

Math 1710 Class 41

Regression Inference, Identifying Which Inference
Dr. Back

Dec. 4, 2009

Regression Inference Questions

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Basic Setup:

- 1) Data (x_i, y_i) , $1 \leq i \leq n$ leads to line of regression

$$\hat{y} = b_0 + b_1x$$

- 2) Assume an ideal line

$$\hat{y} = \beta_0 + \beta_1x$$

- 3) Together with an error process ϵ_i following an $N(0, \sigma)$ law (independent for each i .)

- 4) So that individual observations come from

$$y_i = \beta_0 + \beta_1x_i + \epsilon_i.$$

Regression Inference Questions

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Start with some data (x_i, y_i)



Regression Inference Questions

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Start with some data (x_i, y_i)

(X_i, Y_i)



Regression Inference Questions

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

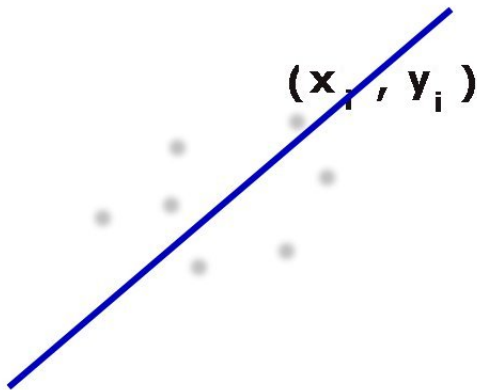
Identifying
Which
Inference

Building
Height

Phila HP and
Crime

These give the line of regression

$$\hat{y} = b_0 + b_1 x_i$$



Regression Inference Questions

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

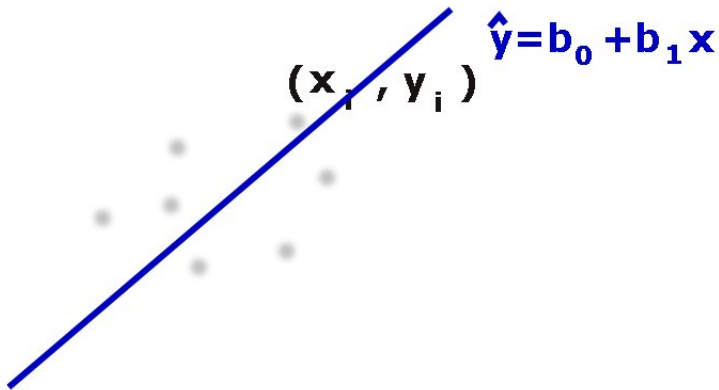
Identifying
Which
Inference

Building
Height

Phila HP and
Crime

These give the line of regression

$$\hat{y} = b_0 + b_1 x_i$$



Regression Inference Questions

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

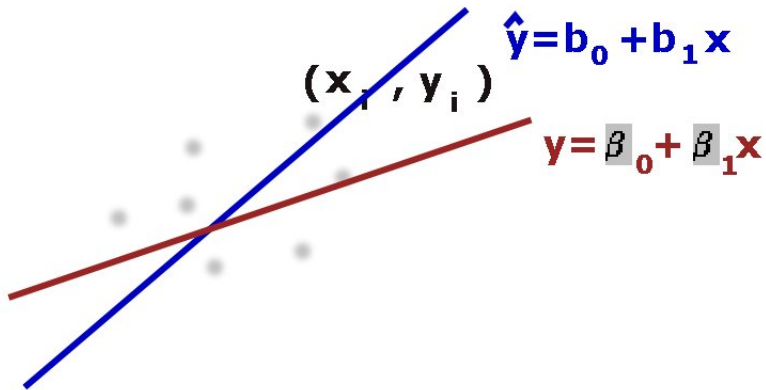
Identifying
Which
Inference

Building
Height

Phila HP and
Crime

There is also an ideal line

$$\hat{y} = \beta_0 + \beta_1 x_i$$



Regression Inference Questions

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

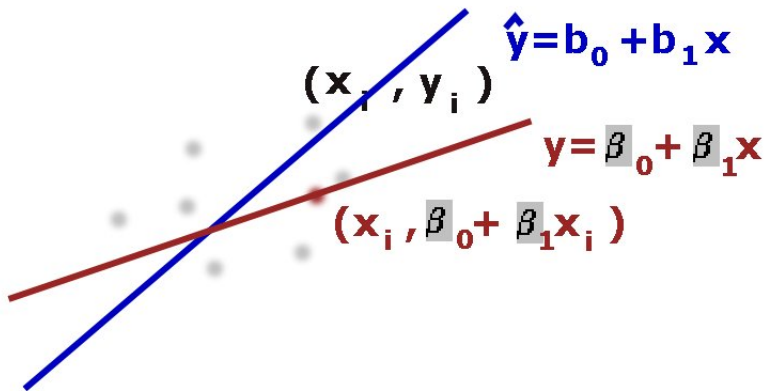
Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

