Today: Section 2.4
Links to websites for today's lecture
(<u>Most</u> of lecture will be on white board)

- Rossman/Chance applet illustrating confidence level

http://www.rossmanchance.com/applets/ConfSim.html

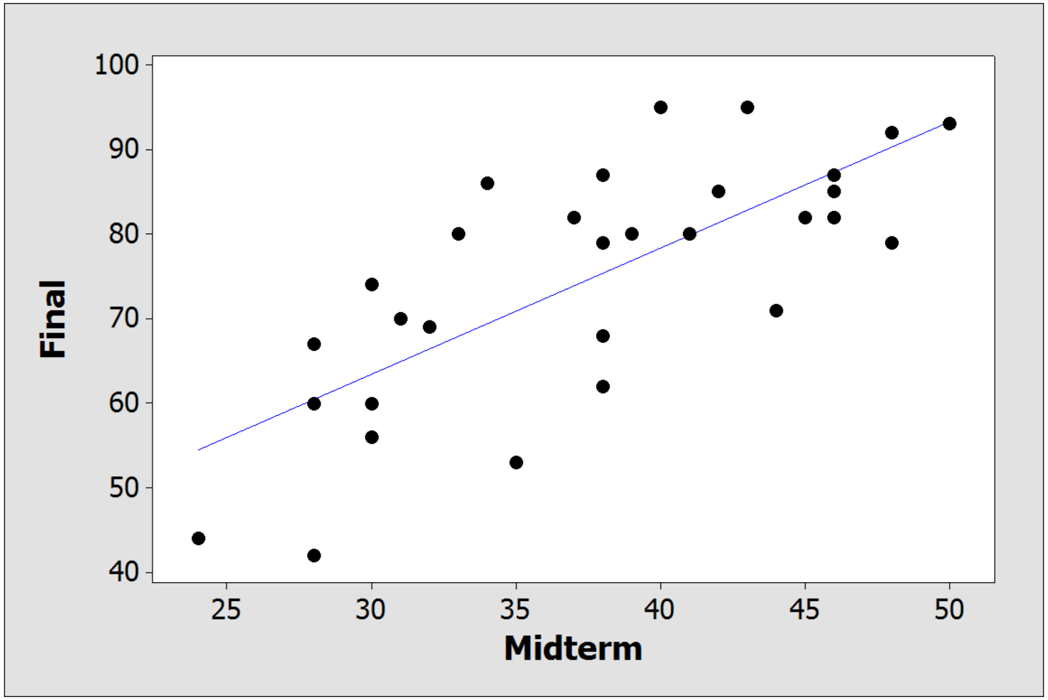
- Sign-reading example, illustrating confidence intervals and prediction intervals in R

http://www.ics.uci.edu/~jutts/110/RSessionHwyDataLecture6.pdf

Example: n = 31 students

X = midterm score (max = 50)

Y = final exam score



$$\hat{Y} = 18.67 + 1.493X$$

Fitting the Model Using R

```
> ExamModel <- lm(Final~Midterm, data=ExamData)
> summary(ExamModel)

Call:
lm(formula = Final ~ Midterm, data = ExamData)

Residuals:
     Min       1Q   Median       3Q      Max 
-18.4621  -6.3571   0.1354   5.7992  16.6279 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  18.6721     9.3311   2.001   0.0548 .  
Midterm       1.4925     0.2413   6.186 9.58e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 9.651 on 29 degrees of freedom
Multiple R-squared:  0.5689,	Adjusted R-squared:  0.554 
F-statistic: 38.26 on 1 and 29 DF,  p-value: 9.582e-07
```

$$\hat{Y} = 18.67 + 1.493X$$

Accuracy of Predictions

Example:
A student has a midterm grade of 41. What grade would we predict for the final?

$$\hat{y} = 18.67 + 1.4925 Midterm$$

$$x = 41 \Rightarrow \hat{y} = 79.86$$

How accurate is that prediction?

Two Forms of Intervals for Regression at a Given Value of x, call it *x**

(1) Confidence Interval for $E(Y|x^*)$ = population mean Y at X = *x**, also written as μ_Y

What is the Y value on the population line when X = *x**?

Where is the average *Y* for everyone in the population with X = *x**?

(2) Prediction Interval for Individual *Y* at X = *x**

Into what range do most of the individual *Y*'s fall for everyone with X = *x**?

