Residuals for GLMs

See pages 54 to 56 of the lecture notes.

Pearson residuals

$$r_i = \frac{w_i e_i}{\sqrt{V(\hat{Y}_i)}}$$

deviance residuals

$$d_i = \begin{cases} w_i\sqrt{D_i} & \text{if } Y_i > \hat{Y}_i \\ -w_i\sqrt{D_i} & \text{if } Y_i < \hat{Y}_i \end{cases}$$

where D_i is shown in table on page 56 for the different types of GLM

The deviance residuals (unlike the Pearson residuals) when squared & summed equal the residual deviance

$$\sum d_i^2 = D(\hat{Y}, Y) \neq \sum r_i^2$$

Note that as we did with ordinary linear models we could standardise by noting that

$$Var(r_i) \approx Var(d_i) \approx \hat{\phi}(1-h_{ii})$$

these are definitely approx. for GLMs

We could also produce deletion residuals (ie excluding the effects of the current observation), but see discussion on page 64 of the notes.

Drop-in-deviance tests for Gamma GLMs

Can we still do these? Yes, but $\phi_{Assumed} \neq 1$, so we need to scale these tests first! (actually equals the CV estimator $\hat{\phi}_{cv}$)

eg Leukaemia data

Mean model $E[g(Y)] = \beta_0 + \beta_1 X_1$

"Variance" model Errors are independent identically Gamma distributed with constant dispersion $\hat{\phi} = \frac{1}{\alpha}$

where $g()$ is the link function, $\log()/\ln()$ or inverse

Y is the surv time, X is the $\log(wbc)$

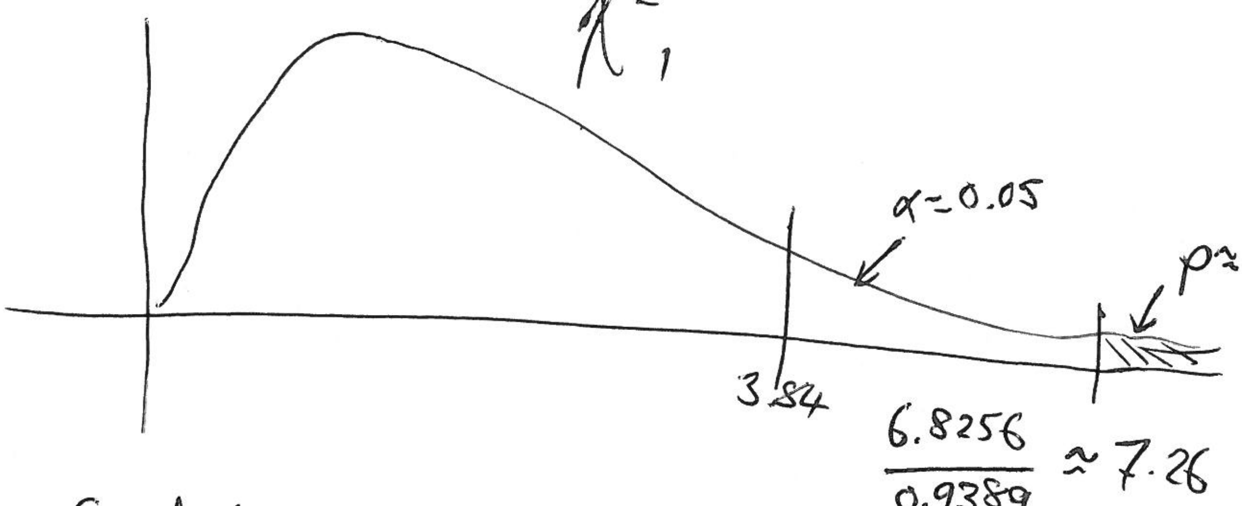
Drop-in-deviance test on the addition of $\beta_1 X_1$ to an intercept only model:

$$H_0 : \beta_1 = 0 \quad vs \quad H_A : \beta_1 \neq 0$$

test statistic $\frac{\Delta D}{\hat{\phi}_{cv}}$ = scaled drop-in-deviance $\sim \chi^2_1$

drop-in-deviance

assumed dispersion



Conclusion as $p << \alpha = 0.05$ reject H_0 & conclude $H_A : \beta_1 \neq 0$

ie $\log(wbc)$ is a significant predictor of surv.

