- Testing for log-linear model:
  - check LNp.3-10 cf. Consider two nested models S (s parameters) $\subset L$ (l parameters) — $1 \le s < l \le k$
  - Test statistic and null distribution D for H_0:S vs. H_1:$L\backslash S$

$-2\log\left[\frac{\mathcal{L}(\hat{\beta}_S)}{\mathcal{L}(\hat{\beta}_{L^*})}\times\frac{\mathcal{L}(\hat{\beta}_{L^*})}{\mathcal{L}(\hat{\beta}_L)}\right] = -2\log[\mathcal{L}(\hat{\beta}_S)/\mathcal{L}(\hat{\beta}_L)]$ likelihood ratio

$$0 \le D_S - D_L \Rightarrow \overset{a}{\sim} \chi^2_{df_S - df_L} \text{ (under } H_0)$$

$df_S - df_L = l-s$

D ↓ 0 when adding more effects; reject if large

where df_S=k−s and df_L=k−l

(cf., same asymptotic properties as for binomial GLM)

check LNp.3-13 cf. ⇔ profile likelihood confidence interval for individual parameter

- check LNp.3-11 cf. An alternative for testing significance of individual parameter: β_i

Wald test statistic for H_0: β_i=c vs. H_1: β_i≠c — Note. for Poisson response, $E(y_i)=Var(y_i)=\mu_i$

reject if $|z_i|$ large

$$z_i = (\hat{\beta}_i - c)/se(\hat{\beta}_i) \Rightarrow \overset{a}{\sim} N(0,1) \text{ (under } H_0)$$

from $(X^T\hat{W}X)^{-1}$ in IRWLS

check LNp.3-12 cf. ⇔ 100(1−α)% confidence interval: $\hat{\beta}_i \pm z(\alpha/2) \times se(\hat{\beta}_i)$

- cf. check LNp.3-32 Pearson X^2 statistic of S (an alternative goodness-of-fit measure):

each y_i ↔ count in a cell (check LNp.4-2)

$$X_S^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{k} \frac{(y_i - \hat{\mu}_{i,S})^2}{\hat{\mu}_{i,S}}$$

O_i: observed; E_i: expected under S

$=[(y_i-\hat{\mu}_i)/\sqrt{\hat{\mu}_i}]^2$ ← (Pearson) standardized residuals

- check LNp.3-33 cf. Pearson X^2 typically close in size to the deviance because:

$X_S^2 \overset{a}{\sim} \chi^2_{k-s}$ under S — asymptotically, i.e., y_i's large, $0 \approx$

$$(G^2=)\,D = 2\sum_{i=1}^{k}[y_i \log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)]$$

check LNp.4-6

$$= \cancel{2}\sum_{i=1}^{k}\{[\cancel{(y_i - \hat{\mu}_i)} + (y_i - \hat{\mu}_i)^2/(\cancel{2}\hat{\mu}_i) + \cdots] - \cancel{(y_i - \hat{\mu}_i)}\}$$

$$\approx \sum_{i=1}^{k} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} = X^2$$

y_i: O_i; $\hat{\mu}_i$: E_i

$f(x) = x\log\left(\frac{x}{x_0}\right) = 0 + (x - x_0) + \frac{1}{2x_0}(x-x_0)^2 + \cdots$

e.g., check LNp.3-33 for using X^2 in place of D in testing

check LNp.3-37 ⇒ Pearson X^2 often used in the same manner as the deviance

- cf. Overdispersion under a model S, i.e., $Var(y_{\mathbf{x},S}) \gg E(y_{\mathbf{x},S})$ → Then, $y_\mathbf{x}$ ≁ Poisson
  - check LNp.3-38 Overdispersion ⇒ large D_S (goodness-of-fit test rejected) ⇍ (not necessary)
  - cf. Other possible reasons causing large D_S should be examined before concluding there is an overdispersion. For example: (1) outliers, (2) wrong $\boldsymbol{X\beta}$ structure, … (too simple)

Then, $\lambda_\mathbf{x}$ is not a parameter, not fixed any more

  - check LNp.3-39 cf. A mechanism that can explain why overdispersion appears:

deficiency in the random part of GLM

$y_\mathbf{x}|\lambda_\mathbf{x} \sim$ Poisson($\lambda_\mathbf{x}$), but $\lambda_\mathbf{x}$ is a random variable

$\lambda_\mathbf{x}$: should be a function of only $\mathbf{x}$; assume fixed in Poisson GLM; e.g., using $1/\lambda_\mathbf{x}$

e.g., might be caused by some important unrecorded covariates

  - Example: tendency to fail for a product may vary from batch to batch even though they have same $\mathbf{x}$.

