Advanced Regression Modelling ARM – ZF

- **Inferenz**: Gibt es signifikanten Zusammenhang? Wie sicher sind wir mit geschätzten Koeffizienten?
- **Modellwahl**: Welche Variablen haben grossen Anteil an Erklärung der Zielgrösse?
- **Vorhersagen**: Vorhersagen mit Genauigkeit?
- **Goodness-of-Fit:** Beschreibt Modell Daten adäquat?
- **Residuenanalyse**: Stimmen Modellannahmen?

Zielgrösse Y (ihre Verteilung)	Modell
Stetig mit konstanter Varianz	Multiple lineare Regression
Binär	Logistische Regression
Anzahl	Poisson Regression
Stetig mit proportionaler Varianz	Gamma Regression
Zensierte Wartezeit stetig	Weibull-/Cox-Regression
Zensierte Wartezeit diskret	Grupp. prop. Hazardmodell

Generalisierte lineare und additive Modelle (GLM, GAM)

→ par(mfrow = c(**2, 3**)); plot(**Y** ~ ., data = **data**)

Logistische Regressionsmodell

- Zielvariable Y ist binär (0/1) und $\pi =$ Wahrscheinlichkeit, dass der Wert 1 angenommen wird ($\pi = P\langle Y = 1\rangle$) ☞ D.h.: Modellieren Y mittels Bernoulli-Verteilung mit unbekanntem Parameter π. π wird aus Daten geschätzt.
- Idee: bedingte Erwartungswert Zielgrösse, $\mathbb{E}\langle Y_i|x_i\rangle = \pi\langle x_i\rangle$ soll erkl. Var. x_i abhängen. $\mathbb{E}\langle Y_i|x_i\rangle = P\langle Y_i = 1|x_i\rangle$. Ansatz $\pi\langle x_i\rangle = \beta_0 + \beta_1 x_i$ funktioniert nicht, da $\pi\langle x_i\rangle$ Werte [0, 1] muss. ☞ brauchen adäquate Transformation auf $[0,1]$
- Suche nicht-linearen Funktion G, welche Wertebereich berücksichtigt: $\pi\langle x_i\rangle = G\langle\beta_0 + \beta_1 x_i\rangle$ bzw. $G^{-1}\langle\pi\langle x_k\rangle\rangle = \beta_0 + \beta_1 x_k$
- Wie G aussehen: Funktion G meistens S-Form. Beliebte G sind unten ersichtlich. $\eta = g\langle\pi\rangle$ ist der lineare Prädiktor.

	$G\langle\eta\rangle$	$g\langle\pi\rangle$	Name
Logisitsche Verteilung	$exp\langle\eta\rangle/\langle 1 + exp\langle\eta\rangle\rangle$	$log\langle\pi/\langle 1-\pi\rangle\rangle$	Logit-Modell; log Regression
Normal-Verteilung	$\Phi\langle\eta\rangle$	$\Phi^{-1}\langle\pi\rangle$	Probit-Modell
Extremwert-Verteilung	$1 - exp\langle -exp\langle\eta\rangle\rangle$	$log\langle -log\langle 1-\pi\rangle\rangle$	Komplementäres Log-Log-Modell

- Link-Funktion $g\langle\cdot\rangle$ ist identisch zur inversen kumulativen Verteilungsfunktion $G^{-1}\langle\cdot\rangle$.
- Beiden ersten Modelle sind theoretisch sehr schwer auseinander zuhalten ausser in ihren äussersten Extremen
- Dritte Link-Funktion **nicht symmetrisch** um $\pi = 0.5$ ☞ **entscheidend**, was 1 (= **Erfolg**) und 0 (=**Misserfolg**) in **ZV** ist.
- **Logistische** Verteilung (Logit-Modell) wird i. A. **bevorzugt**, weil es eine **einfache Interpretation** der Parameter erlaubt.
- Kann anstelle Ereigniseintretenswahr' **Wettverhältnis** (**odds**) für (Nicht-)Erfolg: $log\langle odds\rangle = log\langle\frac{\pi_i}{(1-\pi_i)}\rangle = \beta_0 + \sum_{j=1}^{m}\beta_j x_i^{(j)}$
- Somit haben Koeffizienten β gleiche Bedeutung bei Beschreibung der log-odds wie bei der multiplen lin. Regression.

Die logistische Regression (**Logit-Modell**) ist durch folgende drei Elemente bestimmt:

1. ZV Y_i Bernoulli mit Erfolgswahrscheinlichkeit π_i.
2. Lineare Prädiktor $\eta_i = \beta_0 + \beta_1 x_i^{(1)} + \cdots + \beta_m x_i^{(m)}$
3. Lineare Prädiktor η_i ist mit Erfolgsparameter über logistische Funktion verlinkt: $\eta_i = log\langle\frac{\pi_i}{1-\pi_i}\rangle\ \pi = G\langle\eta\rangle = \frac{exp\langle\eta\rangle}{1+exp\langle\eta\rangle}$

Parameterschätzung: Um geeignete Schätzung herzuleiten ☞ Maximum Likelihood. Verfahren wird für Fall erläutert, dass Zielvariable unabhängig und binomial verteilt, verwenden logistische Link und nur eine erklärende Variable.

- Um Maximum-Likelihood-Schätzung von β_0, β_1 zu erhalten, nehmen wir partiellen Ableitungen von $l\langle\underline{\beta}|\underline{y}^*\rangle$ bezüglich β_0 und β_1 durch Anwendung der Kettenregel. ML-Schätzung erhalten wird durch Gleichung $\sum_k(y_k^* - m_k\hat{\pi}_k)\,x_k^{(j)} = 0,\ j = 0,1$
- Da $\pi_k = G\langle x_k^T\hat{\beta}\rangle$ eine nichtlineare Funktion der unbekannten Parameter $\underline{\beta}$ ist, ist das **Gleichungssystem nicht linear** → iterativ Algorithmen (Gradientenverfahren) wie z.B. der IRLS-Algorithmus müssen verwendet werden.
- Kommt **nicht darauf** an, **ob** wir **binären Daten oder gruppierten** für Anpassung verwenden. Schätzwerte sind gleich
- Anpassung mit binären Zielvariablen: MAC.glm1 <- glm(**Y** ~ **ac**, family = binomial, **MAC**) #Bei Default wird Logit-Modell
- Anpassung mit gruppierten Daten: MAC.glmG1 <- glm(cbind(**Y, m-Y**) ~ **ac**, family = binomial, data = **MAC1**) #angepasst
- **Achtung**: Anpassung ist Resultat nichtlinearen Gleichungssystems ☞ kann vorkommen, dass folgende erscheint: Warning message: glm.fit: algorithm did not converge. Algorithmus hat nicht konvergiert ☞ **kann Schätzwerten nicht trauen**
- Leider nicht möglich Konvergenz Algorithmus sicherzustellen, ausser max. Anzahl Schritte erhöhen glm(maxit = 1000).

Binäre Regression – Besonderheiten

- Punkte liegen auf zwei Kurven wegen binären Werten. Hat nichts mit Verletzungen Modellannahmen zu tun
- fast unmöglich, Ausreisser zu erkennen
- Situation kann durch Datenaggregation erheblich verbessert werden
- Bei binären ZV ist Normal-QQ-Plot oft bedeutungslos. Selbst wenn Residuen nicht normalverteilt, sieht Plot gut aus.
- Leverage-Werte oft sehr klein. Können nicht Faustregel beurteilen, da zu viele Beob. verglichen Anz. unbek. Parameter
- **Beispiel:** Geschätzte Modellgleichung für Erwartungswert binomial Logit-Modell mit einer erkl. Variable weight:
- $\hat{\eta}_i = -4.561 + 0.0051 \cdot weight_i$ → $\hat{\pi}_i = \hat{\mu}_i = \frac{exp\langle\hat{\eta}_i\rangle}{1+exp\langle\hat{\eta}_i\rangle}$, Werte $-4.561, 0.0051$ sind β_0, β_1 Koeffizient aus Summary Output

Koeffizient log. Regression = **Änderung log. Wettverhältnis** → exp(0.0051)–1 #0.005 ← Wenn Erhöhung weight um 1 Einheit, führt dazu dass das Log. Odds Wettverhältnis für zutreffen bspw. Infekt (Wert 1) sich um 0.5 % (0.005*100) erhöht.

sunflowerplot(x = **birth$weight**, y = **birth$Y/birth$m**, number = **birth$m**) | Verteilungsannahme: $(Y_i|X_i) \rightarrow (infect_i|wcc_i)$
x = erkl. Variable, y = Y Anz. zutreffende Argumente, m = Total Anz. | ZV unabhä. bin/bern: Y~Bin(m=**6**, $\pi_i = P(infect_i = 1)$
x <- seq(min(**birth$weight**), max(**birth$weight**), length=50) #Prognose | $\mathbb{E}\langle Y_i\rangle = m_i\pi_i \rightarrow \mathbb{E}\left(\frac{Y_i}{m_i}\right) = \pi_i \equiv \mu_i|\eta_i = log\langle\frac{\pi_i}{1-\pi_i}\rangle = \beta_0 + \beta_1 \cdot wcc_i$
mu.p <- predict(**birth.glm1**, newdata = data.frame(**weight** = x), | Ab welche Wert erkl. Var. wcc_i haben 80% Erfolg
type = "response"); lines(x, mu.p, col = 'blue', lwd = 4, lty = 6) | $wcc_i = \frac{1}{\beta_1}\cdot\left[log\langle\frac{\pi_i}{1-\pi_i}\rangle - \beta_0\right],\ \pi_i = 0.8,\ \beta_k =$summary

Asymptotisch Normalverteilt (Wald Asymptotik): Theorie ML-Schätzer: Jeder ML-Schätzer asymptotisch normalverteilt.

- asymptotisch im Sinne Zentralen Grenzwertsatzes. D.h. Approximation umso besser, je grösser Anzahl Beobachtungen.
- Können deshalb Summary-Output gleich interpretieren wie bei Kleinsten-Quadrate. Testresultate, VI gelten **nur approx**.

Das umfassende Modell – Die einfache Exponentialfamilie

Sowohl das lineare wie auch das logistische Regressionsmodell haben folgende Gemeinsamkeiten:

- Erklärende Variablen $x^{(1)}, x^{(2)}, \ldots, x^{(m)}$
- eine Link-Funktion g, sodass $g\langle\mu\rangle = \underline{x}^T\underline{\beta}$
- eine Zielvariable Y mit Erwartungswert $E(Y) = \mu$
- eine Verteilung für Variabilität in Zielvariablen Y.

- Klassische lin. Reg. Identität $g\langle\mu\rangle = \mu$ als Link und Normalverteilung mit $\mathbb{E}\langle Y_i\rangle = \mu_i$ und konstanter Varianz σ^2 verwendet.

Zufallsvariable Y mit Erwartungswert $\mu = \mathbb{E}\langle Y\rangle$ und Dispersion ϕ hat Verteilung in einfachen Exponentialfamilie, falls

- Dichte oder Wahrscheinlichkeitsfunktion von Y geschrieben werden kann als $f\langle y_i;\mu_i,\phi\rangle = \exp\langle\frac{y_i b\langle\mu_i\rangle - c\langle\mu_i\rangle}{\phi/w_i} + d\langle y_i;\phi,w_i\rangle\rangle$
- $b\langle\,\rangle, c\langle\,\rangle$ **spez. Verteilung** aus **Exp.familie**
- $d\langle\,\rangle$ **Normalisierung** Dichte/Wahrscheinlichkeitsfunktion auf **1**
- Dispersionsparameter ϕ Teil der Varianz von Y
- «Variable» w_i ist für jede Beobachtung i eine bekannte Zahl. w_i kann unter den versch. Beobachtungen unterschiedlich sein. Interpretation als Gewicht. Wird in diesem Modul**teil** verwendet, um Binomialverteilung mit $m_k > 1$ abzudecken.

- Kann zeigen, dass für Erwartungswert von Y gilt $\mu_i = E\langle Y_i\rangle = \frac{c'\langle\mu_i\rangle}{b'\langle\mu_i\rangle}$, wobei $c'\langle\mu_i\rangle, b'\langle\mu_i\rangle$ Ableitung nach μ sind
- Varianz ist Produkt positiven Funktion $V\langle\mu_i\rangle = \frac{1}{b'\langle\mu_i\rangle}$ von Erwartungswert, Dispersion ϕ und Gewicht w_i: $var\langle Y_i\rangle = \frac{\phi}{w_i}V\langle\mu_i\rangle$

Gamma-Verteilung (Verallgemeinerung Exponentialverteilung)

- Zeit zwischen zwei konsekutiven Ereignissen, die Poisson verteilt sind, ist exponentialverteilt.
- beide sind geeignet, um Betragsgrössen wie (Überlebens-)zeiten, Schadenshöhen, Verluste oder Kosten zu modellieren
- Dichte gammaverteilten ZV Y_i ist $f\langle y,\alpha,\beta\rangle = e^{-\beta\cdot y}\cdot y^{\alpha-1}\cdot\frac{\beta^\alpha}{\Gamma\langle\alpha\rangle}$ mit $\alpha,\beta > 0$ auf Definitionsbereich $y \in [0,\infty)$
- Der Parameter α ist **Formparameter** und β ist die **Rate** ($1/\beta$ ein **Skalenparameter**).
- $\Gamma\langle\,\rangle$ ist die Gamma-Funktion, eine Verallgemeinerung der Fakultätsfunktion auf die reellen Zahlen.
- $\alpha = 1,\ \beta = \lambda$ → **Exp.verteilung** mit λ: $T \sim Exp\langle\lambda\rangle$
- $\alpha = \frac{m}{2},\ \beta = \frac{1}{2}$ → **χ^2-Verteilung** mit m Freiheitsgraden.
- m positive ganze Zahl, dann wird Gamma-Verteilung mit $\alpha = m$ und $\beta = \lambda$ auch **Erlang-Verteilung** genannt.
- **Summe von zwei exponentialverteilten Z.V. ist gammaverteilt**. Allgemein: Falls $T_i, i = 1,2\ldots,m$ unabhängig $\sim Exp\langle\lambda\rangle$ dann ist Summe $T_1 + T_2 + \cdots + T_m \sim \Gamma\langle\alpha = m,\beta = \lambda\rangle$. Wenn wir annehmen, dass Wartezeit in einer Warteschlange unabhängig exponentialverteilt mit Parameter λ ist, dann ist Zeit zur Bedienung von m Kunden gammaverteilt.

Zusammenstellung wichtigsten Grössen für einige der populärsten Verteilungen aus der einfachen Exponential-Familie. →

Verteilung	Wertebereich Y	$\mathbb{E}\langle Y\rangle = \mu$	$var\langle Y\rangle$	$b\langle\mu\rangle$	$V\langle\mu\rangle$	ϕ	w
Gauss $\langle\mu,\sigma^2\rangle$	$(-\infty,+\infty)$	μ	σ^2	μ	1	σ^2	1
Binomial $\langle m,\pi\rangle$	$\frac{(0,1,2,\ldots,m)}{m}$	π	$\frac{\pi(1-\pi)}{m}$	$\log\langle\frac{\mu}{1-\mu}\rangle$	$\mu(1-\mu)$	1	m
Poisson $\langle\lambda\rangle$	0,1,2, ...	λ	λ	$\log\langle\mu\rangle$	μ	1	1
Gamma $\langle\alpha,\beta\rangle$	$(0,+\infty)$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	$-\frac{1}{\mu}$	μ^2	$\frac{1}{\alpha}$	1
Invers Gauss	$(0,+\infty)$	μ	$\frac{\mu^3}{\lambda}$	$-\frac{1}{\mu^2}$	μ^3	$\frac{1}{\lambda}$	1

Das generalisierte lineare Modell – GLM

- **1 Verteilungselement:** Verteilung Zielvariable Y_i, gegeben erklärenden Variablen $\underline{x}_i$, ist Mitglied der einfachen Exponentialfamilie mit Erwartungswert $\mathbb{E}\langle Y_i|x_i\rangle = \mu_i$ und Dispersion ϕ. Zielvariable $Y_i, i = 1,\ldots n$, ist unabhängig verteilt.
- **2 Strukturelement:** Erwartungswert μ_i wird zum linearen Prädiktor $\eta_i := \underline{x}_i^T\underline{\beta}$ der erklärenden Variablen mittels einer oft nichtlinearen Funktion $g\langle\,\rangle$ in Beziehung gebracht: $g\langle\mu_i\rangle = \eta_i = \underline{x}_i^T\underline{\beta}$. Funktion $g\langle\,\rangle$ wird Link-Funktion genannt.

