

Math 218 Mathematical Statistics

Prof. D. Joyce, Clark University

27 Mar 2009

Monday and Wednesday. Presentations by you. Exercises 4, 5, 6, or 7 from chapter 10, pages 387–389.

Coming up. Presentations from exercises beginning on page 390: 13, 16, 17, 18, 19, 20, 21, 22, 23. Individuals or pairs of students will select one problem for presentation (no two pairs doing the same problem).

Last time. Estimating the error variance σ^2 of the model. We looked at statistical inferences based on the simple linear regression model.

Summary. The model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ has three parameters: β_0 , β_1 , and σ^2 (the variance of each ϵ_i). Our three estimators are

- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$. It's an unbiased estimator of β_0 with standard deviation $SD(\hat{\beta}_0) = \sigma \sqrt{\frac{\sum x_i^2}{n S_{xx}}}$.
- $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$. It's an unbiased estimator of β_1 with standard deviation $SD(\hat{\beta}_1) = \frac{\sigma}{\sqrt{S_{xx}}}$.
- $S^2 = \frac{SSE}{n-2}$ is an unbiased estimator of σ^2 . $\frac{(n-2)S^2}{\sigma^2} = \frac{SSE}{\sigma^2}$ is a χ^2 distribution with $n-2$ degrees of freedom

where, as always,

$$\begin{aligned} S_{xy} &= \sum (x_i - \bar{x})(y_i - \bar{y}) \\ &= \sum x_i y_i - n \bar{x} \bar{y} \\ S_{xx} &= \sum (x_i - \bar{x})^2 \end{aligned}$$

$$\begin{aligned} S_{yy} &= \sum x_i^2 - n \bar{x}^2 \\ &= \sum (y_i - \bar{y})^2 \\ &= \sum y_i^2 - n \bar{y}^2 \end{aligned}$$

and

$$\begin{aligned} SSE &= \sum e_i^2 = \sum (y_i - \hat{y}_i)^2 \\ SST &= S_{yy} \\ SSR &= \sum (\bar{y} - \hat{y}_i)^2 \end{aligned}$$

Today. Confidence and prediction intervals for simple linear regression. In the model for simple linear regression

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

different values of x produce different predictions for $Y = \beta_0 + \beta_1 x + \epsilon$. After getting n data values, we compute the least squares line

$$y = \hat{\beta}_0 + \hat{\beta}_1 x,$$

and we can determine confidence intervals for the parameters β_0 , β_1 , and σ^2 .

But, we can do more. Suppose we set the predictor variable x to some specified value, x^* . (It would also be reasonable to denote the new values with a subscript of $n+1$ rather than a superscript of $*$.) That introduces a new Y value

$$Y^* = \beta_0 + \beta_1 x^* + \epsilon^*$$

where ϵ^* is a new independent error random variable with the same normal distribution as the other ϵ_i 's, namely, $\text{NORMAL}(0, \sigma^2)$. Therefore, the random variable Y^* is $\text{NORMAL}(\beta_0 + \beta_1 x^*, \sigma^2)$.

Our model gives a predicted value of Y^* :

$$\hat{y}^* = \hat{\beta}_0 + \hat{\beta}_1 x^*.$$

We'll denote the mean of Y^* by μ^* , thus

$$\mu^* = E(Y^*) = \beta_0 + \beta_1 x^*.$$

We don't know what μ^* is since we don't know β_0 and β_1 , but we have a predicted value for it, which is the same as the predicted value for Y^* :

$$\hat{\mu}^* = \hat{\beta}_0 + \hat{\beta}_1 x^*.$$

Although these predicted values are the same, when we use them to determine intervals for what they predict, namely, y^* and μ^* , respectively, we'll get different width intervals since the variances are much smaller for the mean.

A *prediction interval* (PI) for Y^* at α is an interval centered at \hat{y}^* such that the probability that Y^* lies in that interval is $100(1 - \alpha)\%$. It works out to be that its endpoints are

$$\hat{y}^* \pm t_{n-2, \alpha/2} s \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}$$

where $s = \sqrt{\text{MSE}}$ is the estimator of σ as we've saw before.

This interval is closely related to the confidence interval for $\mu^* = \beta_0 + \beta_1 x^*$. At the significance level α , it has endpoints

$$\hat{\mu}^* \pm t_{n-2, \alpha/2} s \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}.$$

If you graph the endpoints of either of these two intervals with x^* on the x -axis, you'll get a pair of hyperbolas, one above the least squares line, the other equally far below it. When x^* is close to \bar{x} , the hyperbolas are close, and that indicates that the prediction interval for Y^* and the confidence interval for μ^* are shorter there. But when x^* is far from \bar{x} , the hyperbolas spread apart, so the intervals there are large. In other words, good predictions can be made near \bar{x} , not so good predictions far from \bar{x} .

The prediction interval for the next observation Y^* is, of course, much larger than the confidence interval for the mean μ^* . Graphically, the prediction hyperbola for Y^* are much further away from the least squares line than the hyperbola for μ^* .

Regression diagnostics. A basic question is: do the data support the hypotheses necessary for the simple linear regression model? A preliminary test we've already seen is to make a scatter plot of the data $(x_1, y_1), \dots, (x_n, y_n)$. If it's obvious that the plot is nonlinear, then maybe the model is not appropriate.

Other tests can be made after fitting the least squares line, and some of these depend on making a scatter plot of the residuals $e_i = y_i - \hat{y}_i$, that is the plot of $(x_1, e_1), \dots, (x_n, e_n)$.

If the hypotheses for the regression model is correct, then the e_i 's are normally distributed with mean 0 and variance close to (but a little less than) σ^2 . They are not independent since $\sum e_i = 0$ and $\sum x_i e_i = 0$ as shown in the text.

If the residual scatter plot shows some pattern, then the simple linear model may not be the best. That's not necessarily a bad thing as the pattern may indicate better models.

For instance, in the tread wear example in the text, there's a clear parabolic form to the residual plot. That suggests a quadratic model might be better. We'll see after we've introduced multiple linear regression that quadratic models can be subsumed in those models, so we'll see this example later.

There are other things that the residual plot can show, but it may take a large n to see them. The model assumes that the errors ϵ_i have the same variance for all i . If in the plot it appears that the residuals are close to 0 at one end, but scattered far from 0 at the other, then a transformation may be needed before applying the model, that is, some linearizing transformation needs to be applied to the y -values to make the variance of the resulting errors more uniform across all the x -values. Examples in the text use the log function and the reciprocal function as transformations.