STA 674

Regression Analysis And Design Of Experiments

Fitting Simple Linear Regression Models – Lecture 12

STA 674, RADOE:
Fitting Simple Linear Regression Models

- Last time: we looked at the formulas the estimation, CIs for the mean and Prediction Intervals for a new observation at a given x.
- This time, look at some SAS code and output, trying to match the output to the intuition I tried to help you gain last time—and wrap up our second (material) module!

STA 674, RADOE:
Fitting Simple Linear Regression Models

- 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:

$$L_m, U_m = (b_0 + b_1 x_m) \mp t_{\alpha/2,n-2} s_m, \text{ with } s_m = \sqrt{\frac{1}{n} + \frac{(x_m - \bar{x})^2}{(n-1)s_x^2}}$$

Statement about a parameter

- Prediction Interval

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

$$L_p, U_p = \left(b_0 + b_1 x_p\right) \mp t_{\alpha/2,n-2} s_p, \text{ with } s_p = \sqrt{1 + \frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{(n-1)s_x^2}}$$

Statement about a new observation (data point)

Simple Linear Regression – Estimation of CI of Conditional Mean

```
/* Add new data point */;
DATA HOOKER2;
     INPUT pressure;
     CARDS;
     28.96
RUN;

DATA HOOKERNEW;
     SET HOOKER HOOKER2;
RUN;

/* Fit regression model and compute predictions */;
/* and 95% confidence intervals */
PROC REG DATA=HOOKERNEW;
     MODEL temperature=pressure;
     OUTPUT OUT=HOOKERPRED1 PREDICTED=pred  LCLM=lclm UCLM=uclm;
RUN;
QUIT;
```

Simple Linear Regression – Estimation of Prediction Interval of New Observation

```
/* Fit regression model and compute predictions */;
/* and 95% prediction intervals */
PROC REG DATA=HOOKERNEW;
     MODEL temperature=pressure;
     OUTPUT OUT=HOOKERPRED2 PREDICTED=pred  LCL=lcl UCL=ucl;
RUN;
QUIT;
```

STA 674, RADOE:

Fitting Simple Linear Regression Models

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

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

The SAS System

Obs	Temperature	Pressure	pred	lclm	uclm
32	.	28.96	211.908	211.152	212.664

Obs	Temperature	Pressure	pred	lcl	ucl
32	.	28.96	211.908	210.094	213.722

Example: Hooker's Data

- 95% confidence interval when pressure=28.96 in: (211.15, 212.66)
- 95% prediction interval when pressure=28.96 in: (210.09, 213.72)

STA 674, RADOE:
Fitting Simple Linear Regression Models

- Last caveat: don't predict outside the range of where data have been collected. **Extrapolation** occurs when we make predictions for values of the predictor outside of the range of the observed data.

Example: Human Growth Rates

- The average human male is 52 cm long at birth and 165 cm tall at 20 yrs, representing a growth rate of 5.65 cm/yr. How tall is the average 50 yr old man?

$$52 + 5.65(50) = 334.5cm$$

or about 12'8½". Moral: **DO NOT EXTRAPOLATE!**