Link-Funktion

Ihre **Wahl** ist zu einem **gewissen Teil recht willkürlich**, sofern sie folgende grundlegende Eigenschaft besitzt:

- Inverse Link-Funktion $g^{-1}\langle\,\rangle$ soll Werte linearen Prädiktor $\eta = \underline{x}^T\underline{\beta}$ auf Raum der möglichen Werte abbilden, die Erwartungswert von Y annehmen kann. Funktion $g^{-1}\langle\,\rangle$ **muss unmögliche Werte vermeiden**. Link-Funktionen:
  - $g\langle\mu\rangle = \mu$, wenn $\mu = \mathbb{E}\langle Y\rangle$ keinen Einschränkungen unterliegt,
  - $g\langle\mu\rangle = \log\langle\mu\rangle$, wenn $\mu = \mathbb{E}\langle Y\rangle > 0$ sein muss,
  - $g\langle\mu\rangle = \text{logit}\langle\mu\rangle := \log\langle\mu/(1-\mu)\rangle$, falls $\mu = \mathbb{E}\langle Y\rangle$ zwischen 0 und 1 liegen muss.

Kanonische Link-Funktion

- Für jede Verteilung aus einfachen Exponentialfamilie gibt spezielle **Link-Funktion**, die im gewissen Sinn **natürlich** ist.
- Führen Erwartungswert μ in kanonischen Parameter $\theta := b\langle\mu\rangle$ über, bzw. wird angenommen, dass Effekte linearen Prädiktors linear auf kanonischen Parameter wirken, $\theta = \eta$. Diese Link-Funktionen nennt man kanonische Link-Funktionen.
- Also entspricht $b\langle\mu\rangle$ gerade dem kanonischen Link für die jeweilige Verteilung. Siehe Tabelle oben Spalte $b\langle\mu\rangle$
- ACHTUNG: kanonische Link für Gammaverteilung erfüllt Anforderungen nicht. Genau Anpassungsresultat überprüfen!

Link-Funktionen, die in R implementiert sind

- Kanonische Link wird mit D bezeichnet: (Spaltenname entspricht Verteilung (= family) in R, Zeilenname dem Namen der Link-Funktion
- z.B. glm(**Y~X**, family=**poisson**(link = **identity**))

	binomial	gaussian	Gamma	inverse.gauss	poisson	quasi
logit	D					●
probit	●					●
cauchit	●					
cloglog	●					●
identity		D	●	●	●	●
inverse		●	!D!	●		●
log	●	●	●	●	D	●
$1/mu^2$				D		●
sqrt					●	●

Poisson-Verteilung/Regression

- Bei Poisson-Verteilung ist Erwartungswert gleich Varianz → Normalverteilung kann nicht angenommen werden, da mit höheren $\mathbb{E}(X)$ dann auch eine höhere Varianz vorhanden ist.
- Bei Schätzungen sollte man ganze natürliche Zahlen nehmen, da die Zielgrösse immer ganze Zahlen (Anzahlen) hat.

- Zielgrösse: $Hurricanes_i = Anzahl\ Wirbelstürme\ im\ Jahr\ i$, $Hurricanes_i \sim \boldsymbol{Pois(\lambda_i)}$, unabhängig mit $\mathbb{E}(Y_i) = \mu_i$
- Struktur: $log(\lambda_i) = \beta_0 + \beta_1 \cdot SOI_i + \beta_2 \cdot NAO_i + \beta_3 \cdot SST_i$, Link: log = kanonischer Link, linearer Prädiktor $log(\mu_i) = \eta_i$
- coef(**pois.log.glm**)[**"SOI"**] #0.0618, → exp(coef(**pois.log.glm**)[**"SOI"**]) #1.063 → 1.063-1 = 0.063 → Wenn SOI um eine Einheit erhöht wird, dann vergrössert sich die erwartete Anzahl Wirbelstürme um 6.3 %.

Poisson-Regression: Annahme: lineare Prädiktor wirkt gemäss einem Potenzgesetz auf Erwartungswert λ ein.

- $0 < \beta$ → Je grösser der Wert in der Variable, desto grösser ist der Faktor (Faktor = x^β)
  - $0 < \beta < 1$ → monoton steigende, konvexe Funktion
  - $1 < \beta$ → monoton steigende, konkave Funktion
- $\beta < 0$ → Je grösser Wert in Variable, desto kleiner (d.h. nahe bei 0) ist Faktor → monoton fallende, konvexe Funktion

Gamma-Regression $\boldsymbol{exp(a+b+c) = exp(a) \cdot exp(b) \cdot exp(c)}$ | $\boldsymbol{(exp(a))^b = exp(a \cdot b)}$

- Zielvariable Y_i ist $Y_i \sim Gamma(\mu, \phi)$ verteilt → $\boldsymbol{ClmAmt_i} \sim Gamma(\mu, \phi)$, wobei $log(\mu_i) = \beta_0 + \beta_1 \cdot \boldsymbol{Bluebook_i} + \ldots + \beta_m \cdot \boldsymbol{DensityUrbans_i}$, log-Link bildet multiplikative Skala ab und stellt sicher, dass $\mathbb{E}(X)$ nur positive Werte annehmen kann
- $\mu_i = e^{\beta_0} \cdot e^{\beta_1 \cdot \boldsymbol{Bluebook_i}} \cdot \ldots \cdot e^{\beta_m \cdot \boldsymbol{DensityUrbans_i}}$ coef(**gamma.log.glm**) #0.0618, → exp(coef(**gamma.log.glm**)) #1.063 → 1.063-1 = 0.063 → Wenn Bluebook um eine Einheit erhöht wird, dann vergrössert sich die erwartete ClmAmt um 6.3 %.

qchisq(c(0.**025**, 0.**975**), **51**) #[33, 72] Gamma ist exp., wenn $\phi = 1$ → Res. deviance (18.05 on **51** df) im Annahmebereich

Statistische Inferenz und Variablenselektion

- p ML-Schätzgleichungen $\underset{\sim}{\beta}$ werden durch Differenzierung Log-Likelihood-Funktion bzgl. p-Parameter hergeleitet. (Ableitungen **score** Funktion). Gleichungssystem wird mit iterativen gewichteten Kleinste-Quadrate-Verfahren (iteratively reweighted least-squares; **IRLS**) gelöst. Gewichte sind abhängig von Lösung, was zu einem iterativen Vorgehen führt.
- Mit $\hat{\underline{\beta}}$ erhält angepasste Werte $\hat{\mu}_i$. (mult. Reg. $\hat{y}_i$) → **schätzen Erwartungswert** und **nicht Responsewerte** predicten

Schätzung des Dispersionsparameters (I. A. muss der Dispersionsparameter ϕ auch geschätzt werden)

- Kann durch Maximierung Likelihood geschehen. Oder wie bei multiplen lin. Regr. durch erwartungstreuen Schätzer.
- Im Falle der Binomial- und der Poisson-Verteilung ist $\phi = 1.$, Für den Gamma-Fall: siehe später

Asymptotische Verteilung und Wald-Typ-Inferenz

- Approx. besser, je grösser Anz. Beob. n ist, weil Approx. nach ZGWS. Stichprobenumfang hängt Datensatz/Modell ab.
- In Praxis heisst das für GLM $\underset{\sim}{\hat{\beta}} \overset{a}{\sim} \mathcal{N}_p\langle \underline{\beta}, \phi^* (\boldsymbol{X^T W X})^{-1}\rangle$ mit $\boldsymbol{W}_{ii} = \frac{f\, w_i}{V(\hat{\mu}_i)(g'(\hat{\mu}_i))^2}$ wobei $\boldsymbol{W}$ Diagonalmatrix und $\phi^* = 1$ bei Binomial- und Poissonverteilung, sonst $\phi^* = \hat{\phi}$ → Inferenz basierend auf diesem Verteilungsresultat = Wald-Typ-Inferenz
- VI **approx**: $\hat{\underline{\beta}} \pm q \cdot se\langle \hat{\underline{\beta}} \rangle$, $se\langle \hat{\underline{\beta}} \rangle = diag(\sqrt{\phi^* (\boldsymbol{X^T W X})^{-1}})$. Falls $\phi^* = 1$ → q Quantil Normalverteilung, bei $\phi^* = \hat{\phi}$ t-Verteilung
- Vertrauensint. Wald: confint.default(**dat.glm**); coef(**dat.glm**)[1]+c(-1,1)*qnorm(**0.975**, 0, 1)*summary(**dat.glm**)$coef[**1**,2]
- family(**USH.glm**)$linkinv(coef(**USH.glm**)) #Retour transformieren Koeffizienten mit Link Funktion | ↖ von Hand Wald VI

Vergleich von geschachtelten Modellen

Mult. lin. Regression F-Test, um statistische Signifikanz einer oder mehrerer erkl. Var. Abzuklären:

- $H_0: \beta_{j1} = 0, \beta_{j2} = 0, \ldots, \beta_{jq} = 0$, Alternative H_A: mindestens ein $\beta_{jk}, k = 1, \ldots, q$, ist ungleich 0. Teststatistik $F := \frac{(SS_E^* - SS_E)/q}{SS_E/(n-p)}$
- $SS_E^* := \sum_{i=1}^n R_i^2$, r_i aus «kleinen» Modell, $(p-q)$ Koeffizienten | $SS_E := \sum_{i=1}^n R_i^2$, r_i aus «grossen» Modell, p Koeffizienten.

Residuen-Devianz – In multiplen linearen Regression werden

- Differenzen zwischen SS_E von versch. Modellen zum Vergleich. SS_E kann auch Schätzung Skalenparameters σ benutzen
- SS_E misst die Genauigkeit des Modells und ist zugleich Ausdruck, der in Methode Kleinsten Quadrate minimiert wird.
- Summe lässt sich auch mit Hilfe der Log-Likelihood ausdrücken und umformulieren, sodass Term nicht von ϕ abhängt.
- Der Term wird Residuen-Devianz genannt: $D\langle \underline{Y}; \hat{\underline{\beta}} \rangle := 2\phi \left(\ell\langle \underline{Y}; \phi \rangle - \ell\langle \hat{\underline{\mu}}; \phi \rangle \right)$

Devianz-Test: GLM-Generalisierung F-Test mit Ersetzung Quadrat-Summen durch Residuen-Devianz. Teststatistik $T_D := \frac{D\langle \underline{Y}; \hat{\widetilde{\beta}}^r \rangle - D\langle \underline{Y}; \hat{\underline{\beta}} \rangle}{\phi^*}$, $D\langle \underline{Y}; \hat{\widetilde{\beta}}^r \rangle$ Devianz reduziertes Modell, $D\langle \underline{Y}; \hat{\underline{\beta}} \rangle$ Devianz volles Modell und ϕ^* 1 oder Schätzung v. ϕ

- I. A. ist Verteilung von T_D unter Nullhypothese zu schwierig herzuleiten. **Asymptotisch** ist Teststatistik T_D unter H_0 χ_q^2-verteilt, q ist Anzahl Koeffizienten, die in H_0 0 gesetzt werden. Näherung ist je besser je grösser Stichprobengrösse n.

Normalverteilte Zielgrössen: Devianz-Teststatistik $T_D := \frac{S(\hat{\widetilde{\beta}}^r) - S(\hat{\underline{\beta}})}{S(\hat{\underline{\beta}})/(n-p)}$ wobei erwartungstreue Schätzer $\phi = \sigma^2$ anstelle ML-

- Nicht identisch mit F-Test (Faktor $1/q$ fehlt), P-Wert stimmen grossen Stichproben überein, weil $\mathcal{F}_{q,n-p} \to \frac{1}{q}\chi_q^2$ falls $n \to \infty$. → Praxis sinnvoll, normalverteilten Zielgrösse F-Test anstelle Devianztest verwenden, da Verteilung F bekannt ist.

anova(**grossesModell**, **reduziertesModell**, test = "Chisq") #Wenn p-Wert kleiner als Sig.niveau, grosses Modell besser. Auf 5% Niveau besteht zwischen beiden Modellen kein signifikanter Unterschied, da P-Wert von 0.0697 grösser als 5% ist. Man kann davon ausgehen, dass kleinere Modell Daten ebenso gut beschreibt. Nehmen kl. Modell: Prinzip Occam's Rasor

Null-Devianz: Null Deviance bezieht sich auf Residuen-Devianz des Modells, das nur aus Achsenabschnitt besteht.

- Muss Differenz zwischen null deviance und residual deviance bilden und durch $\hat{\phi}$ teilen, sofern ϕ unbekannt ist.

- **Beispiel Narkos:** Null Deviance: 15.4334 on 5 degrees of freedom, Residual Deviance: 1.7321 on 4 degrees of freedom
- 1- pchisq((**15.4334 - 1.7321**)/1, df = **5 - 4**) #0.0002143061 #durch 1 teilen, wenn Pois/Binom, sonst durch $\hat{\phi}$
- Da P-Wert kleiner 0.05, muss Null-Hypothese, dass alle Koeffizienten ausser Achsenabschnitt 0 sind, verworfen werden. Folglich hat mindestens eine der erklärenden Variablen signifikanten Einfluss auf Erwartungswert der Zielvariable.

Notwendigkeit für Devianztest: Probleme Wald-Test: hat erheblichen Nachteil in **binären Regression**, wenn wahre Parameterwerte sehr weit von Null entfernt. Standardfehler werden sehr (zu) gross. Je weiter $\hat{\beta}_k$ von 0 entfernt, desto grösser Wald-Teststatistik für $H_0 : \beta_k = 0$. Ab gewissen Punkt wird sie wieder kleiner und geht gegen 0, wenn $\hat{\beta}_k$ unendlich gross wird. Folglich Null-Hypothese nicht signifikant. → **Hauck-Donner-Phänomen**

- Nehmen Devianztest anova(test = "Chisq") oder drop1() Hypothese $H_0 : \beta_k = 0$ oder «profile» Vertrauensintervalle.

Schätzung des Dispersionsparameters – Overdispersion (übermässige Streuung)

- Gem. Modell ist Residuen-Devianz ungefähr χ^2-verteilt. Somit gilt, dass mittleren Residuen-Devianz $\mathbb{E}\langle D/(n-p)\rangle \approx \phi$ und folglich können wir mit $D/(n-p)$ den Parameter ϕ schätzen. Dies ist **keine Maximum-Likelihood-Schätzung**

Im Fall von **gammaverteilten** Zielvariablen:

- können auch $D/(n-p)$ zur Schätzung von ϕ verwenden. Schätzer reagiert empfindlich auf angepasste Werte nahe 0!
- Stabileres Verhalten zeigt Momentenschätzer $CV := \frac{1}{n-p}\sum_{i=1}^n \left(\frac{Y_i - \hat{\mu}_i}{\hat{\mu}_i}\right)^2$ = empirisch Varianz Pearson-Residuen.

Modelle mit $\boldsymbol{\phi = 1}$ (Binomial-, Poisson- und Exponential-Modellen ist $\phi = 1$ → braucht keine Schätzung!)

- Mit $D/(n-p)$ Modellunstimmigkeiten bzgl. Streuung aufdecken, die sich in ungewöhnlichen Werten von D zeigen.
  - wissen: $D \overset{a}{\sim} \chi^2$-verteilt mit $(n-p)$ df, **ausser bei binären Zielvariablen**. Testen auf Nullhypothese H_0: $\phi = 1$
  - Plausibilisierungsverfahren mit statistischem Test: Falls $D > q_{0.95}^{\chi^2_{n-p}}$ dann Overdispersion vorhanden auf 5 % Niveau.
- **Binäre** Zielvariabel gilt asymptotische Theorie nicht → Wenn immer möglich, binäre Daten aggregieren zu binomial ZV!

Bsp Narkos: Residual dev. 1.73 on 4 df, qchisq(0.95, df=**4**) #9.49, D kleiner als q, H_0 nicht abgelehnt, 1-pchisq(**1.73, 4**) Da P-Wert von 0.324 > 0.05, gibt (k)eine Evidenz gegen Null-Hypothese "phi=1" und damit (k)eine Evidenz Overdispersion **Overdispersion**: gibt mehr Variabilität in Daten, als (**Poisson**)-Verteilung zulässt. Dispersionsparameter sollte 1 sein. Bei Overdispersion ist er (deutlich) grösser als 1. Unter Nullhypothese, dass keine Overdispersion gibt, hat Devianz χ^2 Distr.