- Suppose that $\lambda_x \sim$ gamma distribution with $E(\lambda_x)=\mu_x$ and $Var(\lambda_x)=\mu_x/\phi_x$

$T\sim gamma(\alpha,\beta)$, pdf: $\frac{\beta^\alpha}{\Gamma(\alpha)}t^{\alpha-1}e^{-\beta t}$

$E(T)=\alpha/\beta=\mu$, $Var(T)=\alpha/\beta^2=\mu/\phi$, $\beta=\phi$, $\alpha=\mu\phi$

cf. beta-binomial method (LNp.3-42)

- Then (exercise), $y_x \sim$ negative binomial, $y_x = 0, 1, 2, \ldots,$ with $E(y_x)=\mu_x$ but $Var(y_x)=\mu_x\times(1+\phi_x)/\phi_x \geq \mu_x=E(y_x)$

$Y\sim NB(\alpha, \frac{\beta}{1+\beta})$, $r = \alpha$, $p = \frac{\beta}{1+\beta}$; $E(Y)=\frac{r}{p}-r$, $Var(Y)=\frac{r(1-p)}{p^2}$

check LNp.2-21; marginal dist.; # of 0's before observing rth 1; (LNp.2-20) $Y=X-r$, # of trials; ≥ 1

➢ For overdispersion cases, we can model y_x as:

- When mechanism known $\Rightarrow$ can model y_x as a negative binomial response (or other more flexible distribution)

Check LNp.4-14 for its analysis; e.g., might through quasi-likelihood

- When mechanism unknown $\Rightarrow$ can add a dispersion parameter σ^2, which is an unknown constant for all **x**'s, such that:

check LNp3-40~41 cf.; only one more parameter

$$Var(y_x) = \sigma^2\times E(y_x) = \sigma^2\mu_x$$

assume the ratio $\frac{Var(y_x)}{E(y_x)}=\sigma^2$ is irrelevant to x; What if ϕ_x's are constant over x? What if not? negative binomial approach cf.

- $\sigma^2 = 1 \Rightarrow$ the regular Poisson GLM; $\sigma^2 > 1 \Rightarrow$ overdispersion; $\sigma^2 < 1 \Rightarrow$ underdispersion

- Estimation of σ^2:

check LNp 3-41 cf.

$$\hat{\sigma}^2 = \frac{X^2}{k-s} = \frac{\sum_i\left[(y_i-\hat{\mu}_i)^2/\hat{\mu}_i\right]}{k-s}$$

$\because X^2 \overset{a}{\sim} \sigma^2\chi^2_{k-s}$ (D, too); s: dim(B) under S; $=\left(\frac{y_i-\hat{\mu}_i}{\sqrt{\hat{\mu}_i}}\right)^2 = (\text{Pearson residual})^2$

Better than using deviance when y_x's are small. If y_x's are large, not much difference

- Note: the estimation of β is unaffected by $\sigma^2 \Rightarrow$ choosing a dispersion parameter other than 1 has no effect on the estimation procedure (IRWLS) of β (cf., similar to Normal LM and binomial GLM)

weight $\propto [Var(y_x)]^{-1} = 1/\sigma^2\mu_x \propto 1/\mu_x$; same weight when $\sigma^2=1$

- The standard error of $\hat{\beta}_i$ should be adjusted for $\Rightarrow$ scale up the standard error by a factor of $\hat{\sigma}$

$\because Var(\hat{\beta}) = \sigma^2(X^T\hat{W}X)^{-1}$

- The z-statistic (Wald test) and its corresponding confidence interval should use the scaled $se(\hat{\beta}_i)$

LNp.4-7; use N(0,1) or a t-dist. as the null dist. & to get the critical value for C.I.

- When comparing two nested models, an F-test statistic:

check LNp.3-41 cf.

$$\frac{(D_S - D_L)/(df_S - df_L)}{\hat{\sigma}^2_L} \overset{a}{\sim} F_{df_S-df_L,\, df_L} \quad (\text{under } H_0{:}S)$$

$\hat{\sigma}^2_L$: X^2_L/df_L; $\because D_S - D_L \overset{a}{\sim} \sigma^2\chi^2_{df_S-df_L}$ under S

should be used, rather than the chi-square test (LNp.4-7)

Recall. Hauck-Donner effect (LNp.3-12) $se(\hat{\beta}_i)$ over-estimated

- The F-test is more reliable than the z-statistic

based on deviance (likelihood ratio test); based on Wald test

- No goodness-of-fit test is possible. $\because$ no model free estimate of σ^2

- Prediction of μ_x at $\mathbf{x}_0 = (x_{01}, \ldots, x_{0m})^T$ under a log-linear model:

➢ Denote $\mathbf{h}_0 = (h_1(\mathbf{x}_0), \ldots, h_p(\mathbf{x}_0))^T$

$E(y_x)$