We think of the data point (x_i, y_i) as arising by first
plugging in x_i into the ideal line



Regression Inference Questions

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

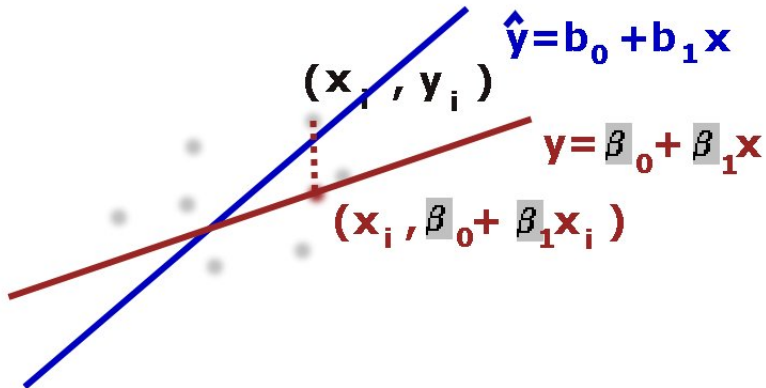
Building
Height

Phila HP and
Crime

Then randomly choosing an offset

$$\epsilon_i \sim N(0, \sigma)$$

to see how far off the ideal line we go.



Regression Inference Questions

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

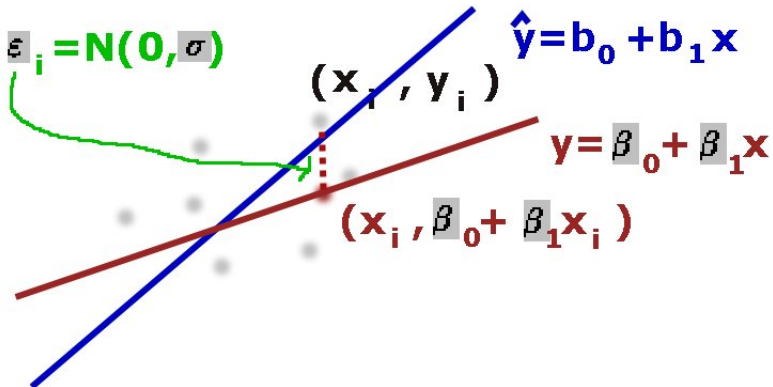
Building
Height

Phila HP and
Crime

Then randomly choosing an offset

$$\epsilon_i \sim N(0, \sigma)$$

to see how far off the ideal line we go.



Regression Inference Questions

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

The errors must be independent and all have the same standard deviation σ in the ideal model.

Regression Inference Questions

Natural Questions in Regression:

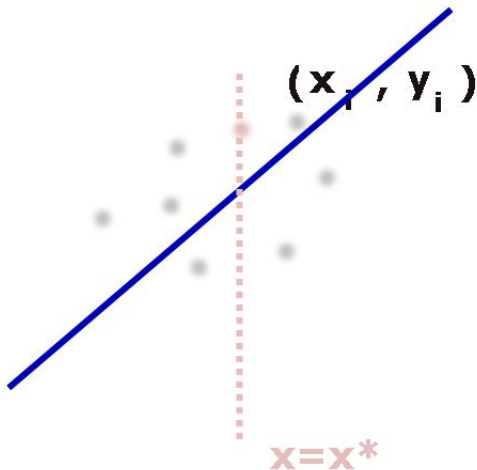
- 1) Estimate β_0 and β_1 .
- 2) Estimate the accuracy of b_0 and b_1 as estimators of β_0 and β_1 .
- 3) Estimate σ , the standard deviation of the error process.

For a given value x^* of x :

- 4) How accurately does the regression estimate $b_0 + b_1x^*$ approximate an actual y observation when $x = x^*$.
- 5) How accurately does the regression estimate $b_0 + b_1x^*$ approximate the average of a lot of y observations when $x = x^*$.

Regression Inference Questions

For $x = x^*$, we can try to predict **one value of y** leading to a **prediction interval when $x = x^*$**



Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

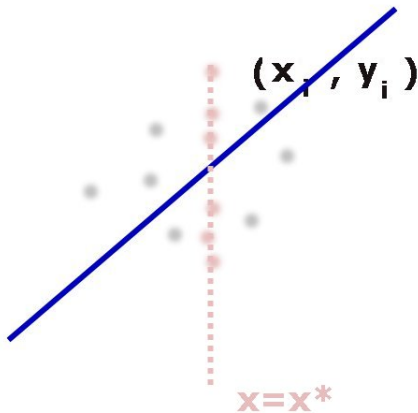
Building
Height

Phila HP and
Crime

Regression Inference Questions

Or we can try to predict the average of many values of y leading to a confidence interval for the mean response μ_y when

$$x = x^*$$



Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

How Many Rooms Can x Crews Clean?