Mögliche Ursachen für Overdispersion

- **Falsche Strukturform** für Modell. **Nicht richtigen erklärenden Variablen** einbezogen oder haben **nicht transformiert** oder **nicht in der richtigen Weise kombiniert**.
- (Kleine Anzahl) **Ausreisser**
- **Zufallskomponente** Modell ist **unzureichend**. Binomialverteilung für Antwort Y ergibt sich, wenn **Erfolgswahr' $\boldsymbol{\pi}$ unabhängig** und für jeden Versuch **innerhalb Gruppe gleich ist**. Wenn Annahmen verletzt → kann Overdispersion sein

Berechnen von Vertrauensintervallen

- Beste Weg formale Schlussfolgerungen darzustellen und Beziehungen in Variablen aufzuzeigen, geschieht mit Konfint.
- «Standard»-Inferenzangaben basiert auf Tatsache, dass geschätzten Koeffizienten $\hat{\underline{\beta}}$ genähert Gauss-verteilt sind
- **Approx.** $100 \cdot (1-\alpha)$ % KI für Linearkombination $\psi = \underline{u}^T\underline{\beta}$ Koeffizienten bestimmen: $\underline{u}^T\hat{\underline{\beta}} \pm q \cdot se\langle \underline{u}^T\hat{\underline{\beta}} \rangle$ mit $se\langle \underline{u}^T\hat{\underline{\beta}} \rangle := \sqrt{\phi^* \underline{u}^T (\boldsymbol{X^T W X})^{-1} \underline{u}}$, q Quantil aus $\mathcal{N}$-/t-Vert. und $u = p$ dim. Vektor festen Komponenten. confint.default(**glm**) #Wald-Typ

Vertrauensintervall erwartete Response-Wert bei $\underline{x}_o$: Schätzung: $\hat{\mu}_o = g^{-1}\langle \underline{x}_o^T\hat{\underline{\beta}} \rangle$. Standardansatz, beruht auf Fehlerfortpflanzungsgesetz. $g^{-1}\langle \underline{x}_o^T\hat{\underline{\beta}} \rangle \pm q \cdot se\langle \hat{\mu}_o \rangle$ mit $se\langle \hat{\mu}_o \rangle := \sqrt{\phi^* \left(G'\langle \underline{x}_o^T\hat{\underline{\beta}} \rangle\right)^2 \underline{x}_o^T (\boldsymbol{X^T W X})^{-1} \underline{x}_o}$, q Quantil $\mathcal{N}$-/t-Verteilung

- **Approximation kann** sehr **unzuverlässig** sein → Vertrauensintervall teilweise ausserhalb Definitionsbereich von μ
- Resp <- predict(**fit.glm1**, se.fit = TRUE, type = "response"); Resp$fit + qnorm(**0.975**)*cbind(-Resp$se.fit, Resp$se.fit)
- Überzeugende praktische **Alternative** ist Vertrauensintervall für $\eta_o = \underline{x}_o^T\underline{\beta}$ zu bestimmen und dessen Endpunkte dann mit inversen Link g^{-1} rücktransformieren: $\underline{x}_o^T\hat{\underline{\beta}} \pm q \cdot se\langle \hat{\eta}_o \rangle$ mit $se\langle \hat{\eta}_o \rangle := \sqrt{\phi^* \underline{x}_o^T (\boldsymbol{X^T W X})^{-1} \underline{x}_o}$
- Transformation Endpunkte mit g^{-1} erhalten Vertrauensintervall für $\mu_o = g^{-1}\langle \eta_o \rangle = g^{-1}\langle \underline{x}_o^T\underline{\beta} \rangle$: $g^{-1}\langle \underline{x}_o^T\hat{\underline{\beta}} \pm q \cdot se\langle \hat{\eta}_o \rangle \rangle$

- pLink <- predict(**fit.glm1**, se.fit = TRUE, type = "link"); h <- pLink$fit + qnorm(**0.975**)*cbind(-pLink$se.fit, pLink$se.fit);
- 1/(1+exp(-h)) #Rücktransformation; family(**MAC1.glm1**)$linkinv(pLink$fit + qnorm(**0.975**) * outer(pLink$se.fit, c(-1, 1)))
- pLink <- predict(**Chal.glm**, newdata = data.frame(**Temp = 31**), type="link", se=T) #Rücktransformation wie oben

Devianz basierte Vertrauensintervalle (VI) – profiling

- Vorgestellten VI beruhen auf Wald-Asymptotik. Diese Näherungen sind nicht immer gut genug. **Besser Devianz-Test:**
- Ein VI für β_k ist verknüpft mit Test $\beta_k = \tilde{\beta}_k$. Sei $\hat{\underline{\beta}}$ die ML-Schätzung von $\underline{\beta}$ und $\hat{\underline{\beta}}^{*k}$ jene von $\underline{\beta}$, wobei β_k auf den Wert $\tilde{\beta}_k$ fixiert ist. Devianz-Teststatistik für $\beta_k = \tilde{\beta}_k$ ist dann: $T_D = \frac{D\langle \underline{y}; \hat{\underline{\beta}}^{*k} \rangle - D\langle \underline{y}; \hat{\underline{\beta}} \rangle}{\hat{\phi}} \overset{a}{\sim} \chi_1^2 \to \tilde{\beta}_k | T_D = \frac{D\langle \underline{y}; \hat{\underline{\beta}}^{*k} \rangle - D\langle \underline{y}; \hat{\underline{\beta}} \rangle}{\hat{\phi}} \leq q_{1-\alpha}^{\chi_1^2}$ ist KI
- Falls ϕ tatsächlich geschätzt werden muss, wird in Praxis χ^2-Verteilung durch entsprechende F-Verteilung ersetzt.
- Nachteil: Berechnung i.A. aufwändiger als Waldtyp VI. Vorteil: Devianz oft gut, auch kleine Stichprobengrösse n
- confint(**fit.glm**) #Bemerkung im Output: Waiting for profiling to be done... → glm(…, maxit = **100**) #falls nicht konvergiert

Binäre Regression – Prognose? Zielvariable kann nur Werte 0 und 1 annehmen!

- Schätzung/Prognose für μ: $\hat{\mu}_o = g^{-1}\langle \underline{x}_o^T\hat{\underline{\beta}} \rangle$, approx VI für erwarteten Responsewert $\underline{x}_o$: $g^{-1}\langle \underline{x}_o^T\hat{\underline{\beta}} \pm q_{1-\alpha/2}^{\mathcal{N}} \cdot se\langle \hat{\eta}_o; \phi = 1 \rangle \rangle$
- Weiteres Problem **class imbalance**, macht Over-/Undersamling. Oversamling sind Gift für Schätzung Wahrscheinlichkeiten μ, weil sie systematisch verzogenen Schätzungen führt. Vielfach reicht bei imbalance klassische ML-Schätzung.
- Falls Stichprobengrösse n sehr klein und imbalance extrem ist, dann alternative Schätzung verwenden

Variablenselektion

- Basiert Informationstheorie, Akaike Information Criterion: $AIC = -2(\max LogLikelihood) + 2 \cdot \#geschätzter\ Parameter$
- GLM AIC durch Devianzen ausdrücken: $AIC = -2 \cdot \ell\langle \hat{\mu}, \hat{\phi} \rangle + 2 \cdot \#geschätzter\ Parameter$, $-2 \cdot \ell\langle \hat{\mu}, \hat{\phi} \rangle = \frac{D}{\hat{\phi}} - 2 \cdot \ell\langle \underline{y}, \hat{\phi} \rangle$
- AIC/BIC? führen ähnlichen Entscheidungen. BIC bestraft jedoch grössere Modelle härter → **nicht** zwingend **Optimum**
- BIC: wenn verstehen will, welche Prädiktoren Beitrag leisten und wenn schlankes, gut interpretierbares Modell will.
- AIC: wenn Modell für Vorhersage von Erw.werten eingesetzt werden soll, verwendeten Prädiktoren weniger zentral sind.
- step(**grosses.glm**, *scope = list(lower = ~1, upper = formula(**grosses.glm**))*) #Default=both|AIC, kleinster Wert = bestes
- step(**grosses.glm**, k = log(nrow(**Datensatz**)), *direction = "backward"*) #für BIC
- Vorwärts: M0 <- glm(**Y ~ 1**, data = **DF**); step(M0, scope = list(lower = ~ 1, upper = ~ **x1+x2+x3+x4**), direction = "forward")
- Rückwärts: MF <- glm(**Y ~ x1 + x2 + x3 + x4**, data = **DF**); step(MF, direction = "backward")
- unter AIC Wert steht finale Modell in R Notation. Variablen mit Minus davor sind enthalten #direction = "both"

Prüfen Modellgüte von GLMs: Sollten Modellannahmen überprüfen bevor GLM als aussagekräftig akzeptieren und es verwenden!

- Überprüfung beruhen in erster Linie auf angepassten Werten und Residuen. **Ziel Residuenanalyse? Aufdecken systematische Abweichung Daten vom Modell und Identifikation ungenügend beschriebener Bereiche / Ausreisser**
- 4 Definition für GLM Residuum, die alle im Fall Normalverteilung identisch mit uns bekannten Definition Residuums sind.
- Einfachste und naheliebendste Definition Residuums ist **Response Residuen**: $R_i \coloneqq Y_i - \hat{\mu}_i$, R_i sind i.A. nicht normalverteilt und ihre Varianz ist in erster Näherung $\phi V\langle\mu_i\rangle(1 - H_{ii})/w_i$, Varianzfunktion $V\langle\mu_i\rangle$ ist i.A. nicht konstant
- Somit Definition sinnvoll **Pearson-Residuen**: $R_i^{(P)} \coloneqq R_i \cdot w_i/\sqrt{V\langle\hat{\mu}_i\rangle}$ → ungefähr konstant Varianz, i.A. nicht normalvertlt.
- Basierend auf IRLS: **Working Residuen**: $R_i^{(W)} \coloneqq R_i g'\langle\hat{\mu}_i\rangle$. → gewichtete Residuen; $\sqrt{\widetilde{w}_i}R_i^{(W)} = \sqrt{\frac{w_i}{V(\hat{\mu}_i)(g'(\hat{\mu}_i))^2}}(Y_i - \hat{\mu}_i)g'\langle\hat{\mu}_i\rangle \equiv R_i^{(P)}$
- Eigenschaft klassisch Residuen: **Unterschiede zwischen beobachteten und angepassten Responsewerten** messen
- Summe quadrierten Residuen zur Schätzung σ^2 verwenden und Unterschiede Summen quadrierten Residuen versch. Modellen zu Modellvergleich einsetzen → $R_i^{(D)} \coloneqq sign\langle y_i - \hat{\mu}_i\rangle\sqrt{d_i}$, d_i =Summanden Residuen-Devianz.
- Wie Erfahrung zeigt, sind Unterschiede zwischen Devianz- und Pearson-Residuen üblicherweise sehr klein.

Tukey-Anscombe-Diagramm (Residuen-Plot): Pearson-Residuen $R_i^{(P)}$ gegen angepassten lineare Prädiktorwerte $\hat{\eta}_i$

- Wird verwendet, um auf **nicht erwartete Nichtlinearitäten** im linearen Prädiktor (oder allenfalls auf **unpassende Link-Funktionen**) und nebenbei auf **Ausreisser** aufmerksam zu werden.
- Abweichungen sichtbarer machen, Glätter hineinlegen, der bei passendem Modell um horizontale Null-Achse schwankt
- Beurteilung kann herausfordern, da bei nicht normalverteilten Zielvariablen irritierende Artefakte auftreten können.

Scale-Location-Diagramm: Wurzel absoluten Werte standardisierten Residuen gegen angepassten linearen Prädiktor.

- Untersucht **konstante Varianz** entlang angepassten Werte und relevanten Strukturen mit Glätter herausfiltern.

Half-Normal QQ-Plot: Indirekt zu prüfen, ob Verteilungsannahme zutrifft

- Falls **Verteilungsannahme** für Daten zutrifft, müssen Devianz-Residuen genähert normalverteilt → um Gerade streuen
- Fokus Abweichungen (Ausreisser rechts im Plot). Deshalb wird half-normal QQ-Plots verwendet.
- Half-normal QQ-Plot werden geordneten absoluten Devianz-Residuen gegen entsprechenden Quantile der Standardnormalverteilung aufgetragen, d.h. gegen die $\frac{n+i+\frac{1}{2}}{2\cdot n+\frac{9}{8}}$-Quantile der Standardnormalverteilung $(i = 1,2,\dots,n)$
- Devianz-Residuen nur genähert normvertlt. Möglich, trotz Verteilungsannahme «korrekt», keine lin. Struktur sichtbar ist!
- Darstellung nützlich um festzustellen, ob allfällige Overdispersion auf kleine Anzahl Ausreisser zurückzuführen ist.
- Bei binären ZV ist Normal-QQ-Plot oft bedeutungslos. Selbst wenn Residuen nicht normalverteilt, sieht Plot gut aus.

- source("ARM/RFn_Plot-glmSim.R"); par(mfrow = c(2, 4)); plot(**turb.glm**); plot.glmSim(**turb.glm**, *SEED = 184*)

Bootstap-Simulation: Idee: Erzeugen Beobachtungen y_i^* gem. Modell unter Verwendung neuer Zufallszahlen. Erzeugt n Zufallszahlen y_i^* entsprechend der verwendeten Verteilungsfamilie mit Erwartungswert $\hat{\mu}_i$ und Streuung $\hat{\phi}$.

- Berechnen GLM-Anpassung mit erkl. Var. aus Datensatz und neu generierten Responsewerten y_i^*. Anschliessend wird glatte Linie TA-Diagramm bestimmt und graue Linie eingezeichnet. Wiederholen diese beiden Schritte $n_{rep} = 19$ mal

- Gibt Fälle, wo rote Glätter starke nichtlineare Strukturen zeigt, aber noch innerhalb grauen Bootstrap-Simulationen liegt.
- Andererseits überdecken Bootstrap-Simulationen jedoch die horizontale Nulllinie nicht! Was heisst das jetzt?
- Können daraus schliessen, dass es **keine** Evidenz gibt, dass Missspezifikation im linearen Prädiktor vorliegt.

Residual gegen Leverage — zu einflussreiche Beobachtungen → Ausreisser und/oder Hebelpunkt

- Messen Hebelwirkung einer Beobachtung mittels der Diagonalelemente H_{ii} der GLM-Hutmatrix
- Beobachtung wird als Leverage Punkt bezeichnet, wenn $H_{ii} > 2p/n$ ist, p Anzahl der Koeffizienten ist.
- Cook's distance misst wieder Einfluss einzelner Beobachtungen und ist definiert als $d_i^{\odot} \coloneqq \frac{1}{p}\frac{R_i^2}{\hat{\phi}V(\hat{\mu}_i)}\frac{H_{ii}}{(1-H_{ii})}$, $1 \le i \le n$
- **Faustregel**: Beobachtungen mit einer Cook's Distance $d_i^{\odot}$ **grösser als 1** gelten als zu einflussreich

(Zeitliche) Korrelation in den Residuen: Residuen z. B. gegen Reihenfolge oder räumliche Anordnung Versuchseinheiten
Alternative: Streudiagramm Paare $(R_{t-1}^{(P)}, R_t^{(P)})$ plot(**fit**, type = "h") #Falls Beob. unab. streuen Punkte wild (weiss Rauschen)

Partielle Residuen-Plots (term plot). Ziel: Entdecken Nichtlinearitäten in erklärenden Grössen

- Auf x-Achse erkl. Variable $x_i^{(k)}$, auf y-Achse Pseudo-Zielvariable minus «Effekt» lin. Prädiktors ohne Komponente $x_i^{(k)}$:
- $z_i - \left(x_i^T\hat{\beta} - \hat{\beta}_k x_i^{(k)}\right) = R_i^{(W)} + \hat{\beta}_k x_i^{(k)}$ → Falls Modell richtig, sollten dargestellten Punkte um Gerade streuen.

