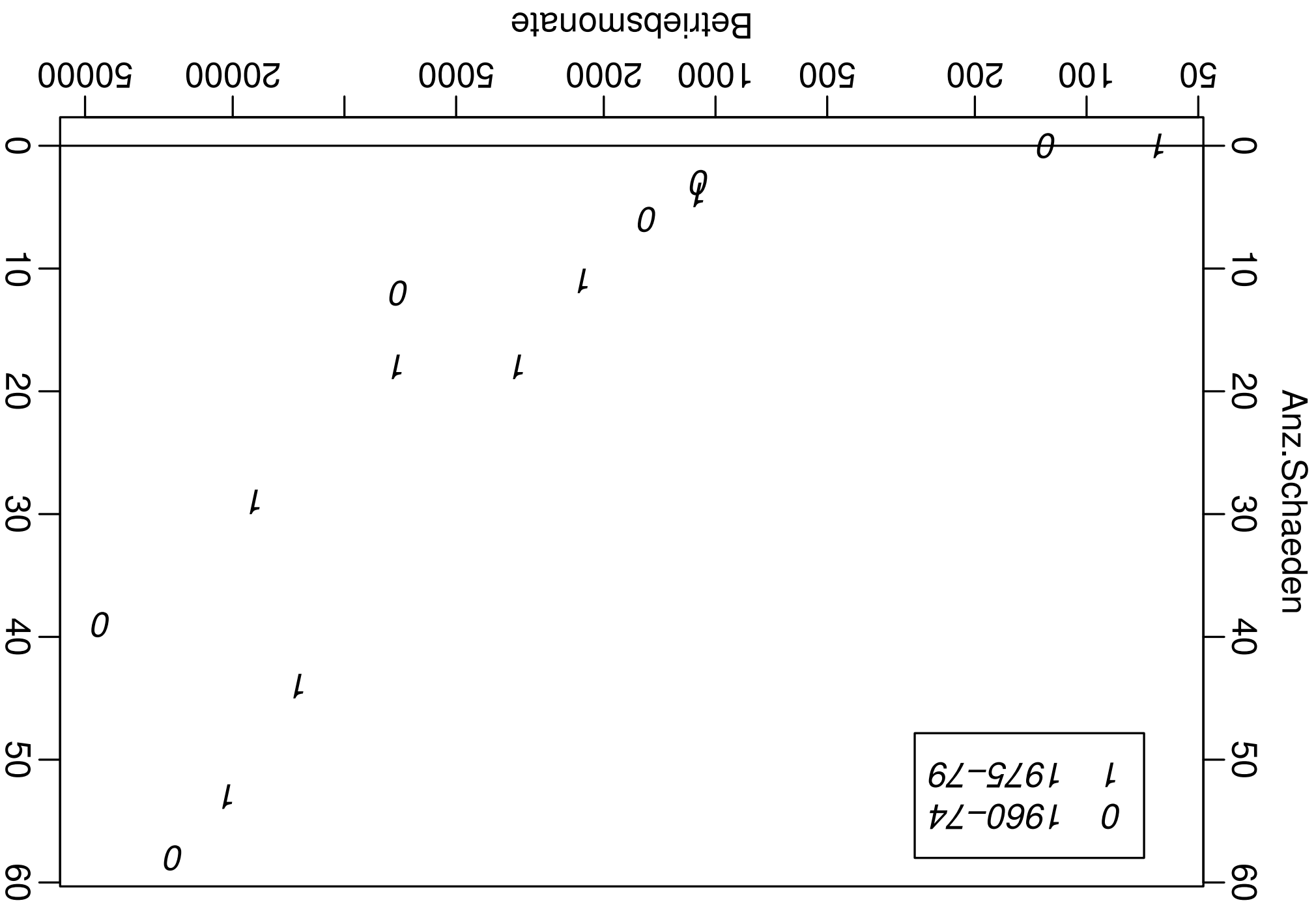
O: 0, 1

C: 60, 65, 70, 75

T: 0, 1

M,

	T	C	O	M	Y
1	0	60	0	127	0
2	0	60	1	63	0
3	0	65	0	1095	3
...	1	70	1	13099	44
14	1	75	1	7117	18



$$c \quad Y_i \sim \mathcal{P}(\lambda_i) \quad \mathcal{E}\langle Y_i \rangle = \lambda_i \quad g \langle \mathcal{E}\langle Y_i \rangle \rangle = n_i = \bar{x}_i^T \bar{\beta}$$

