EDUR 8131
Chat 9

1. Notes 8a Simple Regression

1.1 Regression Equation

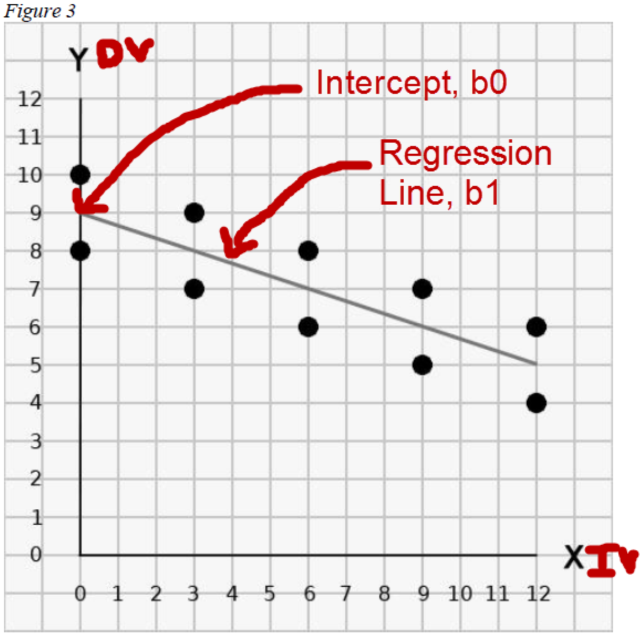
Provides equation for prediction; predicts DV that is quantitative using IVs (multiple) that can be quantitative or qualitative, or both. This course focuses just on quantitative IVs.

Y = b0 + b1 X + e
Y' = b0 + b1 X (prediction equation, used to obtain predicted value of Y)

b0 = intercept, point where regression line crosses Y axis
b1 = indicates how much mean change in Y is expected for a 1-unit change in X
e = error or residual term – deviation between Y and Y', i.e., Y-Y'
Y' = predicted Y using the regression equation

[Note – show scatterplot with regression to illustrate b0, b1, e, and Y' (Notes 8a)]

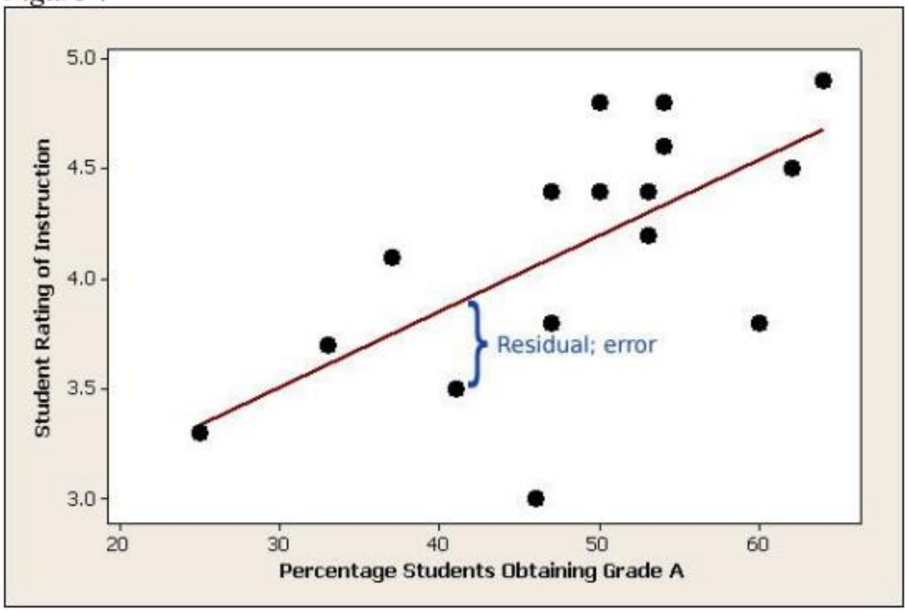
Figure 3



What is the value of b0? Recall b0 is where regression line crosses Y axis when X = 0.
B0 =9