- **Residual Plot:** Da der rote Glätter klar inner-/ausserhalb stochastischen Fluktuation (graue Spaghetti) liegt, gibt es keine/starke Evidenz, dass Einfluss der erklärenden Variablen auf die Erwartungswerte falsch spezifiziert wurde.
- **Half-Normal QQ-Plot:** Da schwarzen Punkte inner-/ausserhalb stochastischen Fluktuation (graue Punkte) liegen, gibt es keine/klare Evidenz, dass die Verteilung inadäquat spezifiziert ist und auch noch Hinweise auf zwei Ausreisser Wegen Overdispersion haben grauen Punkte eine andere Steigung als schwarzen.
  Scale-Location-Plot: Da rote Glätter vollständig inner-/ausserhalb der stochastischen Fluktuation liegt (graue Spaghetti), gibt es keine/klare Evidenz, dass Varianz inadäquat spezifiziert wurde. Wegen Overdispersion liegen grauen Spaghettis auch viel tiefer als rote Glätter!
  a <- 2*length(coef(**glm**))/nrow(**data**); abline(v=a, col=3)
  Y: Ausreiser, X: Hebelpunkte
- **Residuen vs. Leverage:** Keine/Mehrere Beobachtungen haben Cook's Distance grösser als 1 und gibt/sind deshalb (keine) zu einflussreich(en Beoba). Bei den einen ist es bedingt durch Ausreisser, bei den anderen durch Hebelarme.
- **Fazit:** Alle Plots zeigen (keine)klare Evidenz, dass Modell Daten (nicht) adäquat beschreiben kann (Inferenz wertlos)

Behandlung von Unzulänglichkeiten: J.W. Tukey nannte sie First Aid Transformations

- Logarithmus-Transformation für Konzentrationen/Beträge,
- Wurzeltransformation für Zähldaten
- Logit für Anteile (Prozentzahlen/100): $\hat{y} = \boldsymbol{log}\langle\frac{y+0.005}{1.01-y}\rangle$
- Arcus-Sinus-Wurzel $\hat{y} = \boldsymbol{arcsin}\langle\sqrt{y}\rangle$
- Nicht anwenden Zeitvariablen/zu wenig Streuung
- First-Aid/Log Faktor 10

Generalisierte additive Modelle (GAM)

- GAM erweitert GLM indem einige/alle linearen Terme linearen Prädiktor durch **glatte Funktionen** ersetzen: $\eta_i = \beta_0 + \sum_{k=1}^m x_i^{(k)}\beta_k$ → $\eta_i = \sum_{k=1}^m f_k\langle x_i^{(k)}\rangle$. Analyse grafisch mit geschätzten Funktionen: Glatte Kurve partiellen Residuen-Plot
- Schätzen unbekannte Funktionen **Backfitting-Algorithmus** → Zielvariable Y_i durch Pseudo-ZV ersetzen und gewichtete Glättung wie **IRLS**-Algorithmus durchführen. GAM gut geeignet, da Transformation **nicht-parametrisch** schätzen.
- Hauptzweck für Einsatz von GAM ist **Aufdeckung** geeigneter **Transformationen** bei erklärenden Variablen im GLM.

library(gam); **baby**.gam <- gam(**Survival** ~ lo(**Weight**) + lo(**Age**) + lo(**pH**), family = **binomial**, *bf.maxit=100*, data = **baby**)
par(mfrow=c(**2, 3**)) #2*3=Anz.Erkl.Var.; plot(**baby**.gam, se = T, residuals = T) ↓ Bessere Darstellung für Variable EXTRP
lo() geht für Faktorvariablen nicht, diese regulär ohne lo() Modell nehmen. plot(**b**.gam, residuals = T, terms = "lo(**EXTRP**)")
Faustregel 1: Gerade passt (nicht) zwischen gestrichelte Vertrauensband, gem. Faustregel keine Transformation nötig.
Faustregel 2: summary(**baby**.gam) #Anova bei **Nonparametric**: wenn p significant, dann transformieren

baby$tWeight <- ifelse(**baby$Weight < 1000**, 0, **baby$Weight-1000**) #bis zu einem gewissen Schwellenwert Prädiktor
baby$tpH <- ifelse(**baby$pH < 7.27**, 0, **baby$pH-7.27**) #beibehalten und danach diesen transformieren, Hockey Stick
glm4 <- glm(**Survival ~ Weight + tWeight + Age + Apgar1 + Apgar5 + pH + tpH**, family = binomial, data = **baby**)
DaR$HRtr <- ifelse(**DaR$HR < 12.5**, **DaR$HR**, **12**) #Mehrfacher Hockeystick mit HR → HRtr und HRhigh
DaR$HRhigh <- as.integer(**DaR$HR > 19.5**); **DaR$fF** <- as.factor(cut(**DaR$F**, breaks=c(**0,0.1,0.4,0.7,1**))) #kat. Var.
DaR.qP <- glm(**RDR ~ sVH + sPOP + IAR + HRtr + HRhigh + fF + IND**, data = **DaR**, family = quasipoisson)

Modelle für Raten (Rate Models – Poisson-Raten-Regression) → Term $log\langle ex_i\rangle$ heisst **offset** und ist bekannt.

- Wichtige Annahme, dass **sowohl** die **Intensität** (oder Rate) des **Auftretens** von Ereignissen **und** die **Gelegenheit** (oder Exposition) zur Zählung **konstant** sind bei den gleichen erklärenden Variablen für alle entsprechenden Beobachtungen.
- Wenn jedoch z.B. Anzahl Autounfälle pro Jahr modelliert werden, kann sinnvoll sein, dass Unfallrisiko und damit Rate von gefahrenen Kilometern pro Jahr abhängt → **Unfallrate pro gefahrener Kilometer und ZEIT modellieren**
- Rate ρ definiert als erwartete Zählungen pro Expositionseinheit. z.B. Unfälle pro 1'000 gefahrene Km, Festplattenausfälle pro Betriebsstunde. D.h., $\rho \coloneqq \mu/ex$, ex ist **Exposition** und μ erwartete Anzahl innerhalb einer Exposition
- $log\langle\rho_i\rangle \coloneqq log\langle\frac{\mu_i}{ex_i}\rangle = \beta_0 + \beta_1 x^{(1)} + \dots + \beta_m x^{(m)}$ | $log\langle\mu_i\rangle \coloneqq log\langle ex_i\rangle + \beta_0 + \dots + \beta_m x^{(m)} \to \mu_i = ex_i \cdot \exp\langle\beta_0 + \beta_1 x^{(1)} + \dots\rangle$

Ähnlichkeit zwischen Raten und Proportionen im Binomialmodell: (Anzahl Ereignisse) / (Gelegenheiten für Ereignisse)

- Die zwei Modellansätze sind jedoch unterschiedlich. Kann sie **unterscheiden**: Ist es denkbar, dass jede **Expositionseinheit mehr** als **ein Ereignis** erzeugt, oder muss **jede Expositionseinheit** genau **0 oder 1 Ereignis hervorrufen**?
- Falls mehr als ein Ereignis in einer Expositionseinheit möglich → Poisson-Raten besser, weil Annahmen Binom. verletzt
- Z.B. gibt es nichts, was verhindert, dass in 1'000 gefahrenen Kilometern mehrere Unfälle passieren können
- Andererseits wird jedes Baby eine Frühgeburt überleben oder nicht. Dafür sollte Binomialmodell verwendet werden.

- Parameter $\beta_0, \beta_1, \dots, \beta_m$ werden wie zuvor mit **Maximum-Likelihood-Methode** geschätzt.
- Ihre **Interpretation** erfolgt jedoch in **Bezug auf die Rate** und **nicht** auf den **Erwartungswert**.
- β_0 ist Ereignisrate, wenn wir alle erkl. Var. auf 0 setzen, β_1 ist Änderung der Ereignisrate pro Einheit Anstieg von $x_i^{(1)}$
- Die Masseinheit für die Rate ist Zählungen pro Expositionseinheit (z.B. Unfälle pro 1'000 gefahrene Kilometer).

glm1 <- glm(**Y ~ offset(lService) + type + period + year**, family = poisson, data = **Ships**) #offset() für Raten-Modelle

Quasi-Modelle – Quasi-Poissonmodell

- Bei Zähldaten kann sein, dass Poisson-Modell die Daten nicht adäquat beschreibt, weil Dispersion zu gross ist.
- Alternative wo Erwartungswert-Varianz-Beziehung Form $var\langle Y_i\rangle = \phi \cdot \mu_i$ für unbekannte Konstante ϕ. ← Fall $\phi > 1$
- Keine ML-Schätzung, da zu Schätzgleichung keine Likelihood definiert werden kann. Schätzungen asymptotisch normalverteilt. Können keine likelihood Methoden anwenden, wie Devianz-Test, Profiling, AIC. Bootstrap/GAM funktioniert
- Kovarianzmatrix zu Koeffizienten $\hat{\beta}$ wird geschätzt als $\hat{\phi}$-fache der Kovarianzmatrix beim Poisson-Regressionsmodell.

qPois <- glm(**RDR ~ sVH + sPOP + IAR + HR + F + IND**, family = quasipoisson, data = **DaR1**)

GLM — Polytome Zielvariable → für kategoriale Variablen

Kreuz-/Kontingenztabelle & Mosaic-Plot → mosaicplot(table(**wafer$method**, **wafer$quality**))

- **Mosaic-Plots:** Fläche der Rechtecke proportional zur Anzahl Beobachtungen für Merkmalskombination in Stichprobe. Säulenbreite proportional zur relativen Häufigkeit der Stufen in ersten Variable. Höhe Rechtecke in jeder Säule proportional zur relativen Häufigkeit der Stufen aus zweiten Variablen. Anzahl Beobachtungen ist nicht ablesbar.

Multinomiale Logit-Modell für nominale Zielvariablen → Zielvariable kategoriale Variable mit **mehr** als **zwei Stufen**

- Grösse von Interesse: **Wahrscheinlichkeit**, dass Y_i eine bestimmte **Ausprägung** ℓ annimmt, d.h. $\pi_{i\ell} = P(Y_i = \ell)$
- Ziel Beziehung zwischen $\pi_{i\ell}$ und erklärenden Variablen $x^{(k)}$, $k = 1,\dots,m$ finden, wobei garantiert sein soll, dass **Wahrscheinlichkeiten** zwischen **0 und 1** liegen **und Summe** $\pi_{i\ell}$ gleich **1** ist. Modellierung **ähnliche** Idee wie **Logit-Modell**.
- **Schritt 1:** Betrachten logarithmierte Wettverhältnis gegenüber Referenzkategorie (üblich = 1): $\log\langle\frac{P(Y_i=\ell|x_i)}{P(Y_i=1|x_i)}\rangle = \log\langle\frac{\pi_i^{(\ell)}}{\pi_i^{(1)}}\rangle$
- **Schritt 2:** Setzen für jedes log. Wettverhältnis $\ell = 2,\dots,L$ sep. lin. Modell: $\log\langle\frac{\pi_i^{(\ell)}}{\pi_i^{(1)}}\rangle = \eta_i^{(\ell)} = \beta_0^{(\ell)} + \beta_1^{(\ell)}x_i^{(1)} + \dots + x_i^{(m)}\beta_m^{(\ell)}$
- Wahr' modelliert: $\pi_i^{(\ell)} = \exp\langle\eta_i^{(\ell)}\rangle \cdot \pi_i^{(1)}, \ell > 1$ und $\pi_i^{(1)} = \frac{1}{1+\sum_{\ell=2}^{L}\exp(\eta_i^{(\ell)})}$, Referenzkategorie sorgt **Summe** $\pi_{i\ell}$ = **1**
- Für jede Kategorie ℓ der Zielgrösse Y_i wird eigene Abhängigkeit der Wahrscheinlichkeit $\pi_i^{(\ell)}$ von erkl. Variable geschätzt.
- Für zunehmendes $x^{(k)}$ bedeutet positiver Koeffizient $\beta_k^{(\ell)}$ steigende Neigung zur Kategorie ℓ im **Vergleich zur Referenzkategorie**. Für jede Kategorie **separates Modell → viele Parameter**: Gleichungssystem $(L-1)(m+1)$ Parameter
- Sollten genügend Beobachtungen für Anzahl geschätzter Koeffizienten haben: (min. 5-10 Datenpunkte pro Koeffizient!)
- **Wahl Referenzkategorie keine Auswirkung auf Modell** • Beobachtungen können gruppiert sein → Kreuztabellen.

Nicht gruppierte Daten: library(nnet); library(MASS); **waver**.fitsm <- multinom(**quality** ~ **method**, data = **wafer**)
R nimmt jeweils erste Faktorstufe als Referenz. Diese ist dann im summary(**waver**.fitsm) auch nicht direkt ersichtlich.
Bei Gruppierung: **wG**.fitsm <- multinom(cbind(**excellent**, **ok**, **bad**) ~ **method**, data = **waverG**); summary(**wG**.fitsm)
Zielvariable als Matrix übergeben, Faktorstufen/Levels explizit angeben, hier bspw. excellent, etc. Outputs sind identisch!

Interpretation Koeffizient: Wert Koeffizient ist immer im Vergleich zur Referenzausprägung! Log-Odds von ok zu excellent nehmen für Methode2 im Vergleich zu Methode1 um 0.734 ab. Objekte Methode2 haben im Vergleich zu Methode1 eine geringere Neigung ok als excellent zu sein

```
Coefficients:
     (Intercept)  methodmethod2  methodmethod3
ok     1.4028282     -0.7337760     -1.729224
bad    0.5877936     -0.4542601     -1.866200
```

Inferenz: Ob erklärende Grösse Einfluss auf Zielgrösse hat, sollte nicht anhand Standardfehler bestimmen (da L Koeffizienten null sein müssten, wenn kein Einfluss da ist)! → hierarchischer Modellvergleich via Deviazen mit anova()

- testen, ob an allen Standorten gleiche Verteilung verwendet werden könnte (**Homogenitätshypothese**)
- drop1() geht für multinomiale Modelle nicht. Alternativ library(MASS); dropterm(**waver**.fitsm, test = "Chisq")
- **Gleiche Resultat mit Devianztest:** **w1**.fithm <- multinom(**quality** ~ 1, data = **wafer**); anova(**waver**.fitsm, **w1**.fithm)
- Modellvergleich zeigt, beide Modelle signifikant verschieden → Homogenitätshypothese wird auf 5% Niveau verworfen, da P-Wert 4.9e-13 kleiner als 5 %. Saturierte Modell ist also notwendig, um Daten adäquat zu beschreiben. Folglich haben **Methoden** statistisch signifikanten Unterschied auf **Qualität!** (**Methode / Qualität auf erkl. / ZV anpassen!**)

Vorhersage (Prediction): Klassenzugehörigkeit vorhersagen: predict(**waver**.fitsm, newdata = w0, type = "class") ↓ besser w0 <- data.frame(method = c("**method1**", "**method2**", "**method3**")); predict(**waver**.fitsm, newdata = w0, type = "probs")

Modell-Diagnostik: Gibt keine aussagekräftige Residuen Definition. **Keine** Prüfung, ob **Modell Daten adäquat beschreibt**

Modelle ordinale Zielvariable **wafer$quality** <- factor(**wafer$quality**, level = c("**bad**", "**ok**", "**excellent**"), ordered=T)

- Ordnung in Stufen der Zielvariable berücksichtigen. Ordnen Stufen in Zielvariable **natürlich aufsteigende Ordnung**.
- Modell, wie bei logistischen Regression mit Annahme latenten kontinuierlichen Variable T. In Realität kann nur Y_i beobachtet werden, die sich als diskretisierte Version von T_i ergibt mit Schwellenwerten: $\tilde{\alpha}_0 = -\infty < \tilde{\alpha}_1 < \cdots < \tilde{\alpha}_{L-1} = \infty$
- **Schwellenwerte $\tilde{\alpha}$ müssen nicht äquidistant sein** und sind i.d.R. nicht a priori bekannt → mit Daten schätzen
- Weniger Parameter zum Schätzen als bei multinom. Regr. **Residuenanalyse:** Keine, wie bei multinom Logit Modell
- Modell: $\tau_\ell := P(Y_i \geq \ell | \underline{x_i}) = P(T_i > \alpha_\ell | \underline{x_i}) = 1 - \mathcal{F}\left(\alpha_\ell - (\tilde{\beta}_0 + \tilde{\beta}_1 x_i^{(1)} + \cdots + \tilde{\beta}_m x_i^{(m)})\right)$, $\tilde{\beta}_0$ ist unbestimmt, jedoch $\tilde{\beta}_0 = 0$

library(MASS); **w**.polr <- polr(**quality** ~ **method**, data = **wafer**); summary(**w**.polr) #Faktoren mit ordered = T übergeben!