$$g \langle \lambda \rangle = \log \langle \lambda \rangle \Leftrightarrow$$

$$\mathcal{E}\langle Y_i \rangle = \lambda_i = \exp \langle \bar{x}_i^T \bar{\beta} \rangle = e^{\beta_0} \cdot e^{\beta_1 x_{i(1)}} \cdot \dots \cdot e^{\beta_m x_{i(m)}} \\ = \tilde{\beta}_0 \cdot \tilde{\beta}_1 x_{i(1)} \cdot \tilde{\beta}_2 x_{i(2)} \cdot \dots \cdot \tilde{\beta}_m x_{i(m)}$$

e multiplicative Effekte!

f Beispiel:

$$\log \langle \mathcal{E}\langle Y_i \rangle \rangle = \beta_0 + \beta_M \log \langle M_i \rangle + \beta_{T_i} + \beta_{P_i}$$

$$+ \gamma_1 \cdot (C1)_i + \gamma_2 \cdot (C2)_i + \gamma_3 \cdot (C3)_i$$

$\log \langle M \rangle$ : Anz. Havarien proportional zu Anz. Betriebsmonate

g Anzahlen: „gruppierte Daten“

## 13.2 Das grundlegende Modell

a  $g \langle \mathcal{E} \langle Y_i \rangle \rangle = n_i = \bar{x}_i^T \bar{\beta}$

$g$ : Link-Funktion

- b Verteilung von  $Y$ ? Binomial, Poisson, normal, Gamma, ...  
→ Exponentialfamilie!

## c Exponentialfamilie

$$f\langle y; \theta, \phi, \omega \rangle = \exp \left\langle y\theta - b\langle \theta \rangle \right\rangle_{\phi} \omega + c\langle y; \phi; \omega \rangle$$

$\theta$ : kanonischer Parameter.

$\phi$ : Dispersions-Parameter, Stör-Parameter.

$\omega$ : Gewicht bei gruppierten Daten.

$b$ : Welche Verteilung?

$c$ : Normierung auf gesamte  $W_i=1$

d Es gilt:

$$\mu = \mathcal{E}\langle Y \rangle = b'\langle \theta \rangle, \quad \text{var}\langle Y \rangle = b''\langle \theta \rangle = \frac{\mathcal{E}}{\phi} = V\langle \mu \rangle \cdot \frac{\mathcal{E}}{\phi}$$

(mit geeigneter Funktion  $V$ ).

e Normalverteilung:

$$\begin{aligned} \log \langle f \rangle_{y; \mu, \sigma^2} &= -\log \langle \sqrt{2\pi\sigma^2} \rangle - \frac{1}{2} \left( \frac{y - \mu}{\sigma} \right)^2 \\ &= \frac{y\mu - \frac{1}{2}\mu^2}{\sigma^2} - \frac{y^2}{2\sigma^2} - \log \langle \sqrt{2\pi\sigma^2} \rangle \\ \theta &= \mu \\ \phi &= \sigma^2 \\ b \langle \theta \rangle &= \theta^2/2 \\ c \langle y; \phi \rangle &= -y^2/(2\phi) - (1/2) \log \langle 2\pi\phi \rangle \end{aligned}$$