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

x crews working for a building contractor go out each night and clean y rooms.

Understand the relationship?

How Many Rooms Can x Crews Clean?

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

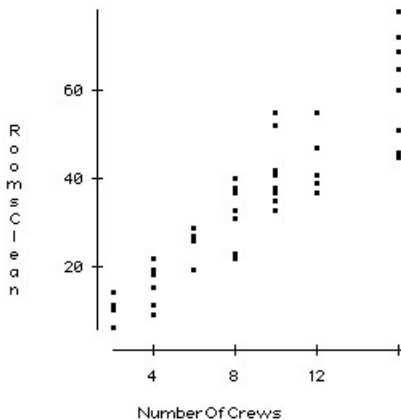
Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Scatterplot



How Many Rooms Can x Crews Clean?

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

NumCrews summary

```
Summary of                               Number Of Crews  
No Selector  
54 total cases of which 1 is missing
```

```
Percentile    25  
  
          Count    53  
          Mean    8.67925  
          Median   8  
          StdDev  4.80294  
          Range   14  
          IntQRange 8  
Lower 5th %tile  4  
Upper 5th %tile 12
```

How Many Rooms Can x Crews Clean?

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

RoomsCleaned Summary

Summary of
No Selector

RoomsClean

54 total cases of which 1 is missing

Percentile 25

Count	53
Mean	33.9057
Median	35
StdDev	19.2026
Range	72
IntQRange	27.5
Lower 5th %tile	18.75
Upper 5th %tile	46.25

How Many Rooms Can x Crews Clean?

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

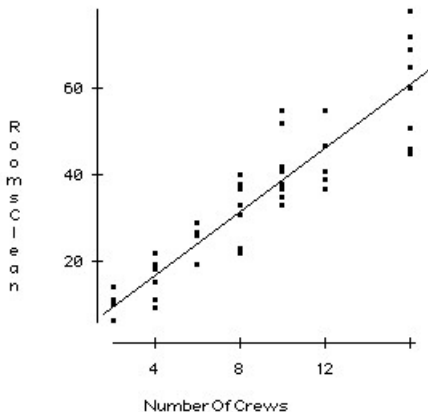
Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Scatterplot with Regression Line



How Many Rooms Can x Crews Clean?

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Regression Display

Dependent variable is: **RoomsClean**
No Selector
54 total cases of which 1 is missing
R squared = 85.7% R squared (adjusted) = 85.4%
s = 7.336 with 53 - 2 = 51 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	16429.7	1	16429.7	305
Residual	2744.8	51	53.8195	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	1.7847	2.096	0.851	0.3986
NumberOfCr...	3.70089	0.2118	17.5	≤ 0.0001

How Many Rooms Can x Crews Clean?

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Regression Display

Dependent variable is: **RoomsClean**
No Selector
54 total cases of which 1 is missing
R squared = 85.7% R squared (adjusted) = 85.4%
s = 7.336 with 53 - 2 = 51 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	16429.7	1	16429.7	305
Residual	2744.8	51	53.8195	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	1.7847	2.096	0.851	0.3986
NumberOfCr...	3.70089	0.2118	17.5	≤ 0.0001

How Many Rooms Can x Crews Clean?

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Regression Display

Dependent variable is: **RoomsClean**
No Selector
54 total cases of which 1 is missing
R squared = 85.7% R squared (adjusted) = 85.4%
s = 7.336 with 53 - 2 = 51 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	16429.7	1	16429.7	305
Residual	2744.8	51	53.8195	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	1.7847	2.096	0.851	0.3986
NumberOfCr...	3.70089	0.2118	17.5	≤ 0.0001

$$\widehat{\text{RoomsCleaned}} = 3.70 \cdot \text{NumCrews} + 1.78$$

How Many Rooms Can x Crews Clean?

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

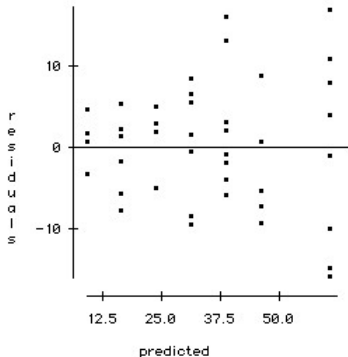
Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Residual Plot



How Many Rooms Can x Crews Clean?

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

There are important deviations from the the assumptions of an ideal linear regression model here.