Statistische Wartezeitanalyse (Analysis of Time-to-Event Data, Event History Analysis, Failure Time Analysis)

- Werden zensierte Beobachtungen weggelassen/ignoriert, kann **systematische Fehler** in **geschätzten Grössen** geben.
- Was wollen wir mit der Analyse von Wartezeitdaten erreichen? Zwei wichtige Modellierungskonzepte:
  - Wahrscheinlichkeit, dass Ereignis **später** als zum **Zeitpunkt** t eintritt → **Überlebensfunktion**
  - Möglichkeit, dass ein Ereignis **zu** einem bestimmten **Zeitpunkt** t eintritt → **Hazard-Rate**
- **(Warte-)zeit interessante Variable**, modelliert $T \geq 0$ reelwertige Zufallsvariable Zeit bis Ereigniseintreten. Zwei Fälle:
  - T **genau gemessen** → T ist **kontinuierlich** ○ nur **bestimmt Zeitpunkt** erfasst (zweimal im Jahr) → T **diskret**
- **Typ I Zensierung:** Experimentstart gleichzeitig. Zeitpunkt t_z Abbruch. Komponenten ohne Ausfall sind zensiert.
- **Typ II Zensierung:** Gleich wie Typ I, jedoch Abbruch, sobald r der insgesamt n Objekte ausgefallen sind.
- **Zufällige Zensierung:** Oft gleichzeitiger Start unmöglich. Zensierung erfolgt, weil Patient weggezogen (**Loss to Follow-up**); Behandlung wegen Nebeneffekt oder sonst abgebrochen (**Drop Out**); oder Studie beendet (**Study Termination**).
- **Beschreibung Wartezeitverteilung:** stetige Verteilungsfun. $\mathcal{F}$. Dazugeh. Dichtefun. f. → $P\langle T \leq t\rangle = \mathcal{F}\langle t\rangle = \int_0^t f\langle x\rangle\, dx$
- **Survivorfunktion** S (Zuverlässigkeitsfunktion): Wahrscheinlichkeit, dass Wartezeit mindestens t beträgt: $S\langle t\rangle := P(T > t) = 1 - \mathcal{F}\langle t\rangle = \int_t^\infty f\langle x\rangle\, dx$ $S\langle t\rangle$ monoton fallend, weil $\mathcal{F}$ monoton steigend ist; also $S\langle 0\rangle = 1$ und $S\langle\infty\rangle = 0$.
- $h\langle t\rangle := \lim_{\Delta t \to 0^+} \frac{P(t \leq T \leq t+\Delta t | T \geq t)}{\Delta t} = \lim_{\Delta t \to 0^+} \frac{P(t \leq T \leq t+\Delta t)}{\Delta t} \frac{1}{P(T \geq t)} = \frac{f\langle t\rangle}{S\langle t\rangle}$ → **Hazardrate (Ausfall-/Sterberate) ist keine Wahrscheinlichkeit**. Tritt Ereignis nicht vor Zeitpunkt t ein, so informiert die Hazardrate über «weiteren Verlauf»:
  - Ist konstant → Objekte «altern» & «verjüngen» nicht ○ Steigt Ausfallrate → Objekte «altern», Verschleiss
  - Abfallende Ausfallrate (selten) → Ausfall, wenn er eintritt, dann meistens früh
- Hazardrate bed. Wahr' Ereigniseintritt in $[a_{t-1}, a_t)$, Bedingung Erreichen Intervall: $\lambda_t := P\langle T \in [a_{t-1}, a_t) | T \geq a_{t-1}\rangle$

Schätzung Hazardrate und Survivorfunktion

Sterbetafel (life-time table) → Eintreten Ereignis beobachtet ($z_i = 1$) oder **nicht beobachtet → zensiert** ($z_i = 0$) ist

- Einfachste, gebräuchlichste Methode **diskreten Wartezeiten**. **Einfluss erkl. Merkmale** explizit **nicht berücksichtigt**.
  - Zeitintervalle t ○ Anzahl Komponenten, die im Zeitintervall dem Risiko unterliegen, dass sie ausfallen können
  - Anzahl Komponenten, die ausgefallen sind ○ Anzahl Komponenten, die zensierte Beobachtungszeit haben.
- n Gesamtzahl gefährdeten Objekte/Subjekte zu Beginn • d_t Anzahl Ereignisse, die t-ten Intervall $[a_{t-1}, a_t)$ eintreffen
- w_t Anzahl Zensierung Intervall t $[a_{t-1}, a_t)$ → erreichen t-te Intervall, aber weder Ereigniseintritt noch nächste Intervall
- n_t Anzahl gefährdeten Objekte, bei denen Ereignis im t-ten Intervall $[a_{t-1}, a_t)$ eintreten könnte → Risikomenge «at risk»
- Schätzen aus Sterbetafeln: Ohne Zensierung $\hat{\lambda}_t = \frac{d_t}{n_t}$, mit Zensierung (Wann gezählt?): Ende t-ten Intervall $\hat{\lambda}_t = \frac{d_t}{n_t}$, Anfang t-ten Intervall $\hat{\lambda}_t = \frac{d_t}{n_t - w_t}$ → übliche Schätzverfahren t $\hat{\lambda}_t = \frac{d_t}{n_t - w_t/2}$. Schätzung Survivorfunktion $\hat{S}\langle t\rangle = \prod_{s=1}^{t}(1 - \hat{\lambda}_s)$

LDT <- data.frame(**Time** = c(**6.8, 17.6, 17.6, …**), **status** = c(**1, 1, 0, …**)) h <- **LDT**$Tdiscrete <- ceiling((**LDT$Time/24**))
dt <- table(factor(h[**LDT$status** == 1], levels = 1:**6**)) #6 = Anz. Intervalle; wt <- table(factor(h[**LDT$status** == 0], levels=1:**6**))
et <- c(0, cumsum(dt[-length(dt)] + wt[-length(wt)])); nt <- nrow(**LDT**) - et; hazard <- dt / (nt - wt/2); S <- cumprod(1 - hazard)

library(discSurv); **RR** <- read.table("**Rossi**.txt", …); **RR4** <- data.frame(**week**=(**RR$week**-1)%/%**4** + 1, **arrest**=**RR$arrest**)
h.lab1 <- paste("**W**",seq(**1,49**,by=**4**), sep=""); h.lab2 <- paste("**W**",seq(**4,52**,by=**4**), sep="") #auf je vier Wochen aggregieren
h.lab <- paste("[",h.lab1,",", h.lab2,"]", sep="") # ↓ Spalte "dropouts" ≡ w_t ; Spalte "atRisk" ≡ Nenner $(n_t - w_t/2)$
IT <- lifeTable(dataShort = **RR4**, timeColumn = "**week**", eventColumn = "**arrest**", intervalLimits = **h.lab**); IT$Output

Verteilung Survivorfunktions-Schätzung: Keine zensierte Beobachtung ($w_t = 0$) : $\hat{S}_t = \frac{n_t}{n} = \frac{n - d_1 - d_2 - \cdots - d_{t-1}}{n}$. Anz. Ereignis, die bis Zeitpunkt a_t noch nicht eingetroffen, ist binomialverteilt, $N_t \sim \mathcal{B}(n, S_t)$. Folglich gilt $\mathbb{E}\langle\hat{S}_t\rangle = S_t$, $Var\langle\hat{S}_t\rangle = \frac{S_t(1-S_t)}{n}$

- Genähertes 95%–Vertrauensintervall: $\hat{S}\langle t\rangle \pm 2 \cdot \sqrt{\frac{\hat{S}(t)(1-\hat{S}(t))}{n}}$, genähertes 95%-VI Hazardfunktion: $\hat{\lambda}_t \pm 2 \cdot \sqrt{\frac{\hat{\lambda}_t(1-\hat{\lambda}_t)}{n_t}}$, $\hat{\lambda}_t = \frac{d_t}{n_t}$
- Falls **zensierte Beobachtungen** ($w_t \geq 0$): Schätzer $\hat{\lambda}_t = \frac{d_t}{n_t - w_t/2}$ ist **nicht konsistenter Schätzer**.
- Verzerrung ignoriert und Varianzschätzung analog Schätzer korrigiert: $\hat{\lambda}_t \pm 2 \cdot \sqrt{\frac{\hat{\lambda}_t(1-\hat{\lambda}_t)}{n_t - w_k/2}}$, genähertes, **konservativ** 95%-VI
- Varianzschätzung Suvivorfunktionsschätzung oft **Greenwoods** Formel $Var\langle\hat{S}\langle t\rangle\rangle = \hat{S}\langle t\rangle^2 \sum_{k=1}^{t} \frac{\hat{\lambda}_k}{(1-\hat{\lambda}_k)\left(n_k - \frac{w_k}{2}\right)}$

- **Geglättete Hazardrate:** Schätzung Hazardfunktion nach Sterbetafel kann sehr volatil sein vor allem in einzelnen Zeitintervallen, in denen Risikomenge n_t klein wird. Dies ist vor allem bei grossen t der Fall.
- Da es sich bei **gewissen** Daten eigentlich um GLM handelt mit erkl. Var. «Nummer Zeitintervall», können GAM einsetzen, um geglättete Hazardrate zu bestimmen. Glätten mit Hilfe gam hilft bei vielen Zeitintervallen weiter.

Bestimmung wöchentlichen Hazardfunktion mit Sterbetafelmethode und mit geglätteter Sterbetafelmethode GAM
IT<-lifeTable(dataShort = **RR**, timeColumn = "**week**", eventColumn = "**arrest**"); h <- IT$Output; h$tl<-1:nrow(h)
library(mgcv); h.gam <- gam(cbind(events, round(atRisk) - events) ~ s(tl), family = binomial, gamma = 0.7, data = h)
h.gam.p <- predict(h.gam, type = "response", se = TRUE); h <- data.frame(h, sfit = h.gam.p$fit, se.sfit = h.gam.p$se.fit);
plot(h$tl[], h$hazard, type = "l"); lines(htl, hhazard+2*h$seHazard, lty = 2); lines(htl, hhazard-2*h$seHazard, lty = 2)
lines(htl, hsfit, col="red"); lines(htl, hsfit+2*h$se.sfit, col="red", lty=2) #GAM-Fit, noch mit -2*

Vorhersage auf Link-Ebene (besser): gam.link <- predict(h.gam, type = "link", se = T); h1.fit <- gam.link$fit
h1.se <- gam.link$se.fit; h1 <- data.frame(fit=h1.fit, lci=h1.fit - 2*h1.se, uci= h1.fit + 2*h1.se); h1.trans <- exp(h1)/(1+exp(h1)) #Rücktrans lines(h$tl, h1.trans$fit); lines(h $tl, h1.trans$lci); lines(h $tl, h1.trans$uci)

Kaplan-Meier-Schätzung Survivorfunktion für stetige Wartezeit

- **Keine Zensierung:** Schätzung $S\langle t\rangle$ wie emp. Verteilungsfun. (stetigen) Messgrösse: $\hat{S}_n\langle t\rangle = \frac{\#\,Beobachtungen > t}{n} = \frac{\#\{t_i > t\}}{n}$
- $\hat{S}_n\langle t\rangle$ konsistent Schätzung unzensiert. $n \cdot \hat{S}_n\langle t\rangle$ ist $\mathcal{B}(n, S_t)$-verteilt. ZGWS approx. VI für $S\langle t_o\rangle$: $\hat{s}_n\langle t_o\rangle \pm 2 \cdot \sqrt{\frac{\hat{S}_n(t_o)(1-\hat{S}_n(t_o))}{n}}$
- **Zensierten Beobachtung**: (t_i^*, z_i) t_i^* enthalten Wartezeiten und z_i hält fest, ob «+» zensiert ($z_i = 0$) oder nicht ($z_i = 1$)
- Positive Halbachse mit disjunkten Intervallen $I_k = [a_{k-1}, a_k)$, wobei $a_0 = 0$ ist. Survivorfunktion wird geschätzt mit $\hat{S}\langle t\rangle = \prod_{k=1}^{t}(1 - \hat{\lambda}_k)$, $\hat{\lambda}_k$ = Wahr', Ereignis im Intervall zu haben. $\hat{\lambda}_k = 1$ falls kein Ereignis im Zeitintervall beobachtet wurde.
- Stetigen Fall Intervall ganz klein gemacht, sodass höchstens eine Ereigniszeit vorkommt: $\hat{S}_{KM}\langle t\rangle = \prod_{i:t_i^* \leq t}\left(1 - \frac{d_i}{n_i}\right) = \prod_{i:t_i^* \leq t}\frac{n_i - d_i}{n_i}$
- KM-Schätzer ist wie emp. Schätzer stückweise horizontal und hat nur an Stellen t_i^* Sprünge (von **un**zensierten Beob.)
- KM ist **asymptotisch** normalverteilt: $\hat{S}_{KM}\langle t\rangle \sim \mathcal{N}\langle S\langle t\rangle, \sigma_{KM}^2\langle t\rangle\rangle$. Asymptotische Varianz wir mit Greenwood's Formel geschätzt: $\sigma_{KM}^2\langle t\rangle \equiv se_{GF}^2\langle t\rangle := \hat{S}_{KM}^2\langle t\rangle \cdot \sum_{i:\, t_{(i)}^* \leq t} \frac{d_i}{n_i(n_i - d_i)}$, wobei $t_{(k)}^* \leq t < t_{(k+1)}^*$. Genähertes 95%-VI für $S\langle t\rangle$ bei festem t: $\hat{S}_{KM}\langle t\rangle \pm q_{0.975}^{\mathcal{N}} se_{GF}\langle t\rangle$. Ansatz **kann** dazu führen, dass Vertrauensintervall **nicht innerhalb** [0, 1] zu liegen kommt.

library(survival); KM <- survfit(Surv(**Time**, **status**) ~ 1, data = **LDT**); summary(KM) #status = zensiert mit 0, beobachtet = 1
plot(KM, conf.int = T, las = 1, mark.time = TRUE, xlab = "Lebensdauer", ylab = "Geschätzte Survivorfunktion S")

Parametrische Modelle für kontinuierliche Wartezeiten

Exponential-Verteilung $\mathcal{E}\langle\lambda\rangle$ → Modellierung Ausfallzeiten Objekten, die nicht altern → gedächtnislose Verteilung!

- Dichtefunktion $f\langle t\rangle = \lambda \cdot e^{-\lambda t}$ • Erwartungswerst $\mathbb{E}\langle X\rangle = \frac{1}{\lambda}$ • Varianz $var\langle X\rangle = \frac{1}{\lambda^2}$
- Survivor-Funktion Verweildauer T: $S\langle t\rangle = P\langle T > t\rangle = e^{-\lambda t}$ • Verteilungsfunktion $F\langle t\rangle = 1 - e^{-\lambda t}$ mit $\lambda > 0$ für $t \geq 0$
- Hazard- (Ausfall-) Rate Exponential-Verteilung ist zu jedem Zeitpunkt t **konstant** gleich Parameter λ, da $h\langle t\rangle = \frac{f\langle t\rangle}{S\langle t\rangle} = \lambda$
- **Vorzeitige Erneuerung** noch nicht ausgefallener Objekte bei exponentialverteilter Lebensdauern ist **sinnlos**!