**Binomial-Verteilung:** Zielgröße  $Y_k = \tilde{Y}_k / m_k$ .

$$\begin{aligned} \log \langle P \langle Y = y \rangle \rangle &= \log \left\langle \binom{m}{y} \right\rangle + \langle y \rangle \log \langle \pi \rangle + m \log \langle 1 - \pi \rangle - \langle y \rangle \log \langle 1 - \pi \rangle \\ &= \left\langle y \log \left\langle \frac{1 - \pi}{\pi} \right\rangle + \log \langle 1 - \pi \rangle + m \log \langle 1 - \pi \rangle \right\rangle + \left\langle \binom{m}{y} \right\rangle \end{aligned}$$

$$\theta = \log \langle \pi / (1 - \pi) \rangle$$

$$m = m$$

$$\log \langle \binom{m}{y} \rangle = \langle \phi; y \rangle$$

$$\langle \theta \rangle = \log \langle 1 + e^\theta \rangle \quad \langle \theta \rangle' = \langle \theta \rangle \quad \pi = \langle \theta \rangle' \quad \langle \theta \rangle'' = \langle \theta \rangle = \pi(1 - \pi)$$

0-1-Variable:  $m = 1$ .

! Link-Funktion.

Inverse Link-Funktion  $h$  soll unmögliche Werte vermeiden:

$$\begin{aligned}
 g\langle n \rangle &= \mu, & \text{wenn } \mathcal{E}\langle Y \rangle &\text{ beliebig} \\
 g\langle n \rangle &= \log\langle n \rangle, & \text{wenn } \mathcal{E}\langle Y \rangle &> 0, \\
 g\langle n \rangle &= \log\langle n \rangle = \log\left\langle \frac{n-1}{n} \right\rangle, & \text{wenn } 0 &\leq \mathcal{E}\langle Y \rangle \leq 1
 \end{aligned}$$

! „kanonische Link-Funktion“:  $n = g\langle n \rangle = \theta = \theta = (b')^{-1}\langle n \rangle$   
Wähle  $g = (b')^{-1}!$ Normalverteilung  $g\langle n \rangle = \mu$ Poissonverteilung  $g\langle n \rangle = \log\langle n \rangle$ Binomialverteilung  $g\langle n \rangle = \log\langle n \rangle$ 

Vorteile: Existenz und Eindeutigkeit, einfachere Schätzgleichungen

### 13.3 Schätzungen und Tests

b **Likelihood**. Schätzung:  $\bar{s}^{(j)} \langle \bar{\beta} \rangle = 0 \rightarrow \hat{\beta}_j$

Normalgleichungen für gewichtete Kleinste Quadrate  
Gewichte und "Residuen" hängen von  $\beta$  ab!

**Algorithmus**: "iteratively reweighted least squares"

→ Es kommt vor, dass das Progr. keine geeigneten Startwerte findet!

e Dispersionsparameter: Schätzung nach Max.Lik.,

mit Korrekturfaktor wegen Freiheitsgraden.

f Verteilung der geschätzten Parameter: Asymptotik

⇒ Normalverteilung (z-Test).

$$\hat{\beta} \approx \mathcal{N}(\bar{\beta}, \mathbf{V}/n) \quad \mathbf{V} = \dots$$

```

summary(glm(...))

Call: glm(formula = y ~ TYPE + factor(C) + OPER + log(MONTHS),
          family = poisson, data = d.ship)

Deviance Residuals: ...
Coefficients:
(Intercept) -6.6109 1.2744
TYPE -0.6569 0.3262
factor(C) 1 -0.5556 0.1470
factor(C) 2 0.1242 0.1038
factor(C) 3 0.2965 0.1129
OPER 0.4585 0.1359
log(MONTHS) 1.0825 0.1550
Value Std. Error z_appr. Pr(>|z|) Signif.
*** 0.000 -5.19
* 0.044 -2.01
*** 0.000 -3.78
0.231 1.20
0.009 2.63
** 0.001 3.37
*** 0.000 6.99
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion Parameter for Poisson family taken to be 1 )
Null Deviance: 267.3 on 13 degrees of freedom
Residual Deviance: 3.434 on 7 degrees of freedom
Number of Fisher Scoring Iterations: 3

```

## h Residuen-Devianz

Vergleich des gefitteten Modells mit „Maximal möglichem“ Modell, d.h. für jede Beobachtung ein Parameter:

$$D\langle \bar{y}; \bar{\mu} \rangle = 2(\ell(M) - \ell(\bar{\beta})) = \sum_i 2\omega_i \left( y_i(\theta_i) - \theta\langle \bar{x}_i^T \bar{\beta} \rangle - b\langle \theta_i \rangle + b\langle \theta\langle \bar{x}_i^T \bar{\beta} \rangle \rangle \right)$$

$\tilde{\theta}_i$  = Parameterwert, der am besten zu  $y_i$  passt.

