Logistic Regression

1.1 Introduction

Only partially translated at this time

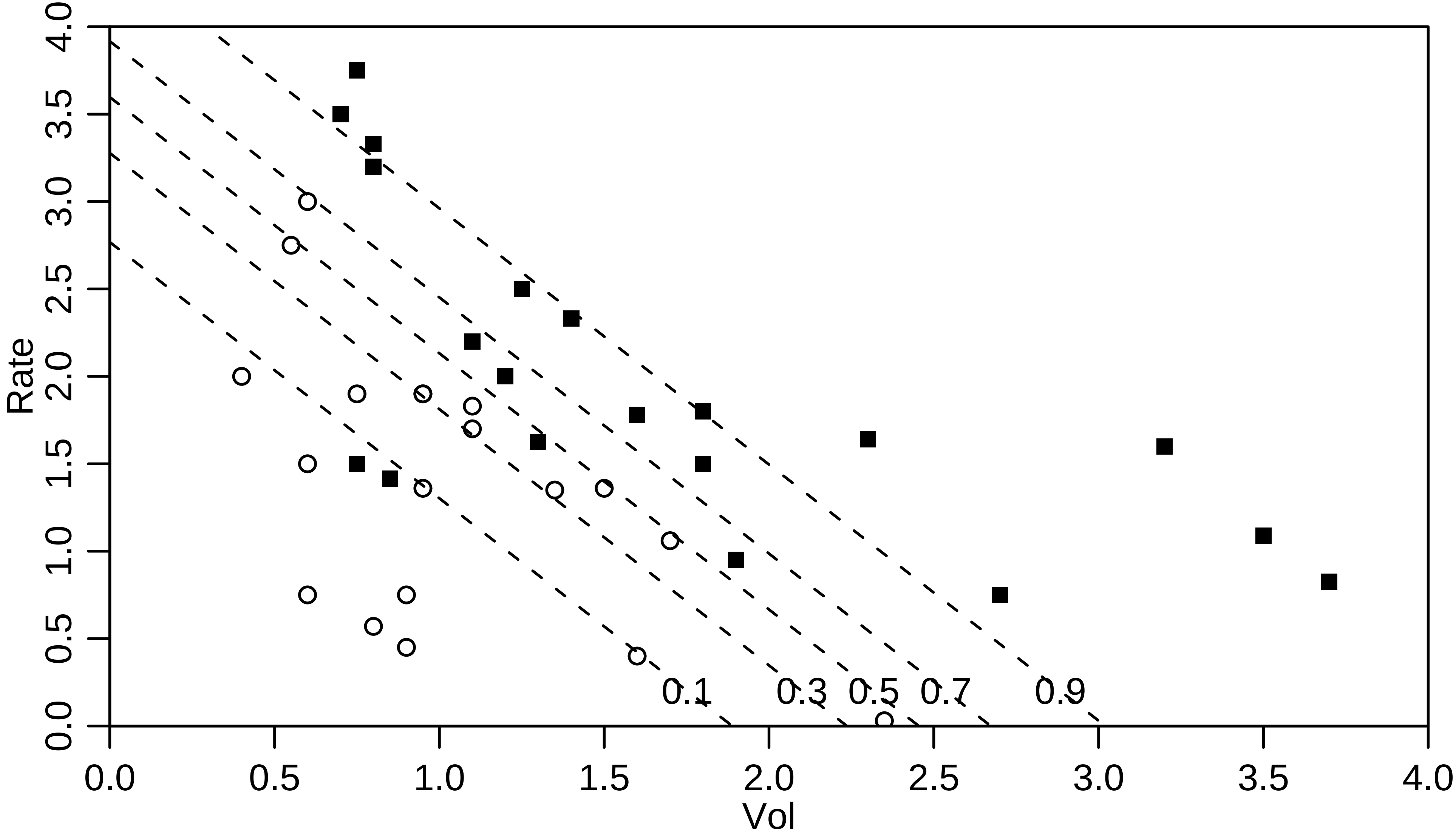
b Example: Shrinked blood vessels

Y: shrinked: yes (1) / no (0)

erkl.: Breath Volume (`Vol`) and Frequency (`Rate`)

Ziel: $P\langle Y = 1 \mid \texttt{Vol}, \texttt{Rate}\rangle$ modellieren!

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


Rate
Vol
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0.1
0.3
0.5
0.7
0.9

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

d Why is an ordinary linear regression inadequate?

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

- What is the error term E_i ?
  $\mathcal{E}\langle Y_i\rangle = \beta_0 + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \ldots + \beta_m x_i^{(m)}$
  We have $P\langle Y_i = 1\rangle = \mathcal{E}\langle Y_i\rangle$. $\longrightarrow$ Same form o.k.

- But: Estimated values may become < 0 and > 1!
  $\longrightarrow$ Transformation of Y_i? 2 values remain 2 values!
  $\longrightarrow$ Transformation of $\mathcal{E}\langle Y_i\rangle = P\langle Y_i = 1\rangle$!

1.1

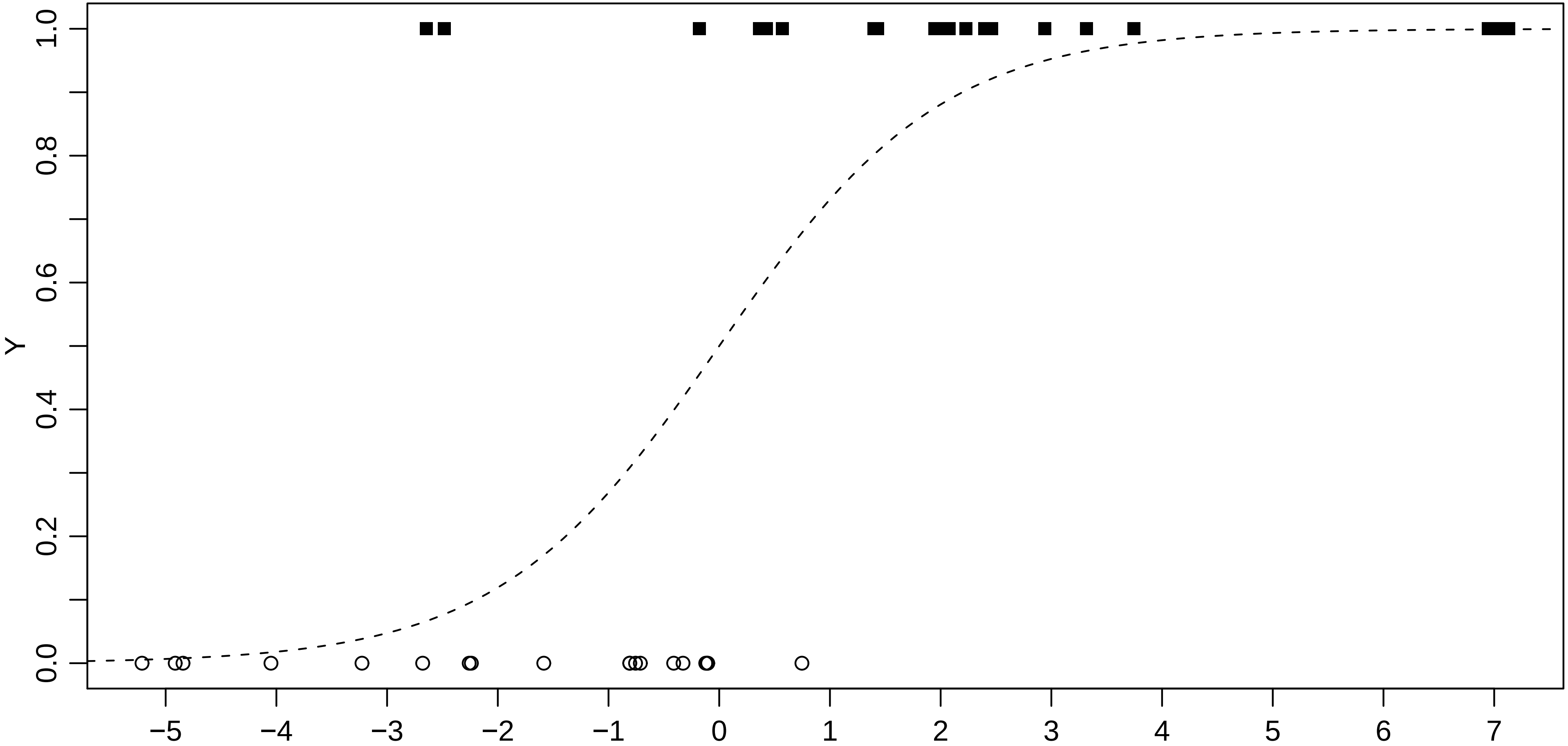
e **Modell.** Logit-Funktion $g\langle\pi\rangle = \log\left\langle\frac{\pi}{1-\pi}\right\rangle = \frac{\log\langle\pi\rangle}{\log\langle 1-\pi\rangle}$