What is value of b1, slope showing association between X and Y?
B1 = -.333 (-1/3)

Figure 4

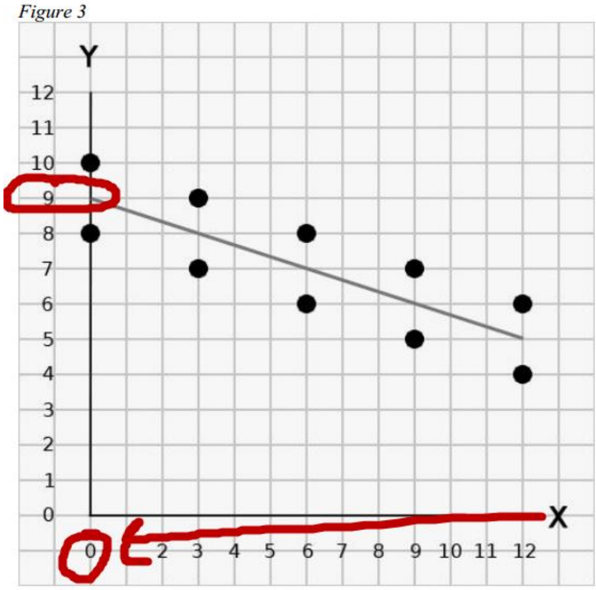


1.2 Literal Interpretation of Coefficients and Predicted Values

b0 = predicted value of Y when X = 0.00

b1 = for each 1 unit increase in X, the mean of Y is expected to change by b1

Figure 3



Examples

(a) X and Y Interpretation

b0 = 9.00

b1 = -0.33

Prediction Equation

$Y' = b0 + b1\ X$

$Y' = 9.00 + -0.33\ (X)$

Interpretation of Coefficients

b0 = 9.00: Y is expected to be 9.00 when X is 0.00

b1 = for each 1 point increase in X, Y is expected to change by -.33

(b) Math Achievement and Test Anxiety

IV = Test Anxiety

DV = Math Achievement

Prediction Equation

(Math Achievement)’ = b0 + b1 (Test Anxiety)

(Math Achievement)’ = 9.00 + -0.33 (Test Anxiety)

Interpretation of Coefficients

b0 = 9.00:

Math achievement is predicted to be 9.00 when test anxiety is 0.00

b1 = -0.33:

For each 1 point increase in test anxiety, math achievement changes by -0.33.

If Test Anxiety is 12, what is the predicted value of Math Achievement?

(Math Achievement)’ = 9.00 + -0.33 (Test Anxiety)

(Math Achievement)’ = 9.00 + -0.33 (12)

= 9.00 + (-3.96)

= 5.04

If Test Anxiety is 5, what is the predicted value of Math Achievement?

(Math Achievement)Y’ = 9.00 + -0.33 (Test Anxiety)

(Math Achievement)’ = 9.00 + -0.33 (5)

= 9.00 + (-1.65)

= 7.35

(c) EDUR 8131 Test 2 Score and Hours Studied Since Prior Test

IV = Hours studied since prior test

DV = EDUR 8131 Test 2 Score

Prediction Equation

Test 2 Score’ = b0 + b1 (Hours Studied)

Test 2 Score’ = 55.00 + 2.50 (Hours Studied)

Interpretation of Coefficients

b0 = 55.00:

Remember, b0 is the predicted value of the DV (Test 2 scores) when the IV is 0.00.

Test 2 Score’ = 55.00 + 2.50 (Hours Studied = 0.00)

Test 2 Score’ = 55.00 + ~~2.50 (Hours Studied = 0.00)~~

Test 2 Score’ = 55.00

Test 2 score is predicted to be 55.00 for those who do not study (hours studied = 0.00)

b1 = 2.50:

For each hour studied, test 2 scores are expected to increase 2.50 points

If one studied 11 hours, what is the predicted score on EDUR 8131 Test 2?

Test 2 Score’ = 55.00 + 2.50 (Hours Studied)

Test 2 Score’ = 55.00 + 2.50 (11)

Test 2 Score’ = 55.00 + 27.50

= 82.5

If one studied 17 hours, what is the predicted score on EDUR 8131 Test 2?