! **Poisson-Regression:**  $\tilde{\theta}_i = \log(y_i)$

$$D\langle \bar{y}; \bar{\mu} \rangle = \sum_i 2 \left( y_i(\log\langle y_i \rangle - \log\langle \mu_i \rangle) - e^{\log\langle y_i \rangle} + e^{\log\langle \mu_i \rangle} \right) = \sum_i 2 \left( y_i \frac{\log\langle \mu_i \rangle}{\log\langle y_i \rangle} - y_i + \mu_i \right)$$

! Vergleich von Modellen. Likelihood-Ratio-Test

$$\tilde{D}\langle \bar{y}; \hat{\mu}^{(K)}, \bar{\mu}^{(G)} \rangle = D\langle \bar{y}; \hat{\mu}^{(K)} \rangle - D\langle \bar{y}; \hat{\mu}^{(G)} \rangle = 2(\ell^{(G)} - \ell^{(K)})$$

Gesamt-Test: null deviance – residual deviance.

$$\tilde{D}\langle \bar{y}; \hat{\mu}^{(0)}, \bar{\mu} \rangle = D\langle \bar{y}; \hat{\mu}^{(0)} \rangle - D\langle \bar{y}; \hat{\mu} \rangle = 2(\ell^{(G)} - \ell^{(0)})$$

### 1. Likelihood-Quotienten-Test: **Devianz-Differenz**

$H_0$ : Modell  $K$  mit  $p$  Parametern

$H_1$ : Modell  $G$  mit  $r > p$  Parametern

$$\text{Teststatistik } 2 \cdot \log \frac{L_G}{L_K} = 2(\ell(G) - \ell(K))$$

Verteilung unter  $H_0$ :  $\chi^2_{r-p}$

### 2. Vergleich mit maximalem Modell: **Residuen-Devianz** $D(\hat{y}; \hat{\mu})$

$H_0$ : Angepasstes Modell mit  $p$  Parametern

$H_1$ : Maximales Modell  $m$  mit  $n_k$  Parametern

$$\text{Teststatistik } D(\hat{y}; \hat{\mu}) = 2(\ell(M) - \ell(\hat{\mu}))$$

Verteilung unter  $H_0$ :  $\chi^2_{n-p}$  (Gruppierten Daten!)

### 3. Gesamt-Test: Vergleich von **Null Devianz** und Residuumdevianz

$H_0$ : Null Modell mit einem Parameter

$H_1$ : Angepasstes Modell mit  $p$  Parametern

$$\text{Teststatistik } D(\widehat{y}; \widehat{\mu}) - D(\bar{y}; \bar{\mu}) = 2(\ell(\widehat{\mu}) - \ell(\bar{\mu}))$$

Verteilung unter  $H_0$ :  $\chi^2_{p-1}$

```

> r.ship1 <- glm(Y~TYPE + factor(CONS) + OPER + log(MONTHS),
data=d.ship, family=poisson)
> summary(r.ship1, corr=F)
Call: glm(formula = Y ~ TYPE + factor(CONS) + OPER + log(MONTHS),
family = poisson, data = d.ship)

Coefficients:
(Intercept) -6.6109 1.2744
TYPE -0.6569 0.3262
factor(CONS)1 -0.5556 0.1470
factor(CONS)2 0.1242 0.1038
factor(CONS)3 0.2965 0.1129
OPER 0.4585 0.1359
log(MONTHS) 1.0825 0.1550

Value Std. Error z_appr. Pr(>|z|) Signif
***          0.000    -5.19
*           0.044    -2.01
***          0.000    -3.78
***          0.231    1.20
**           0.009    2.63
***          0.001    3.37
***          0.000    6.99

(Dispersion Parameter for Poisson family taken to be 1 )
Null Deviance: 267.3 on 13 degrees of freedom
Residual Deviance: 3.434 on 7 degrees of freedom
Number of Fisher Scoring Iterations: 3
> 1-pchisq(3.434,7) 0.8421659

