STAT 231 — Lecture 30

Bartosz Antczak *Instructor: Michael Wallace* *November 20, 2017*

Last Time

Just a shit ton of equations.

30.1 Building a Prediction Model

Continuing with our STAT 230/231 grades example, we define Y as a *new* potential observation for x:

$$Y = \mu(x) + R$$

where $R \sim G(0, \sigma)$.
We're now interested in the error in the point estimator $Y - \widetilde{\mu}(x)$. We see quite easily that $Y - \widetilde{\mu}(x) = R + [\mu(x) - \widetilde{\mu}(x)]$. We see the following characteristics:

- $E[Y - \widetilde{\mu}(x)] = 0$
- $Var(Y - \widetilde{\mu}(x)) = Var(Y) + Var[\widetilde{\mu}(x)]$

The corresponding interval

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

where $P(T \leq a) = (1 + p)/2$ with $T \sim t_{n-2}$ is called the $100p\%$ **prediction interval**.

30.1.1 Gaussian Response Model

The general form for a **Gaussian Response Model** is $Y_i \sim G(\mu(\overline{x}_i), \sigma)$, where $\overline{x}_i = (x_{i1}, \ldots, x_{ik})$ is a vector. The Gaussian response model can also be written

$$Y = \mu(\overline{x}_i) + R_i$$

We also see

$$E[Y_i] = \mu(\overline{x}_i) + \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}$$