Test 2 Score’ = 55.00 + 2.50 (Hours Studied)

Test 2 Score’ = 55.00 + 2.50 (17)

Test 2 Score’ = 55.00 + 42.5

= 97.5

If one studied 0 hours, what is the predicted score on EDUR 8131 Test 2?

Test 2 Score’ = 55.00 + 2.50 (Hours Studied)

Test 2 Score’ = 55.00 + 2.50 (0)

= 55.00

Another example for interpretation of coefficients

DV = money earned
IV = hours worked per week

Money earned’ = b0 + b1(hours worked)

Coefficient values b0 = \$0.00 b1 = \$10 Money earned' = b0 + b1(hours worked) Money earned' = \$0.00 + \$10(hours worked) Literal Interpretation b0 = if you don't work any hours, you money earned would be \$0.00 (no work, no pay) b1 = for each additional hour worked, money earned increases by \$10

1.3 Predicted Values vs. Expected Change

Predicted values obtained from prediction equation:

Prediction Equation

Y' = b0 + b1 (Hours Studied)

Y' = 55.00 + 2.50 (Hours Studied)

Example:

Y' = 55.00 + 2.50 (Hours Studied = 10)

Y' = 55.00 + 2.50 (10) =

80.00 – predicted test score on Test 2

Expected change is obtained from the slope coefficient:

b1 = 2.50 (Hours Studied)

Example:

How much increase is expected for someone who studies 10 additional hours?

Expected Change = 2.50 (Hours Studied)

= 2.50 (10)

= 25 point increase

How much increase is expected for someone who studies an additional 7 hours?

Expected Change = 2.50 (Hours Studied)

= 2.50 (7)

= 17.5 point increase

1.4 Obtaining Regression Estimates
Data taken from this site:

http://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/mlr/frames/frame.html

SPSS Data file here:

http://www.bwgriffin.com/gsu/courses/edur8131/data/cricket_data.sav

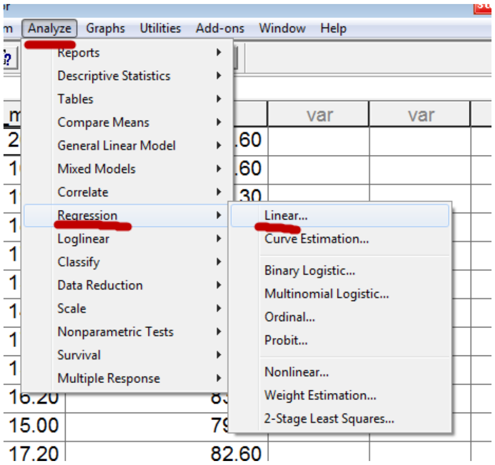
Cricket data

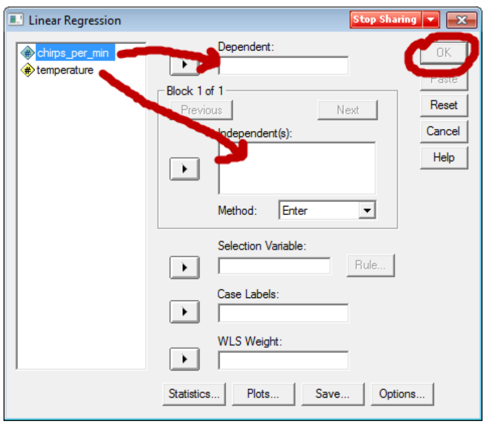
Chirps per minute	temperature
20	88.6
16	71.6
19.8	93.3
18.4	84.3
17.1	80.6
15.5	75.2
14.7	69.7
17.1	82
15.4	69.4
16.2	83.3
15	79.6
17.2	82.6
16	80.6
17	83.5
14.4	76.3

(a) What is the regression equation and coefficient estimates for the above data?