Confidence Intervals for Mean Response

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

For a given value x^* of x :

How accurately does the regression estimate $b_0 + b_1x^*$ approximate the average μ_y of a lot of y observations when $x = x^*$.

Confidence Intervals for Mean Response

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

For a given value x^* of x :

How accurately does the regression estimate $b_0 + b_1x^*$ approximate the average μ_y of a lot of y observations when $x = x^*$.

The line of regression can be written as

$$y = b_1(x - \bar{x}) + \bar{y}$$

so

$$\mu_y = b_1(x^* - \bar{x}) + \bar{y}$$

.

Confidence Intervals for Mean Response

For a given value x^* of x :

How accurately does the regression estimate $b_0 + b_1x^*$ approximate the average μ_y of a lot of y observations when $x = x^*$.

The line of regression can be written as

$$y = b_1(x - \bar{x}) + \bar{y}$$

so

$$\mu_y = b_1(x^* - \bar{x}) + \bar{y}$$

We already know $Var(b_1)$ and the CLT tells us

$$Var(\bar{y}) = \frac{\sigma^2}{n}$$

which is naturally estimated by $\frac{s^2}{n}$.

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Confidence Intervals for Mean Response

For a given value x^* of x :

How accurately does the regression estimate $b_0 + b_1x^*$ approximate the average μ_y of a lot of y observations when $x = x^*$.

The line of regression can be written as

$$y = b_1(x - \bar{x}) + \bar{y}$$

so

$$\mu_y = b_1(x^* - \bar{x}) + \bar{y}$$

If b_1 and \bar{y} were known to be independent, we'd be done and have our $SE(b_1)$ formula.

That isn't guaranteed, but a little calculation shows the covariance of b_1 and \bar{y} is 0 which is good enough to establish:

Confidence Intervals for Mean Response

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

For a given value x^* of x :

How accurately does the regression estimate $b_0 + b_1x^*$ approximate the average μ_y of a lot of y observations when $x = x^*$.

Confidence interval for mean response μ_y when $x = x^*$ has

$$SE_{\hat{\mu}_y} = \sqrt{\frac{s^2}{n} + SE^2(b_1)(x^* - \bar{x})^2}.$$

Confidence Intervals for Mean Response

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

For a given value x^* of x :

How accurately does the regression estimate $b_0 + b_1x^*$ approximate the average μ_y of a lot of y observations when $x = x^*$.

Confidence interval for mean response μ_y when $x = x^*$ has

$$SE_{\hat{\mu}_y} = \sqrt{\frac{s^2}{n} + SE^2(b_1)(x^* - \bar{x})^2}.$$

Again a rescaled t distribution with $n - 2$ df is used to form the CI for the mean response.

Regression Inference Condiitons

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

**Regr Inf
Conditions**

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

straight enough

Regression Inference Condiitons

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

straight enough

Actually, this isn't really part of the underlying model from a more advanced point of view. But it is true that regression is usually not very useful if this is violated.

Regression Inference Condiitons

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

**Regr Inf
Conditions**

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

straight enough
independent errors (no pattern in residuals)

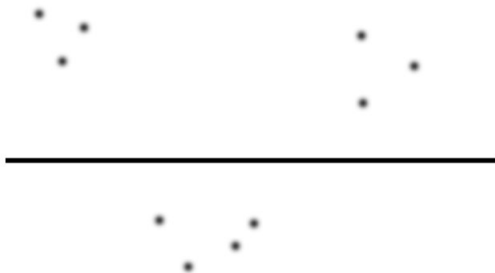
Regression Inference Conditions

straight enough

independent errors (no pattern in residuals)

This is extremely important to look for!

For example the classic curved residual plot picture



Regression Inference Conditions

straight enough

independent errors (no pattern in residuals)

This is extremely important to look for!

Or any x-range of the residual plot which seems to systematically have a greater share of y-values of the same sign than might be expected based on “random” fluctuations.



Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

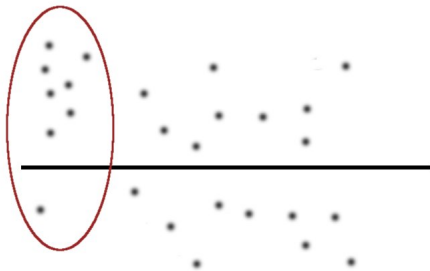
Regression Inference Condiitons

straight enough

independent errors (no pattern in residuals)

This is extremely important to look for!

Or any x-range of the residual plot which seems to systematically have a greater share of y-values of the same sign than might be expected based on “random” fluctuations.