Weibull-Verteilung $\mathcal{W}\langle\lambda, \alpha\rangle$ → Wurde 1939 von W. Weibull Beschreibung Materialermüdungserscheinungen erfunden

- Verteilungsfunktion $F\langle t\rangle = 1 - \exp\langle-(\lambda t)^\alpha\rangle$ mit $\alpha, \lambda > 0$
- Dichte $f\langle t\rangle = \lambda\alpha(\lambda t)^{\alpha-1} \cdot \exp\langle-(\lambda t)^\alpha\rangle$
- λ ist Skalenparameter und α Formparameter
- Falls $\alpha = 1$, erhalten **Exponential**-Verteilung
- Falls $\alpha = 2$, erhalten **Rayleigh**-Verteilung
- Survivorfunktion $S\langle t\rangle = \exp\langle-(\lambda t)^\alpha\rangle$
- Hazardrate $h\langle t\rangle = \lambda\alpha(\lambda t)^{\alpha-1}$, sie ist monoton steigend falls $\alpha > 1$, monoton fallend falls $\alpha < 1$ und konstant falls $\alpha = 1$

- **Weibull-Diagramm: komplementäre log-log-transformierte Survivorfunktion** einer Weibullverteilung ist lineare Funktion in $\log\langle t\rangle$: $\log\langle-\log\langle S\langle t\rangle\rangle\rangle = \alpha(\log\langle\lambda\rangle + \log\langle t\rangle)$ → Plot $\log\langle-\log\langle\hat{S}\langle t\rangle\rangle\rangle$ vs. $\log\langle t\rangle$ Klärung Zeiten Weibull
- $\hat{S}\langle t\rangle$ wird mit Kaplan-Meier-Methode geschätzt. • Aufgetragen werden nur unzensierten Beobachtungen.
- Falls Daten von Weibull-Verteilung stammen, müssen Punkte bis auf zufällige Abweichungen auf einer Geraden liegen.

library(survival); source("addFn4survAnal.R"); KM <- survfit(Surv(**Time**, **status**) ~ 1, type = "kaplan-meier", data = **LDT**)
plot(KM, xlab = "Zeit in Stunden bis zum Ausfall", ylab = *expression(hat(S)[KM](t))*, conf.int = TRUE)
Überprüfen Daten exp. verteilt: plot.weibull(Surv(**Time**, **status**) ~ 1, data = **LDT**, *line* = *F*, xlab = "log(**Std**)"); h <- (log(-log(KM$surv)) - log(KM$time)); (h <- h[is.finite(h)]); abline(a = mean(h), b = 1, lty = 1, col = 4) #log(λ) von Hand geschätzt.

Überprüfung Verteilungsannahme: plot.weibull(Surv(**Time**, **status**) ~ 1, data = **diab**) #Weibull, mit scale=1 Exp Verteil.

plot(KM, fun = "cloglog", col = 2:3, mark.time = T, xlab = *"log(Stunden)"*, ylab = *expression(log(-log(hat(S)[KM](t))))*)

Gumbel-Verteilung $\mathcal{G}\langle\mu, \sigma\rangle$ → Lage-Skalen-Verteilung

- Ist Zeit Weibull-verteilt, so ist logarithmierte Zeit Gumbel verteilt: $T \sim \mathcal{W}\langle\alpha, \lambda\rangle \to Y = \log\langle T\rangle \sim \mathcal{G}\langle\mu = -\log\langle\lambda\rangle, \sigma = \frac{1}{\alpha}\rangle$
- Lageparameter μ, Skalenparameter σ auf ganzen reellen Achse mit Verteilungsfunktion $F\langle y\rangle = 1 - \exp\langle-\exp\langle\frac{y-\mu}{\sigma}\rangle\rangle$

QQ-Plot für Gumbel-Verteilung

- Tragen geordneten logarithmierten Wartezeiten (nur jene die beobachtet wurden) gegen entsprechenden Quantile Standard-Gumbel-Verteilung auf. Quantile Standard-Gumbel-Verteilung berechnen mit $F_G^{-1}\langle F_n\langle y_k\rangle\rangle = log\langle -log\langle 1 - F_n\langle y_k\rangle\rangle\rangle$
- $F_n\langle y_k\rangle$ ist empirische Verteilungsfunktion der Wartezeiten. Tragen im Streudiagramm $\log\langle t\rangle$ gegen $\log\langle -\log\langle \hat{S}_{KM}\rangle\rangle$ auf.
- Um zensierten Beobachtungen zu berücksichtigen wird $1 - F_n\langle y_k\rangle = S_n\langle y_k\rangle$ mit Kaplan-Meier-Methode geschätzt.
- Folglich ist QQ-Plot für Gumbel-Verteilung gerade dem an der Winkelhalbierenden gespiegelten Weibull-Diagramm.

Verteilung	$\mathbb{E}\langle T\rangle$	$var\langle Y\rangle$	Median	$h\langle t\rangle$	$S\langle t\rangle$
Weibull $\langle \boldsymbol{\alpha}, \boldsymbol{\lambda}\rangle$	$\frac{1}{\lambda}\Gamma\langle 1+\frac{1}{\alpha}\rangle$	$\frac{1}{\lambda^2}\left(\Gamma\langle 1+\frac{2}{\alpha}\rangle - \Gamma^2\langle 1+\frac{1}{\alpha}\rangle\right)$	$\frac{1}{\lambda}\log\langle 2\rangle^{\frac{1}{\alpha}}$	$\lambda\alpha\langle\lambda t\rangle^{\alpha-1}$	$\exp\langle -\langle\lambda t\rangle^{\alpha}\rangle$
Gumbel $\langle \boldsymbol{\mu}, \boldsymbol{\sigma}\rangle$	$\mu - \gamma\sigma$	$\frac{\pi^2}{6}\sigma^2$	$\mu + \sigma\log\langle\log\langle 2\rangle\rangle$	kompliziert	$\exp\langle -\exp\langle\frac{Y-\mu}{\sigma}\rangle\rangle$

Loglogistische Verteilung – Wartezeit T ist loglogistisch verteilt, falls $Y := \log\langle T\rangle$ logistisch verteilt ist

- Dichte $f\langle t\rangle = \lambda\cdot\alpha\cdot\frac{\langle\lambda\cdot t\rangle^{\alpha-1}}{\langle 1+\langle\lambda\cdot t\rangle^{\alpha}\rangle^2}$
- Survivorfunktion $S\langle t\rangle = \frac{1}{1+\langle\lambda\cdot t\rangle^{\alpha}}$
- Hazardrate $h\langle t\rangle = \frac{\lambda\cdot\alpha\cdot\langle\lambda\cdot t\rangle^{\alpha-1}}{1+\langle\lambda\cdot t\rangle^{\alpha}}$
- Wettverhältnisse mit Survivorfunktion: $\frac{S\langle t\rangle}{1-S\langle t\rangle} = \frac{1}{\langle\lambda\cdot t\rangle^{\alpha}}$ somit $\log\langle\frac{S\langle t\rangle}{1-S\langle t\rangle}\rangle = -\alpha\cdot\log\langle\lambda\rangle - \alpha\cdot\log\langle t\rangle$

Inferenz am Beispiel der Exponential-Verteilung – Maximum-Likelihood Schätzung bei zensierten Beobachtungen

- Sei t_i beobachteten Zeiten, n_u Anzahl unzensierten Beobachtungen und z_i gibt an, ob **Eintreten Ereignis beobachtet wurde** ($z_i = 1$) oder **zensiert** ($z_i = 0$) ist. Bei zensierten Beobachtungen wird Beitrag Log-Likelihood mit $\log\langle S\langle t_i\rangle\rangle$ erfasst
- $\ell\langle\lambda\rangle = \sum_{i=1}^{n}\{z_i\log\langle f\langle t_i\rangle\rangle + (1-Z_i)\log\langle S\langle t_i\rangle\rangle\} = n_u\log\langle\lambda\rangle - \lambda\sum_{i=1}^{n_u} t_i \rightarrow 0 = \frac{\partial\ell\langle\lambda\rangle}{\partial\lambda} = \frac{n_u}{\lambda} - \sum_{i=1}^{n_u} t_i$, ML-Schätzer $\hat{\lambda} = \frac{n_u}{\sum_{i=1}^{n_u} t_i}$
- Wie unzensierten Schätzung Kehrwert Mittelwert, ausser statt Anzahl Beoab n → Anzahl unzensierten Beob n_u geteilt.
- Schätzung $\hat{\lambda}$ ist asymptotisch normalverteil mit $\mathbb{E}\langle\hat{\lambda}\rangle = \lambda$, Varianz $var\langle\hat{\lambda}\rangle = \frac{\lambda^2}{\mathbb{E}\langle n_u\rangle}$. Genäherte 95%-VI Wald: $\hat{\lambda} \pm q^{\mathcal{N}}_{0.975}\cdot\frac{\hat{\lambda}}{\sqrt{n_u}}$
- Approximation bei **kleinen Stichproben schlecht**, u.a., da asympt. Varianz vom unbekannten Parameter abhängt.

Log-basiertes Vertrauensintervall:

- Abhilfe: Transformieren $\hat{\lambda}$ so, dass asympt. Varianz unabhängig vom unbekannten Parameter ist: $\log\langle\hat{\lambda}\rangle$ betrachten
- Wird approximatives VI gerechnet und Endpunkte zurücktransformiert: $\hat{\lambda}\cdot\exp\langle\pm q^{\mathcal{N}}_{0.975}\cdot\frac{1}{\sqrt{n_u}}\rangle$ → gute Näherung 95%-VI
- Aufgrund asympt. Resultat $\log\langle\hat{\lambda}\rangle \sim \mathcal{N}\langle\log\langle\lambda\rangle, \frac{1}{\mathbb{E}\langle N_u\rangle}\rangle$ können genähertes 95 %-VI für $\log\langle\lambda\rangle$ durch $\log\langle\hat{\lambda}\rangle \pm q^{\mathcal{N}}_{0.975}\cdot\frac{1}{\sqrt{n_u}}$

Parameter schätzen der Exponentialverteilung mit ML-Schätzer. Wie gross ist erwartete Überlebenszeit?
library(survival); **LDT**.E <- survreg(Surv(**Time**, **status**) ~ 1, data = **LDT**, dist = "**exponential**"); exp(-coef(**LDT**.E))
#(oben) Rücktransformation **Schätzwert für** λ; exp(-confint(**LDT**.E))[,2:1] #Rücktrans. VI für λ → Reihenfolge Intervall Schätzwert **erwartete Lebenszeit** (= $1/\lambda$) → exp(coef(**LDT**.E)); exp(confint(**LDT**.E)) #gleich, einfach **ohne das Minus Wahrscheinlichkeit**, dass mehr als *26 Wochen* vergehen: exp(-**26***exp(-coef(**LDT**.E))) #Transformation aufgrund Gumbel
LDT.W <- survreg(Surv(**Time**, **status**) ~ 1, data = **LDT**, dist = "**weibull**") #**Anpassung** in survreg(...) erfolgen via **Gumbel-Modell!** → Schätzungen für Exp./Weibull **müssen rücktransformiert werden!** summary(**LDT**.W) # ↓ P-Wert «log(scale)»
Log(scale) ist auf 5% Niveau (nicht) signifikant von 0 verschieden; d.h. scale ist auf 5% Niveau (nicht) signifikant von 1 verschieden. Da α = 1/scale folgt, dass α auf 5% Niveau (nicht) signifikant von 1 verschieden ist. Somit wird Nullhypothese, dass Daten mit Exponential-Verteilung modelliert werden können, (nicht) verworfen. D.h. keinen Sinn frühzeitig Austausch
plot(**LDT**.KM, conf.int = F, col = 2, lty = 2, las = 1); h.tp <- seq(**0, 3, 0.5**); lines(h.tp, pexp(h.tp, rate = exp(-coef(**LDT**.E)), lower.tail = F)); lines(h.tp, pweibull(h.tp, scale = exp(coef(**LDT**.W)), shape = 1/**LDT**.W$scale, lower.tail = F), col = 3)

- KM: Weil grössten Lebensdauerdaten zensiert sind, können keine mittlere Lebenszeit schätzen.
- Mit Exponential-/Weibullmodell ist möglich erwartete Lebensdauer zu schätzen (siehe Tabelle oben Median Gumbel):
- **Bsp** (Werte Summary: Intercept 4.503, Scale fixed at 1) $Median = \exp\langle\mu + \sigma\log\langle\log\langle 2\rangle\rangle\rangle = \exp\langle 4.503 + 1\cdot\log\langle\log\langle 2\rangle\rangle\rangle$

Beschleunigte Ausfallzeit-Modelle (accelerated failure time models – AFT) → Einfluss erklärenden Variablen berücksichtigen

- Lineare Prädiktor (Linearkombination erkl. Var.) soll auf Lage logarithmierten Zielvariable einwirken: $\log\langle T_i\rangle := Y_i = \beta_0 + \underline{x}_i^T\underline{\beta} + \sigma\cdot W_i$, wobei W entweder standard Gumbel, standard logistisch oder standard normalverteilt ist.
- Sei nun $W_i^{\circ} := \beta_0 + \sigma\cdot W_i$ (**Referenzverteilung**), kann obiges Modell geschrieben werden als $\log\langle T_i\rangle =: Y_i = \underline{x}_i^T\underline{\beta} + W_i^{\circ}$, wobei W_i° Gumbel, loglogistisch oder normalverteilt mit jeweils Lageparameter $\mu = \beta_0$ und Skalenparameter σ ist.
- Rücktrans. $\mathcal{W}$-, loglogistischen, Lognormalvert. $T_i^{\circ} := \exp\langle W_i^{\circ}\rangle \rightarrow T_i = \exp\langle Y_i\rangle = \exp\langle\underline{x}_i^T\underline{\beta}\rangle\cdot\exp\langle W_i^{\circ}\rangle = \exp\langle\underline{x}_i^T\underline{\beta}\rangle\cdot T_i^{\circ}$
- Der exponenzierte lineare Prädiktor $\underline{x}_i^T\underline{\beta}$ wirkt also multiplikativ auf eine «Referenz»-Wartezeitverteilung T_i° ein.

Survivorfunktion von T: Beschleunigungsfaktor $\gamma := \exp\langle -\underline{x}^T\underline{\beta}\rangle$. Survivorfunktion von T gegeben x lässt mit Survivorfunktion von T° ausdrücken: $S\langle t|\underline{x}\rangle = S_0\langle t\cdot\gamma|\underline{x}\rangle$ → Der exponenzierte lineare Prädiktor ändert also Skala von der Zeit t.

- **Erklärenden Variablen beschleunigen/verlangsamen** Ausfall-/Wartezeit; d.h. Zeit bis Ausfall, abhängig $\gamma < 1$, $\gamma > 1$
- Gibt für AFT-Modelle verschiedene Schreibweisen. $\log\langle T_i\rangle := Y_i = \beta_0 + \underline{x}_i^T\underline{\beta} + \sigma\cdot W_i$ heisst **log-lineare Darstellung**.

Hazardrate: Hazardrate $h_0\langle t\rangle$ Zufallsvariablen T° ist unbeeinflusst von erkl. Var. $\underline{x}$ und Parametervektor $\underline{\beta}$

- Für Hazardrate $h\langle t|\underline{x}\rangle$ von $T = e^{\underline{x}^T\underline{\beta}}\cdot T^{\circ}$ gilt dann $h\langle t|\underline{x}\rangle = -\frac{d\log\langle S\langle t|\underline{x}\rangle\rangle}{dt} = h_0\langle t\cdot\gamma|\underline{x}\rangle\cdot\gamma$, mit $\gamma := e^{-\underline{x}^T\underline{\beta}}$
- Darstellung zeigt, dass Beschleunigungsfaktor γ sowohl **multiplikativ** auf **Hazardrate** als **multiplikativ** auf **Zeit** einwirkt.

Weibull-Regressionsmodell (gleiches Modell kann auch mit Gumbel-Verteilung beschrieben werden)

- Weibull beliebt, weil Survivorfunktion $S\langle t|\underline{x}\rangle = e^{-\left(\lambda\cdot t\cdot e^{-\underline{x}^T\underline{\beta}}\right)^{\alpha}} = e^{-\langle\lambda\cdot t\cdot\gamma_{\underline{x}}\rangle^{\alpha}}$ explizit darstellen und mathematisch «einfach» sowie es nicht nur ein **beschleunigtes Ausfallzeiten-Modell** ist, sondern es hat auch **proportionale Hazard-Eigenschaft**.