```

## 13.4 Übergrosse Streuung

- a Ablehnung des Modells (residual deviance):  $\Rightarrow$  **over-dispersion**.  
b Neues Modell:  $\phi > 1$

Keine entsprechende Verteilung  $\rightarrow$  „Quasi-Modelle“, **Quasi-Likelihood**.

c  $V\langle \mu \rangle = \phi \mu(1 - \mu)$  resp.  $V\langle \mu \rangle = \phi \mu$

für übermässig streuende Binomial- respektive Poisson-Zielgrössen.

$\leftarrow$  Dispersionsparameter schätzen statt fixieren.

$$\hat{\phi} = \frac{1}{n-d} \sum \frac{\omega_i(y_i - \hat{\mu}_i)^2}{V\langle \hat{\mu}_i \rangle}$$

d Parameterschätzer  $\hat{\beta}$  bleiben gleich

Konfidenzintervalle um den Faktor  $\sqrt{\hat{\phi}}$  breiter  
e Was, wenn  $\hat{\phi} > 1$ ?  $\phi > 1$  ist unplausibel!

## 13.5 Residuen-Analyse

- Rohe Residuen oder **response residuals**:  $R_i = Y_i - \widehat{\mu}_i$

- Pearson-Residuen:  $R_i^{(P)} = R_i / \sqrt{V\langle \widehat{\mu}_i \rangle}$

- Arbeits-Residuen (**working residuals**):  $R_i^{(W)} = R_i \cdot g'\langle \widehat{\mu} \rangle$ .

- Devianz-Residuen:  $R_i^{(D)} = \text{sign}\langle y_i - \widehat{\mu}_i \rangle \sqrt{d_i}$

- Poisson-Regr.:  $D\langle \bar{y}, \widehat{\mu} \rangle = \sum_i 2 \left( y_i \frac{\log\langle \widehat{\mu}_i \rangle}{\log\langle y_i \rangle} - y_i + \widehat{\mu}_i \right) =: \sum_i d_i^2$

## Residuen-Analyse:

- Linearität: Arbeitsresiduen  $R_i^{(W)}$  oder adjustierte Beobachtungen  $\tilde{y}_i = \bar{x}_i^T \bar{\beta} + r_i^{(W)}$  vs  $\hat{\eta}_i$ .
- Residuenvarianz:  $\phi$
- Hutmatrix:  $\tilde{W}^{1/2} \tilde{X} (\tilde{X} \tilde{W} \tilde{X})^{-1} \tilde{X}^T \tilde{W}^{1/2}$  (?)
- Residuenplot mit glatten Kurven:
  - Tukey-Anscombe-Plot
  - Beobachtungen vs fitted values
  - Partielle Residuen-Plots

## Verallgemeinerte Lineare Modelle

### Merkmale

- Verallg. lineare Modelle umfassen als Verteilungen der Zielgröße:
  - Normalverteilung
  - Bernoulli- und Binomial-Verteilung,
  - Poisson-Verteilung,
  - Exponential- und Gamma-Verteilung.
- Theorie und Algorithmus für alle diese Modelle:
  - Iteratively Reweighted Least Squares
  - Maximum Likelihood-Schätzung
  - Likelihood ratio tests zum Vergleich von Modellen, Deviance
- Residuen sind weniger nützlich, da „künstliche“ Strukturen auftreten.  
Man braucht sie trotzdem!

## 14 Geordnete diskrete Zielgrößen

### 14.1 Modelle

a Anwendungen:

- Beurteilung von sehr schlecht bis sehr gut,
- gruppierte Häufigkeiten,
- quantitative, klassierte Grösse, etc.

b **Beispiel: Lokale Anaesthetie des Armes.**

Welche erklärenden Variablen beeinflussen den **Erfolg**?