Regression Inference Condiitons

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

**Regr Inf
Conditions**

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

straight enough
independent errors (no pattern in residuals)
equal variance of errors

Regression Inference Conditions

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

straight enough

independent errors (no pattern in residuals)

equal variance of errors

The standard but still mind-blowing word “heteroscedasticity” refers to a failure of this condition.

Your textbook’s formulation of “Does the plot thicken?” is a million times friendlier sounding.

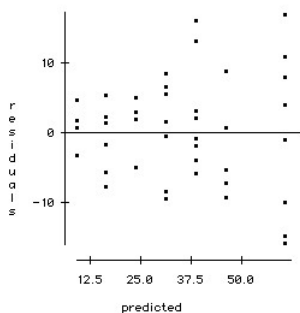
Regression Inference Condiitons

straight enough

independent errors (no pattern in residuals)

equal variance of errors

The cleaning residual plot clearly fails this test.



Regression Inference Condiitons

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

**Regr Inf
Conditions**

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

straight enough
independent errors (no pattern in residuals)
normality of errors

Regression Inference Condiitons

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

straight enough
independent errors (no pattern in residuals)
normality of errors

This might be assessed by a histogram or normal probability plot.

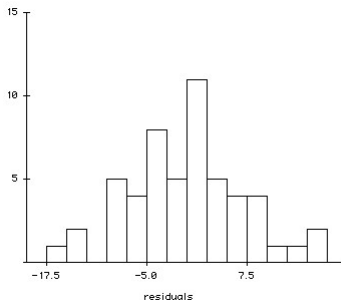
Regression Inference Conditions

straight enough

independent errors (no pattern in residuals)

normality of errors

This might be assessed by a histogram or normal probability plot.



Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

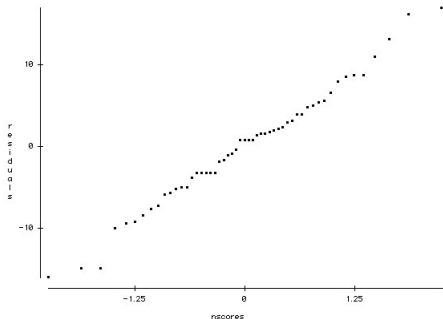
Regression Inference Condiitons

straight enough

independent errors (no pattern in residuals)

normality of errors

This might be assessed by a histogram or normal probability plot.



Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

NumCrews=10 PI's and CIMR's

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

When $x^* = 10$:

- $\hat{y} = 1.78 + 3.70 \cdot 10 = 38.78$.
- $SE_{\mu_y} = \sqrt{\frac{7.336^2}{53} + 0.212^2 \cdot (10 - 8.68)^2} = \sqrt{1.0154 + .0783} = \sqrt{1.0937} = 1.046$
- 95% CI for mean
response = $38.78 \pm 2.009 \cdot 1.046 = (36.68, 40.88)$
- $SE_{pi} = \sqrt{7.336^2 + 1.0937} = 7.417$
- 95% PI for single
value = $38.78 \pm 2.009 \cdot 7.417 = (23.88, 53.68)$

NumCrews=16 PI's and CIMR's

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

When $x^* = 16$:

- $\hat{y} = 1.78 + 3.70 \cdot 16 = 60.98.$

- $SE_{\mu_y} = \sqrt{\frac{7.336^2}{53} + 0.212^2 \cdot (16 - 8.68)^2} = \sqrt{1.0154 + 2.4082} = \sqrt{3.4236} = 1.850$

- 95% CI for Mean

$$\text{Response} = 60.98 \pm 2.009 \cdot 1.850 = (57.26, 64.70)$$

- $SE_{pi} = \sqrt{7.336^2 + 3.4236} = 7.566$

- 95% PI for Single

$$\text{Value} = 60.98 \pm 2.009 \cdot 7.566 = (45.78, 76.18)$$

NumCrews=4 PI's and CIMR's

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

When $x^* = 4$:

- $\hat{y} = 1.78 + 3.70 \cdot 4 = 16.58.$

- $SE_{\mu_y} = \sqrt{\frac{7.336^2}{53} + 0.212^2 \cdot (4 - 8.68)^2} = \sqrt{1.0154 + .9844} = \sqrt{1.9998} = 1.414$

- 95% CI for Mean

$$\text{Response} = 16.58 \pm 2.009 \cdot 1.414 = (13.74, 19.42)$$

- $SE_{pi} = \sqrt{7.336^2 + 1.9998} = 7.471$

- 95% PI for Single

$$\text{Value} = 16.58 \pm 2.009 \cdot 7.471 = (1.57, 31.59)$$

Cleaning Inaccuracies in PI's and CIMR's

Math 1710
Class 41

V1

When 10 crews were working

Summary of
cases selected according to
54 total cases of which 46 are missing

RoomsClean
10 Crews

Percentile	25
Count	8
Mean	41.625
Median	39.5
StdDev	7.92712
Range	22
IntQRRange	11
Lower ith %tile	36
Upper ith %tile	47