$$g\langle P\langle Y_i = 1\rangle\rangle = \frac{\log\langle P\langle Y_i = 1\rangle\rangle}{\log\langle P\langle Y_i = 0\rangle\rangle} = \eta_i$$

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

η: „linearer Prädiktor".

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


Y
0.0
0.2
0.4
0.6
0.8
1.0
−5
−4
−3
−2
−1
0
1
2
3
4
5
6
7

g **Diskriminanzanalyse:**

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

Logistische Regression:

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

1.1

h Further Applications:

- Toxikology: Toxic matter deadly for mice? What concentration?
- Medicine: Treatment successful?
- Failure of (technical) devices,
- Bugs in (technical) products,
- Occurence of characteristics in animals or plants,
- client scoring,

General: 2 Groups.

1.2 Considerations about the Model

a Same flexibility as linear regression.
Frequently: factors (nominal variables) as explanatory v.

b Example: Assessment of work situation.

Y_i happy (1), unhappy (0)
$X_i^{(j)}$ Region, Age, Gender, Race

Only 1 factor $\longrightarrow$ $2 \times k$-cross table

	NE	Mid-Atl.	S	Midwest	NW	SW	Pacific	total
unzufrieden	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

Higher dimension $\longrightarrow$ log-linear models

c **Gruppierte Daten:** m_ℓ Beob. Y_i zu gleichen $\underline{x}_i = \widetilde{\underline{x}}_\ell$:

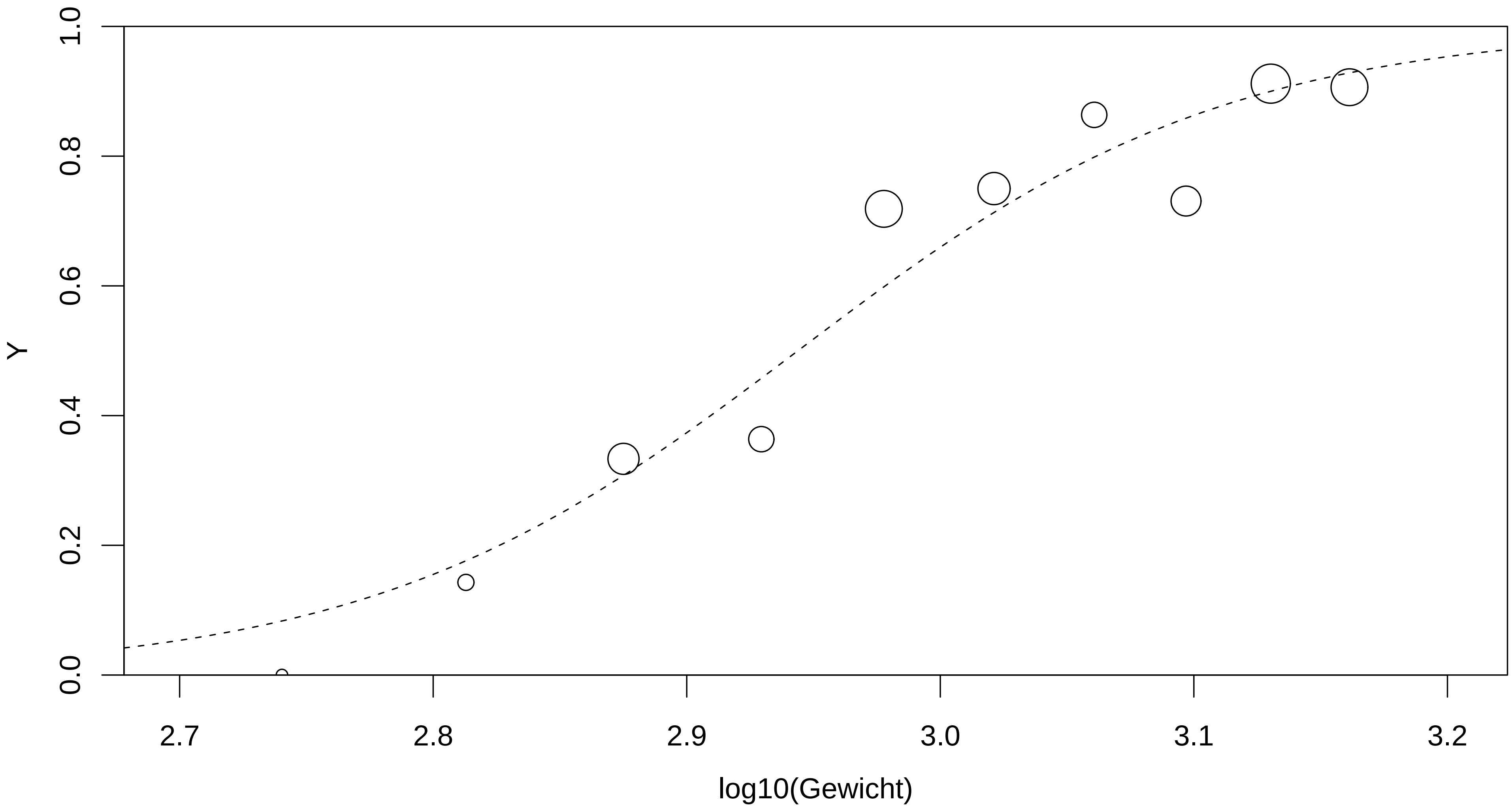
$\widetilde{Y}_\ell = \sum_{i\,:\,\underline{x}_i = \widetilde{\underline{x}}_\ell} Y_i \qquad \widetilde{Y}_k \sim \mathcal{B}\langle m_k, \pi_k\rangle \qquad \mathcal{E}\langle \widetilde{Y}_\ell / m_\ell\rangle = \pi_\ell$

