Analysis of Residuals[1]

STA302 Fall 2017

[1] See last slide for copyright information.

Residual means left over: $e_i = y_i - \widehat{y}_i$

- Vertical distance of y_i from the regression hyper-plane
- An error of "prediction."
- Big residuals merit further investigation.
- Big compared to what?
- They are normally distributed.
- Consider standardizing.
- Maybe detect outliers.
- Plots can also be informative.

Residuals are like estimated error terms

$$e_i = y_i - \widehat{y}_i \quad \Leftrightarrow \quad y_i = \widehat{y}_i + e_i$$

$$\begin{aligned} y_i &= \widehat{y}_i + e_i \\ &= b_0 + b_1 x_{i,1} + \cdots + b_k x_{i,k} + e_i \\ &= \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k} + \epsilon_i \end{aligned}$$

Normal distribution of ϵ_i implies normal distribution of e_i, but the e_i are not independent, and they do not have equal variance.

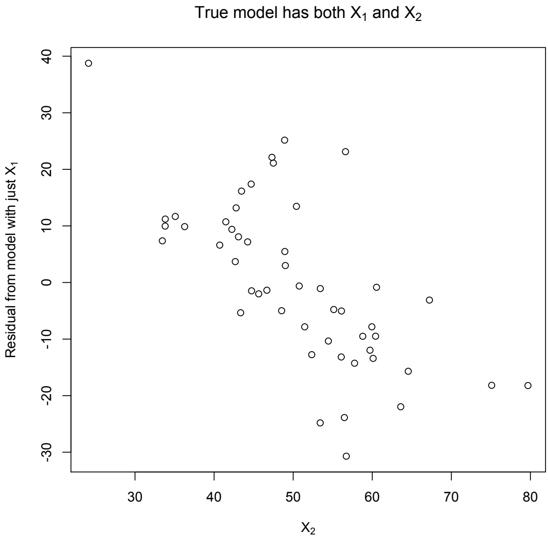
Data = Fit + Residual

$$y_i = \widehat{y}_i + e_i$$

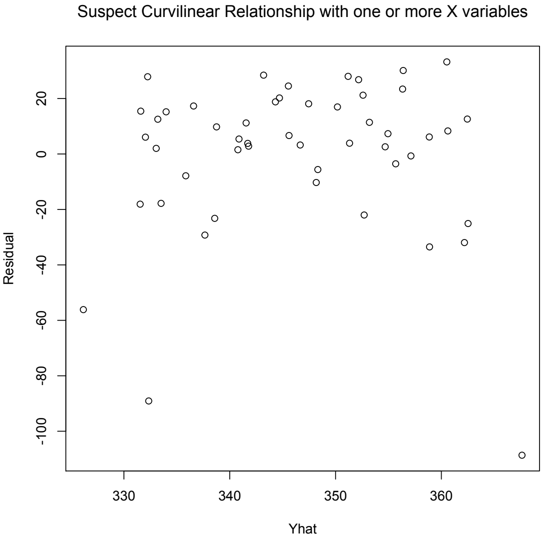
Plotting residuals can be helpful

- Against predicted y.
- Against explanatory variables not in the equation.
- Against explanatory variables in the equation.
- Look for serious departures from normality.

Plot Residuals Against Explanatory Variables Not in the Equation

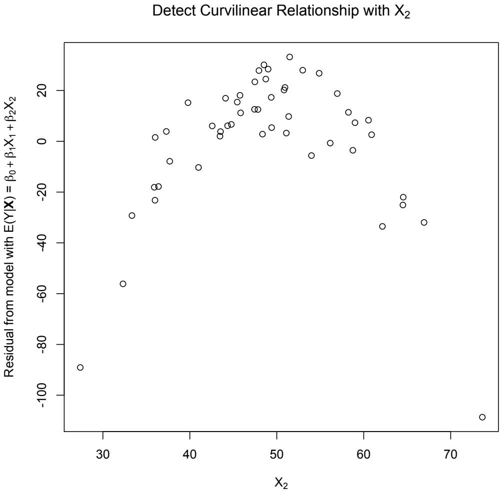


Plot Residuals Against $\widehat{y}$



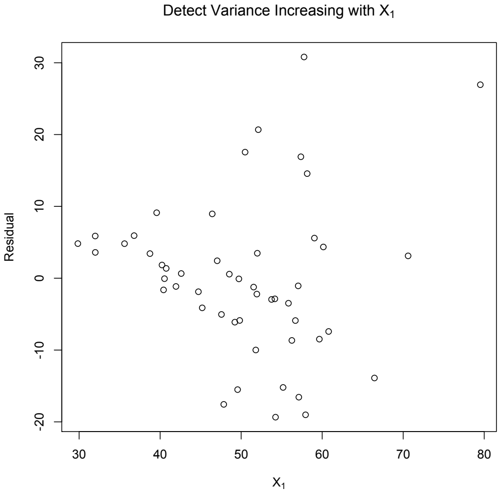
Plot Residuals Against Explanatory Variables in the Equation

