## Simple Linear Regression in StatCrunch

Hello! In this video I will be going through the steps involved in solving a typical linear regression problem using StatCrunch. Given a data set, we will draw a scatter diagram and then find the correlation coefficient, the critical value for r, and the equation of the regression line. Then, using our regression line, we will find the residual for a given value.

Ok so let's say you have a problem like this:

"The accompanying data represent the square footage and selling price (in thousands of dollars) for a random sample of homes for sale in a certain region." (Problem 4.2.29)

Square Footage, x	Selling Price (\$000s), y	모
2286	392.7	
3116	364.4	
1100	185.5	
2074	353.5	
3146	626.7	
2799	374.3	
4028	614.4	
2256	384.2	
2508	409.2	
1712	298.9	
1720	260.8	
3845	694.9	

#### a) Draw a scatter diagram of the data.

Part a) says "Draw a scatter diagram of the data." The first thing we do is open up our data in StatCrunch.

	tLab Data S					
StatCr	unch Ap	plets - Ed	lit - Data	• Stat •	Graph 🕶	Help 🕶
Row	Square Footage	Selling Price (\$	var3	var4	var5	var
1	2286	392.7				
2	3116	364.4				
3	1100	185.5				
4	2074	353.5				
5	3146	626.7				
6	2799	374.3				
7	4028	614.4				
8	2256	384.2				
9	2508	409.2				
10	1712	298.9				
11	1720	260.8				
12	3845	694.9				
10						

We can make a scatter plot using the "Regression" function under "Stat -> Regression -> Simple Linear".

	ntLab Data Set								
StatC	runch - Appl	ets ▼ Edit ▼	Data 🕶	Stat -	Graph 🕶	Help	•		
Row 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	Square Footage Sc 2286 3116 1100 2074 3146 2799 4028 2256 2508 1712 1720 3845	Hing Price (\$ 392.7 364.4 185.5 353.5 626.7 374.3 614.4 384.2 409.2 298.9 260.8 694.9	var3	Tables Z Stat T Stat Propor Varian Regree ANOV Nonpa Goodn	s s tion Stats ce Stats ssion A rametrics ress-of-fit ol Charts aple		Polyn	ole Linear	•

Next we select our X and Y variables. In this question, we are interested in predicting the price of a home using its square footage. This means that Square Footage will be our explanatory or independent variable, also known as X. Since we are trying to predict the Selling Price, Selling Price will be our dependent or response variable, which is our Y.

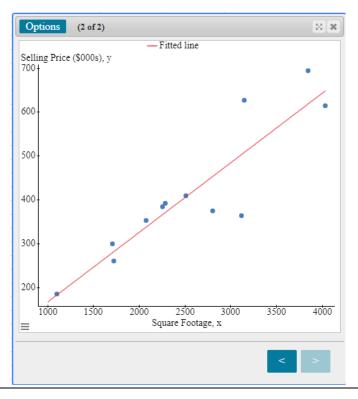
ple Linear Regression		
X variable:	¥	_
Square Footage, x	*	
Y variable: Selling Price (\$000s), y	<b>*</b>	
Where:		
optional	Build	
Group by:		
optional	Ŧ	
Perform:		

For now, we don't have to change anything else. So just hit "Compute!"

And we have our results window:

Depende Independ Selling P Sample s R (correl R-sq = 0. Estimate	inean nt Va lent V rice ize: ation .8161 of en	coefficient)	ng Price (\$00 are Footage, 8.213062 + 0 = 0.9033882	x 0.1 37	5890628	Squ	are Fo	ootage, x	53 8
Parame	eter	Estimate	Std. Err.		Alterna	tive	DF	T-Stat	P-valu
Intercep	t	8.213062	64.0254	42		≠ <b>0</b>	10	0.12827814	0.900
Slope		0.15890628	0.0238531	09		≠ <b>0</b>	10	6.6618689	<0.000
Analysis		ariance tabl SS	e for regres MS		n model: F-stat	D 1	alue		
							uiuc		
Model	1		214059.27	44	1.380497	<0.	0001		
Error	10	48232.735	4823.2735						
Total	11	262292.01							
				_				<	>

To find our scatter plot, click the arrow on the bottom of the results box. And there we go, we have a scatter plot with our regression line running through it.



b) Find the correlation coefficient, r.