CI for μ_Y When $X = x^*$

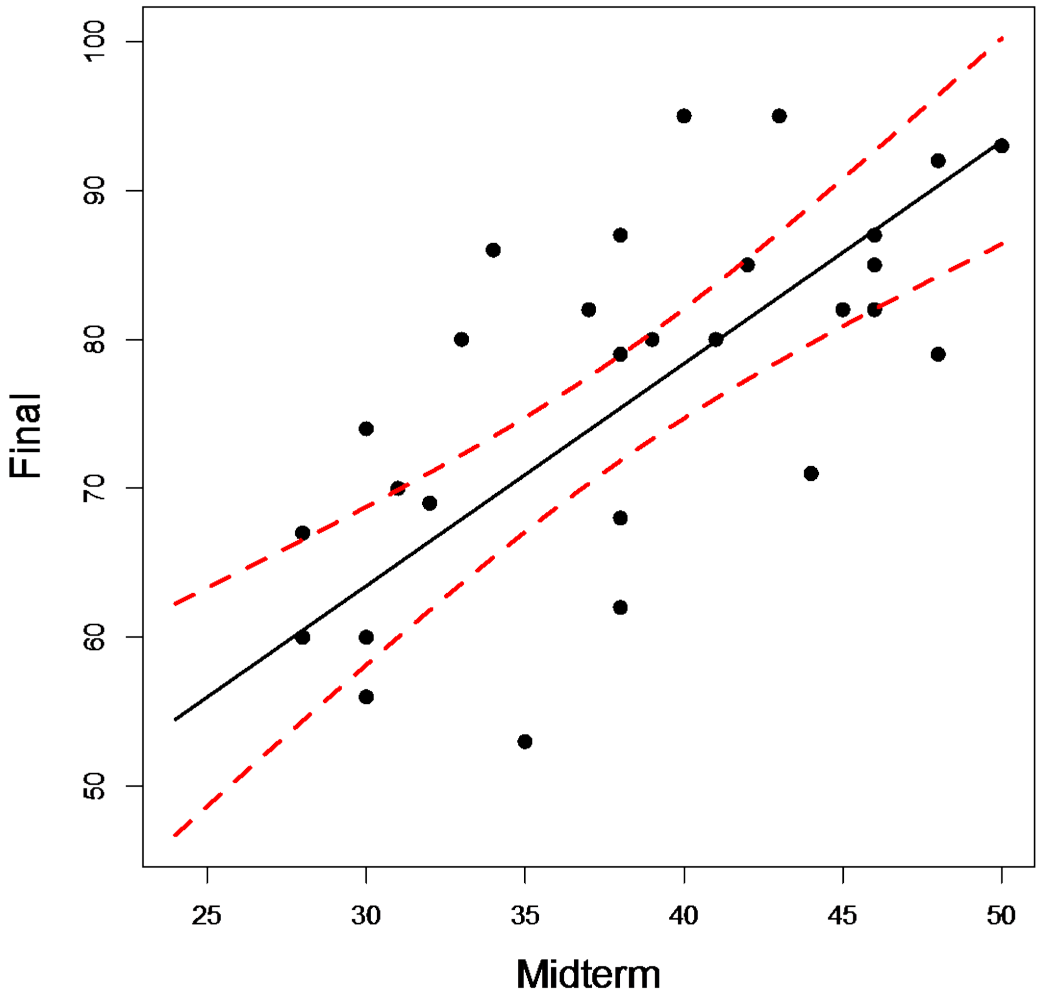
$$\hat{y} \pm t^* \hat{\sigma}_\varepsilon \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x - \bar{x})^2}}$$

This is called "SE.fit" in R

Called SE of $\hat{\mu} = SE_{\hat{\mu}}$ in book

Graphical CI for μ_Y (with *R*)