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Cleaning Inaccuracies in PI's and CIMR's

Math 1710
Class 41

V1

When 16 crews were working

Summary of
cases selected according to
54 total cases of which 44 are missing

RoomsClean
16 Crews

Percentile	25	
	Count	10
	Mean	59.7
	Median	62.5
	StdDev	12.0005
	Range	33
	IntQRange	23
	Lower ith %tile	46
	Upper ith %tile	69

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Cleaning Inaccuracies in PI's and CIMR's

Math 1710
Class 41

V1

When 4 crews were working

Summary of
cases selected according to
54 total cases of which 48 are missing

RoomsClean
4 Crews

Percentile	25
Count	6
Mean	15.6667
Median	16.5
StdDev	4.96655
Range	13
IntQRange	8
Lower ith %tile	11
Upper ith %tile	19

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Cleaning Inaccuracies in PI's and CIMR's

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Crews	\hat{y}	ActualMean	95% CIMR	95%PIIV	ActualStdv
4	17	15	(14, 19)	(2, 32)	5
10	39	42	(37, 41)	(24, 54)	8
16	61	60	(57, 65)	(46, 76)	12

Cleaning Inaccuracies in PI's and CIMR's

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Crews	\hat{y}	ActualMean	95% CIMR	95%PIIV	ActualStdv
4	17	15	(14, 19)	(2, 32)	5
10	39	42	(37, 41)	(24, 54)	8
16	61	60	(57, 65)	(46, 76)	12

For example the respective margins of error in the PIIV's are all around 15. But these are about 3, 2, and 1.25 times the actual standard deviations we see in the data.

Cleaning Inaccuracies in PI's and CIMR's

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Crews	\hat{y}	ActualMean	95% CIMR	95%PIIV	ActualStdv
4	17	15	(14, 19)	(2, 32)	5
10	39	42	(37, 41)	(24, 54)	8
16	61	60	(57, 65)	(46, 76)	12

For example the respective margins of error in the PIIV's are all around 15. But these are about 3, 2, and 1.25 times the actual standard deviations we see in the data.

Because of the “plot thickening”, our 4 crew PIIV's are too big and our 16 crew PIIV's are too small.

Cleaning Inaccuracies in PI's and CIMR's

Math 1710
Class 41

V1

Regression
Inference
Questions

Crews	\hat{y}	ActualMean	95% CIMR	95%PIIV	ActualStdv
4	17	15	(14, 19)	(2, 32)	5
10	39	42	(37, 41)	(24, 54)	8
16	61	60	(57, 65)	(46, 76)	12

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

For example the respective margins of error in the PIIV's are all around 15. But these are about 3, 2, and 1.25 times the actual standard deviations we see in the data.

Using the CLT style $\frac{\sigma}{\sqrt{n}}$ formula, we see suggestions of a similar problem with the CIMR's.

Identifying Which Inference

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

**Identifying
Which
Inference**

Building
Height

Phila HP and
Crime

If only considering the methods we've studied, then the type (Categorical vs. Quantitative) of each individual observation (and whether or not there are two groups) narrows one down to just a few choices:

Identifying Which Inference

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Individual Observations	Num. Vars.	More	Method
Categorical (2 choice)	1		1-Proportion
Categorical (2 choice)	2		2-Proportion
Quantitative	1		1 Sample Mean
Quantitative	2	Independence	2 Sample Mean
Quantitative	2	Matching	Matched Pairs
Quantitative	2	$(x_i, y_i), H_0$	Regression Slope
Categorical	2		Independence
Categorical	2	Different sample each population.	Homogeneity
Categorical	1	Also given a probability dist.	Goodness of Fit

Identifying Which Inference

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

**Identifying
Which
Inference**

Building
Height

Phila HP and
Crime

We view tests for *homogeneity* as a 2 categorical variable situation since *which_population?* is also a natural categorical variable.

Identifying Which Inference

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

**Identifying
Which
Inference**

Building
Height

Phila HP and
Crime

The remaining overlaps are often differentiated by the nature of the null hypothesis (e.g. $\mu_1 = \mu_2$ vs. *vars. associated*) and even often by comparison of units!

Building Height

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

**Building
Height**

Phila HP and
Crime

Height vs. stories for 60 randomly chosen tall buildings in North America.