$\longrightarrow$ Logistische Regression: $g\langle\pi_\ell\rangle = \eta_\ell$

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

Erklärende Variable: Geburtsgewicht. Klassen von je 100 g

	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


Y
0.0
0.2
0.4
0.6
0.8
1.0
2.7
2.8
2.9
3.0
3.1
3.2
log10(Gewicht)

e **Transformierte Beobachtungen.**

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

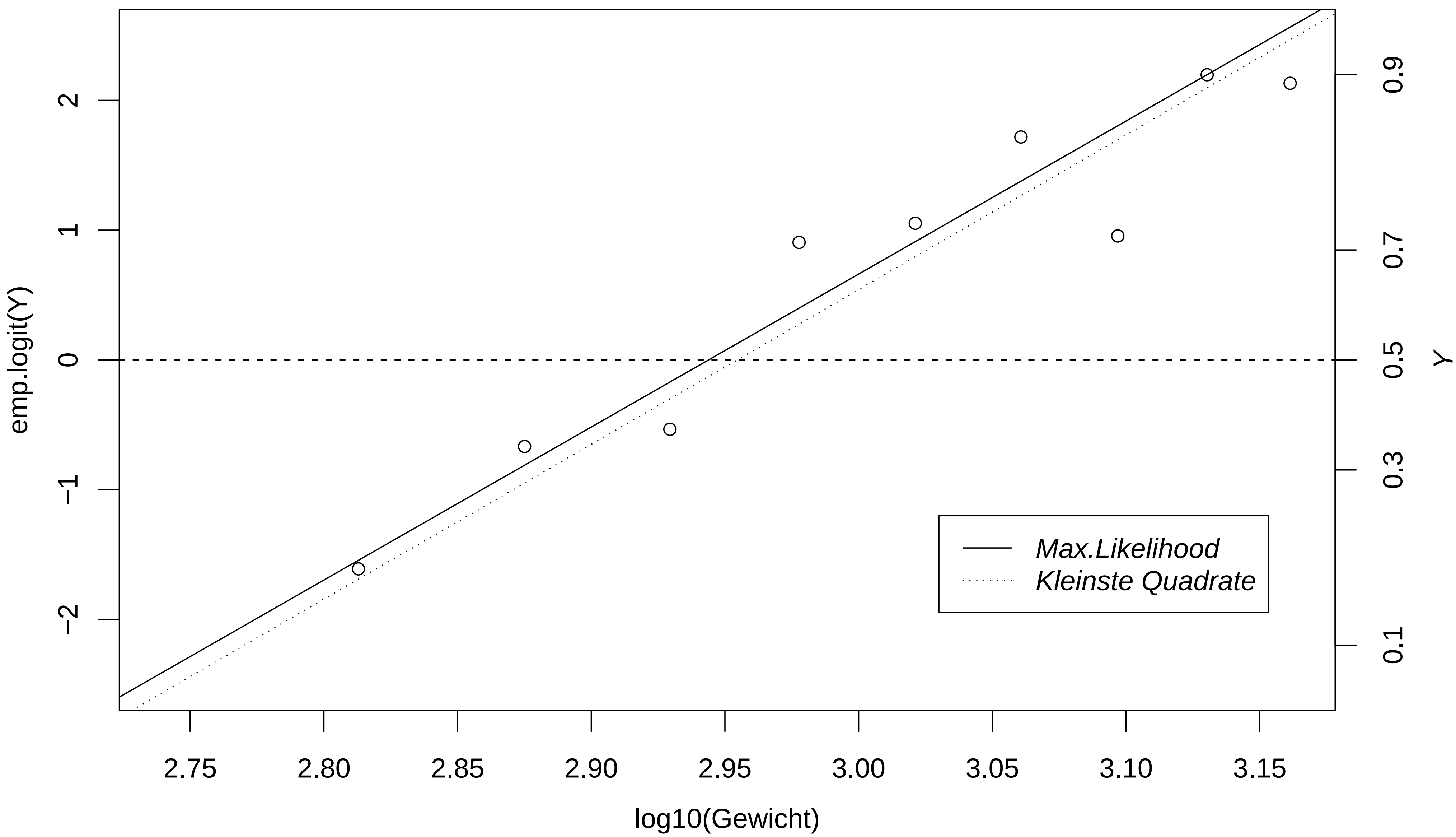
$g\langle \widetilde{Y}_\ell / m_\ell \rangle \approx$ linearer Prädiktor.

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

Abhilfe: **Empirische Logits**

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

$\longrightarrow$ Gewöhnliche multiple Regression mit Z_ℓ? $\longrightarrow$ Näherung.


emp.logit(Y)
−2
−1
0
1
2
Y
0.1
0.3
0.5
0.7
0.9
log10(Gewicht)
2.75
2.80
2.85
2.90
2.95
3.00
3.05
3.10
3.15
Max.Likelihood
Kleinste Quadrate

f **Interpretation of Coefficients?** Need following concepts:

odds

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

$\pi = 1/4$: odds 1:3 (failure is 3 $\times$ more frequent)

log(odds)$= g\langle Y_i = 1\rangle\rangle$, g: Logit-Funktion.
log(odds)$= \eta \longrightarrow$ Wahrsch. $\pi = g^{-1}\langle\eta\rangle = \frac{\exp\langle\eta\rangle}{1+\exp\langle\eta\rangle}$. G^{-1}: „logistische Funktion".

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

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

$$\log\left\langle \frac{\text{odds}\langle \underline{x}_1\rangle}{\text{odds}\langle \underline{x}_2\rangle} \right\rangle = \log\langle \text{odds}\langle \underline{x}_1\rangle\rangle - \log\langle \text{odds}\langle \underline{x}_2\rangle\rangle$$
$$= \eta_1 - \eta_2 = (\underline{x}_1 - \underline{x}_2)\underline{\beta}$$

Koeffizient β_j: Vergrösserung von $x^{(j)}$ um 1 erhöht odds ratio um Faktor e^{β_j}.

h Beispiel Ader-Verengung:

Wert für Vol $= 0.5$, Rate $= 1.75$

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

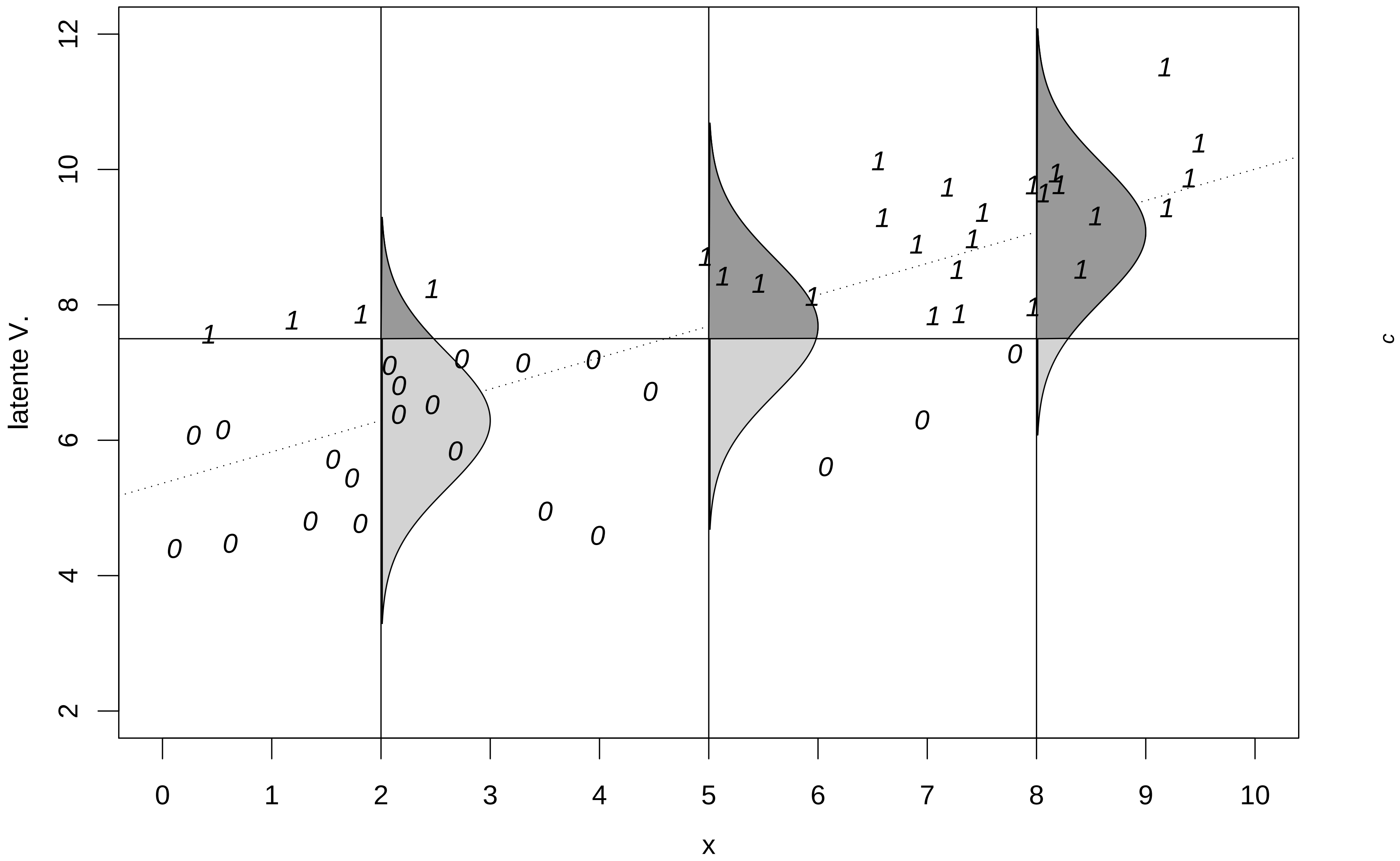
$\longrightarrow$ odds $= 0.0578$, $\quad g^{-1}(-2.85) = 0.0546$

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

$\longrightarrow$ odds $= 0.0578 \cdot 48.4 = 2.80$, $\quad 2.80/3.80 = 0.73$

1.2

i Model with Latent Variable = Schwellenwert-Modell.



$$\begin{aligned} Z_i &= \underline{x}_i^T \widetilde{\beta} + E_i \\ \pi_i = P\langle Y_i = 1\rangle = P\langle Z_i \geq c\rangle &= P\left\langle E_i \geq c - \underline{x}_i^T \widetilde{\beta} \right\rangle \\ &= 1 - F\left\langle c - \left(\beta_0 + \sum_j \beta_j x_i^{(j)}\right)\right\rangle \end{aligned}$$

F: kumulative Verteilungsfunktion des Zufallsfehlers E_i

$\underline{\beta} = [\widetilde{\beta}_0 - c, \widetilde{\beta}_1, \ldots, \widetilde{\beta}_m] \quad \Rightarrow \quad P\langle Y_i = 1\rangle = g^{-1}\langle \underline{x}_i^T \underline{\beta}\rangle$ mit $g^{-1}\langle \eta \rangle = 1 - F\langle -\eta \rangle$

E_i	$\sim$	logistische Vt.:	logistische Regression
E_i	$\sim$	Normal-Vt.:	Probitmodell
E_i	$\sim$	Extremwertvt.:	Komplementäres log-log Modell

1.3 Estimation and Tests

a Method of Maximal Likelihood. There are programs!

b Log-Likelihood:

$$\ell\ell\langle \underline{\widetilde{y}}; \underline{\beta}\rangle = \log\left\langle \prod_\ell P\langle \widetilde{Y}_\ell = y_\ell\rangle \right\rangle = \sum_\ell \log\left\langle \binom{m_\ell}{y_\ell} \pi_\ell^{y_\ell} (1-\pi_\ell)^{m_\ell - y_\ell} \right\rangle$$
$$= \sum_\ell \log\left\langle \binom{m_\ell}{y_\ell} \right\rangle + \sum_\ell (y_\ell \log\langle \pi_\ell\rangle + (m_\ell - y_\ell)\log\langle 1-\pi_\ell\rangle)$$

mit $\text{logit}\langle \pi_\ell\rangle = \underline{x}_i^T \underline{\beta}$

Ungrupp. Daten: $m_\ell = 1$. $\ell\ell\langle \underline{\widetilde{y}}; \underline{\beta}\rangle = \sum_{y_i=1} \log\langle \pi_i\rangle + \sum_{y_i=0} \log\langle 1-\pi_i\rangle$.

1.3

c* Schätzung:

$$\begin{aligned}\partial\ell\langle\underline{\widetilde{y}};\underline{\beta}\rangle/\partial\beta_j &= \sum_\ell y_\ell\frac{\partial\log\langle\pi_\ell\rangle}{\partial\beta_j} + (m_\ell - y_\ell)\frac{\partial\log\langle 1-\pi_\ell\rangle}{\partial\beta_j}\\ &= \sum_\ell\left(y_\ell\frac{1}{\pi_\ell} - (m_\ell - y_\ell)\frac{1}{1-\pi_\ell}\right)\frac{\partial\pi_\ell}{\partial\beta_j}\\ &= \sum_\ell\frac{y_\ell(1-\pi_\ell) - (m_\ell - y_\ell)\pi_\ell}{\pi_\ell(1-\pi_\ell)}\cdot\frac{dg^{-1}\langle\eta_\ell\rangle}{d\eta_\ell}\widetilde{x}_\ell^{(j)}\\ &= \sum_\ell(y_\ell - m_\ell\pi_\ell)\,\widetilde{x}_\ell^{(j)}\end{aligned}$$

da $dg^{-1}\langle\eta\rangle/d\eta = \exp\langle\eta\rangle/(1+\exp\langle\eta\rangle)^2 = \pi(1-\pi)$.

