STA 674

Regression Analysis And Design Of Experiments

Fitting Simple Linear Regression Models – Lecture 11

STA 674, RADOE:
Fitting Simple Linear Regression Models

- Last time: introduction to a notion peculiar to LR: the two different types of estimation, CIs for the mean and Prediction Intervals for a new observation at a given x.
- This time, look at the formulas these different procedures, and try to get some intuition about them.

STA 674, RADOE:
Fitting Simple Linear Regression Models

Different questions: given a new value of the explanatory variable, x_j:

- What is a plausible range of values for $\mu_{y|x_j}$?

  Called a **confidence interval for $\boldsymbol{\mu_{y|x_j}}$**

- What is a plausible range of values for a new observation at x_j?

  Called a **prediction interval for y at $\boldsymbol{x_j}$**

STA 674, RADOE:
Fitting Simple Linear Regression Models

For the first: given a new value of the explanatory variable, x_m (for a value we want the *m*ean of):

- What is the most likely value of the conditional mean $\mu_{y|x_m}$?
- What is a plausible range of values for $\mu_{y|x_m}$?

STA 674, RADOE:
Fitting Simple Linear Regression Models

- **Fitted Value** (conditional mean)

The best estimate of the mean of y at x_m (i.e., $y_m = \beta_0 + \beta_1 x_m$) is:

$$\hat{y}_m = b_0 + b_1 x_m$$

- **Confidence Interval for the Mean** (of y given x_m)

A $(1-\alpha)100\%$ confidence interval for the fitted value of y at x_m has endpoints:

$$L_m, U_m = (b_0 + b_1 x_m) \mp t_{\alpha/2, n-2} s_m$$

where

$$s_m = s_e \sqrt{\frac{1}{n} + \frac{(x_m - \bar{x})^2}{(n-1)s_x^2}}$$

CI gets bigger if standard error sm is big...and/or...if xm is further away from the mean...and/or...the varianace of x is small (i.e., the x values are close together...so the regression line may not as be as well constrained as with a higher variance)

STA 674, RADOE:
Fitting Simple Linear Regression Models

For the second (of the two types of estimation, prediction): given a new value of the explanatory variable, x_p (for a value we want a *p*rediction for):

- What is the most likely value of a new observation at x_j? xp at xj = xm at xj
- What is a plausible range of values for a new observation at x_j?

STA 674, RADOE:
Fitting Simple Linear Regression Models

- Predicted Value (new observation)

The best estimate of a new observation of y at x_p (i.e., $y_p = \beta_0 + \beta_1 x_p + \epsilon$) is:

$$\hat{y}_p = b_0 + b_1 x_p$$

- Prediction Interval

A $(1-\alpha)100\%$ prediction interval for the new observation of y at x_m has endpoints:

$$L_p, U_p = \left(b_0 + b_1 x_p\right) \mp t_{\alpha/2, n-2} s_p$$

where

$$s_p = s_e \sqrt{\boxed{1 +} \frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{(n-1)s_x^2}}$$

1+ is added to standard error of sm...so sp is bigger...also...as n increases, sm decreases to eventually 0...but sp will never reach 0 because of the 1+...never certain of prediction!