STAT 231 — Lecture 29

Bartosz Antczak *Instructor: Michael Wallace* *November 17, 2017*

Last Time

We started chapter 6. We looked at the STAT 230/231 grade comparisons example. We looked at determining some α, β such that the line $y = \alpha + \beta x$ "best" describes this grades comparison. Let Y be the STAT 231 grade of some randomly selected student. We can assume the model $Y \sim G(\mu, \sigma)$. To make an estimate for μ, we observe that the relationship for this particular data is linear, so we assume our model is $Y_i \sim G(\alpha + \beta x_i, \sigma)$. This is referred to as a simple **linear regression model**.

29.1 Likelihood Function for our Linear Regression Model

For our model $Y_i \sim G(\alpha + \beta x_i, \sigma)$, where every i is independent, then

$$L(\alpha, \beta) = \prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(y_i - \alpha - \beta x_i)^2\right) \tag{29.1}$$

$$= \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2\right) \tag{29.2}$$

We see that our estimates are

$$\widehat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \sum_{i=1}^{n}\frac{(x_i - \overline{x})}{S_{xx}}y_i$$

and

$$\widehat{\alpha} = \overline{y} - \widehat{\beta}\overline{x}$$

which are also the least square estimates!

Theorem

A distribution of our point estimator for β is

$$\tilde{\beta} \sim G\left(\beta, \frac{\sigma}{\sqrt{S_{xx}}}\right)$$

This means that we can obtain the result

$$\frac{\tilde{\beta} - \beta}{\sigma/\sqrt{S_{xx}}} \sim G(0, 1)$$

which can be used as a pivotal quantity! However, σ^2 is unknown, so we estimate it using

$$s_e^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \widehat{\alpha} - \widehat{\beta}x_i)^2 = \frac{1}{n-2}(S_{yy} - \widehat{\beta}S_{xy})$$

We also see that

$$\frac{(n-2)S_e^2}{\sigma^2} \sim \chi^2(n-2)$$

which results in

$$\frac{\tilde{\beta} - \beta}{S_e/\sqrt{S_{xx}}} \sim t_{n-2}$$

Because of this, we can construct a $100p\%$ confidence interval for β:

$$\widehat{\beta} \pm a\frac{s_e}{\sqrt{S_{xx}}}$$

where $P(T \leq a) = (1+p)/2$ and $T \sim t_{n-2}$.
We can also to determine our $p-$value for some hypothesis $H_0 : \beta = \beta_0$.

$$p - \text{value} = 2\left[1 - P\left(T \leq \frac{|\widehat{\beta} - \beta_0|}{s_e/\sqrt{S_{xx}}}\right)\right]$$

Confidence Interval for our Mean

A $100p\%$ confidence interval for $\mu(x) = \alpha + \beta x$ (i.e., the mean response at x) is

$$\widehat{\mu}(x) \pm as_e\sqrt{\frac{1}{n} + \frac{(x-\overline{x})^2}{S_{xx}}} = (\widehat{\alpha} + \widehat{\beta}x) \pm as_e\sqrt{\frac{1}{n} + \frac{(x-\overline{x})^2}{S_{xx}}}$$

where $P(T \leq a) = (1+p)/2$ with $T \sim t_{n-2}$.

Confidence Interval for Alpha

$$\widehat{\alpha} \pm as_e\sqrt{\frac{1}{n} + \frac{(\overline{x})^2}{S_{xx}}}$$