Schätzgleichung:

$$\sum_\ell(y_\ell - m_\ell\,\widehat{\pi}_\ell)\,\underline{\widetilde{x}}_\ell = \underline{0}$$

f **Beispiel Ader-Verengung.**

```
Call: glm(formula = Y ~ Vol + Rate, family = binomial,
        data = d.adern)
Deviance Residuals: ...
Coefficients:
             Value Std. Error z_appr. 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
```

g **Residuen-Devianz**

$$D\langle \underline{y}\,;\underline{\widehat{\pi}}\rangle = 2\left(\ell\ell^{(M)} - \ell\ell\langle \underline{\widetilde{y}}\,;\underline{\widehat{\beta}}\rangle\right)\,.$$

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

$$\begin{aligned}\ell\ell^{(M)} \;&=\; \sum\nolimits_\ell \Big(\log\left\langle \binom{m_\ell}{y_\ell}\right\rangle + y_\ell \log\langle y_\ell\rangle \\ &\quad + (m_\ell - y_\ell)\log\langle m_\ell - y_\ell\rangle - m_\ell \log\langle m_\ell\rangle\Big)\,.\end{aligned}$$

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

$$\widetilde{D}\langle \underline{y};\underline{\widehat{\pi}}^{(K)},\underline{\widehat{\pi}}^{(G)}\rangle = D\langle \underline{y};\underline{\widehat{\pi}}^{(K)}\rangle - D\langle \underline{y};\underline{\widehat{\pi}}^{(G)}\rangle = 2(\ell\ell^{(G)} - \ell\ell^{(K)})$$

asymptotisch chiquadrat-verteilt, wenn das kleine Modell stimmt.

1.3

i **Residuen-Devianz** vergleicht geschätztes Modell mit max. Mod.

$\longrightarrow$ „Anpassungstest"

Achtung: Geht nur bei nicht zu kleinen m_ℓ $\longrightarrow$ grupp. Daten.

j Kleinstes Modell: π_i für alle Beobachtungen gleich.

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

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

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

$\longrightarrow$ Gesamt-Test für das Modell. (H_0: alle βs $=0$!)

1.4 Residuen-Analyse

a Rohe Residuen (response residuals)

$$R_\ell = \widetilde{Y}_\ell / m_\ell - \widehat{\pi}_\ell \;, \qquad \widehat{\pi}_\ell = g^{-1}\langle \underline{\widetilde{x}}_\ell^T \underline{\widehat{\beta}} \rangle$$

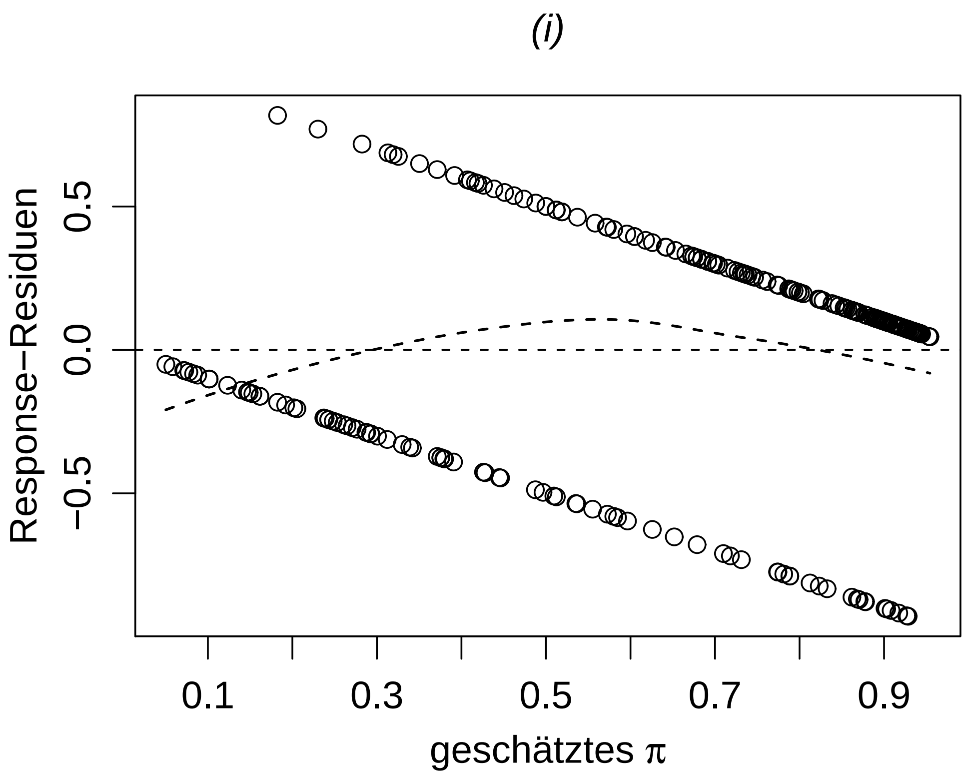
Pearson residuals: $R_\ell^{(P)} = R_\ell \Big/ \sqrt{\widehat{\pi}_\ell (1 - \widehat{\pi}_\ell)/m_\ell}$

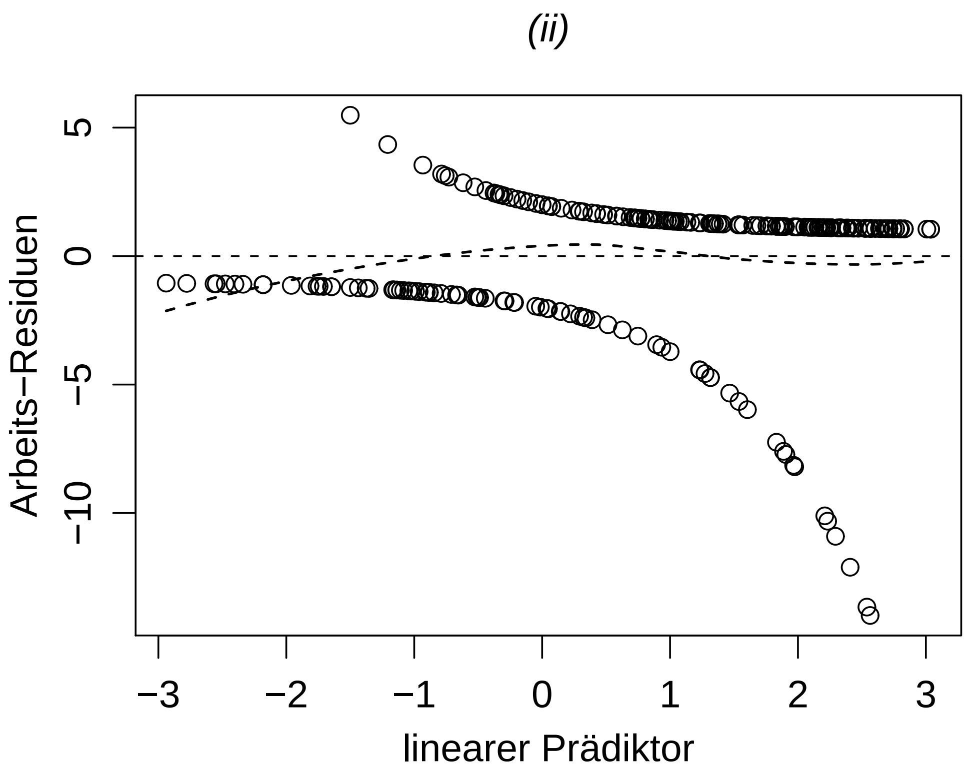
Deviance residuals: Beitrag der i-ten Beobachtung zur Devianz