Next, for part b), Find the correlation coefficient, r. Basically, we are being asked how strongly these two variables are related.

In order to find r, we click on the left arrow to go back to the original results box. The correlation coefficient R is the 5<sup>th</sup> item down. For this data set, r is about .903. Be sure to pay attention to how the question wants you to round your answer.

#### c) Is there a linear relationship between these variables?

Next we want to know for part c): "Is there a linear relationship between these variables?" Is square footage actually a good predictor for the selling price of a house?

In order to figure this out, we need to compare the absolute value of our r to the critical value for this sample size.

First let's find the critical value. If you are using MyStatLab, you will be given a table of Critical Values

- . . . . . . . . . . . .

Critical Values for C	orrelation Coefficient
п	
3	0.997
4	0.950
5	0.878
6	0.811
7	0.754
8	0.707
9	0.666
10	0.632
11	0.602
12	0.576
13	0.553
14	0.532
15	0.514
16	0.497
17	0.482
18	0.468
19	0.456
20	0.444
21	0.433
22	0.423
23	0.413
24	0.404
25	0.396
26	0.388
27	0.381

for the Correlation Coefficient that looks something like this:

In order to find our critical value, we need to know how many observations are in our data set. In this case we had 12 observations. 12 houses. This is our "n" in this new table. The critical value associated with 12 is .576.

Next, we take the absolute value of our correlation coefficient.  $|\mathbf{r}|$  is |.903| is just .903.

Now we compare the absolute value of r to the critical value. Since .903 is bigger than .576, we say that there is in fact a linear relationship between these variables.

d) Find the least-squares regression line treating square footage as the explanatory variable.

For part d), we want to find the least-squares regression line, treating square footage as the explanatory variable. Since we know that there is a linear relationship between these two variables, it makes sense to find a linear regression line for them.

R-sq = 0 Estimate	ation .8161 of er	12 coefficient) 11054	8.213062 + 0 = 0.9033883 deviation: 69	7	- 1 -		8-,	
Parame Param		Estimates:	Std. Err.	Alterna	tive	DF	T-Stat	P-va
Intercep	ot	8.213062	64.0254	2	≠ 0	10	0.12827814	0.9
Slope		0.15890628	0.02385310	9	≠ 0	10	6.6618689	<0.00
Analysis			e for regress					
	DF	SS	MS	F-stat	P-1	alue		
Source	DI			44.200407	<0	0001		
Source Model		214059.27	214059.27	44.380497		0001		
	1	214059.27 48232.735		44.380497		0001		

This is also found in the first Results screen. It is the third item down. Something to be aware of: Some homework might want you to report the line in the form  $y_hat=mx+b$ , but StatCrunch gives you the line in the form y = b + mx. So be sure to keep your slope and y intercept straight when you are inputting answers.

Here for example, if we wanted the line reported in y = mx+b form rounded to 3 decimals, we would write:

 $y_hat = .159x + 8.213$ 

### e) Interpret the slope of the regression line.

Next, we are asked to interpret the slope of the regression line. Many problems will ask you to also "interpret" the slope of your line. There is a format to follow when answering. The basic format for a positive slope is:

For every 1 (units of x) increase in (x-variable), we can expect a (slope) (units of y) increase in (y-variable).

So for our example, our answer would be "For every 1 square foot increase in the Square Footage of a house, we can expect a .159 thousand dollar increase in the Selling Price of the house."

#### f) What is the predicted selling price of a home that is 1433 square feet?

Ok now, we want to know "What is the predicted selling price of a home that is 1433 square feet?" So, in order to find the predicted value by hand, we would have to plug the value 1433 into our regression equation and solve for y. However, StatCrunch will do this for us. Go to "Options -> Edit" to get back to our Simple Linear Regression menu.

