

# Statistical Stories

## Outliers

Bob Stine

Dept of Statistics, Wharton

# Why Stories?

- ⑤ Don't we have enough to cover already?
- ⑤ Thinking is hard!
- ⑤ Short stories
- ⑤ Choice of topics
- ⑤ Real data

# Three Outlier Stories

## ⑤ Comparisons

- ⑤ Two sample tests
- ⑤ What's the right answer

## ⑤ Leverage point in regression

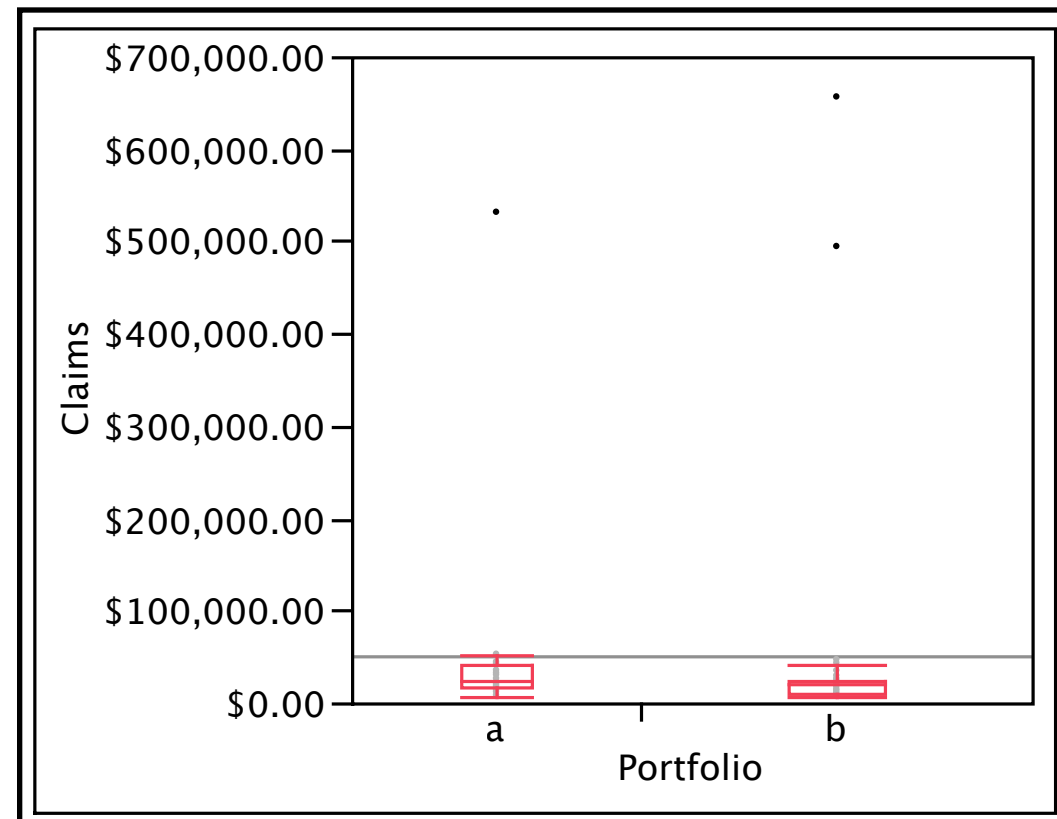
- ⑤ Simple linear regression
- ⑤ What's the right question?

## ⑤ Leverage point and transformation

- ⑤ Simple regression with a transformation
- ⑤ What is an outlier anyway?

# Disasters

- ④ Compare losses in two portfolios of loans.
- ④ What should we do about outliers in this two-sample comparison?
- ④ Comparison  $b - a$   
With,  $t = 0.16$   
Without,  $t = -2.4$
- ④ Logs, rank statistic?