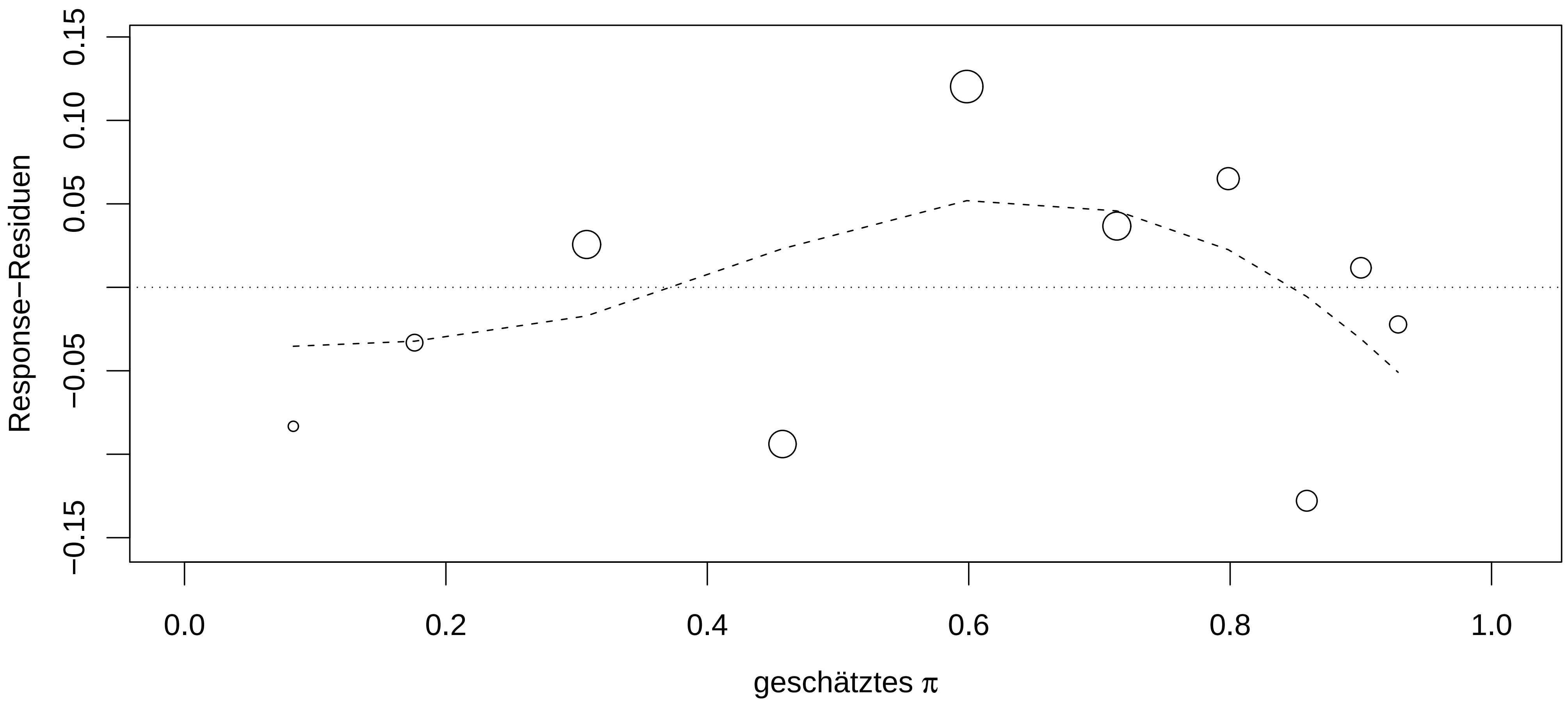
Working residuals:
Berechnung der logist. Regr. via iterativ gewichtete Kl.Qu. (vgl. nichtlin. Regr.)
$\longrightarrow$ lineare Näherung $\longrightarrow$ Residuen : „working residuals".

b Grafische Darstellungen:
Q-Q- (normal) plot meist unnütz!

c **Tukey-Anscombe-Diagramm**:
Rohe Res. / geschätzte π_i oder Arbeitsres. / lin. Prädiktor braucht Glättung.



