Estimation and Prediction

DASC 512

Using our OLS Model

Two common uses of an OLS model are:

- Forecasting future behavior
- Characterizing average behavior

Each of these use cases has the same expected value $\hat{y}$ for a given x, but they use different confidence intervals

- Prediction intervals
- Estimation intervals

Standard Error for Estimation/Prediction

Estimation

- The standard deviation for the estimator $\hat{y}$ for the mean value of y at a specific level of x, (x_p), is

$$\sigma_{\hat{y}} = \sigma\sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}}$$

Standard Error for Estimation/Prediction

Prediction

- The standard deviation for the prediction error for the predictor $\hat{y}$ of an individual new y value at a specific value of x, $\left(x_p\right)$ is

$$\sigma_{y-\hat{y}} = \sigma\sqrt{1 + \frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{SS_{xx}}}$$

Standard Error for Estimation/Prediction

<u>Comparison</u>

- Estimation: $$\sigma_{\hat{y}} = \sigma \sqrt{\frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{SS_{xx}}}$$

- Prediction: $$\sigma_{y-\hat{y}} = \sigma \sqrt{1 + \frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{SS_{xx}}}$$

Note: we won't know σ, so we'll use s to estimate it.

Standard Error for Estimation/Prediction

Comparison

- Estimation: $$s_{\hat{y}} = s\sqrt{\frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{SS_{xx}}}$$

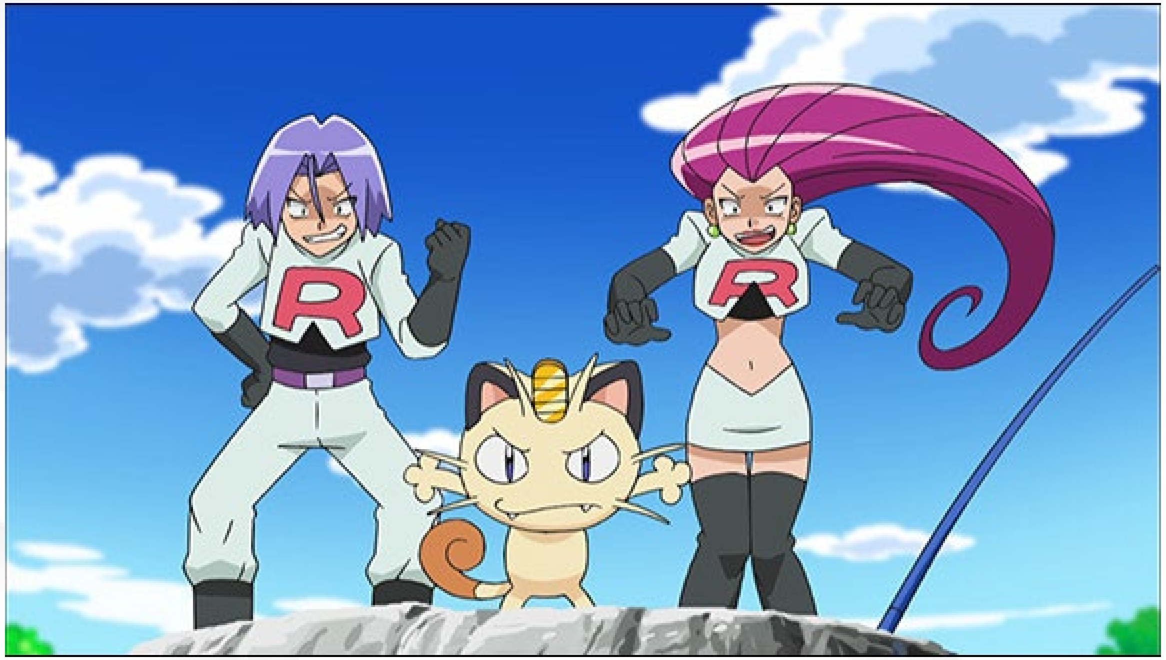
- Prediction: $$s_{y-\hat{y}} = s\sqrt{1 + \frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{SS_{xx}}}$$

Prediction/Estimation Intervals

- Estimation: $\hat{y} \pm t_{ISF}\left(\frac{\alpha}{2}, \nu = n - 2\right) \times s_{\hat{y}}$

- Prediction: $\hat{y} \pm t_{ISF}\left(\frac{\alpha}{2}, \nu = n - 2\right) \times s_{y-\hat{y}}$

Back to Team Rocket



Equations Recap (1 of 2)

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

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$s^2 = MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{n - 2}$$

$$= \frac{SS_{yy} - \hat{\beta}_1 SS_{xy}}{n - 2}$$

$$SS_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$SS_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$SS_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Equations Recap (2 of 2)

$$s_{\hat{\beta}_1} = \frac{s}{\sqrt{SS_{xx}}}$$

$$t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}$$

$$s_{\hat{y}} = s\sqrt{\frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{SS_{xx}}}$$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} \times SS_{yy}}} = \sqrt{\frac{SS_{xx}}{SS_{yy}}}\hat{\beta}_1$$

$$r^2 = \frac{SS_{yy} - SSE}{SS_{yy}} = 1 - \frac{SSE}{SS_{yy}}$$

$$s_{y-\hat{y}} = s\sqrt{1 + \frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{SS_{xx}}}$$

But can’t Python do this all for me?

Next week on DASC 512…

Multiple Linear Regression

Higher-Order Models

Model Building