Chirps' = b0 + b1 (temp)

SPSS Commands to obtain estimates





SPSS Results

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.309	3.109		-.099	.922
	temperature	.212	.039	.835	5.475	.000

a. Dependent Variable: chirps_per_min

So what is the resultant regression equation with estimates?

Chirps' = b0 + b1 (temp)

Chirps' = -.309 + .212 (temp)

(b) Literal Interpretation

DV = cricket chirps per minute

IV = temperature F

b0 = -.309:

Predicted chirps per minute is -.309 when temp = 0.00

b1 = .212:

for each 1 degree increase in temp, mean chirps per minute are predicted to increase by .212

What is the general interpretation for the slope?

Note that general interpretation is simply a more relaxed interpretation of results.

There is a positive association between chirps per minute and temperature, as temperature increases so too do the number of chirps per minute.

(c) How many chirps are expected when temperature is 70?

Chirps’ = b0 + b1 (temp)

Chirps’ = -.309 + .212 (temp)

Chirps’ = -.309 + .212 (70) = 14.531

(d) How many chirps are expected when temperature is 85?

Chirps’ = b0 + b1 (temp)

Chirps’ = -.309 + .212 (temp)

Chirps’ = -.309 + .212 (85) = 17.711

(e) If temperature increases by 15 degrees, what change is expected in number of chirps per minute?

Expected Change in DV = b1 * (change in IV)

Expected Change in DV = .212 * (15) = 3.18

Recall two predicted values above:

Temp = 85, then predicted Chirps’ = -.309 + .212 (85) = 17.711

Temp = 70, then predicted Chirps’ = -.309 + .212 (70) = 14.531

Note difference in temperature = 85 – 70 = 15

Difference in predicted chirps = 17.711 – 14.531 = 3.18

(f) If temperature decreases by 28 degrees F, what change is expected in number of chirps per minute?

Expected Change in DV = b1 * (change in IV)

Expected Change in DV = .212 * (-28) = -5.936

(g) How well does this regression model, Chirps’ = -.309 + .212(temp), fit the observed data?

1.5 Model Fit

How well does predicted (expected) chirps fit with, or match, observed chirps? (Note that the difference between expected and observed chirps is known as error or residual.)

Recall from chi-square we examine the difference between expected and observed counts to assess goodness-of-fit.

Standard error of estimate (SEE) is one measure of model fit that examines size of discrepancy between observed and expected values of the DV; SEE is just the standard deviation of the residuals (errors). Problem is that SEE can range from 0.00 to infinity (i.e., it has no standardized index for judging value), so difficult to judge goodness-of-fit for regression model using SEE.

We need standardized measure of model fit that works across all linear regression equations.

(a) Multiple R = correlation of the actual Y and predicted Y', correlation between observed and predicted (i.e., Y and Y')

[Excel - calculate predicted values, SPSS correlation table]

SPSS Linear Regression Model Fit Table

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.835(a)	.697	.674	.97152

a Predictors: (Constant), temperature

SPSS Regression Model Fit Table

Multiple R = .835

Correlations

		chirps_per_min	predicted_chiprs
chirps_per_min	Pearson Correlation	1	.835(**)
	Sig. (2-tailed)		.000
	N	15	15
predicted_chiprs	Pearson Correlation	.835(**)	1
	Sig. (2-tailed)	.000	
	N	15	15

** Correlation is significant at the 0.01 level (2-tailed).

(b) Multiple R^2 = multiple R squared

For the current model R^2 = ?

For the current model $R^2 = .835^2 = .697$

Interpretations

Proportional reduction in error = this tells us we can reduce our error in predicting chirps by .697 or about 70% by using this regression model over just using the mean number of chirps as a predicted value

Proportion of Variance Explained or Predicted = our regression model predicts about 70% (actual 69.7%) of the variance in chirps. Thus, multiple R^2 is an estimate of how much variability in DV can be predicted by regression equation.

(c) Adjusted R^2 = simply R^2 adjusted downward given the sample size.

See SPSS
adjusted R^2 = .674

Same interpretation as R^2 above.

