Using the Regression: Confidence Intervals for Coefficients

Overview: Often a goal of regression analysis is to estimate unknown physical constants by their corresponding regression coefficients. For example, consider the data set containing spring lengths and loads. According to Hook's Law these measurements should obey the model

$$\text{length}_i \quad = \quad \beta_0 + \beta_1 \text{weight}_i + \epsilon_i$$

where β_1 is the spring constant. Thus the least squares estimate $\hat{\beta}_1$ of β_1 provides a very good - optimal with respect to some criteria - estimate of the spring constant.

Confidence Intervals (CI's): In science and engineering we typically estimate a quantity by a point estimate plus/minus an error margin. In statistics we do this by constructing a confidence interval. Recall that a confidence interval is an interval estimate of the form (θ_L, θ_H) where θ_L and θ_H are functions of the data chosen so that the interval (θ_L, θ_H) has a specified probability of containing/capturing the true value of the unknown constant θ. We call these intervals **confidence intervals** or **CI**'s. The probability of success or success rate is called the **confidence level**.

Regression Coefficient CI's: Recall that the least squares regression coefficients are normally distributed and that fitting the regression model provides an estimate of their standard deviation or standard error **if the regression assumptions are met**. Also recall that whenever we work with a normal RV and use an estimate of its standard deviation (as opposed to its **true** standard deviation) we use the t distribution. Thus we have the following formula for a 100(1-α)% confidence interval for a regression coefficient β_i:

$$(\hat{\beta}_i - t_{\alpha/2,n-2}\hat{\sigma}_{\hat{\beta}_i}, \hat{\beta}_i + t_{\alpha/2,n-2}\hat{\sigma}_{\hat{\beta}_i})$$

where $\hat{\sigma}_{\hat{\beta}_i}$ is the estimated standard deviation (standard error) of β_i (square root of corresponding diagonal of $\hat{\sigma}^2(X'X)^{-1}$) and $t_{\alpha/2,n-2}$ satisfies $P(t_{n-2} \geq t_{\alpha/2,n-2}) = \alpha/2$. For example, if we fit a regression model using data consisting of 10 observations and we want a 95% confidence interval for β_1, then α, the **failure rate**, is 0.05 and so $\alpha/2 = 0.025$. Since $n = 10$ we have $n - 2 = 8$ degrees of freedom. Using the t-table in the back of the book (Table V, page 711) we see that $t_{0.025,8} = 2.306$. Alternatively, we can compute $t_{0.025,8}$ using the Minitab routine

```
Graph -> Probability Distribution Plot -> View Probability
```

and then selecting t distribution, `shaded area`, and `Probability`. Using this routine we again see that $t_{0.025,8} = 2.306$.

Example 1: See next page...

Example 1: The data for this example consists of lengths of a spring subjected to loads of varying weight. According to Hook's Law there should be a linear relationship between length and weight. Taking into account errors in measurement this implies the simple linear regression model

$$\text{length}_i \quad = \quad \beta_0 + \beta_1 \text{weight}_i + \epsilon_i \tag{1}$$

Note that β_1 is the spring constant. Construct a 95% confidence interval for the spring constant as follows:

i. Open the spring length vs weight data set at the bottom of the course website and fit the regression model (1).

ii. Determine the value of $\hat{\beta}_1$ and $\hat{\sigma}_{\hat{\beta}_1}$ from the `Coef` and `SE Coef` columns of the coefficient table.

iii. Compute a 95% confidence interval for β_1, i.e, the spring constant. Compute $t_{\alpha/2,n-2}$ using Minitab.

iv. What is the probability the resulting CI contains the true value of β_1?

Example 2: Estimating Absolute Zero: Recall the *Ideal Gas Law* from Physics/Chemistry: $PV = nRT$. Using this law we will estimate absolute zero, the temperature at which the molecules of a gas have no kinetic energy and thus zero pressure and/or volume. Note that for a fixed pressure our gas law reduces to $T = \beta_0 + \beta_1 V$. Since β_0 corresponds to the temperature when volume is zero, given a set of bivariate observations $\{(v_i, t_i)\}$ we can estimate absolute zero using the least squares estimate $\hat{\beta}_0$ of β_0.

i. Download the data set

`http://www.rose-hulman.edu/~inlow/absolute_zero.MTW`

which gives tempererature and length measurements. Length is the length of a volume of gas in a tube of constant diameter separated from the atmosphere by a plug of mercury; see diagram on board. Thus length is proportional to volume here.

ii. Fit the following simple linear model using Minitab

$$T_i \quad = \quad \beta_0 + \beta_1 V_i + \epsilon_i$$

and compute the residuals.

iii. Analyze the residuals. Is the fit of the model good? Are the regression assumptions met?

iv. Compute a 95% confidence interval for β_0. Note that this is also a 95% CI for the temperature at absolute zero.