Edit gression results: Save Save Save Save Sele: Selling Price (\$000s), y able: Square Footage, x Dos), y = 8.213062 + 0.15890628 Square Footage, x Print Download efficient) = 0.90338837 standard deviation: $69.449791$ Delete rarameter estimates:Deficient) = 0.90338837 standard deviation: $69.449791$ DeleteParameterEstimate Std. Err.Alternative 4 0 10DFT-StatP-valIntercept8.213062 $64.02542$ $\neq 0$ 10 $0.12827814$ $0.90$ Slope $0.15890628$ $0.023853109$ $\neq 0$ 10 $6.6618689$ $<0.00$ Analysis of variance table for regression model: SourceDeSMSF-statP-valueModel1214059.27214059.2744.380497 $<0.0001$ Error1048232.7354823.2735 $\_$ Total11262292.01 $\_$ $\_$		-		-				_	_		
gression results:Save ble: Selling Price (\$000s), y able: Square Footage, x Dos), y = 8.213062 + 0.15890628 Square Footage, x PrintDownload efficient) = 0.90338837 standard deviation: $69.449791$ Deletearameter esomates:Parameter EstimateStd. Err.Alternative $4$ DFT-StatP-valInterce resonates:Parameter EstimateStd. Err.AlternativeDFT-StatP-valInterce tableStd. Err.AlternativeDFT-StatP-valMSF-statP-valModel10.158906280.0001ErrorDI deleteAnalysis of variance table for regression model:SourceDFSMSF-statP-valueModel1214059.27214059.2744.380497<0.0001Error104823.2735Delete	Options		(1 of 2)								2
Copy Printable: Square Footage, x Dos), y = 8.213062 + 0.15890628 Square Footage, x PrintDownload Refresh Deleteefficient) = 0.90338837 54 standard deviation: $69.449791$ Deletestandard deviation: $69.449791$ Deletestandard deviation: $69.449791$ Deletestandard deviation: $69.449791$ Deletestandard deviation: $69.449791$ Intercept8.213062 $64.02542$ $\neq 0$ 100.128278140.90Slope0.158906280.023853109 $\neq 0$ 106.6618689sourceDFSSMSModel1214059.27214059.2744.380497 $<0.0001$ Error1048232.7354823.2735Total11262292.01 $<$ Predicted values:	Edit		gressi	on	results:						
Copy $pos$ , y = 8.213062 + 0.15890628 Square Footage, x         Print $efficient$ ) = 0.90338837         Refresh $efficient$ ) = 0.90338837         Download $efficient$ ) = 0.90338837         Refresh $standard deviation: 69.449791$ Delete $standard deviation: 69.449791$ Parameter       Estimate       Std. Err.       Alternative       DF       T-Stat       P-val         Intercept       8.213062       64.02542 $\neq 0$ 10       0.12827814       0.90         Slope       0.15890628       0.023853109 $\neq 0$ 10       6.6618689       <0.00	Save						), y				
Print Download Refresh Delete9.909.909.909.90Refresh Deleteefficient) = $0.90338837$ standard deviation: $69.449791$ <b>DFT-StatP-val</b> Delete rarameter estimates: <b>Std. Err.AlternativeDFT-StatP-val</b> Intercept $8.213062$ $64.02542$ $\neq 0$ 10 $0.12827814$ $0.90$ Slope $0.15890628$ $0.023853109$ $\neq 0$ 10 $6.6618689$ $<0.00$ Analysis of variance table for regression model: <b>SourceDFSSMSF-statP-val</b> Model1 $214059.27$ $214059.27$ $44.380497$ $<0.0001$ $<0.0001$ Error10 $48232.735$ $4823.2735$ $<0.0001$ Total11 $262292.01$ $<0.001$ $<0.0001$ Predicted values: $<0.001$ $<0.0001$	Сору						5800628	Sou	are Fo	ootage x	
Refresh DeleteA standard deviation: $69.449791$ DeleteParameterEstimateStd. Err.AlternativeDFT-StatP-valIntercept $8.213062$ $64.02542$ $\neq 0$ 10 $0.12827814$ $0.90$ Slope $0.15890628$ $0.023853109$ $\neq 0$ 10 $6.6618689$ $<0.00$ Analysis of variance table for regression model:SourceDFSSMSF-statP-valueModel1 $214059.27$ $214059.27$ $44.380497$ $<0.0001$ Error10 $48232.735$ $4823.2735$ $=$ $=$ Total11 $262292.01$ $=$ $=$ $=$ Predicted values:							5050020	oqui		Jourge, A	
Refresh Deletestandard deviation: 69.449791Deletestandard deviation: 69.449791ParameterEstimateStd. Err.AlternativeDFT-StatP-valIntercept $8.213062$ $64.02542$ $\neq 0$ 10 $0.12827814$ $0.90$ Slope $0.15890628$ $0.023853109$ $\neq 0$ 10 $6.6618689$ $<0.00$ Analysis of variance table for regression model:SourceDFSSMSF-statP-valueModel1 $214059.27$ $214059.27$ $44.380497$ $<0.0001$ Error10 $48232.735$ $4823.2735$ $=$ $=$ Predicted values:Predicted values:Predicted values:Predicted values:Predicted values:	Dowr	nloa	ad efficient	nt)	= 0.9033883	37					
Source of SS         MS         F-stat         P-val           Intercept         8.213062         64.02542 $\neq 0$ 10         0.12827814         0.90           Slope         0.15890628         0.023853109 $\neq 0$ 10         6.6618689         <0.00           Analysis of variance table for regression model:         Source         DF         SS         MS         F-stat         P-value           Model         1         214059.27         214059.27         44.380497         <0.001	Refre	sh	· ·	ard	deviation: 6	9.4	49791				
Parameter       Estimate       Std. Err.       Alternative       DF       T-Stat       P-val         Intercept $8.213062$ $64.02542$ $\neq 0$ 10 $0.12827814$ $0.00$ Slope $0.15890628$ $0.023853100$ $\neq 0$ 10 $6.6618689$ $<0.00$ Analysis       Transmeter       For regression       F-stat       P-val       P-val         Model       1 $214059.27$ $214059.27$ $44.380497$ $<0.000$ Model       1 $26292.03$ $4823.2735$ $< 5.7$ $< -5.7$ Predicter $< 1000$ $26292.01$ $< 1000$ $< 5.7$ $< -5.7$			sum afes.								
Slope       0.15890628       0.023853109 $\neq$ 0       10       6.6618689       <0.00				e	Std. Err.		Alterna	tive	DF	T-Stat	P-val
Source         DF         SS         MS         F-stat         P-value           Model         1         214059.27         214059.27         44.380497         <0.0001	Intercep	t	8.2130	62	64.025	42		≠ 0	10	0.12827814	0.90
Source         DF         SS         MS         F-stat         P-value           Model         1         214059.27         214059.27         44.380497         <0.0001	Slope		0.158906	28	0.0238531	09		≠ 0	10	6.6618689	<0.00
Source         DF         SS         MS         F-stat         P-value           Model         1         214059.27         214059.27         44.380497         <0.0001	Analysis	ofv	ariance ta	abl	e for regres	sio	n model:				
Error       10       48232.735       4823.2735         Total       11       262292.01	_				_				alue		
Total 11 262292.01 Predicted values:	Model	1	214059.2	27	214059.27	44	1.380497	<0.	0001		
Predicted values:	Error	10	48232.7	35	4823.2735						
	Total	11	262292.0	01							
	Drodicto	d va	huos								
				s.e.	(Pred. v)	95	5% C.I. f	or m	lean	95% P.	I. for
										<	>
< >											

