MS Module 20: Residuals and Standardized Residuals (overview 3rd edition)

(The attached PDF file has better formatting.)

(Readings from the third 3rd edition of the Devore, Berk, and Carlton text.)

Reading: §12.6 Investigating Model Adequacy: Residual Analysis.

The residual is the observed value minus the fitted value. The standard deviation is the same for all points in a distribution, so the procedure to standardize the distribution is the same at all points; see table 12.2. this assumption simplifies the statistical model, though it is rarely precise; points with higher expected values often have higher standard deviations.

For classical regression analysis, the variance of the error term is the same at all points. But the fitted value is more strongly drawn to the observed value at outlying (influential) points, so the variance of the residual is greatest at the mean and lower at outlying points. Know expression 12.11.

You will grasp this relation most quickly by forming regression equations with Excel's regression add-in.

- Chose X-values of {1, 2, 3, 4, 5, 6, 7, 8, 9}.
- Chose any nine Y-values, and run the regression add-in.
- The add-in computes β_1 and the residuals at each point.

Now change the Y value by *one unit* at one of the points.

- If you change the Y value at X=5, β_1 doesn't change but the residual at X=5 changes by 1/N.
- If you change the Y value at X=9, β_1 changes at lot but the at X=9 changes by less than 1/N.

This relation seems counterintuitive at first, but it makes sense. Changing the Y value at X=5 doesn't change the regression line; changing the Y value at X=9 makes the regression line follow the new value, reducing the residual.

The standardized residual is the residual divided by its standard deviation, which differs for each residual. The arithmetic is similar to that for confidence intervals and prediction intervals, but the signs differ: a larger width of the prediction interval is associated with a smaller standard deviation of the residual.

Final exam problems may give the set of *x* values, the residuals at two points, and the standardized residual at one of these points, and it may ask to derive the standardized residual at the other point.

Know the diagnostic plots and their uses, and difficulties with the regression analysis that they indicate.

Figure 12.24 "Scatterplots and residual plots" shows how (a) violation of the linearity assumption and (b) violation of the constant variance assumption are visualized. Figure 12.25 shows plots with no violations. Figure 12.26 shows three more plots with potential violations. These plots help you understand the concepts.

Skip the section "Remedies for Assumption Violations." These remedies are not tested on the final exam.

Review end of chapter exercises 68, 69 (the final exam tests the calculations, not the plots).

Exercise 75, based on Frank Anscombe, "Graphs in Statistical Analysis," shows the importance of graphs. Anscombe's article is posted on the discussion forum; this article is worth reading, though it is not tested on the final exam.

Take heed: The standardized residuals computed by the Regression add-in in Excel's Analysis Pack differ from the standardized residuals in the textbook. See the discussion at the web page:

https://stats.stackexchange.com/questions/166533/how-exactly-are-standardized-residuals-calculated

(The discussion is attached as a PDF file to this posting.)