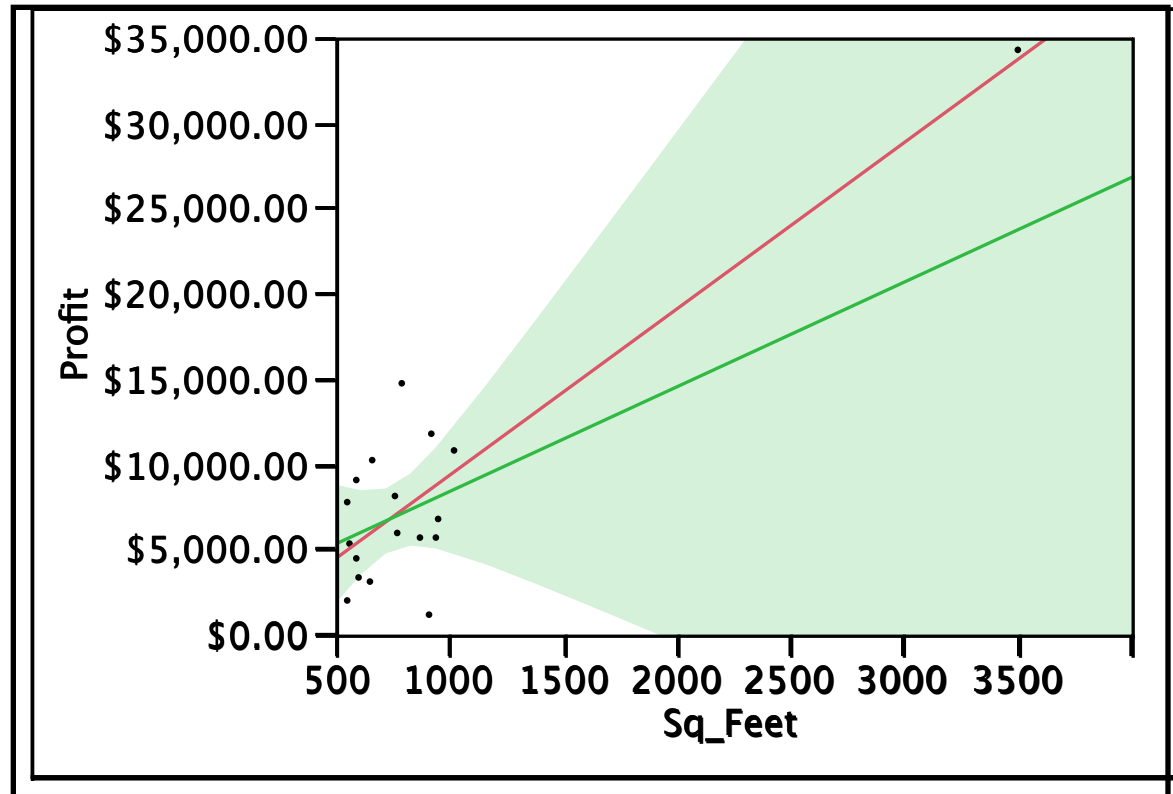
# Cottages

⑤ Cost estimates for construction cottage projects.  
Build another big one?

⑤ With outlier  
 $r^2 = 80\%$ ,  $t = 7.5$

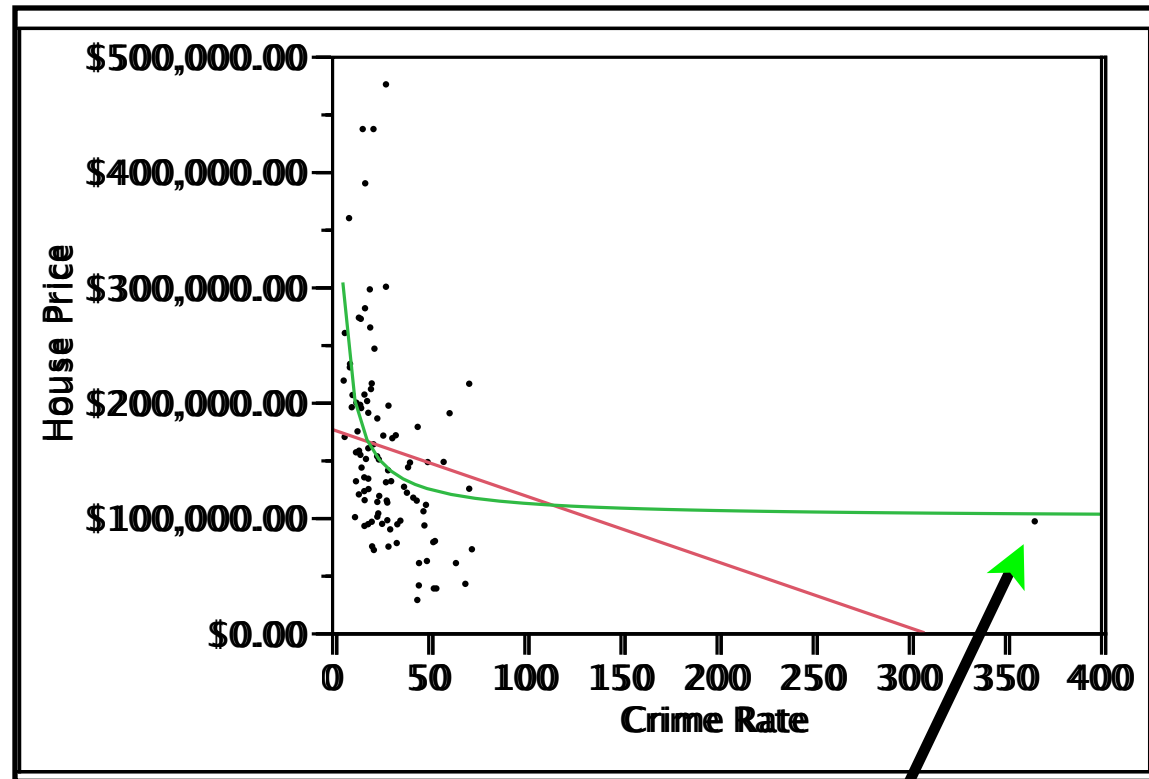
⑤ Without  
 $r^2 = 8\%$ ,  $t = 1.1$

⑤ What question do we need to ask?



# Philadelphia

- Real estate values (in 2000) and crime rates per 100,000 in Philadelphia region
- Which area is that outlier?
- Outlier pulls the slope toward zero
- Should we be fitting lines?



Fit does NOT include  
this point

# Stories

- ① They may not remember what you'd like for them to remember!
- ① Get them thinking!

Thanks