STAT 350: Lecture 14

Reading: Chapter 3, Chapter 9.
Model Assessment and Residual Analysis:

Standardized / studentized residuals

Recall that

$$\hat{\epsilon} \sim MVN(0, \sigma^2(I - H))$$

where $H = X(X^T X)^{-1} X^T$. It follows that

$$\hat{\epsilon}_i \sim N(0, \sigma^2(1 - h_{ii}))$$

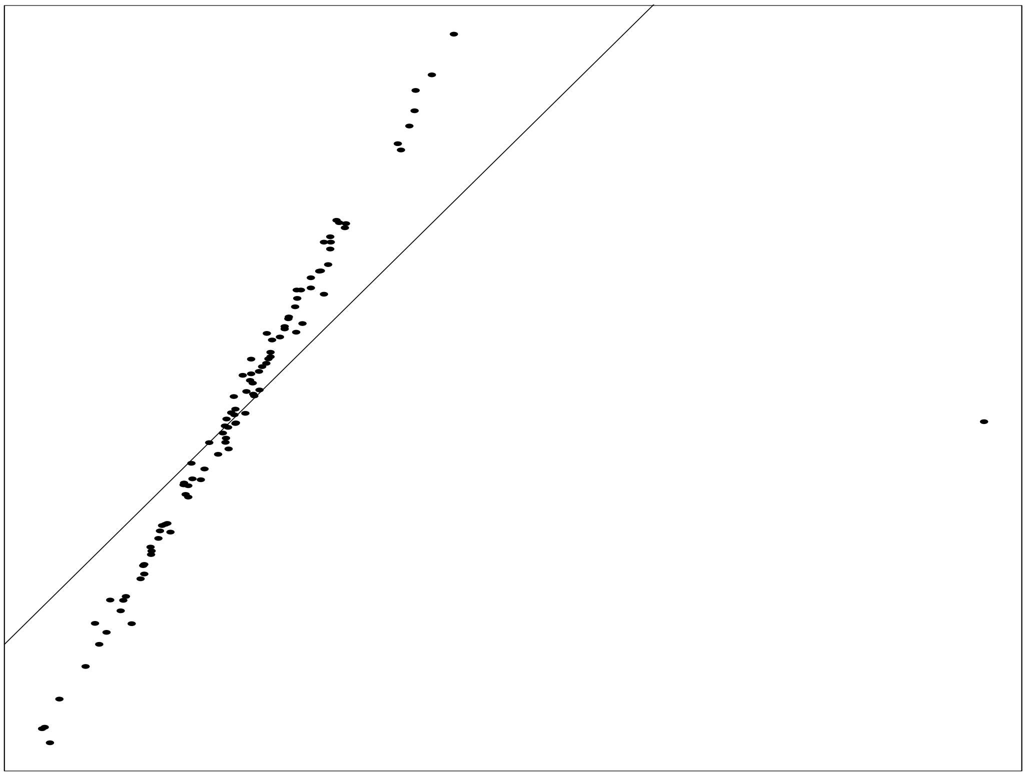
where h_{ii} is the iith diagonal entry in H. The quantity

$$\frac{\hat{\epsilon}_i}{\hat{\sigma}}$$

is called a **standardized** residual. It ignores the leverages which are small for most observations. In judging the quality of a model, however, observations with large leverages are particularly important so it is now standard to adjust for them. We sometimes use

$$\frac{\hat{\epsilon}_i}{\hat{\sigma}\sqrt{1 - h_{ii}}} \sim N(0, 1)$$

which is called an **internally studentized** residual.
Critique: when the model is wrong a bad data point can inflate $\hat{\sigma}$ leaving the internally studentized residual small.



Suggestion:

$$Y_i - \hat{Y}_{i(i)}$$

where $\hat{Y}_{i(i)}$ is the fitted value using all the data **except** case i. This residual is called a "PRESS prediction error for case i". The acronym PRESS stands for Prediction Sum of Squares.