SPSS Results from Chirps and Temperature

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.835[a]	.697	.674	.97152

a. Predictors: (Constant), temperature

Standard error of estimate = standard deviation of residuals = .97152

1.6 Inference in Regression

(a) Overall model fit

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.835[a]	.697	.674	.97152

a. Predictors: (Constant), temperature

Overall Model Fit Null Hypothesis

Ho: $R^2 = 0.00$

Alternative hypothesis

H1: $R^2 \neq 0.00$

If $R^2 = 0.00$, what does this tell us about the regression model?

Regression is unable to predict any variance in DV; extremely poor fit.

This null tests whether R^2 is greater than what would be expected by random chance coincidence. If Ho: $R^2 = 0.00$ is rejected, this tells us that the regression model is predicting more variance in the DV than would be expected by chance alone.

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	28.287	1	28.287	29.970	.000[a]
	Residual	12.270	13	.944		
	Total	40.557	14			

a. Predictors: (Constant), temperature

b. Dependent Variable: chirps_per_min

We test Ho: $R^2 = 0.00$ using what is known as an F-test; ANOVA F-test.

If the F value is larger than the critical (we learn how to find this when we study ANOVA), or if the F ratio's p-value (called Sig. above) is less than alpha, then reject Ho: $R^2 = 0.00$ and conclude regression equation is predicting more variance in the DV than would be expected by coincidence or random error. In short, this test informs us whether our regression model as a whole is predicting more variance in the DV than would be expected by chance.

(b) Regression Coefficients

SPSS Results

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.309	3.109		-.099	.922
	temperature	.212	.039	.835	5.475	.000

a. Dependent Variable: chirps_per_min

Tested with t-ratio just like with a t-test.
Same decision rule with p-values and alpha (α) applies:

If $p \leq \alpha$ reject Ho; if $p > \alpha$ fail to reject Ho.

Ho: $\beta_0 = 0.00$

Since $p > .05$ ($p = .922$ and .922 is greater than .05) fail to reject Ho and conclude that the intercept does not differ from 0.00.

Ho: $\beta_1 = 0.00$

Since $p < .05$ ($p = .000$ and is less than .05) we reject Ho and conclude that temp and chirps are related positively

1.7 APA Style Presentation with Cricket Data

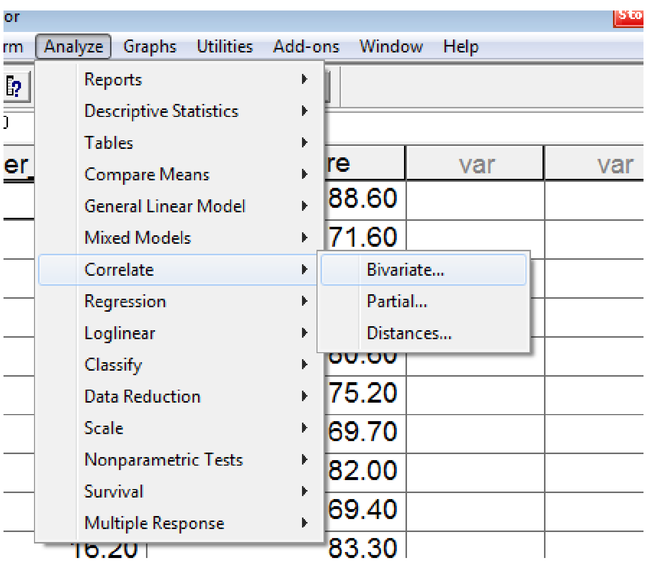
Table 1: Correlations and Descriptive Statistics for Cricket Chirps per Minute and Temperature

