STAT 213
Confidence and Prediction Intervals in Regression

Colin Reimer Dawson

Oberlin College

February 19-21, 2018

Notes

Outline

Confidence and Prediction Bands

Notes

Outline

Confidence and Prediction Bands

Notes

Intervals at a particular X

- A confidence interval for the slope is useful, but if our goal is a predictive model, we want to be able to make statements about Y values at particular X values.
- I should be able to estimate
  1. What the mean Y value is at that X *in the population*
  2. Where the particular Y is likely to be for *this one new observation*
- Note: These are different things, in the same way that a 95% confidence interval does *not* tell us where 95% of the *individual cases* are.

Notes

Confidence and Prediction Intervals for a Linear Model

(Population) linear model:

$$\begin{aligned} Y &= \beta_0 + \beta_1 X + \varepsilon \\ &= f(X) + \varepsilon \end{aligned}$$

1. A **confidence interval** (for a particular X) is an estimate (with a margin of error) of $f(X)$.
2. A **prediction interval** (for a particular X) is an estimate about Y

Notes

Confidence vs. Prediction Intervals

Which is wider? The prediction interval is wider, b/c it has uncertainty about ε *plus* the uncertainty about $f(X)$

Notes

A Subtlety Re: Prediction Intervals

Interpreting Prediction Intervals

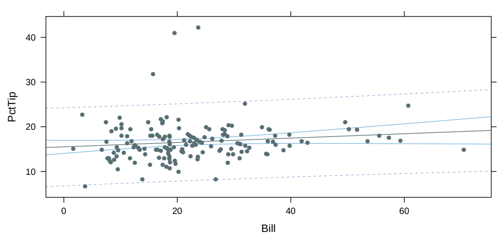
A coverage level of 95% for a prediction interval does *not* mean that, having fit a model from a *particular* sample, we will make successful predictions 95% of the time going forward. The worse our line, the lower the %.

What we can say is that the *average* success rate across *all possible samples* is 95%

Notes

Confidence and Prediction Bands

Intervals for all x in the range are called "confidence / prediction bands".



Why the hourglass shape? More leverage at extreme X^*: bigger change in line from one sample to the next

Notes

Calculating Confidence and Prediction Intervals

Both types of intervals are of the form

$$(1-\alpha) \text{ interval } = \text{Point Estimate} \pm t_{n-2}^{*(1-\alpha/2)} \cdot SE$$

Confidence Interval:

$$\hat{f}(X^*) \pm t_{n-2}^{*(1-\alpha/2)} \cdot \sqrt{\hat{\sigma}^2_{\hat{f}(X^*)}}$$

where $\hat{\sigma}^2_{\hat{f}(X^*)} = \hat{\sigma}^2_{\varepsilon} h(X^*)$ and $h(X^*) = \frac{1}{n} + \frac{(x* - \bar{x})^2}{\sum(x_i - \bar{x})^2}$ is the leverage at X^*.

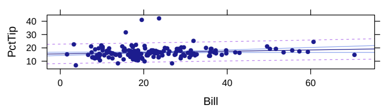
Prediction Interval:

$$\hat{Y}* \pm t_{n-2}^{*(1-\alpha/2)} \cdot \sqrt{\hat{\sigma}^2_{\hat{f}(X^*)} + \hat{\sigma}^2_{\varepsilon}}$$

Notes

R code for a confidence/prediction bands plot:

```
library("mosaic"); library("Lock5Data")
data("RestaurantTips")
xyplot(PctTip ~ Bill, data = RestaurantTips,
       panel = panel.lmbands,  # Note, no quotes
       level = 0.90,           # The confidence level
       ## OPTIONAL: band.lty= what kind of lines to use
       ##    format: c(conf.linetype, pred.linetype), where
       ##    1 = solid, 2 = dashed, 3 = dotted
       band.lty = c(1,2),
       ## OPTIONAL: band.col: what color lines to use
       ##    format: c(conf.color, pred.color)
       band.col = c("royalblue", "blueviolet")
       )
```



Notes

We can get intervals for specific X values as follows:

```
tip.model.using.bill <- lm(PctTip ~ Bill, data = RestaurantTips)
## Creates a new function with the given name
f.hat <- makeFun(tip.model.using.bill)
## Use it like a regular function
##   First arg name: name of predictor variable
##      (= the desired x value to get the interval for)
##   interval="confidence" or interval="prediction"
##      controls which interval type to return
##      (or leave this out to just get the pt estimate)
##   level=confidence.level controls the confidence level
f.hat(Bill = 40, interval = "confidence", level = 0.90)

          fit      lwr      upr
     1 17.46215 16.45974 18.46455

f.hat(Bill = 40, interval = "prediction", level = 0.90)

          fit     lwr      upr
     1 17.46215 10.1786 24.74569
```


Notes

Notes