But: $Y_i - \hat{Y}_{i(i)}$ must be compared to other residuals or to σ so we suggest **Externally Studentized Residuals** which are also called **Case Deleted Residuals**:

$$\frac{\hat{\epsilon}_{i(i)}}{\text{est'd SE not using case } i} = \frac{Y_i - \hat{Y}_{i(i)}}{\text{Case } i \text{ deleted SE of numerator}}$$

Apparent problem: If $n = 100$ do I have to run SAS 100 times? NO.

FACT 1:

$$Y_i - \hat{Y}_{i(i)} = \frac{\hat{\epsilon}_i}{1-h_{ii}}$$

Jargon: h_{ii} is the **leverage** of point i. If h_{ii} is large then

$$\left|\frac{\hat{\epsilon}_i}{1-h_{ii}}\right| >> |\hat{\epsilon}_i|$$

and point i influences the fit strongly.
FACT 2:

$$\mathrm{Var}\left(\frac{\hat{\epsilon}_i}{1-h_{ii}}\right) = \frac{\sigma^2}{1-h_{ii}} \left(= \frac{\sigma^2(1-h_{ii})}{(1-h_{ii})^2}\right)$$

The Externally Standardized Residual is

$$\frac{\hat{\epsilon}_i/(1-h_{ii})}{\sqrt{\mathrm{MSE}_{(i)}/(1-h_{ii})}} = \frac{\hat{\epsilon}_i}{\sqrt{\mathrm{MSE}_{(i)}(1-h_{ii})}}$$

where

$$\mathrm{MSE}_{(i)} = \text{ estimate of } \sigma^2 \text{ not using data point } i$$

Fact:

$$\mathrm{MSE} = \frac{(n-p-1)\mathrm{MSE}_{(i)} + \hat{\epsilon}_i^2/(1-h_{ii})}{n-p}$$

so the externally studentized residual is

$$\hat{\epsilon}_i\sqrt{\frac{n-p-1}{\mathrm{ESS}(1-h_{ii})-\hat{\epsilon}_i^2}}$$

Distribution Theory of Externally Standardized Residuals

1. $\hat{\epsilon}_{(i)}/\sqrt{\mathrm{Var}(\hat{\epsilon}_i)} \sim N(0,1)$

2. $$\frac{(n-p-1)\mathrm{MSE}_{(i)}}{\sigma^2} \sim \chi^2_{n-p-1}$$

3. These two are independent.

4. SO:

$$\begin{aligned} t_i = & \frac{(n-p-1)\mathrm{MSE}_{(i)}}{\sigma^2} \sim \chi^2_{n-p-1} \\ & \sim t_{n-p-1} \end{aligned}$$

Example: Insurance Data

Cubic Fit:

Year	$\hat{\epsilon}_i$	Internally Studentized	PRESS	Externally Studentized	Leverage
1975	1.17	0.59	1.54	0.56	0.24
1980	-1.09	-1.15	-6.20	-1.19	0.82

Note the influence of the leverage.
Note that edge observations (1980) have large leverage.
Quintic Fit:

Year	$\hat{\epsilon}_i$	Internally Studentized	PRESS	Externally Studentized	Leverage
1978	0.82	1.79	1.43	3.48	0.43
1980	0.08	1.02	4.79	1.03	0.98

Notice 1978 residual is unacceptably large.
Notice 1980 leverage is huge.

Formal assessment of Externally Standardized Residuals

1. Each residual has a t_{n-p-1} distribution. For example, for the quintic, $t_{10-7,0.025} = 3.18$ is the critical point for a 5% level test.

2. But there are 10 residuals to look at. This leads to a multiple comparisons problem. The simplest multiple comparisons procedure is the Bonferoni method: divide α by the number of tests to be done, 10 in our case giving $0.025/10 = 0.0025$. The corresponding critical point is

$$t_{3,0.0025} = 7.45$$

   so none of the residuals are significant.