$Y$  suc.deg Erfolg in 4 Klassen:

1: schmerzvoll, ... 4: kein Schmerz

$X$  medic

Medikamentdosis

anäst2: Anzahl Einstiche (Intervall-Skala)

Anästesist/in (Faktor),

Mondphase (Faktor) ...

	anest2	suc.deg
...	...	...
1	A0	1
3	A2	3
4	A2	3
8	A0	1
9	A2	4

	suc.deg
1	1
2	2
3	3
4	4

	A0	A2	A1
1	28	4	6
2	18	10	1
3	23	13	6
4	25	36	8

c Latente Variable,  $Z$  kontinuierlich

$$Y_i = 0 \iff Z_i \leq \alpha_1$$

$$Y_i = k \iff \alpha_k > Z_i \geq \alpha_{k+1}$$

$$Y_i = k^* \iff \alpha_{k^*} > Z_i$$

$k^* + 1$  Schwellenwerte:  $\alpha_1 > \alpha_1 > \dots > \alpha_{k^*}$ .

$$P\langle Y_i \leq k \rangle = P\langle Z_i \leq \alpha_k \rangle \quad k = 1, \dots, k^* - 1$$

## Latentes Modell

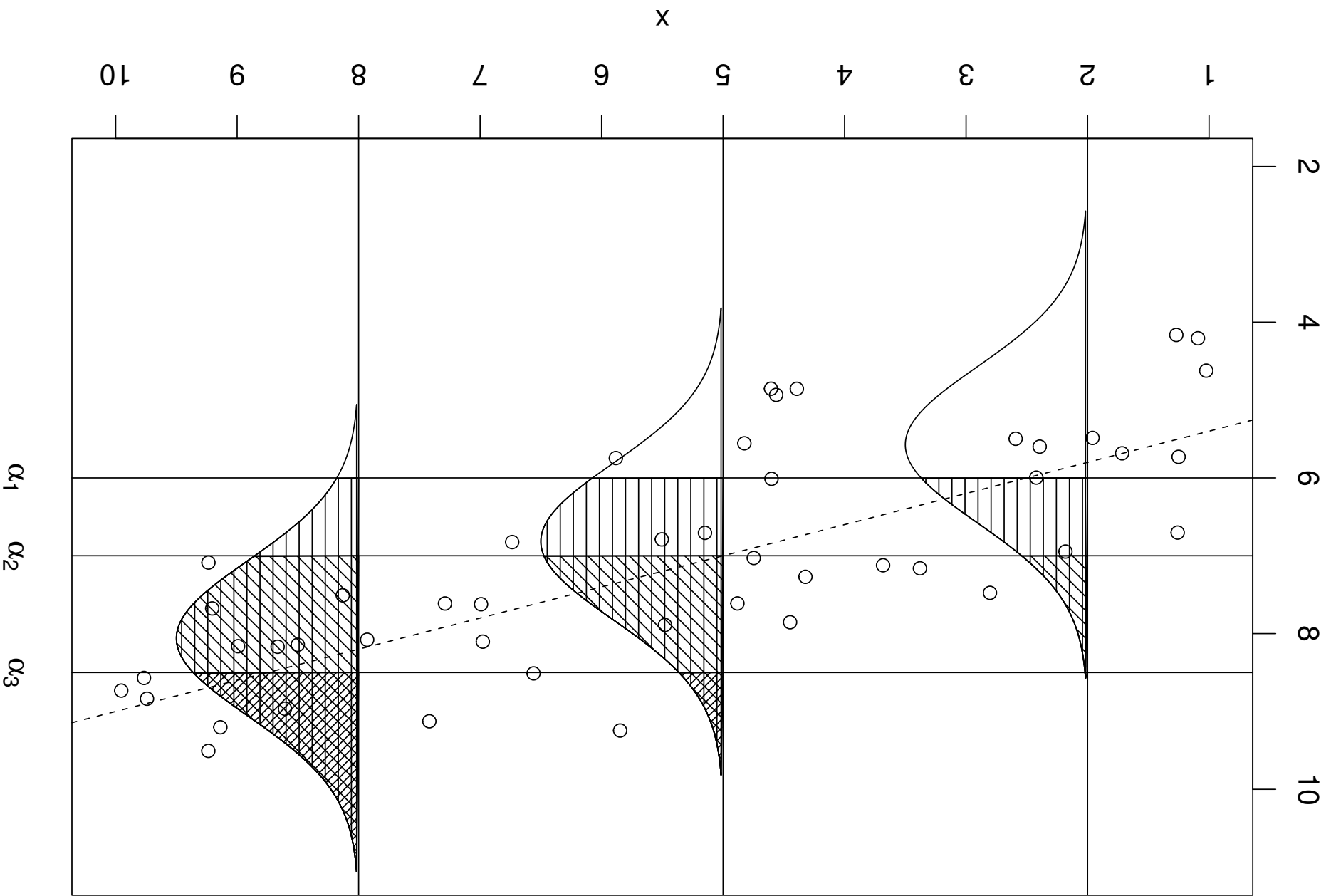
1. Annahme:

Latente Variable hat z.B. logistische Verteilung

2. Annahme:

Multiple lineare Regression für latente Variable

latente V.



## d Modell

$$\begin{aligned}
 Z_i &= \beta_0 + \sum_{j=1}^J x_i^{(j)} \beta_j + E_i \\
 \gamma_k &:= P\langle Y_i \geq k \rangle = P\langle Z_i \geq \alpha_k \rangle = P\langle E_i \geq \alpha_k - (\beta_0 + \bar{x}_T \bar{\beta}) \rangle \\
 &= 1 - F_E \langle \alpha_k - (\beta_0 + \bar{x}_T \bar{\beta}) \rangle = F_{-E} \langle \bar{x}_T \bar{\beta} - (\alpha_k - \beta_0) \rangle \\
 g\langle \gamma_k \rangle &= \bar{x}_T \bar{\beta} - (\alpha_k - \beta_0)
 \end{aligned}$$

f Schwellenwerte  $\alpha_k$  nicht gleich-abständig. Schätzen!

g Kumulatives Modell:  $P\langle Y_i \geq k \rangle = P\langle Y_i = k^* \rangle + \dots + P\langle Y_i = k \rangle$

h Gruppierung:

$$\tilde{Y}_k = \text{Anzahl}\{i \mid Y_i = k\} \sim M_{k^*} \langle m, \pi \rangle$$

- Multinomialverteilung = mehrparametrische Exponentialfamilie  
genauer: betrachte für gegebenes *vcc*

$$\text{den Vektor der relativen Häufigkeiten } \bar{Y} = \tilde{Y} / m$$

- Erwartungsvektor:  $\pi_1, \dots, \pi_{k^*}$

- Wahrscheinlichkeiten

$$P \langle \bar{Y} = \bar{y} \rangle = \frac{m!}{i_1! \dots i_{k^*}!} \pi_{(1)}^{i_1} \pi_{(2)}^{i_2} \dots \pi_{(k^*)}^{i_{k^*}}$$

$\pi_k = P \langle Y_i = k \rangle$  aus dem Modell der latenten Variablen.

← Multivariates generalisiertes lineares Modell

! Wettverhältnisse (odds)

$$\text{odds}\langle Y \geq k \mid \bar{x} \rangle = \frac{P\langle Y_i < k \mid \bar{x}_i \rangle}{P\langle Y_i > k \mid \bar{x}_i \rangle} = e^{\alpha_k} e^{\beta_1 x_{(1)}} \cdots e^{\beta_m x_{(m)}}$$

Odds Ratio

$$\frac{\text{odds}\langle Y \geq k \mid \bar{x}_1 \rangle}{\text{odds}\langle Y \geq k \mid \bar{x}_2 \rangle} = \exp\langle -(\bar{x}_1 - \bar{x}_2)^T \bar{\beta} \rangle$$

sind für alle Schwellenwerte  $\alpha_k$  gleich! $\Leftrightarrow$  proportional odds model.