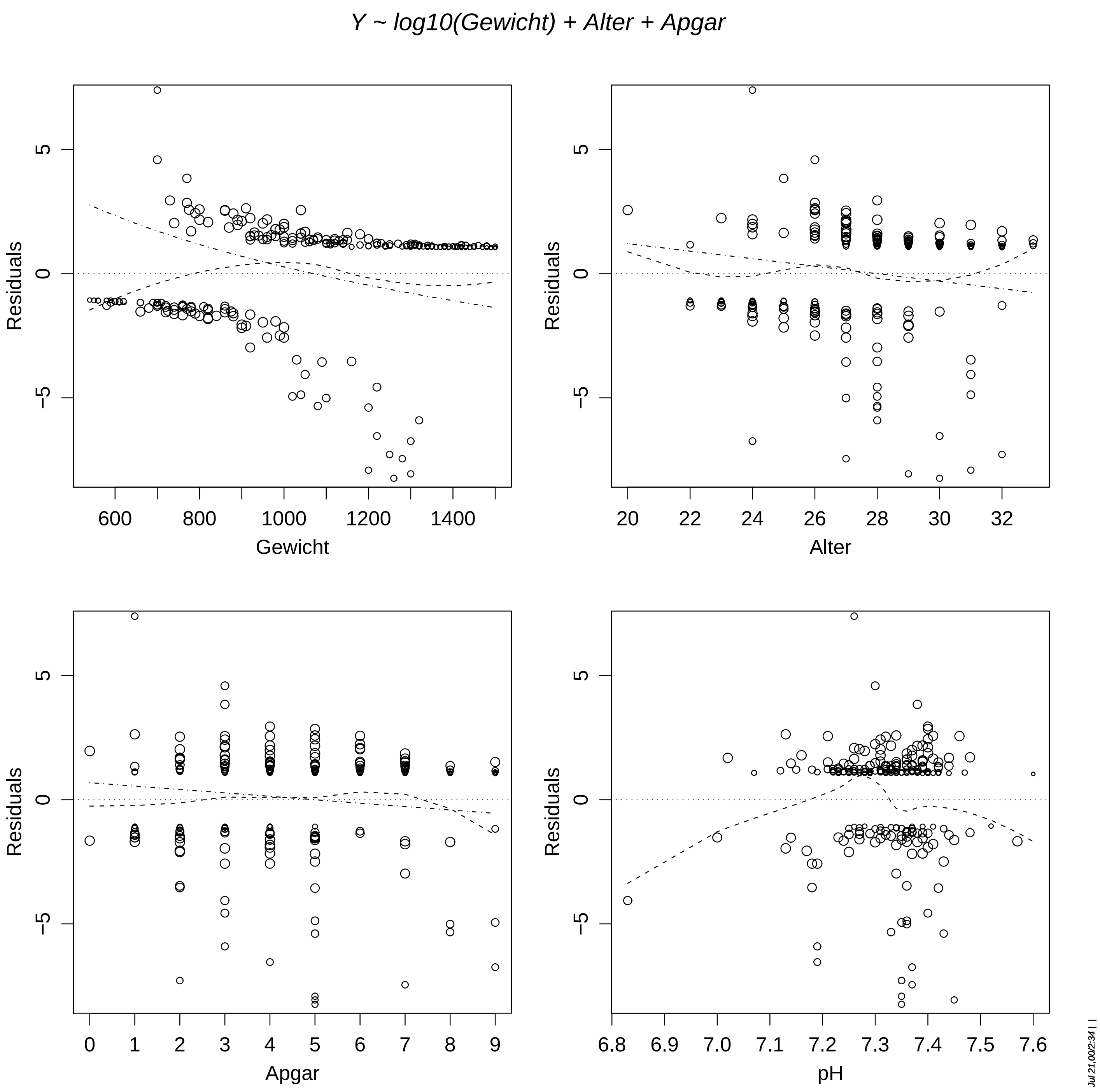



Response−Residuen
−0.15
−0.05
0.05
0.10
0.15
0.0
0.2
0.4
0.6
0.8
1.0
geschätztes π

1.4

e „Partial residual plots":

„Effekte" von $x_i^{(j)}$ ($= \widehat{\beta}_j x_i^{(j)} -$ Konst.) plus geeignete Residuen gegen $x_i^{(j)}$.

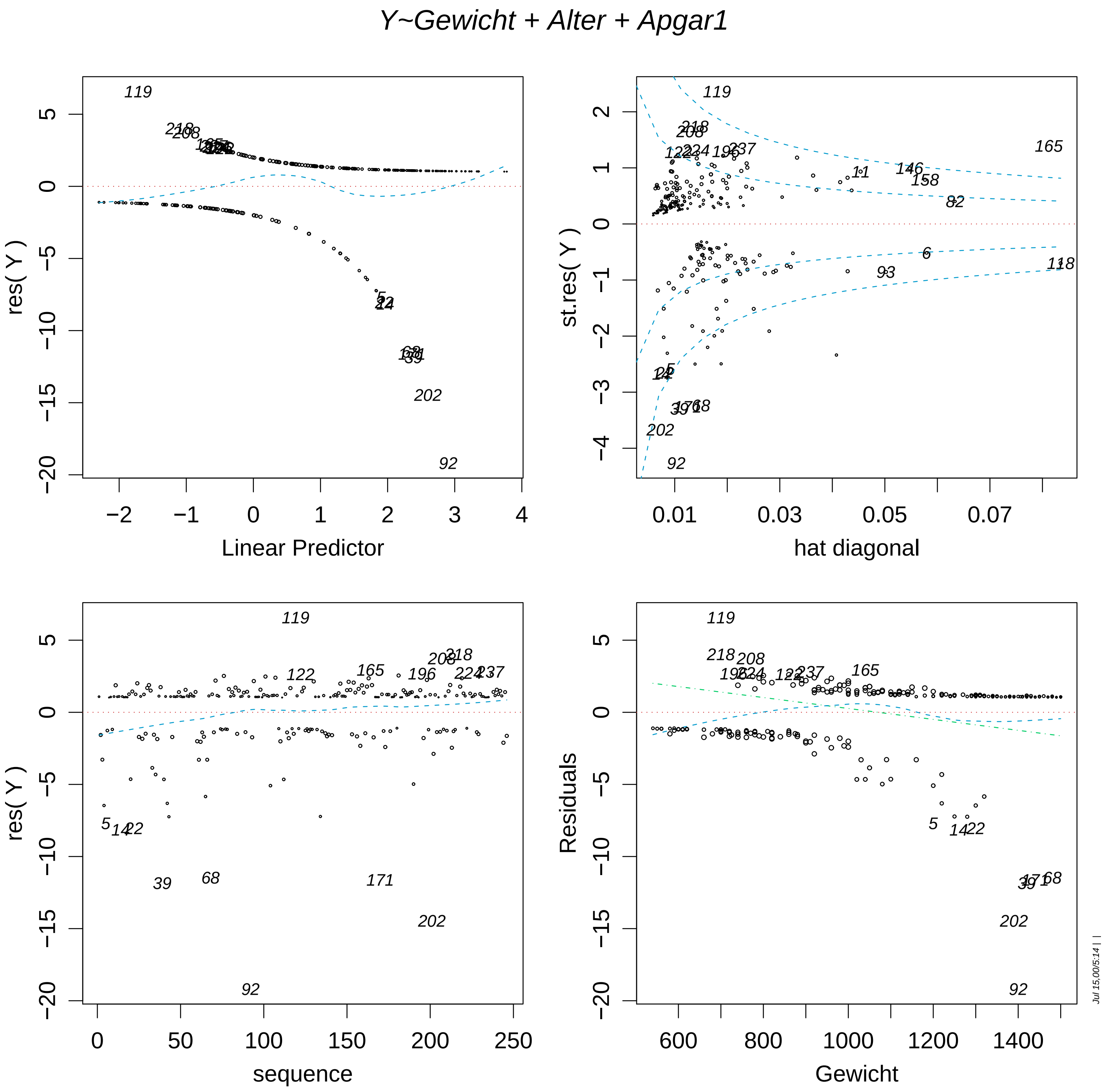

Y ~ log10(Gewicht) + Alter + Apgar
Residuals
5
0
−5
600
800
1000
1200
1400
Gewicht
20
22
24
26
28
30
32
Alter
0
1
2
3
4
5
6
7
8
9
Apgar
6.8
6.9
7.0
7.1
7.2
7.3
7.4
7.5
7.6
pH
Jul 21,00/2:34 | |

