

READING:

EXERCISES:

ASSIGNED:

PRODUCER:

IMG CREDIT: ALEX RIEGERT-WATERS



Leverage

- “Potential” Impact on Linear Regression Fit
- Formula: $h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}$
- Sum of Leverages: $\sum h_i = \sum \frac{1}{n} + \frac{\sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} = 1 + 1 = 2$
- Average Leverage or Typical Leverage: $\bar{h} = \frac{\sum h_i}{n} = \frac{2}{n}$
- Less Data or Farther from \bar{x} = More Leverage

Example: Butterfly Ballots

- 2000 Election
 - Bush (Republican)
 - Gore (Democrat)
 - Buchanan (Third Party)
- Unusual Results in Palm Beach
 - Nonstandard “Butterfly Ballot”
 - Accidental Voting?

Confusion over Palm Beach County ballot

Although the Democrats are listed second in the column on the left, they are the third hole on the ballot.

(REPUBLICAN)	
GEORGE W. BUSH - PRESIDENT	3 →
DICK CHENEY - VICE PRESIDENT	
(DEMOCRATIC)	
AL GORE - PRESIDENT	5 →
JOE LIEBERMAN - VICE PRESIDENT	
(LIBERTARIAN)	
HARRY BROWNE - PRESIDENT	7 →
ART OLIVIER - VICE PRESIDENT	
(GREEN)	
RALPH NADER - PRESIDENT	9 →
WINONA LA DUKE - VICE PRESIDENT	
(SOCIALIST WORKERS)	
JAMES HARRIS - PRESIDENT	11 →
MARGARET TROWE - VICE PRESIDENT	
(NATURAL LAW)	
JOHN HAGELIN - PRESIDENT	13 →
NAT GOLDHABER - VICE PRESIDENT	

Punching the second hole casts a vote for the Reform Party.

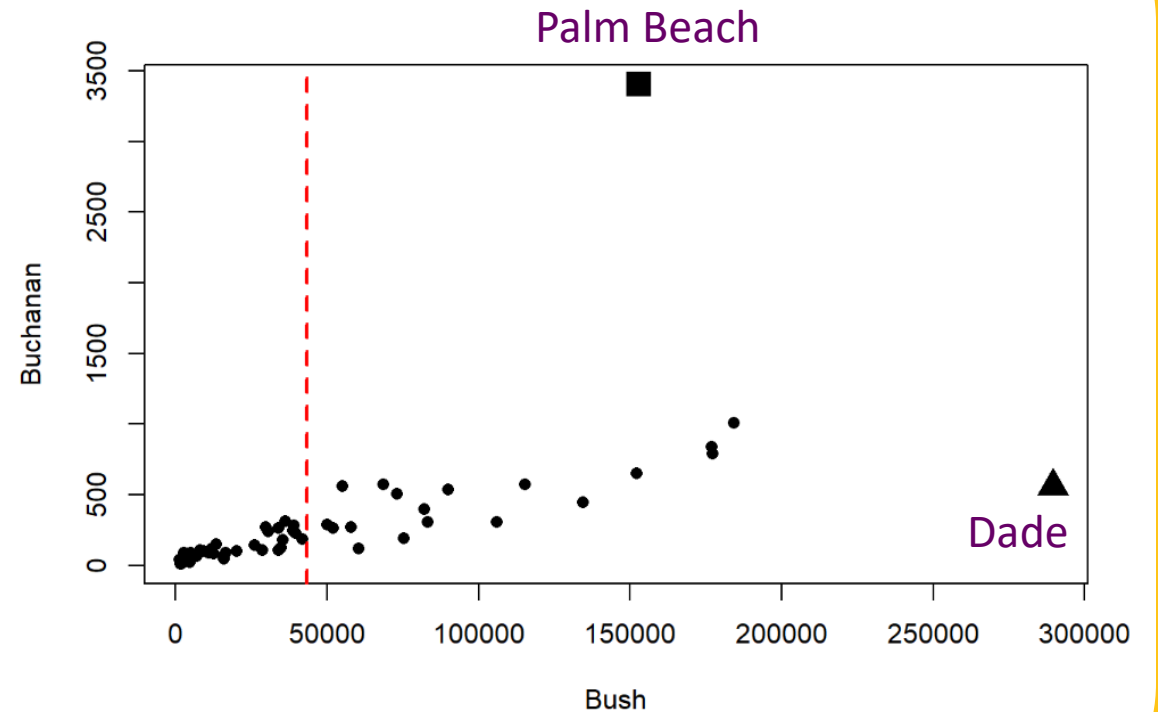
(REFORM)	
PAT BUCHANAN - PRESIDENT	4 ←
EZOLA FOSTER - VICE PRESIDENT	
(SOCIALIST)	
DAVID McREYNOLDS - PRESIDENT	6 ←
MARY CAL HOLLIS - VICE PRESIDENT	
(CONSTITUTION)	
HOWARD PHILLIPS - PRESIDENT	8 ←
J. CURTIS FRAZIER - VICE PRESIDENT	
(WORKERS WORLD)	
MONICA MOOREHEAD - PRESIDENT	10 ←
GLORIA LA RIVA - VICE PRESIDENT	
WRITE-IN CANDIDATE	
To vote for a write in candidate, follow the directions on the long stub of your ballot card.	