Building Height

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

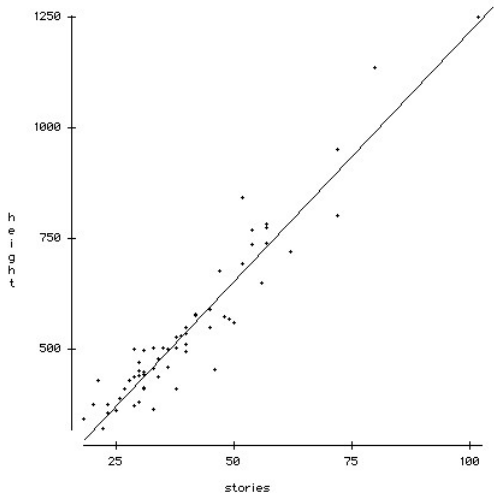
Cleaning
Inaccuracies

Identifying
Which
Inference

**Building
Height**

Phila HP and
Crime

Scatterplot Heights vs. Stories



Building Height

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Regression Display

stories

Dependent variable is: **height**

No Selector

R squared = 90.4% R squared (adjusted) = 90.2%

s = 58.33 with $60 - 2 = 58$ degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	1.84852e6	1	1.84852e6	543
Residual	197311	58	3401.91	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	90.3096	20.96	4.31	≤ 0.0001
stories	11.2924	0.4844	23.3	≤ 0.0001

Building Height

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

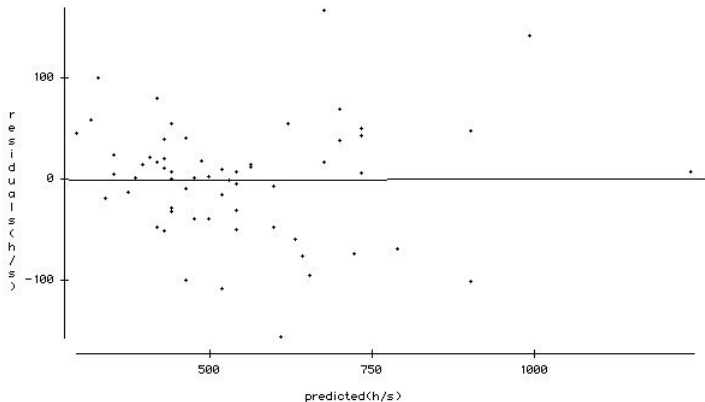
Cleaning
Inaccuracies

Identifying
Which
Inference

**Building
Height**

Phila HP and
Crime

Residual Plot



Building Height

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

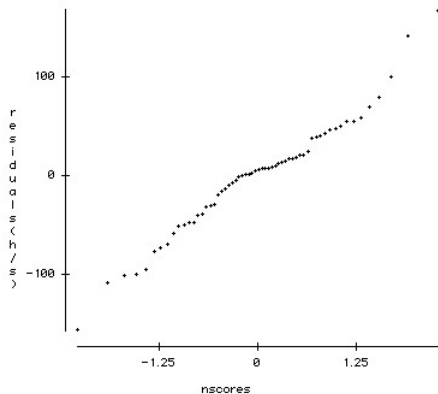
Cleaning
Inaccuracies

Identifying
Which
Inference

**Building
Height**

Phila HP and
Crime

Normal Probability Plot



Phila Housing Prices and Crime

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

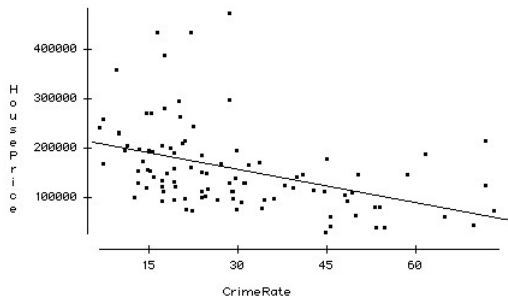
Housing Prices vs. Crime Rate for suburbs of Philadelphia, not including the center city.

Phila Housing Prices and Crime

Math 1710
Class 41

V1

Scatterplot Housing Prices vs. Crime Rate (crimes per 1000)



Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Phila Housing Prices and Crime

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Regression Display (Not Center City)

Dependent variable is:

cases selected according to

111 total cases of which 13 are missing

R squared = 18.4% R squared (adjusted) = 17.6%

s = 7.886e4 with 98 - 2 = 96 degrees of freedom

HousePrice

Not Center City

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	1.34831e11	1	1.34831e11	21.7
Residual	5.97038e11	96	6.21914e9	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	225234	1.64e4	13.7	≤ 0.0001
CrimeRate	-2288.69	491.5	-4.66	≤ 0.0001