Goodness of fit test for Gamma GLMs

$$H_0 : \phi = \phi_{Assumed} \quad \text{vs} \quad H_A : \phi \neq \phi_{Assumed}$$

as calculated by R – the CV estimator

The alternative estimated ϕ is the same as before

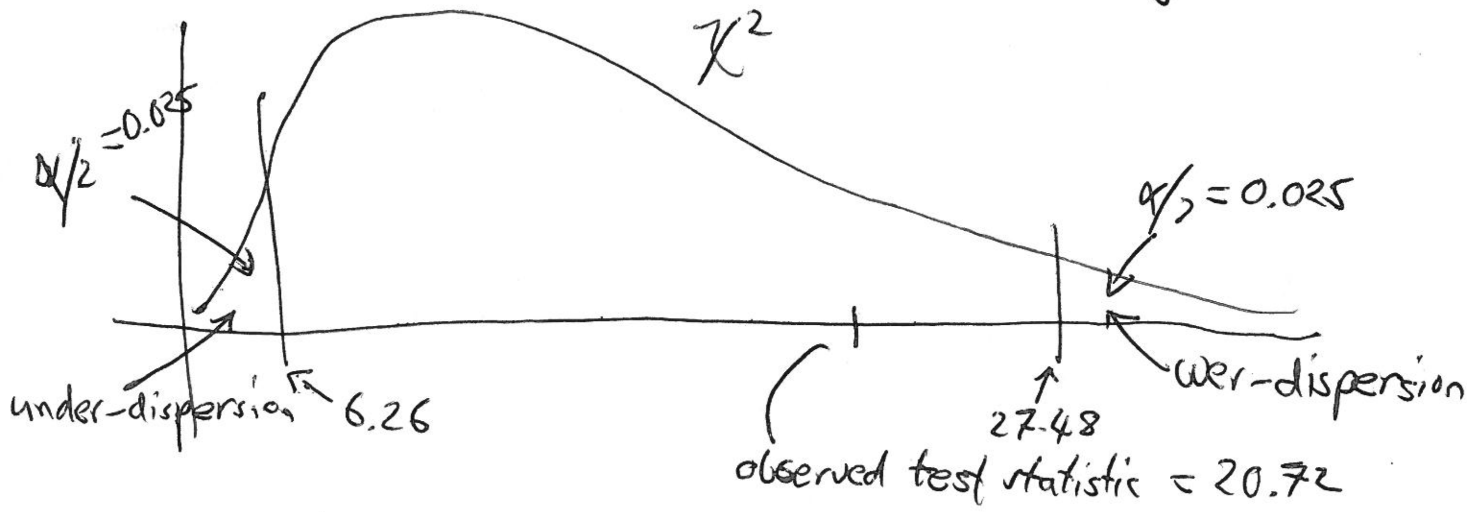
$$\hat{\phi}_{Alt} = \frac{\sum d_i^2}{n-p} \quad \begin{matrix} \leftarrow \text{residual deviance} \\ \leftarrow \text{residual df} \end{matrix}$$

But, to use this in the goodness of fit test, we must first scale!

ie test statistic

$$\frac{\sum d_i^2}{\hat{\phi}_{CV}} \quad \begin{matrix} \leftarrow \text{residual deviance} \\ \leftarrow \text{model\$dispersion} \end{matrix}$$

$$\sim \chi^2_{n-p} \leftarrow \text{residual df}$$



as $6.26 < 20.72 < 27.48$ do not reject H_0
ie model is not under or over-dispersed

Note caveat on page 57 & check fitted values not close to 0.