3. For the cubic model

$$t_{5,0.0025} = 4.77$$

   and again all the residuals are judged ok.

Reparametrization

Suppose X is the design matrix of a linear model and that X_1 is the design matrix of the linear model we get by imposing some linear restrictions on the model using X. A good example is on assignment 3 but here is another. Consider the one way layout, also called the K sample problem. We have K samples from K populations with means ν_i for $i = 1, \ldots, K$. Suppose the ith sample size is $n + i$. This is a linear model, provided we are able to assume that all the population variances are the same.

The resulting design matrix is

$$X = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ & & \vdots & & \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ & & \vdots & & \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

where there are n_1 copies of the first row, $[1, 0, \ldots, 0]$ and then n_2 copies of $[0, 1, 0, \ldots, 0]$ and so on. Under the restriction $\nu_1 = \cdots = \nu_K$ there is just one parameter, say η, and the design matrix X^* is just a column of $n_1 + \cdots + n_K$ 1s. Notice that X^* is **not** a submatrix of the original design matrix X. However, if A is the $K \times 1$ matrix with all entries equal to 1 we have $X^* = XA$ so that the column space of X^* is a subspace (of dimension 1) of the K dimensional column space of X. The extra sum of squares principle can thus be used to test the null hypothesis $H_o : \nu_1 = \cdots = \nu_K$ by fitting the two models and computing, using the notation of ANOVA from STAT 330:

$$F = \frac{[\sum(Y_{ij} - \bar{Y}_{\cdot\cdot})^2 - \sum(Y_{ij} - \bar{Y}_{i\cdot})^2]/(K-1)}{\sum(Y_{ij} - \bar{Y}_{i\cdot})^2/(\sum n_i - K)}$$

The numerator of this sum simplifies algebraically to

$$\sum_i n_i(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2/(K-1)$$

so that this extra Sum of Squares F test is the usual F test in this problem; this is universal.

It is actually quite common to reparametrize the full model in such a way that the null hypothesis of interest is of the form $\beta_{i_1} = \beta_{i_2} = \cdots = 0$. For the 1 way ANOVA there are two such reparametrizations in common use.

The first of these defines a grand mean parameter $\bar{\nu} = E(\bar{Y}_{\cdot\cdot}) = \sum n_i\nu_i / \sum n_i$ and individual "effects" $\alpha_i = \nu_i - \bar{\nu}$. This new model has $K+1$ parameters apparently and the corresponding design matrix, X_1, would not have full rank; its rank would be K although it would have $K+1$ columns. As such the matrix $X_1^T X_1$ would be singular and we could not find unique least squares estimates. The problem is that we have defined the parameters in such a way that there is a linear restriction on them, namely, $\sum n_i\alpha_i = 0$. We get around this problem by dropping α_K and remembering in our model equations that $\alpha_K = -\sum_{i=1}^{K-1} n_i\alpha_i / n_K$.

If you now write out the model equations with $\bar{\nu}$ and the α_i as parameters you get the design matrix

$$X_2 = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ & & \vdots & & \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ & & \vdots & & \\ 1 & -\frac{n_1}{n_k} & -\frac{n_2}{n_K} & \cdots & -\frac{n_{K-1}}{n_K} \end{bmatrix}.$$

Students will have seen this matrix in 330 in the case where all the n_i are the same and the fractions in the last n_K rows of X_2 are all equal to -1. Notice that the hypothesis $\nu_1 = \cdots = \nu_K$ is the same as $\alpha_1 = \cdots = \alpha_{K-1} = 0$.

The other reparametrization is "corner-point coding" where we define new parameters by $\beta_1 = \nu_1$ and $\beta_i = \nu_i - \nu_1$. For this parameterization the null hypothesis of interest is $\beta_2 = \cdots = \beta_K = 0$. The design matrix is

$$X_3 = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ & & & \vdots & \\ 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ & & & \vdots & \\ 1 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$