!\*"komplementäre Log-Log-Funktion"

$$g(\gamma) = \log \langle -\log \langle 1 - \gamma \rangle \rangle, \quad 0 < \gamma < 1$$

Zuverlässigkeits- und Überlebenszeit-Studien: Weibull-Verteilung.

Logarithmierte Ausfall- oder Überlebenszeiten: Gumbel-Verteilung.

Proportional hazards, Cox-Regression.

Für zensierte Daten brauchbar!

## 14.2 Schätzungen und Tests

c R-Funktion polr

```
call: polr(formula = ordered(suc.deg) ~ medic + age + bmi.gr +
napplc + anest2 + expnce + moon, data = d.anest.ln)
Coefficients:
```

	Value	Std. Error	t value
medic	-0.088959000	0.031276473	-2.8442785
age	0.005610906	0.008526959	0.6580196
bmi.gr21to28	-0.825546649	0.384865654	-2.1450255
bmi.grgt28	-0.697811976	0.539368704	-1.2937569
napplc	0.896180981	0.218809357	4.0957160
anest21	1.118096186	0.596838858	1.8733636
anest22	1.475949108	0.399248382	3.6968193
expnce	0.181347910	0.182372903	0.9943797
moonneutral	0.165380216	0.346301994	0.4775607
moonvoll	-0.911484405	0.481794990	-1.8918511

Intercepts:  
Value Std. Error t value  
1|2 -3.1091 1.3802 -2.2526  
2|3 -2.0922 1.3664 -1.5312  
3|4 -0.8735 1.3656 -0.6396

Residual Deviance: 418.8633  
AIC: 444.8633

d drop1

Single term deletions

Model:

ordered(suc.deg) ~ medic + bmi.gr + napplic +

anest2 + moon

DF AIC LRT Pr(&gt;Chi)

<none>	442.54				
medic	1 448.71	8.17	0.004264	**	
bmi.gr	2 443.69	5.15	0.076107	.	
napplic	1 458.29	17.74	2.526e-05	***	
anest2	2 452.15	13.61	0.001111	**	
moon	2 445.24	6.70	0.035081	*	

Vergleich mit gewöhnlicher Regression: ähnliches Resultat.

## 14.3 Multinomiale Zielgrößen

b  $Y_i$  ungeordnete Kategorien  $\rightarrow$  Multinomiales Logit-Modell

$$\log \frac{P\langle \tilde{Y}_\ell = k \mid x_\ell \rangle}{P\langle \tilde{Y}_\ell = 0 \mid x_\ell \rangle} = \beta_{0k} + \sum_j \beta_{jk} x_{\ell(j)}.$$

$\beta_{jk} > 0$ : Präferenz für Zielkategorie  $k$  vor der Referenzkategorie 0.

```
c R: library(nnet) ! multinom(...)
```

```
call:
```

```
multinom(formula = Wersoll ~ Alter.ljn + Geschlecht,  
data = d.umweltumfrage, weights = Freq)
```

```
Coefficients:
```

```
(Intercept)  Alter.ljn  Geschlechtw  
Staat      -0.40372    0.0026276  -0.19487  
beide     -1.20102   -0.0022776  -0.24129
```

```
Std. Errors:
```

```
(Intercept)  Alter.ljn  Geschlechtw  
Staat      0.13697    0.0029070  0.097437  
beide      0.19513    0.0042460  0.140090
```

```
Residual Deviance: 3913.0
```

```
AIC: 3925.0
```

d  $k_* \cdot m$  statt  $k_* + m$  Parameter!!

e Schätzung:

- getrennte Schätzung der  $k_*$  Logit-Modelle
- simultane Lösung der  $k_*$  Schätzgleichungen

f Anwendung: **Diskriminanzanalyse** mit mehr als 2 Kategorien

## Geordnete und nominale Zielgrößen

### Merkmale

- Viele nominale Variable sind geordnet.  
Zielgrößen dieser Art werden mit kumulativen Logits (Probits, ...) adäquat modelliert.
- Vorstellung der latenten Variablen  
und der Schwellenwerte dient zur Begründung.
- Ungeordnete nominale Zielgrößen → Multinomiale Regression