Example: Butterfly Ballots

```
library(Stat2Data)
library(mosaic)

data("PalmBeach")

plot(Buchanan~Bush, data= PalmBeach, pch=16)
abline(v=mean(PalmBeach$Bush), col="red", lty=2, lwd=2)
points(x=PalmBeach$Bush[which(PalmBeach$County=="PALM BEACH")],
       y=PalmBeach$Buchanan[which(PalmBeach$County=="PALM BEACH")],
       cex=2, pch=15)
points(x=PalmBeach$Bush[which(PalmBeach$County=="DADE")],
       y=PalmBeach$Buchanan[which(PalmBeach$County=="DADE")],
       cex=2, pch=17)
```



Example: Butterfly Ballots

```
#Calculate Average Number of Votes for Bush
avg.x = mean(PalmBeach$Bush)
devsquared= (PalmBeach$Bush - avg.x)^2
#Calculate Leverage
Leverage = 1/nrow(PalmBeach) + devsquared/sum(devsquared)
#Save Leverage Into Data
PalmBeach$Leverage= Leverage
#Print Data in Order of Leverage
head(PalmBeach[order(PalmBeach$Leverage,decreasing=TRUE),],5)
```

##	County	Buchanan	Bush	Leverage
## 13	DADE	561	289456	0.29747301
## 52	PINELLAS	1010	184312	0.10761608
## 6	BROWARD	789	177279	0.09859725
## 29	HILLSBOROUGH	836	176967	0.09820784
## 50	PALM BEACH	3407	152846	0.07085197

Very Unusual Leverage
“Triple the Typical Leverage”

$$3 * \frac{2}{n} = 3 * \frac{2}{67} = 0.09$$

Standardized Residual

- Old Formula:

$$stdres_i = \frac{y_i - \hat{y}_i}{\hat{\sigma}_\epsilon}$$

- New Formula:

$$stdres_i = \frac{y_i - \hat{y}_i}{\hat{\sigma}_\epsilon \sqrt{1 - h_i}}$$

- Further Adjust for Leverage Since Some Points Impact More
- Points with More Leverage Tend to Higher Standardized Residuals

Studentized Residual

- Old Formula:

$$studres_i = \frac{y_i - \hat{y}_i}{\hat{\sigma}_{(i)}}$$

- New Formula:

$$studres_i = \frac{y_i - \hat{y}_i}{\hat{\sigma}_{(i)}\sqrt{1 - h_i}}$$

- We Calculate Standard Error of Regression After Removing Point
- Line Doesn't Change, Then Standardized Equals Studentized

Cook's Distance

- Formula:
$$D_i = \frac{stdres_i^2}{k+1} \left(\frac{h_i}{1-h_i} \right) \quad \text{for SLR} \rightarrow k = 1$$
- Amount of Influence Depends On:
 - Leverage
 - Distance from Line of “Best” Fit
- Large Cook's D Indicates Point Strongly Influences Regression

Rules of Thumb

- We Can Calculate All Four Measures On Each Observation
- Table Showing Accepted Definitions of
 - “Moderately Unusual”
 - “Very Unusual”

Statistic	Moderately Unusual	Very Unusual
Leverage	Above $2(^2/n)$	Above $3(^2/n)$
Standardized Residual	Beyond ± 2	Beyond ± 3
Studentized Residual	Beyond ± 2	Beyond ± 3
Cook's D	Above 0.5	Above 1