Phila Housing Prices and Crime

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

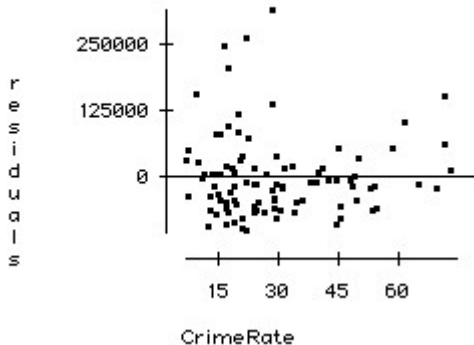
Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Residual Plot (Not Center City)



Phila Housing Prices and Crime

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

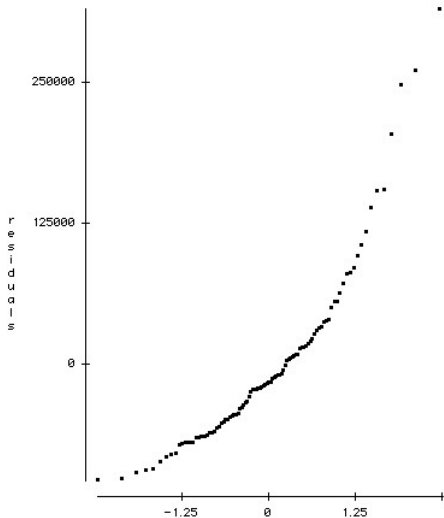
Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Normal Probability Plot (Not Center City)



Phila Housing Prices and Crime

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Housing Prices vs. 1000 People per Crime for suburbs of Philadelphia, now including the center city.

Phila Housing Prices and Crime

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

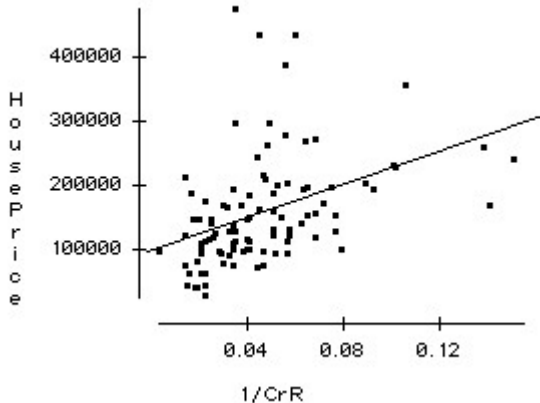
Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Scatterplot Housing Prices vs. Crime Rate
(1000 people per crime)



Phila Housing Prices and Crime

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Regression Display (1000 people per crime)

Dependent variable is: **HousePrice**
No Selector
111 total cases of which 12 are missing
R squared = 17.4% R squared (adjusted) = 16.5%
s = 79.16e3 with 99 - 2 = 97 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	127.948e9	1	127.948e9	20.4
Residual	607.758e9	97	6.26555e9	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	97920.6	15.46e3	6.33	≤ 0.0001
1/CrR	1.30138e6	288e3	4.52	≤ 0.0001

Phila Housing Prices and Crime

Math 1710
Class 41

V1

Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

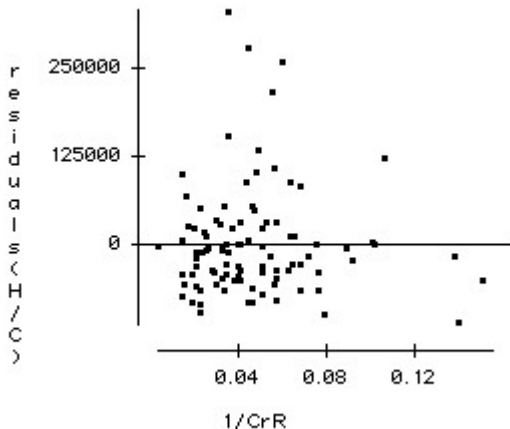
Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime

Residual Plot (1000 people per crime)

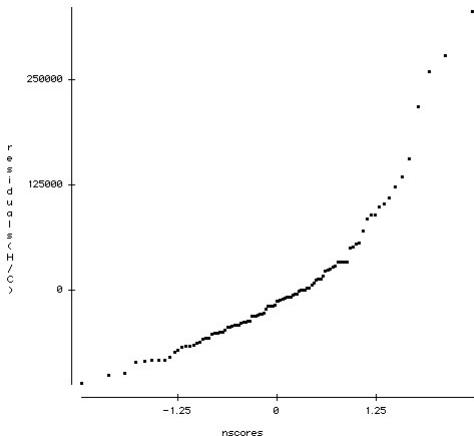


Phila Housing Prices and Crime

Math 1710
Class 41

V1

Normal Probability Plot (1000 people per crime)



Regression
Inference
Questions

Cleaning
Crews
Example

Question 5

Regr Inf
Conditions

Calculations in
3 Cases

Cleaning
Inaccuracies

Identifying
Which
Inference

Building
Height

Phila HP and
Crime