regr

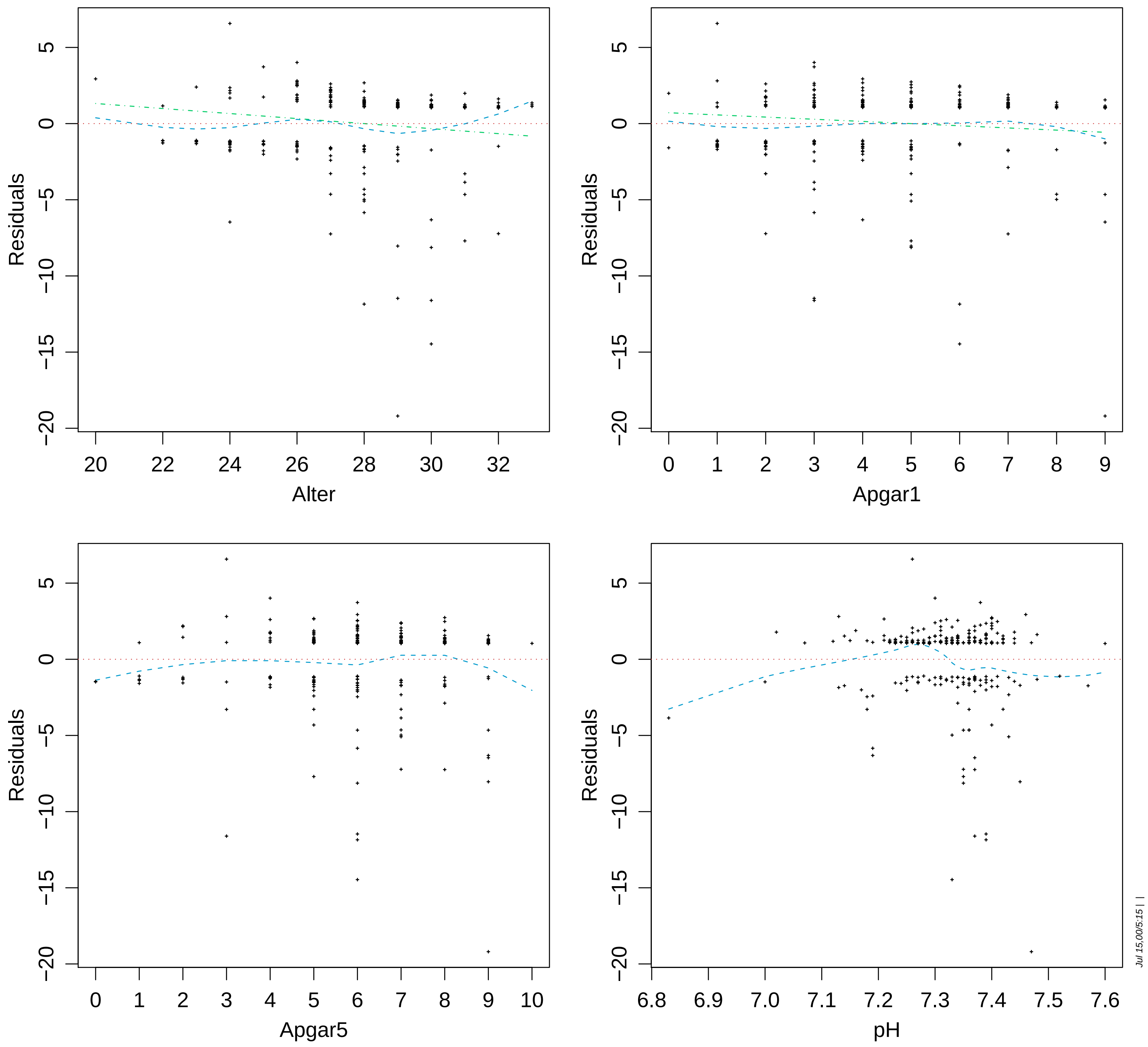
```
regr(formula = Survival ~ Weight + Age + Apgar1, data = t.d,
    family = binomial)
Terms:
                  coef stcoef t.ratio df    Chi2 p.value
(Intercept) -8.484190     NA      NA  1      NA      NA
Weight       0.003791 1.0065  2.2780  1 22.535   0.000
Age          0.165297 0.4519  1.1254  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
```


Y~Gewicht + Alter + Apgar1
res(Y)
Linear Predictor
st.res(Y)
hat diagonal
res(Y)
sequence
Residuals
Gewicht
Jul 15,00/5:14 | |

Y ~ Gewicht + Alter + Apgar1

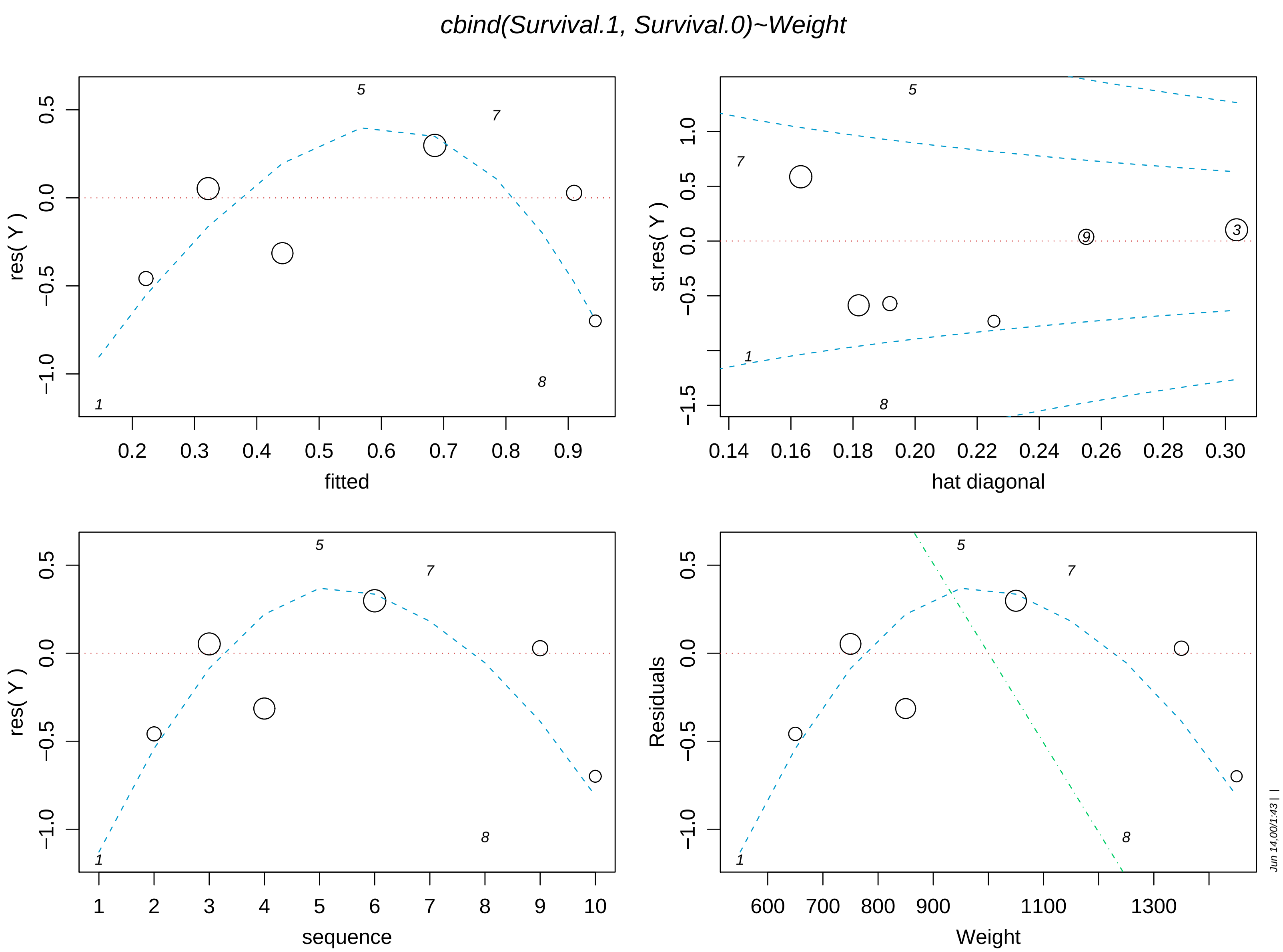


```
Call:
regr(formula = cbind(Survival.1, Survival.0) ~ Weight,
     data = t.d, family = binomial)

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

         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
res(Y)
fitted
st.res(Y)
hat diagonal
res(Y)
sequence
Residuals
Weight
Jun 14,00/1:43 | |