Section 4.6 – Testing the Significance of the Least-Squares Regression Model

Requirement 1 for Inference on the Least-Squares Regression Model

For any particular value of the explanatory variable *x*, the mean of the corresponding responses in the population depends linearly on *x*. That is,

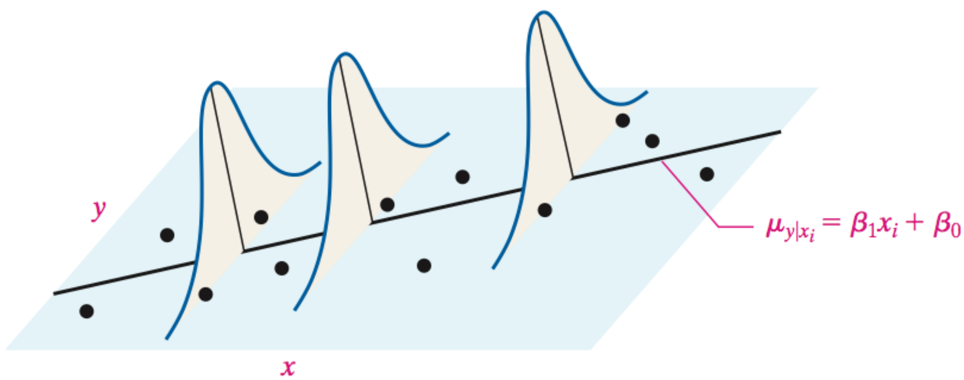
$$\mu_{y|x} = \beta_1 x + \beta_0$$

for some numbers β_0 *and* β_1 , where $\mu_{y|x}$ represents the population mean response when the value of the explanatory variable is *x*.

The response variables are normally distributed with mean $\mu_{y|x} = \beta_1 x + \beta_0$ and standard deviation *σ*.

When doing inference on the least-squares regression model, we require (1) for any explanatory variable, *x*, the mean of the response variable, *y*, depends on the value of *x* through a linear equation, and (2) the response variable, *y*, is normally distributed with a constant standard deviation, *σ*. The mean increases/decreases at a constant rate depending on the slope, while the standard deviation remains constant.

A large value of *σ*, the population standard deviation, indicates that the data are widely dispersed about the regression line, and a small value of *σ* indicates that the data lie fairly close to the regression line



The least-squares regression model is given by $y_i = \beta_1 x_i + \beta_0 + \varepsilon_i$ where

y_i is the value of the response variable for the i^{th} individual

β_0 *and* β_1 are the parameters to be estimated based on sample data

$\beta_1 x_i$ is the value of the explanatory variable for the i^{th} individual

ε_i is a random error term with mean 0 an variance, the error terms are independent.

i = 1, …, *n*, where *n* is the sample size (number of ordered pairs in the data set)

The standard error of the estimate, s_e, is found using the formula

$$s_e = \sqrt{\frac{\sum\left(y_i - \hat{y}_i\right)^2}{n-2}} = \sqrt{\frac{\sum residuals^2}{n-2}}$$

Example

Compute the standard error of the estimate for the drilling data which is presented

Depth at Which Drilling Begins, x (in ft)	Time to Drill 5 Feet, y (in min)
35	5.88
50	5.99
75	6.74
95	6.1
120	7.47
130	6.93
145	6.42
155	7.97
160	7.92
175	7.62
185	6.89
190	7.9

Solution

Step **1:** The least squares regression line to be $\hat{y} = 0.116x + 5.5273$

Step **2, 3:** The predicted values as well as the residuals for the 12 observations

Depth, x	Time, y	$\hat{y}$	$y-\hat{y}$	$(y-\hat{y})^2$
35	5.88	5.9333	-0.0533	0.0028
50	5.99	6.1073	-0.1173	0.0138
75	6.74	6.3973	0.3427	0.1174
95	6.1	6.6293	-0.5293	0.2802
120	7.47	6.9193	0.5507	0.3033
130	6.93	7.0353	-0.1053	0.0111
145	6.42	7.2093	-0.7893	0.6230
155	7.97	7.3253	0.6447	0.4156
160	7.92	7.3833	0.5367	0.2880
175	7.62	7.5573	0.0627	0.0039
185	6.89	7.6733	-0.7833	0.6136
190	7.9	7.7313	0.1687	0.0285
				$\sum \text{residuals}^2 = 2.7012$

Step **4**: We find the sum of the squared residuals by summing the last column of the table:

Step **5**: The standard error of the estimate is then given by

$$s_e = \sqrt{\frac{\sum residuals^2}{n-2}} = \sqrt{\frac{2.7012}{10}} \underline{= 0.5197}$$