Plot versus X_1 showed nothing



Plot Residuals Against Explanatory Variables in the Equation

Can show non-constant variance



Outlier detection

- Big residuals may be outliers. What's "big?"
- Consider standardizing.
- Could divide by square root of sample variance of $e_1, \ldots, e_n$.
- Semi-Studentized: Estimate $Var(e_i)$ and divide by square root of that: $\frac{e_i}{\sqrt{s^2\,(1-h_{i,i})}}$
- In R, this is produced with `rstandard`.

Studentized deleted residuals
The idea

- An outlier will make s^2 big.
- In that case, the standardized (Semi-Studentized) residual will be too small – less noticeable.
- So calculate $\widehat{y}$ for each observation based on all the other observations, but not that one.
- Predict each observed y based on all the others, and assess error of prediction (divided by standard error).
- Big values suggest that the expected value of y_i is not what it should be.

Apply prediction interval technology

$$t = \frac{y_0 - \mathbf{x}_0'\mathbf{b}}{\sqrt{s^2(1 + \mathbf{x}_0'(X'X)^{-1}\mathbf{x}_0)}} \sim t(n-k-1)$$

- Note that y_i is now being called y_0.
- If the "prediction" is too far off there is trouble.
- Use t as a test statistic.
- Need to change the notation.

Studentized deleted residual

$$e_i^* = \frac{y_i - \mathbf{x}_i' \mathbf{b}_{(i)}}{\sqrt{s_{(i)}^2 (1 + \mathbf{x}_i' (X_{(i)}' X_{(i)})^{-1} \mathbf{x}_i)}} \sim t(n-k-2)$$

- In R, this is produced with `rstudent`.
- There is a more efficient formula.
- Use e_i^* as a test statistic of $H_0 : E(y_i) = \mathbf{x}_i' \boldsymbol{\beta}$.
- If H_0 is rejected, investigate.
- We are doing n tests.
- Type I errors are very time consuming and disturbing.
- If independent, probability of no false positives would be $(1-\alpha)^n \rightarrow 0$.
- But they are not independent.
- How about a Bonferroni correction?

Bonferroni Correction for Multiple Tests

- The curse of a thousand t-tests.
- If the null hypotheses of a collection of tests are all true, hold the probability of rejecting one or more to less than $\alpha = 0.05$.
- Based on Bonferroni's inequality:

$$Pr\left\{\cup_{j=1}^{r} A_j\right\} \leq \sum_{j=1}^{r} Pr\{A_j\}$$

- Applies to any collection of r tests.
- Assume all r null hypotheses are true.
- Event A_j is that null hypothesis j is rejected.
- Do the tests as usual, obtaining r test statistics.
- For each test, use the significance level α/r instead of α.

Use the significance level α/r instead of α

Bonferroni Correction for r Tests

Assuming all r null hypotheses are true, probability of rejecting at least one is

$$
\begin{aligned}
Pr\left\{\cup_{j=1}^{r} A_j\right\} &\leq \sum_{j=1}^{r} Pr\{A_j\} \\
&= \sum_{j=1}^{r} \alpha/r \\
&= \alpha
\end{aligned}
$$

Just use critical value(s) for α/r instead of α.

Advantages and disadvantages of the Bonferroni correction

- Advantage: Flexibility — Applies to any collection of hypothesis tests.
- Advantage: Easy to do.
- Disadvantage: Must know what all the tests are before seeing the data.
- Disadvantage: A little conservative; the true joint significance level is less than α.

Application to Studentized deleted residuals

- There are $r = n$ tests, one for each observed $i = 1, \dots, n$.
- Use the critical value $t_{\frac{\alpha}{2n}}(n-k-2)$.
- Even for large n it is not overly conservative.

Copyright Information