- Hazardfunktion Weibullverteilung ist $h\langle t|\underline{x}_i\rangle = \lambda\alpha\langle\lambda t\rangle^{\alpha-1}\cdot e^{-\underline{x}_i^T\underline{\beta}\cdot\alpha} = \lambda\alpha\left(\lambda t\gamma_{\underline{x}}\right)^{\alpha-1}$ und damit ist Verhältnis Weibull-Hazarddraten bei zwei verschiedenen Zuständen der **erkl. Var.**, gekennzeichnet durch $\underline{x}_j$ und $\underline{x}_k$, $\frac{h\langle t|\underline{x}_i\rangle}{h\langle t|\underline{x}_k\rangle} = \left(\frac{e^{-\underline{x}_i^T\underline{\beta}}}{e^{-\underline{x}_k^T\underline{\beta}}}\right)^{\alpha} = \left(\frac{\gamma_{\underline{x}_i}}{\gamma_{\underline{x}_k}}\right)^{\alpha}$
- Eigenschaft **Verhältnis** Hazardraten **unabhängig** von **Wartezeit** t ist, heisst **Proportionale-Hazard-Eigenschaft**.
- Unterscheiden **erkl. Vektoren** $\underline{x}_i, \underline{x}_k$ nur darin, dass beim einen Behandlung A vorgenommen $x_i^{(1)} = 1$ und anderen B $x_k^{(1)} = 0$, dann ist bei positiven Koeffizienten β_1 Hazardrate Gruppe Behandlung A immer um Faktor $\left(e^{\beta_1}\right)^{\alpha} = e^{\beta_1\cdot\alpha}$ kleiner (wegen Minuszeichen, das nicht mehr Faktor ist) als jene Gruppe B. Folglich Wartezeit Eintritt Ereignis Behandlung A länger als B. Weil **Faktor unabhängig Wartezeit**, **Behandlung A** immer **bessere** unabhängig Wartezeit.
- proportionale Hazard-Eigenschaft führt weiter dazu, dass Hazardrate konzeptionell auch als $h\langle t|\underline{x}\rangle = h_0\langle t\rangle\cdot\left(e^{-\underline{x}^T\underline{\beta}}\right)^{\alpha} = h_0\langle t\rangle\cdot e^{-\underline{x}^T\underline{\beta}\cdot\alpha}$ gilt, $h_0\langle t\rangle = \lambda\alpha\langle\lambda t\rangle^{\alpha-1}$ ist **Referenz-Hazardrate** (baseline). Hazardrate Referenzzustandes in erkl. Var. $\underline{x}_j = 0$. $S\langle t|\underline{x}\rangle = S_0\langle t\rangle^{e^{-\underline{x}^T\underline{\beta}\cdot\alpha}}$ Survivorfunktion, $S_0\langle t\rangle$ Survivorfunktion der Referenz-Hazardrate $exp\langle -\langle\lambda\cdot t\rangle^{\alpha}\rangle$
- **Weibull-Regressionmodell** in log-lineare Darstellung mit **Gumbelparameter** $\log\langle T_i\rangle = Y_i = \beta_0 + \underline{x}_i^T\underline{\beta} + \sigma\cdot W_i$, W_i Standard-Gumbel $G\langle\mu = 0, \sigma = 1\rangle$ verteilt. Weibull-Regressionsmodell Weibull-Parameter $h\langle t|\underline{x}, \alpha, \lambda, \underline{\theta}\rangle = \lambda\alpha\langle\lambda t\rangle^{\alpha-1}\cdot e^{\underline{x}^T\cdot\underline{\theta}}$
- Weibull-$(\alpha, \lambda, \underline{\theta})$ Gumbel $(\beta_0, \underline{\beta}, \sigma)$ Beziehungen $\beta_0 = -\log\langle\lambda\rangle$ oder $\alpha = \frac{1}{\sigma} = e^{-\log\langle\sigma\rangle}$, | $\underline{\beta} = -\frac{\underline{\theta}}{\alpha}$ oder $\lambda = e^{-\beta_0}$ λ von Baseline | $\sigma = \frac{1}{\alpha}$ oder $\underline{\theta} = -\frac{\underline{\beta}}{\sigma}$. In Verteilungen: $\log\langle T_i\rangle \sim \mathcal{G}\langle\mu_i = \beta_0 + \underline{x}_i^T\underline{\beta}, \sigma = \sigma\rangle$ und $T_i \sim \mathcal{W}\langle\lambda_i = \exp\langle -\left(\beta_0 + \underline{x}_i^T\underline{\beta}\right)\rangle, \alpha = \frac{1}{\sigma}\rangle$

Maximum-Likelihood-Schätzung: Unbekannten Parameter $\beta_0, \underline{\beta}, \sigma$ Gumbelparameter werden maximalen Likelihood geschätzt. Falls $r_i := y_i - \left(\beta_0 + \underline{x}_i^T\underline{\beta}\right)$ ist, dann lautet entsprechende Log-Likelihood $\ell\langle\beta_0, \underline{\beta}, \sigma|y, z, \underline{x}\rangle = \sum_{i=1}^{n}\{z_i\cdot\log\langle f\langle r_i\rangle\rangle + (1-z_i)\cdot\log\langle S\langle r_i\rangle\rangle\}$ wobei y_i gleich $\log\langle t_i\rangle$ ist und z_i angibt, ob die **Beobachtung zensiert ($z_i = 0$) ist oder nicht ($z_i = 1$).**

- (Allgemein gilt:) Die ML-Schätzungen $\hat{\beta}_0, \underline{\hat{\beta}}, \log\langle\hat{\sigma}\rangle$ sind asympt. normalvert.
  - mit Erwartungswerten $\beta_0, \underline{\beta}, \log\langle\sigma\rangle$, respektive
  - Varianzen, gegeben durch inverse Informationsmatrix

Mo.W <- survreg(Surv(**time**, **status**) ~ **x**, data = **Mo**, dist = "weibull"); summary(**Mo**.W); h.**Mo** <- data.frame(**x** = **1000**))
#Prognose Weibull-Parameter; (h.a <- 1/**Mo**.W$scale) #$\hat{\alpha}$; (h.li <- predict(**Mo**.W, newdata = h.**Mo**, type = "response", se = T)) #Prognose $\frac{1}{\lambda\langle 1000\rangle}$ und StdFhlr; h.li$fit * gamma(1+1/h.a) #Schätzung $E(X)$: $E(Weibullvert) = \frac{1}{\lambda}\cdot\Gamma\left(1+\frac{1}{\alpha}\right)$
1/h.li$fit #$\lambda(x)$; 1/(h.li$fit + 1.96*c(1,-1)*h.li$se.fit) # VI für $\lambda(x)$ | Scale im Summary, hier $Scale = 0.711$ → ($\alpha = \frac{1}{scale} = 1.41$)

- Intercept ist $\hat{\beta}_0 = 5.3$. Folglich ist $\exp\langle -\hat{\beta}_0\rangle$ geschätzte Parameter λ in Referenzweibullverteilung → $\exp\langle -5.3\rangle = 0.0049$

Baselin-Hazard-R $h_0\langle t\rangle = \lambda\alpha\langle\lambda t\rangle^{\alpha-1} = \alpha\cdot\lambda^{\alpha}\cdot t^{\alpha-1} = 1.41\cdot 0.0049^{1.41}\cdot t^{1.41-1} = 0.0008\cdot t^{0.41}$↓Exponent 1/**Mo**.W$scale-1

- Weil Exponent zu t positiv und kleiner 1 ist Baseline-Hazardrate monoton steigend, jedoch abflachend wachsendem t.
- Weil Exponent zu t positiv und grösser 1 ist, ist Baseline-Hazardrate monoton steigend und wird steiler mit wachsend t.
- Weil Exponent zu t negativ ist, ist Baseline-Hazardrat monoton fallend und wird flacher mit wachsendem t.
- Weil Exponent zu t gleich 1 ist, ist Baseline-Hazardrate konstant, also eine Gerade → Exponential-Verteilung
- Stufe «yes» bei fin mit 0.3042 signifikant positiv, verschiebt Lage nach oben und Wartezeit wird grösser, wenn Variable Beobachtung «yes» ist.
- Koeffizient wPrio (Wurzel Anzahl Verurteilungen) signifikant negativ, verschiebt sich Lage log. Wartezeit nach unten, also wird Wartezeit kleiner je mehr Verurteilungen vorausgegangen sind.
- Einfluss erkl. Var. auf Hazardrate $h\langle t|x\rangle$: $h\langle t|x\rangle = h_0\langle t\rangle\cdot\exp\langle -\eta\cdot\alpha\rangle$, $\eta = \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_m x_m$ wobei x = erkl. Var.

```
survreg(formula = Surv(week, arrest) ~ ., data = RR)
              Value   Std. Error        z        p
finyes       0.3042        0.139   2.1825  0.02908
wPrio       -0.2096        0.089  -2.3540  0.01857
```

Modellvergleich: Statistisch Vergleich von zwei ineinander geschachtelten Modellen erfolgt Likelihood-Quotienten-Test.

- Diese Teststatistik ist äquivalent zum Devianztest. Unter Null-Hypothese ist Teststatistik D asymptotisch χ_q^2-verteilt mit q gleich Anzahl in Null-Hypothese festgelegten Parameter. Umsetzung in R gleich wie bei anderen Modellen, anova()

Bsp: loglogistisches vs. weibull AFT: Welches Modell aufgrund max Log-Likelihood wählen?
Mo.LL <- survreg(Surv(**time**, **status**) ~ **x**, data = **mot**, dist = "loglogistic"); summary(Mo.LL) #gleiches Modell mit dist weibull Grösseres Max Log-Likelihood bessers Modell. **Vergleich nur**, wenn **beide Modelle gleich** viele **Parameter** besitzen.

Variablenselektion: Wie bei GLM AIC zur Variablenselektion benutzt. Man wählt Modell, dass das AIC minimiert. step()

Modelleignung: Weil AFT-Modelle Lage-Skalen-Verteilungsfamilien beruhen, wirken erklärenden Variablen durch linearen Prädiktor auf Lage der Verteilung der Zielvariablen ein. Aber Achtung! $Lage \neq Erwartungswert;\ E\langle\log\langle T\rangle\rangle = \mu - \gamma\sigma$

- Deshalb ergibt sich folgende naheliegende Definition für Residuen: $r_i := y_i - \left(\hat{\beta}_0 + \underline{x}_i^T\underline{\hat{\beta}}\right)$
- Falls Parameter bekannt wären, dann Residuen Gumbel verteilt mit $\mu = 0$ und σ unbekannt. → Gumbel-QQ-Plot
- Wenn Beobachtungen rechts-zensiert sind, dann Schätzung $\underline{x}_i^T\underline{\hat{\beta}}$ üblicherweise grösser als $y_i = \log\langle t_i\rangle$ sein.
- Folglich sind entsprechenden Residuen tendenziell negativ und können als negative Ausreisser in Erscheinung treten.

T-A-/Gumbel-QQ-Plot: source("addFn4survAnal.R"); par(mfrow = c(1,2), las=1); plotGumbelRes(**Mo.LL**, resType = "raw")

Cox-Snell-Residuen: Sind Stat. Wartezeit-Analyse beliebt: $r_i^{(cs)} := -\log\langle\hat{S}_{KM}\langle r_i|\underline{x}_i\rangle\rangle$, r_i Residuen aus Regressionsmodell

- CS-Res Weibull-Regr.: $r_i^{(cs)} := -\log\langle S\langle r_i|\underline{x}_i\rangle\rangle = \exp\langle\frac{y_i - \hat{\beta}_0 + \underline{x}_i^T\underline{\hat{\beta}}}{\hat{\sigma}}\rangle = e^{\frac{r_i}{\hat{\sigma}}}$ → Originalskala zurücktransformierten $\hat{\sigma}$ skalierten Gumbelresiduen. Können $\log\langle r_i^{cs}\rangle$ mit Gumbel-QQ-Plot überprüfen, ob (approximativ) exponentialverteilt mit $\lambda = 1$.
- Da $\hat{S}_{KM}\langle r_i|\underline{x}_i\rangle$ immer im Intervall [0, 1] liegt, sind alle r_i^{cs} nicht negativ und entsprechend schwierig zu interpretieren.
- Sind keine Residuen, da sie aus Transformation der Differenz zwischen beobachteten und angepassten Werten sind.

Devianz-Residuen: Martingal-Residuum $r_i^{(M)} := z_i - r_i^{(cs)}$, Wert Intervall $(-\infty, 1]$. **Zensierte** Beobachtungen immer **negativ**!

- Verbesserte, symmetrisierte Martingale Residuen: Devianz-Residuen $r_i^{(D)} := \text{sign}\langle r_i^{(M)}\rangle\cdot\sqrt{-2\cdot\left(r_i^{(M)} + z_i\cdot\log\langle z_i - r_i^{(M)}\rangle\right)}$
- R multipliziert $r_i^{(D)}$ mit (−1). Devianz-Residuen sollten symmetrisch um 0 verteilt sein. Summieren sich nicht auf 0 wie in klass. lin. Reg. Wenn nur schwach bis moderat Zensierung, dann sollten $r_i^{(D)}$ aussehen wie iid normalverteilt qqnorm()

Tukey-Anscombe Devianzresiduen/Gumbel-QQ-Diagramm: plotGumbelRes(**Mo.LL**, smooth = TRUE)

- **Tukey-Anscombe-Plot** streuen Punkte (nicht) gut um horizontale Null-Linie, wie Glätter zeigt. Auch Streuung Punkte bleibt über linearen Prädiktor-Werte gleich. Ob dies statistisch signifikant ist oder relevant kann nicht beurteilt werden.
- **Gumbel QQ-Plot** streuen Punkte gut um Linie "0 + 1*x", welche Exponentialverteilung mit $\lambda = 1$ repräsentiert (log Cox-Snell Residuen müssen so verteilt sein). Allenfalls haben wir einen "leichten" Ausreisser links unten. → Gesamtfazit…

Cox-Regressionsmodell → bei kontinuierlichen Wartezeiten

- Hazardrate $h_c\langle t|\underline{x}\rangle$ ist Produkt Baseline-Hazardrate $h_o\langle t\rangle$ und erkl. Var.: $h_c\langle t|\underline{x}\rangle = h_o\langle t\rangle \cdot \exp\langle \underline{x}^T\underline{\theta}\rangle$
- Baseline-Hazardrate $h_o\langle t\rangle$ ist unabhängig von erkl. Var. $\underline{x}$, jedoch Form nicht wie Weibull-Reg.modell explizit spezifiziert.
- Funktion $h_o\langle t\rangle$ kann als Hazardrate für Beobachtung betrachtet werden, bei der **alle erkl. Var. die Werte 0 haben**.
- **Effekte konstant über Zeit** auf Hazardrate einwirken: Unterscheiden sich erkl. Vektoren $x_1^{(1)}, x_2^{(1)}$ nur durch Behandlung A und B, dann ist bei positiven Koeffizienten θ_1 Hazardrate für verbleibende Dauer B immer Faktor e^{θ_1} grösser als A.

library(survival); **D**.cox <- coxph(Surv(**lzeit**, **tod**) ~ **sex** + **diab**, data = **D**); length(unique(**D$lzeit**)); which(table(**D$lzeit**)>1)
Vergleich Koeffizientenschätzungen **Cox- mit Weibull**-Regressionsmodell → transformieren Gumbelparametrisierung (-coef(**D.weib**)[-1])/**D.weib**$scale; coef(**D.cox**) #können nun verglichen werden mit: summary(**D.cox**)$coefficients[,3]
$\hat{h}\langle t\rangle = h_o\langle t\rangle \cdot \exp\langle 0.14 \cdot sexW\rangle \cdot \exp\langle 0.53 \cdot diabYes\rangle$ #Werte aus 1. Block, sig. positiver Koeffizient = kürzer Wartedauer

- 2 Block $e^{\hat{\theta}_k}$, Faktor mit Hazardrate multipliziert, wenn andere Variablen gleich und k-te Variable um 1 Einheit ändert.
- **Harrell's concordance index** (C-Index) Masszahl **Anpassungsgüte** (Wartezeit-) Modell: Konkordanz betrachtet alle möglichen Paare von Personen, bei dem eine gestorben und andere mind. so lange lebt. Falls vorhergesagte Lebensdauer grösser für Person, die tatsächlich länger lebte, so ist Vorhersage für Paar konkordant mit tatsächlichen Ausgang. C-Index Anteil konkordanten Paare unter allen Paaren. Werte [0, 1]. Je näher 1, desto besser beschreibt Modell Daten.

Variablenselektion und Modellvergleich step(); anova() → basieren auf «max. partial Log-Likelihoods».

Maximum Partial Likelihood: Um Parametervektor $\underline{\theta}$ zu schätzen, neue Schätzmethodik → maximum partial likelihood

- Schätzansatz ignoriert Baseline-Hazardrate. Berücksichtigt nur Reihenfolge der beobachteten (**unzensierten**) Wartezeiten. **Damit Schätzer zulässig, dürfen keine beobachteten Wartezeitwerte mehrmals vorkommen!**
- Da sonst Wartezeiten nicht rangiert werden können. Trotz Modifikation Likelihoodansatz, können Resultate aus allgemeinen Max-Likelihood-Theorie übernommen werden. **Nicht möglich, damit Vorhersagen zu machen**.