And then now we find the section labeled: "Prediction of Y:" In the "X value(s):" box, enter our value of

	Prediction of Y:         X value(s):       1433         Level:       0.95	
	Transformation:	L
	X: None 🔻	E
	Y: None	
1433.	Use original units in graphs	

Then click "Compute!" again. Now on our Results page, we have a new section at the bottom labeled "Predicted values:" The value we want for this question is listed under "Pred. Y" Which is about 235.9 (or 236).

X value	Pred. Y	s.e.(Pred. y)	95% C.I. for mean	95% P.I. for new
1433	235.92577	33.328355	(161.66557, 310.18597)	(64.285988, 407.56555)
100	233.32311	55.528555	(101.00557, 510.16557)	(04.203300, 407.30333)

So for our question, the predicted selling price for a house that is 1433 square feet is about 236 thousand dollars.

# g) We observe a 1433 square foot home selling for \$210 thousand. What is the residual of this value?

Next for our final part, if we observe a 1433 square foot home selling for \$210 thousand, what is the residual of this value?" Recall that residuals are found by subtracting the predicted value from the observed value. In other words: Residual = Observed – Predicted.

We found the predicted value in the previous question. It was 236 thousand. We know our observed value is 210 thousand. So our residual for this observation is 210 - 236 = -26 thousand.

Alright, so we have answered all parts of our question. I hope you found this video helpful.

If you are a UNT student, there will be some links to other resources in the video description. Thank you for watching!