Final = 18.67 + 1.49*Midterm



Prediction Interval for Individual *Y*'s When *X* = *x**

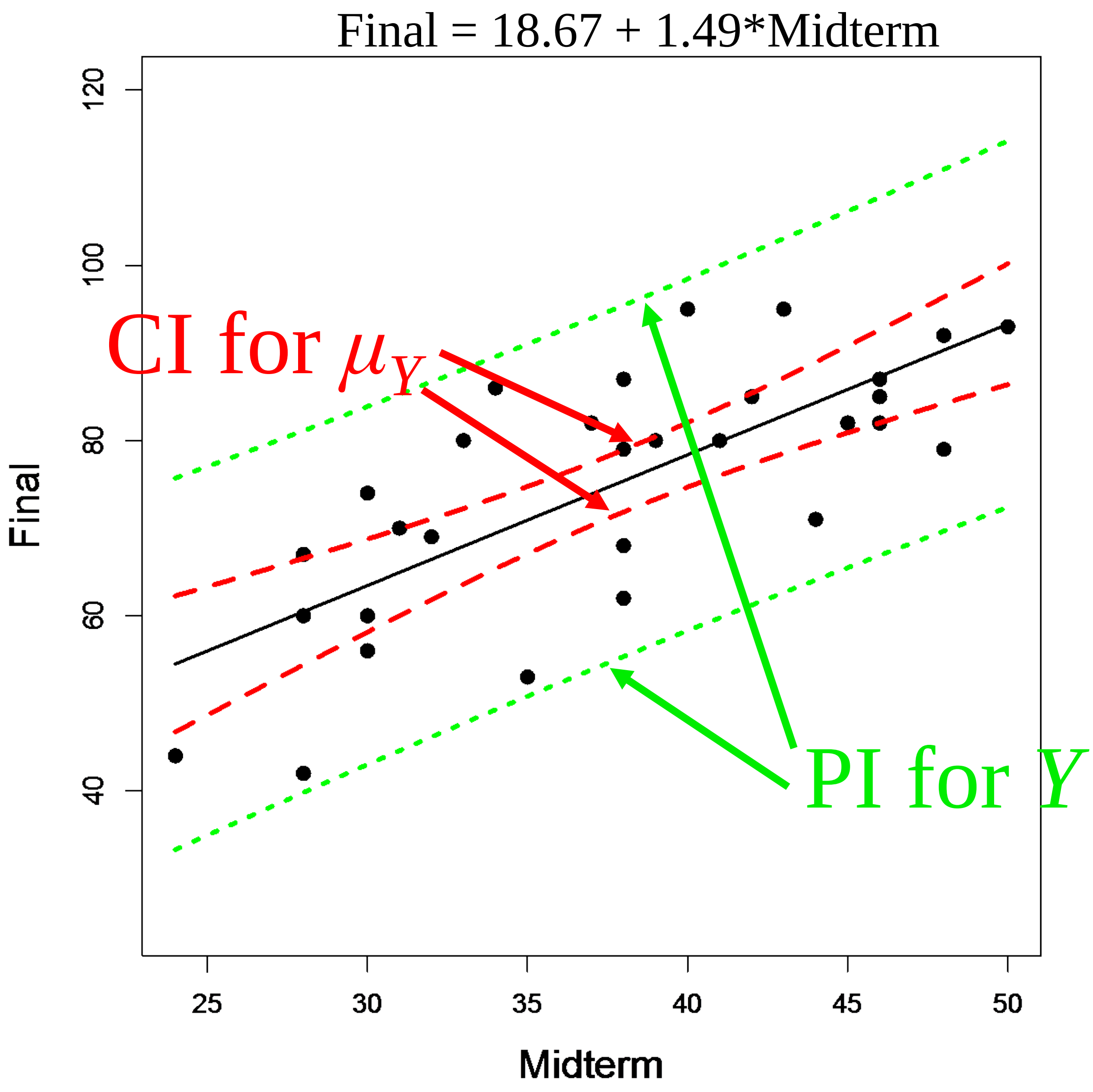
$$\hat{y} \pm t^* \hat{\sigma}_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x - \bar{x})^2}}$$

Accounts for extra ε at a point in

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Called SE of $\hat{Y} = SE_{\hat{Y}}$ in book

$$\hat{y} \pm t^* \hat{\sigma}_{\varepsilon} \sqrt{1 + \frac{1}{n} + \frac{\left(x^* - \bar{x}\right)^2}{\sum \left(x - \bar{x}\right)^2}}$$



Final = 18.67 + 1.49*Midterm
CI for μ_Y
PI for Y
Final
Midterm
120
100
80
60
40
25
30
35
40
45
50

Prediction and Confidence Intervals from R

x^*

```
> predict(ExamModel, list(Midterm=41), se.fit=T, interval="c")
$fit
       fit      lwr      upr
1 79.8646 76.02267 83.70652

$se.fit
[1] 1.878482

$df
[1] 29

$residual.scale
[1] 9.651273
> predict(ExamModel, list(Midterm=41), se.fit=T, interval="p")
$fit
       fit      lwr      upr
1 79.8646 59.75512 99.97408

$se.fit
[1] 1.878482
```

Confidence interval

Prediction interval

Another method in R
Define newx first

```
> data1=read.csv("MidtermFinal.csv")
> model=lm(Final~Midterm,data=data1)

> newx=data.frame(Midterm=41)

> predict.lm(model,newx,interval="confidence")
       fit      lwr      upr
1 79.8646 76.02267 83.70652
```

Confidence interval

```
> predict.lm(model,newx,interval="prediction")
       fit      lwr      upr
1 79.8646 59.75512 99.97408
```

Prediction interval

$$\text{fit} = \hat{y}$$