Example: Butterfly Ballots

```
#Fit Model
vote.mod = lm(Buchanan~Bush, data=PalmBeach)
#Extract Standard Error of Regression
sigma.reg = summary(vote.mod)$sigma
#Save Actual Residuals
PalmBeach$Residual = vote.mod$residuals
#Calculate and Save Standardized Residual
PalmBeach$StdRes = PalmBeach$Residual/(sigma.reg*sqrt(1-PalmBeach$Leverage))
#Calculate and Save Cook's Distance
PalmBeach$CookD=cooks.distance(vote.mod)
#Find Very Unusual Cook's Distance
PalmBeach[which(PalmBeach$CookD>1),]
```

##	County	Buchanan	Bush	Leverage	Residual	StdRes	CookD
## 13	DADE	561	289456	0.29747301	-907.4953	-3.059180	1.981366
## 50	PALM BEACH	3407	152846	0.07085197	2610.1926	7.651072	2.231935

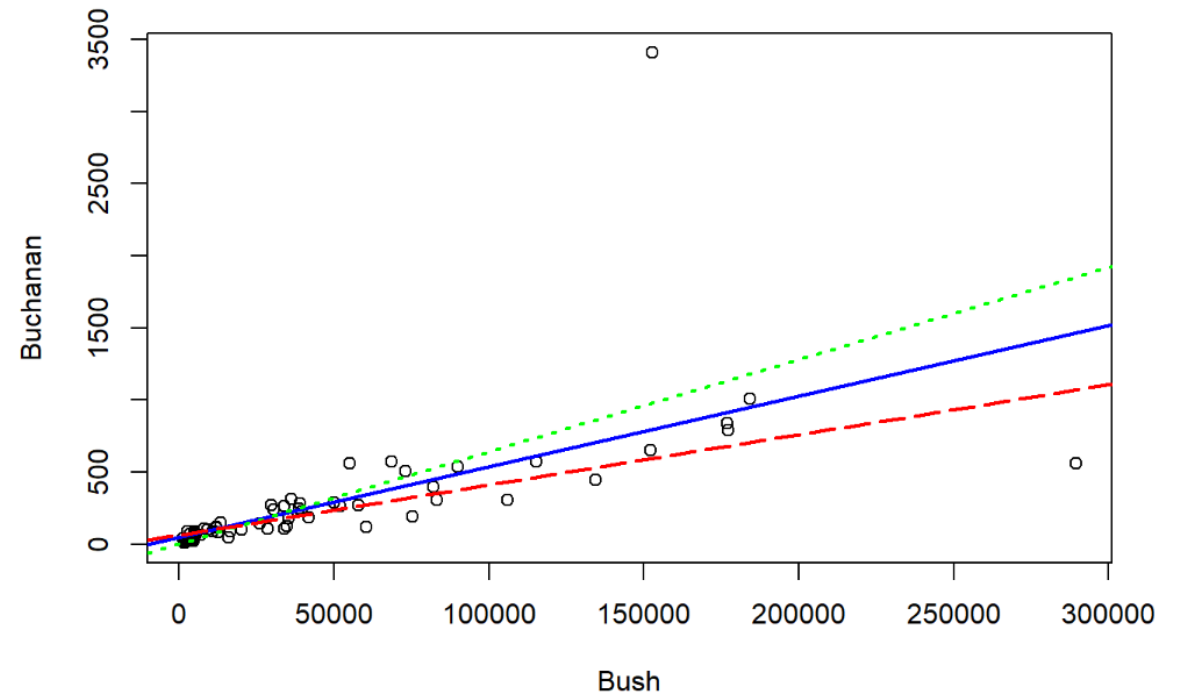
Example: Butterfly Ballots

```
vote.mod = lm(Buchanan~Bush, data=PalmBeach)

noDade = PalmBeach[which(PalmBeach$County!="DADE"),]
vote.mod.noDade = lm(Buchanan~Bush, data=noDade)

noPB = PalmBeach[which(PalmBeach$County!="PALM BEACH"),]
vote.mod.noPB = lm(Buchanan~Bush, data=noPB)

plot(Buchanan~Bush, data=PalmBeach)
abline(vote.mod, col="blue", lwd=2)
abline(vote.mod.noDade, col="green", lwd=2, lty=3)
abline(vote.mod.noPB, col="red", lwd=2, lty=5)
```



Thank You

Make Reasonable Decisions

