Stat 400, section 12.3 Inferences about the Slope Parameter β_1

notes by Tim Pilachowski

In virtually all of our inferential work thus far (beginning with the Central Limit Theorem in section 5.4), the notion of sampling variability has been pervasive. In particular, properties of sampling distributions of various statistics have been the basis for developing confidence interval formulas and hypothesis-testing methods. The key idea here is that the value of any quantity calculated from sample data – the value of any statistic – will vary from one sample to another.

From section 12.1 – Definition: The Simple Linear Regression Model
There are parameters β_0, β_1, and σ^2, such that for any fixed value of the independent variable *x*, the dependent variable is a random variable related to *x* through the **model equation** $Y = \beta_0 + \beta_1 x + \in$.
The quantity $\in$ in the model equation is a random variable, assumed to be normally distributed with $E(\in) = 0$ and $V(\in) = \sigma^2$.

In the text's Example 12.10, the authors generated a sample of random deviations $\tilde{\in}_1, \tilde{\in}_2, \ldots, \tilde{\in}_{14}$ from a normal distribution with mean 0 and standard deviation 35 and then added $\tilde{\in}_i$ to an assumed $\beta_0 + \beta_1 x$ to obtain 14 corresponding *y* values. This entire process was repeated 19 more times. The variation in values of the estimated slope is illustrated in a dotplot, and the 20 estimated regression equations are plotted in a graph (Figure 12.13).

Here's the point. The slope β_1 of the population regression line is the true average change in the dependent variable *y* associated with a 1-unit increase in the independent variable *x*. The slope of the least squares line, $\hat{\beta}_1$, gives a point estimate of β_1. In the same way that a confidence interval for μ and procedures for testing hypotheses about μ were based on properties of the sampling distribution of $\overline{X}$, further inferences about β_1 are based on thinking of $\hat{\beta}_1$ as a statistic and investigating its sampling distribution.

The values of the x_i's are either chosen before the experiment is performed, or fixed in the course of the experiment, so only the Y_i's are random.

The estimators (statistics, and thus random variables) for β_1, β_0 and σ^2 are obtained by replacing y_i with the random variable Y_i in their respective formulas.

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(Y_i - \bar{Y})}{\sum (x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \frac{\sum Y_i - \hat{\beta}_1 \sum x_i}{n}, \quad \sigma^2 = \frac{\sum (Y_i)^2 - \hat{\beta}_0 \sum Y_i - \hat{\beta}_1 \sum x_i Y_i}{n-2}$$

The denominator of $\hat{\beta}_1$, S_{xx}, depends only on the x_i's and not on the Y_i's, so it is a constant.
Since $\sum (x_i - \bar{x})\bar{Y} = \bar{Y} \sum (x_i - \bar{x}) = \bar{Y} * 0 = 0$, the formula for the slope estimator $\hat{\beta}_1$ can be written as

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x}) Y_i}{S_{xx}},$$

a linear function of the independent random variables $Y_1, Y_2, \ldots, Y_n$, each of which is normally distributed.

Proposition:

1. The mean value of $\hat{\beta}_1$ is $E(\hat{\beta}_1) = \mu_{\hat{\beta}_1} = \beta_1$. In other words, $\hat{\beta}_1$ is an unbiased estimator of β_1.
2. The variance and standard deviation of β_1 are $V(\hat{\beta}_1) = (\sigma_{\hat{\beta}_1})^2 = \dfrac{\sigma^2}{S_{xx}}$, $\sigma_{\hat{\beta}_1} = \sqrt{\dfrac{\sigma^2}{S_{xx}}} = \dfrac{\sigma}{\sqrt{\sum (x_i)^2 - \dfrac{(\sum x_i)^2}{n}}}$.

   Replacing σ by its estimator s gives the estimated standard error of $\hat{\beta}_1$.
3. Since it is a linear function of independent normal random variables, the estimator $\hat{\beta}_1$ has a normal distribution.