Wertgleiche Wartezeiten: Schätzverfahren funktioniert i.A. nur, wenn **keine wertgleichen Wartezeiten** (tied event times)

- Häufigste Ausweg **Breslow's Approx.** verwenden. Funktioniert gut, sofern **Ties nur selten**. Hat **viele Ties** kann Approximation zu **unzuverlässigen Resultaten** führen. Bessere Approximation ist **Efron** → rechnerisch **effizienter**

Exakter Cox-Regressionsansatz I → nicht in R implementiert, da in Praxis selten genutzt

- Annahme: Ties durch **grobe Rundung Wartezeit**. Folglich gäbe wahre Reihenfolge Wartezeiten. Weil diese unbekannt, alle Reihenfolgen berücksichtigt und Likelihood-Terme hinzufügen. 5 Zeiten gleich → $5! = 120$ mögliche Reihenfolge **Möglichkeiten** bei mehr Ties **explodieren** förmlich und **Berechnung** wird auch mit heutigen Mitteln **zu zeitintensiv**.

Exakter Cox-Regressionsansatz II – diskreter Cox-Regressionsansatz → ties = "exact"

- Annahme: Ties durch Diskretheit Wartezeiten verursacht. Folglich kann keine wahre Reihenfolge geben. Ansatz beruht auf anderen Modell, welches kein proportional Hazardmodell ist. Schätzmethodik beruht «partiellen Likelihood-Ansatz».

coxph aus Package survival hat **Efron's** Ansatz als Standardeinstellung. Argument ties kann "breslow" oder "exact" wählen.

- Schätzverfahren bei Cox-Regression geht davon aus, dass **keine wertgleichen Wartezeiten** beobachtet werden
- Wird stark gerundet oder stammen Wartezeiten von diskreten Zeitskala, dann versagt eigentliche Schätzverfahren
- Bei **wenigen** wertgleichen **Wartezeiten** kann man **Problem** mit Approximationen **umgehen**. **Bei vielen nicht!**

Regressionsmodelle bei diskreten Wartezeiten

- Wenn Wartezeit **diskrete Zeitskala** gibt **zwei exakte Vorschläge**. Sind **rechenintensiv → nur kleinen Stichproben**
- Diskret Wartezeit mit **GLM** → ist einfacher, Datenaufbereitung aufwändiger. Möglich zeitveränderlich erkl. Var. einbinde
- Einfluss erkl. Var. auf Hazardfunktion $\lambda\langle t\rangle$ wie in binomialen Regression (GLM): $g\langle\lambda\langle t|\underline{x}\rangle\rangle = \beta_0 + \beta_1 x^{(1)} + \cdots + \beta_m x^{(m)} = \beta_0 + \underline{x}^T\underline{\beta}$ wobei $g\langle.\rangle$ Linkfunktion ist. Modellierung $\lambda\langle t|\underline{x}\rangle$ nicht nur bei einem t, sondern für alle Intervalle $t = 1, 2, \ldots, q$.
- Einfacher Ansatz: **nur Achsenabschnitt** β_0 von **Wartezeit** t **abhängig** machen (mit $\beta_0 \to \gamma_t$): $g\langle\lambda\langle t|\underline{x}\rangle\rangle = \gamma_t + \underline{x}^T\underline{\beta}$
- **Interpretation** dieses Ansatzes **hängt** von Wahl **Linkfunktion** ab! (Logit-Modell oder Komplementäres Log-Log-Modell)
- Modellansatz modelliert Effekte der erkl. Variablen auf Hazardfunktion → es wird ganze Survivor-Funktion modelliert

Logit-Modell für diskrete Wartezeiten: Logit-Modell modelliert Hazardfunktion wie folgt: $\log\langle\frac{\lambda\langle t|\underline{x}\rangle}{1-\lambda\langle t|\underline{x}\rangle}\rangle = \gamma_t + \underline{x}^T\underline{\beta}$

- variieren nur Achsenabschnitte γ_t über Wartezeit t.
- Weiteren Koeffizienten $\underline{\beta}$ sind unabhängig von t.
- Linke Seite = log. Wettverhältnis (log-odds) Hazardfunktion. Mit Wahr' lautet (bedingte) Wettverhältnis: $\frac{P(T=t|T\geq t,\underline{x})}{1-P(T=t|T\geq t,\underline{x})}$
- $\underline{x} = \underline{0}$, dann Wettverhältnis Hazardwahrscheinlichkeit = exponenzierten Parameter γ_t: $\frac{\lambda\langle t|\underline{x}=0\rangle}{1-\lambda\langle t|\underline{x}=0\rangle} = \exp\langle\gamma_t\rangle$
- **Parameter** $\gamma_t,\ t = 1,\ldots,q$ **legen** über Hazard-Wettverhältnis, **Baseline-Hazardfunktion** fest. **Erkl. Var. skalieren** (transformierte) **Baseline** einheitlich über Zeitintervalle $t = 1,\ldots,q$ mit Faktor: $\frac{\lambda\langle t|\underline{x}\rangle}{1-\lambda\langle t|\underline{x}\rangle} = \exp\langle\gamma_t\rangle \cdot \exp\langle\underline{x}^T\underline{\beta}\rangle,\ t = 1,\ldots,q$
- Odds-Ratio Hazardwahrscheinlichkeiten ändert um Faktor $\exp\langle\beta_j\rangle$, wenn sich erkl. Variable x^j um eine Einheit ändert.
- Baseline-Hazard hängt von Parametrisierung in erkl. Var. ab, weil sie der Hazardfunktion entspricht, bei der $\underline{x} = \underline{0}$ ist!
- Quotient Wettverhältnis Hazardwahr' bei **versch. Wartezeiten** t und s: $\exp\langle\gamma_t - \gamma_s\rangle$ unabhängig von $\underline{x}$; hängt also nur von Baseline-Hazards ab. Quotienten Wettverhältnis Hazardwahr' bei **verschiedenen Werten** x_1 und x_2 der erkl. Var.: $\exp\langle(\underline{x}_1 - \underline{x}_2)^T\underline{\beta}\rangle$ unabhängig von Wartezeit t; hängt nur von $(\underline{x}_1 - \underline{x}_2)$ ab.
- 2 Survivorfun. Versch. Prädiktoren kreuzen nie: $\lambda\langle t|\underline{x}_1\rangle > \lambda\langle t|\underline{x}_2\rangle$ → Survfun $\underline{x}_1$ alle Intervalle t kleiner als jene bei $\underline{x}_2$

Gruppiertes proportionales Hazardmodell → link = cloglog

- Proportionales Hazardmodell, jedoch (stark) gerundete Wartezeiten → Zeitachse unterteilt in disjunkte Intervalle
- Äquivalent: $\log\langle-\log\langle 1-\lambda\langle t|\underline{x}\rangle\rangle\rangle = \gamma_t + \underline{x}^T\underline{\theta}$ der mit komplement. log-log Link aus binären Regression übereinstimmt.
- Was Einfluss erkl. Var. $\underline{x}$ betrifft, wird **diskrete Modell gleiche Analyse erlauben wie proportionale Hazard-Modell**.
- $\lambda\langle t|\underline{x}\rangle$ des diskretisierten proportionalen Hazardmodells kann **nicht als Produkt** einer Baseline-Hazard-Wahrscheinlichkeit (≠ Hazardrate) und des Einflusses der erklärenden Variablen dargestellt werden.
- Quotient logarithmierten Survivorfunktionen ist unabhängig von Wartezeit: $\frac{\log\langle S\langle t|\underline{x}_1\rangle\rangle}{\log\langle S\langle t|\underline{x}_2\rangle\rangle} = \exp\langle(\underline{x}_1 - \underline{x}_2)^T\underline{\theta}\rangle$
- Obwohl Ansatz nicht proportionale Hazardeigenschaft besitzt, heisst es «**gruppiertes proportionale Hazardmodell**», weil aus proportionalen Hazardmodell von kontinuierlichen Wartezeiten bei Annahme (starken) Rundungen ableitet.

Schätzung bei diskreten Wartezeiten | proportional continuation ratio model → link = logit

- ML-Schätzung Regressionsmodell diskrete Wartezeit führt zu äquivalent Lösung wie GLM. Müssen Daten aufbereiten
- Jede **Wartezeitbeobachtung** ist die Realisierung **zweier Zufallsvariablen**: Eine Zufallsvariable ist **Wartezeit** T_i und die andere Zufallsvariable ist **Indikator** Z_i, der angibt, ob **Ereignis beobachtet wurde** ($Z_i = 1$) oder **nicht** ($Z_i = 0$).
- Zensierung Ende jeweiligen Intervall. Annahme: **Zensierung nicht informativ** (Zensierung hängt nicht von Parametern Warteprozess ab) → Zensierungselemente können in Likelihood zu **Konstanten** c_i zusammengefasst werden.
- Mit Hazardwahrscheinlichkeiten $\lambda\langle t_i|\underline{x}\rangle$ gilt dann für Likelihood $L_i = c_i \cdot \lambda\langle t_i|\underline{x}\rangle^{z_i} \cdot \left(1-\lambda\langle t_i|\underline{x}_i\rangle\right)^{1-z_i} \cdot \prod_{j=1}^{t_i-1}\left(1-\lambda\langle t_i|\underline{x}_j\rangle\right)$
- ML-Schätzung Parameter im diskreten Wartezeit-Regressionsmodell hat gleiche asymptotische Verteilung wie jene, die auf binären Repräsentation der Übergänge zwischen Wartezeitkategorien (d.h. GLM-Anpassung) beruht.

library(discSurv); long <- dataLong(**data**, timeColumn = "**Dauer**", eventColumn = "**Beob**", timeAsFactor = T)
gphm <- glm(**y** ~ -1 + timeInt + **Kundengruppe** + **Region**, data = long, family = binomial(link="cloglog")); summary(gphm)
drop1(gphm, test = "Chisq") Vorteil von drop1() gegenüber summary(): → erkl. Faktorvariabeln

- informiert, welche Variablen signifikant sind. Können damit klären, ob timeInt (und damit Hazardwahr') konstant sind.
- P-Werte zuverlässiger, da beruhen Teststatistiken/Devianztests → Asymptotik wirkt schneller (bzgl. Stichprobengrösse).
- gibt für Variable an, ob Einfluss auf ZV signifikant ist unter Berücksichtigung, dass andere Variablen auch Einfluss haben

Darstellung Hazardwahr' mit KI: h.q <- length(dummy.coef(**gphm**)$timeInt); h.hc <- dummy.coef(**gphm**)$timeInt
h.hcse <- summary(**gphm**)$coefficients[1:h.q, 2]; plot(1-exp(-exp(h.hc)), type="l", panel.first=abline(v=(1:4)*13/2), ylab = "Hazard-Wahr'", xlab = "Dauer in '…'"); lines(1-exp(-exp(h.hc+2*h.hcse)), lty=2, col=2); lines(1-exp(-exp(h.hc-2*h.hcse)))
Alternative mit predict: newdata = Referenz-Hazardwahr'. R findet heraus, welche Levels Variablen sonst noch haben
h.new <- data.frame(timeInt=**as.factor(1:52)**, **fin**="**no**", **race**="**black**", **wexp**="**no**", **paro**="**no**", **mar**="**married**", educ="**2**", **fAge** = "**(0,21]**", **wPrio** = **0**); h.p <- predict(**gphm**, newdata=h.new, se.fit=T, type="link") #h.hc=h.p$fit und h.hcse=h.p$se.fit

- **Variablenselektion:** wie GLM AIC: step(**g.glm**, scope = list(lower = ~ -1 + timeInt, *upper = formula(**g.glm**)*))
- Bei Wechselwirkungen/Interaktionen: Sobald Variable timeInt mit einer erkl. Var. (z.B. V1) in Wechselwirkung tritt, bedeutet dies, dass unterschiedliche Baseline-Hazards haben für Population, die durch V1 beschrieben wird.

- **Additive Modelle**: wenn Hazardwahr' aus vielen Intervallen zusammensetzen (Faktorvariable timeInt hat viele Stufen).
- Wenn viele Stufen gibt, ist dies einzige Möglichkeit Hazardwahr' zu schätzen, weil sonst Design-Matrix zu gross wird.
- gam() aus library(mgcv): + numerisch geschickter implementiert, beruht auf Splines-Schätzungen der unbekannten Transformation. +/- wählt automatisch optimalen Glättungsparameter, führt oft zu einer zu glatten Funktion
- gam() aus library(gam): + Schätzungen unbekannten Transformation mit loess. Kann zu lokalen Minima konvergieren. (-) Glättungsparameter von Hand wählen. Ziehe leicht Unterglättung vor, weil von Auge etwas «weiter glätten» kann.
long$tNr <- as.integer(as.character(long$timeInt)); f.gam <- gam(y ~ lo(tNr) + **fin** + **mar** + lo(**wPrio**) + lo(**age**), data = long, family = binomial(link = "cloglog")); par(mfrow = c(**2, 3**)); plot(f.gam, se = T) #Residuen wie unten analysieren.

- **Devianz-Residuen** sind nicht für Überprüfung Modelleignung geeignet, weil Pseudo-Beobachtungen y_i aus n Beobachtungen abgeleitet wurden. Ursprünglichen Daten bestehen aus n unabhängigen Beobachtungen der Form (t_i, z_i, x_i).
  - Bei GLM wurde als Gütemass Devianz D eingeführt als (-2) mal Log-Likelihood, $D = -2 \cdot \ell.\ D = \sum_{i=1}^n d_i$
  - «Elemente» d_i der Devianz als Grundlage Devianz Residuen für disk. Zeiten: $r_i^{(D)} := \sqrt{d_i}$ ohne Vorzeichen (GLM mit)
  - Zensierung ist in $r_i^{(D)}$ nicht vollständig berücksichtigt, da Zensierungsmechanismus, der in Zensierungskonstanten $\log\langle c_i\rangle$ festgehalten ist, weggelassen wurde. Definition für Residuen ungewöhnlich, da alle **Residuen positive Werte**.

library(discSurv); haz <- predict(**glm**, type = "response"); dR <- devResid(dataLong = **DL**, hazards = haz)
qqnorm(dR$Output$DevResid) #Darstellung der berechneten Devianz Residuen. Werden Praxis nicht weiter einsetzen

- **Adjustierte Devianz-Residuen**: mit einer **Transformation** adjustierten Residuen.
  - genähert **normalverteilt**, werden Normal-QQ-Plot zur Beurteilung, ob Modell Daten adäquat beschreibt, eingesetzt.
  - Beachten: Beobachtungen, bei denen eine **Hazardwahrscheinlichkeiten 0** geschätzt ist, zu ganz **groben Ausreissern** in den adjustierten Devianz-Residuen **führen**. adR <- adjDevResid(dataLong = **DL**, hazards = haz); qqnorm(adR$Output$AdjDevResid); qqline(adR$Output$AdjDevResid) #Ist Normalverteilung ersichtlich? → i.O.

- **Martingal-Residuen:** sind bei disk. Wartezeiten analog kontinuierlichen Wartezeiten definiert: $r_i^{(M)} = z_i - \sum_{s=1}^{t_i}\hat{\lambda}\langle s|\underline{x}_i\rangle$
  - Mit Hilfsgrössen y_{is} kann Martingal-Residuum schreiben: $r_i^{(M)} = \sum_{s=1}^{t_i} y_{i,s} - \hat{\lambda}\langle s|\underline{x}_i\rangle$. Folglich entspricht Martingal-Residuum Differenz zwischen Übergangsindikator y_i und geschätzten Eintrittswahrscheinlichkeit.
  - Martingal-Residuen nur Logit-Modell Summe 0, $\sum_{i=1}^n r_i^{(M)} = 0$, Wertebereich Residuen $[-1, 1]$, asymmetrisch um 0.
  - Falls Modell Daten adäquat beschreibt und entsprechend alle erkl. Var. enthält, dann sind Martingal-Residuen zufällig und unkorreliert mit erkl. Var.. Deshalb werden sie auch eingesetzt, um Effekte der erkl. Var. zu überprüfen.

mR <- martingaleResid(dataSetLong=**DL**, hazards=haz); plot(mR, covariates=c("**fin**", "**erkl. Var X**"), dataSetLong = DL)
Faktorvar. = Boxplot, sonst Streudiagramm + Glätter. Boxplot mean rechnen (da Median in Boxplot). aggregate(as.vector(mR), by = list(**fin** = **data$fin**), FUN = mean) #data=data normal, nicht lang, sum(as.vector(mR)) #Werte=Null → gut.