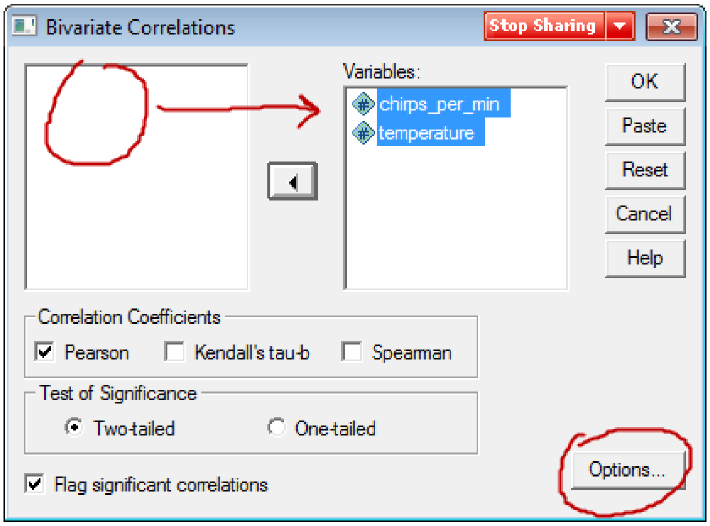
	1	2
1. Number of Cricket Chirps	---	
2. Temperature	.835*	---
Mean	16.65	80.04
SD	1.70	6.71

Note: n = 15

* $p < .05$

[Remember, use Correlate->Bivariate command to obtain correlations otherwise the regression command descriptive statistics option will report only 1-tailed p-values which are too small.
Also, under Options with correlation in SPSS, choose "Exclude cases listwise" so the sample size will be the same for both correlation and regression.]





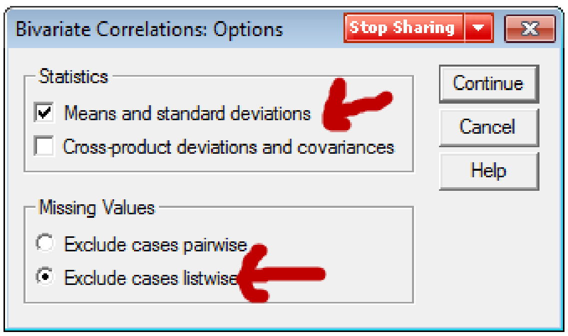


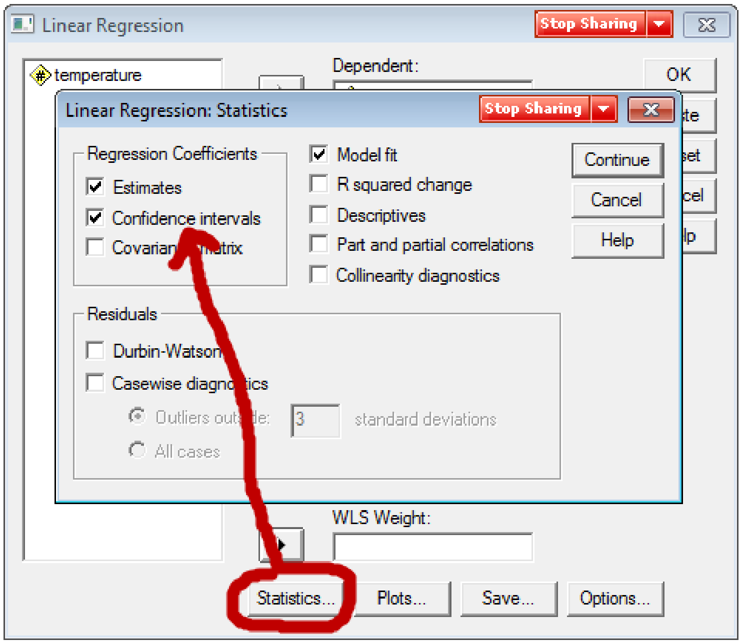
Table 2: Regression of Number of Cricket Chirps on Temperature

Variable	b	se b	95% CI	t
Temperature	0.212	0.039	0.128, 0.296	5.475*
Intercept	-0.309	3.109	-7.025, 6.407	-0.099

