

Tutorial 10

STAT3015/4030/7030 Generalised Linear Modelling

The Australian National University

Week 10, 2017

Overview

1 Summary

Studentised residuals

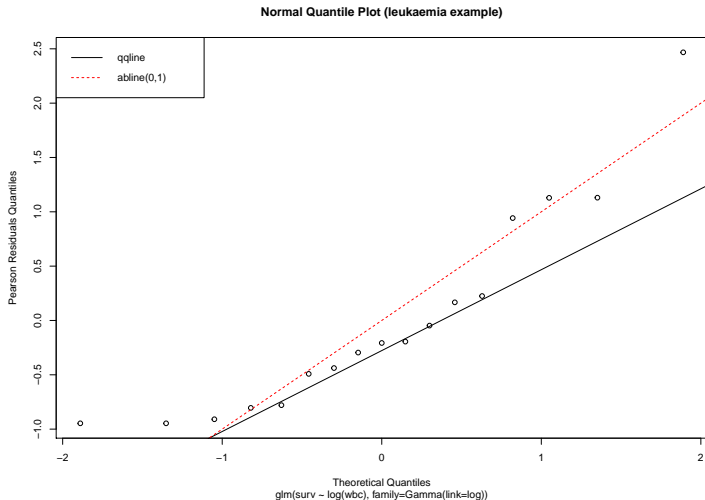
- $Var(r_i) \approx Var(d_i) \approx \hat{\phi}(1 - h_{ii})$
- Since the variance are approximate values, these Studentised residuals are rarely used in this course.
- Example in “R Example: Residual Plots” and Page 64 of the brick.

```
residuals(model)/sqrt(summary(model)$dispersion*  
(1-influence(model)$hat))
```

Deletion residuals

- $r_i^* = \frac{Y_i - \hat{Y}_{i,-i}}{\sqrt{V(\hat{Y}_{i,-i})}}$
- Similar idea to the so-called PRESS residuals of multiple linear regression
- Relevant code in “R Example: Outliers”
- Used for assessing outliers

Normal QQ plot



Normal QQ plot

- Both lines are only approximate theoretical reference lines.
- If we can reasonably expect the scales of the X and Y axes, i.e., the assumption of normality is fine, then `abline(0,1)` will work.
- `help(qqline)` tells us the added line passing through (by default) the first and third quartiles.
- Pearson residual ($r_i = \frac{Y_i - \hat{Y}_i}{\sqrt{V(\hat{Y}_i)}}$) has an asymptotic normal distribution as $n \rightarrow \infty$.
→ we will revisit this in Pearson Chi-square test later.

Half-normal plot

- A useful technique for judging whether some cases in a set of positive observations are unusually extreme is the **half-normal plot**.
- Plot the **sorted values** against $\Phi^{-1}(\frac{n+i}{2n+1})$ which represent the quantiles of the upper half of a standard normal distribution.
- Not looking for a straight line relationship since we do not necessarily expect a positive normal distribution for quantities like the **leverages**.
- We are looking for outliers, which will be apparent as points that diverge substantially from the rest of the data.

Half-normal plot

- A useful technique for judging whether some cases in a set of positive observations are unusually extreme is the **half-normal plot**.
- Plot the **sorted values** against $\Phi^{-1}(\frac{n+i}{an+1})$ which represent the quantiles of the upper half of a standard normal distribution.
- Not looking for a straight line relationship since we do not necessarily expect a positive normal distribution for quantities like the **leverages**.
- We are looking for outliers, which will be apparent as points that diverge substantially from the rest of the data.

Half-normal plot

- A useful technique for judging whether some cases in a set of positive observations are unusually extreme is the **half-normal plot**.
- Plot the **sorted values** against $\Phi^{-1}\left(\frac{n+i}{2n+1}\right)$ which represent the quantiles of the upper half of a standard normal distribution.
- Not looking for a straight line relationship since we do not necessarily expect a positive normal distribution for quantities like the **leverages**.
- We are looking for outliers, which will be apparent as points that diverge substantially from the rest of the data.

Half-normal plot

- A useful technique for judging whether some cases in a set of positive observations are unusually extreme is the **half-normal plot**.
- Plot the **sorted values** against $\Phi^{-1}(\frac{n+i}{2n+1})$ which represent the quantiles of the upper half of a standard normal distribution.
- Not looking for a straight line relationship since we do not necessarily expect a positive normal distribution for quantities like the **leverages**.
- We are looking for outliers, which will be apparent as points that diverge substantially from the rest of the data.

Half-normal plot

- Examples of using half-normal plots to seek outliers are contained in “Extra Binomial GLM” and “Extra Poisson GLM” on Wattle and also in Question 3 of this week’s tutorial.
- Those previously learned leverage barplots and Cook’s distance plots are still useful in GLM settings. See “R Examples: Outliers” example.

Confidence/Prediction intervals

- Normally we want CI/PI of Y or “response”.
- For relevant codes of calculating CI and PI see “R Example: Predictions using GLMS” on Wattle.
- Note CIs in Q2 of this week’s tutorial are intervals for estimated parameters.