The denominator S_{xx} is a measure of how spread out the x_i's are about $\bar{x}$. Making observations at x_i values that are quite spread out results in a more precise estimator of the slope parameter β_1, a result of a smaller variance. Values of x_i all close to one another imply a highly variable estimator. However, we need to be careful if the x_i's are spread out too far: a linear model may not be appropriate throughout the range of observations.

Theorem: The assumptions of the simple linear regression model imply that the standardized variable $T = \dfrac{\hat{\beta}_1 - \beta_1}{S/\sqrt{S_{xx}}}$ has a *t* distribution with *n* – 2 degrees of freedom.

In a manner similar to the earlier development of confidence intervals, we can formulate a way to construct a 100(1 – α)% confidence interval for the value of the parameter β_1.

test statistic ± critical value times standard error

$$\hat{\beta}_1 \pm t_{\alpha/2,\, n-2}\left(\frac{s}{\sqrt{S_{xx}}}\right)$$

(Go to Example A-1 below.)

As before, the null hypothesis in a test about β_1 will be an equality statement. The null value is denoted by $\beta_{10} = \beta_{1\,\text{nought}}$ (*not* “beta sub ten”). The test statistic has a *t* distribution with *n* – 2 degrees of freedom when H_0 is true, so the type I error probability is controlled at the desired level α by using an appropriate *t* critical value.

The most commonly encountered pair of hypotheses about β_1 is $H_0 : \beta_1 = 0$ versus $H_\text{a} : \beta_1 \neq 0$. When the null hypothesis is true, $\mu_{Y \cdot x} = \beta_0$, independent of *x*. That is, there is no significant relationship/connection/correlation between *X* and *Y*.

A test of these two hypotheses is often referred to as the **model utility test** in simple linear regression. Unless *n* is quite small, $H_0 : \beta_1 = 0$ will be rejected, and the utility of the model will be confirmed, precisely when the coefficient of determination r^2 is reasonably large. If the conclusion is “fail to reject the null hypothesis”, the simple linear regression model should not be used for further inferences (estimates of mean value or predictions of future values).

(Go to Example A-2 below.)

The decomposition of the total sum of squares $\text{SST} = \sum (y_i - \bar{y})^2$ into a part SSE, which measures unexplained variation, and a part SSR, which measures variation explained by the linear relationship, is strongly reminiscent of one-way ANOVA. In fact, the null hypothesis $H_0 : \beta_1 = 0$ can be tested against $H_\text{a} : \beta_1 \neq 0$ by constructing an ANOVA table and rejecting the null hypothesis if $f \geq F_{\alpha,1,n-2}$.

The *F* test (ANOVA) gives exactly the same result as the model utility *t* test because $t^2 = f$ and $\left(t_{\alpha/2,\,n-2}\right)^2 = F_{\alpha,1,\,n-2}$. Virtually all software packages that have regression options include such an ANOVA table in the output, and include a calculation of the *P*-value (reject the null if $p < \alpha$).

(Go to Example A-3 below.)

Example A: A paper in *Measurement and Evaluation in Counseling and Development* (Oct 90, pp. 121–127) discussed a survey instrument called the *Mathematics Anxiety Scale for Children* (MASC). Suppose the MASC was administered to ten fifth graders with the following results:

MASC Score	67	37	70	40	35	65	40	35	30	40
Math grade (%)	75	85	60	90	80	75	70	90	95	80

From Lectures 12.1 and 12.2, we have

$$S_{xy} \approx -1075,\ S_{xx} \approx 2064.9,\ S_{yy} \approx 1000,\ \hat{\beta}_1 \approx -0.52061,\ \hat{\beta}_0 \approx 103.8958,\ \hat{\beta}_1 \approx -0.52061$$

$$\text{SSE} \approx 440.3482,\ \text{SST} \approx 1000,\ \text{SSR} \approx 559.6518,\ s^2 \approx 55.04353,\ r^2 \approx 0.559652\,.$$

1. Construct a 95% confidence interval for the slope of the true linear regression line, β_1.

2. Conduct a *t*-test to determine the significance of the data and results. $(\alpha = 0.05)$

3. Repeat the hypothesis test of question 2, this time using ANOVA.

Source of Variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	***f***
Regression		559.6518		
Error		440.3482		
Total		1000		