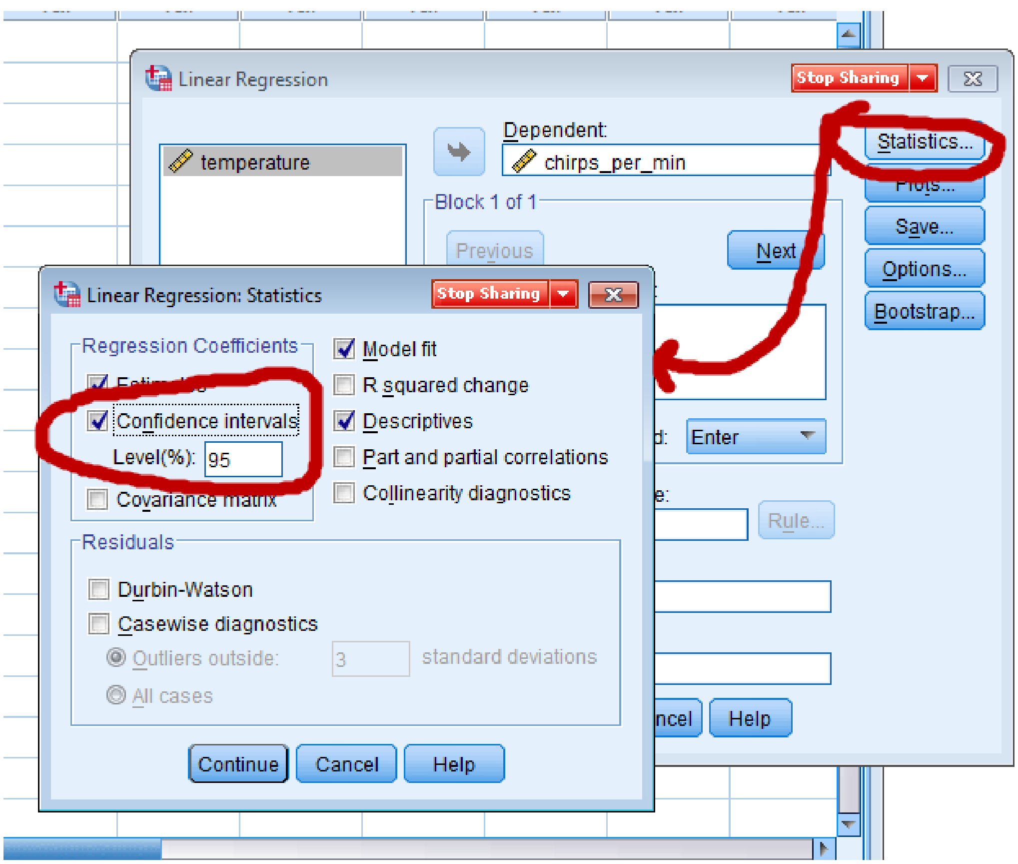
Note: R^2 = .697; adj. $\underline{R}^2$ = .674; F =29.97*; df = 1,13; MSE = 0.94; n = 15
*p < .05

Regression results show a statistically significant association between temperature and number of cricket chirps per minute. As temperature increases, the number of chirps per minute also tends to increase.

Note, to obtain confidence intervals with regression, use the following command:



Newer version SPSS commands for CI



SPSS Results

Bivariate Correlation Command Output

Descriptive Statistics

	Mean	Std. Deviation	N
chirps_per_min	16.6533	1.70204	15
temperature	80.0400	6.70733	15

Correlations

		chirps_per_min	temperature
chirps_per_min	Pearson Correlation	1	.835**
	Sig. (2-tailed)		.000
	N	15	15
temperature	Pearson Correlation	.835**	1
	Sig. (2-tailed)	.000	
	N	15	15

**. Correlation is significant at the 0.01 level (2-tailed).

Liner Regression Command Output

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.835[a]	.697	.674	.97152

a. Predictors: (Constant), temperature

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	28.287	1	28.287	29.970	.000[a]
	Residual	12.270	13	.944		
	Total	40.557	14			

a. Predictors: (Constant), temperature

b. Dependent Variable: chirps_per_min

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-.309	3.109		-.099	.922	-7.025	6.407
	temperature	.212	.039	.835	5.475	.000	.128	.296

a. Dependent Variable: chirps_per_min