Self-Assessment

Weeks 2 and 3: Regression with One Quantitative Predictor; Model Fit

Answers

1. Below is a regression analysis with two variables:

DV = violent crime rate per 100,000 in each US state
IV = percent of US state population living in single-parent households

Descriptive statistics for the two variables are provided first. Violent crime rate ranges from a low of 82 (North Dakota) to a high of 1206 (Florida), with a mean crime rate of 566.67. Percent of population in single-parent households ranges from a low of 8.4% (North Dakota) to a high of 14.9% (Louisiana), with a mean of 11.11%.

Statistics

		violent crime rate per 100,000	percent of population in single parent family
N	Valid	50	50
	Missing	0	0
Mean		566.66	11.1100
Median		509.50	10.9000
Mode		208	10.80
Std. Deviation		295.877	1.47513
Variance		87543.372	2.176
Range		1124	6.50
Minimum		82	8.40
Maximum		1206	14.90

Regression results are presented below.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.649(a)	.421	.409	227.514

a Predictors: (Constant), percent of population in single parent family

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1805011.036	1	1805011.036	34.871	.000(a)
	Residual	2484614.184	48	51762.796		
	Total	4289625.220	49			

a Predictors: (Constant), percent of population in single parent family
b Dependent Variable: violent crime rate per 100,000

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-878.861	246.895		-3.560	.001	-1375.278	-382.445
	percent of population in single parent family	130.110	22.033	.649	5.905	.000	85.809	174.411

a Dependent Variable: violent crime rate per 100,000

The prediction equation for this regression model is

Predicted Y = b0 + b1 X

Predicted Crime Rate per 100,000 = -878.86 + 130.11 (Percent in Single-parent Households)

(a) What is the literal interpretation for values obtained for b0 and b1?

> **b0 = states with 0.0% single-parent households have a predicted violent crime rate of -878.86 per 100,000 population.**
>
> **b1 = for each 1 percentage point increase in percent of single-parent homes, the violent crime rate is expected to increase by 130.11 per 100,000 population**

(b) What is the general interpretation for b1?

> **There is a positive relation between percent of households that are single-parent and violent crime rate across states; the greater the percentage of single-parent households, the greater will be violent crime rates.**

(c) What percent of the crime rate variance can be predicted by knowing the percent of households that are single parents?

> **R^2 = .421 and adjusted R^2 = .409 so about 41% to 42% of variance in violent crime rate is predicted by percent of families in single-parent homes.**

(d) Does this model predict more variance in crime rates than would be expected by chance?

> **Yes since R^2 = .421 is statistically significant at the .05 level – see replies below.**
>
> (d1) Which is the statistic used to measure how much variance is predicted, and what amount was predicted for this model?
>
> > **R^2 and adjusted R^2, R^2 = .421 and adjusted R^2 = .409**
>
> (d2) What is the symbolic and written null hypothesis assessed in question (d)?
>
> > **Symbolic: Ho: R^2 = 0.00**
> > **Written: The model predicts no variance in violent crime rate across the states**

(d3) What is the test statistic used to test the null hypothesis found in question (d2)?

F-ratio, which for this model is F = 34.87

(d4) What p-value is reported in the regression analysis for the test statistic sought in question (d3)?

p = 0.000 (or more appropriately reported as p > .001).

(e) Is the slope, b1, for this model statistically significant at the .05 level (explain your response)?

Yes because p-value for the slope is less than .05, p = .000

(f) What is the interpretation for the 95% confidence interval for b1?

One may be 95% confidence that the population slope relating percent of single-parent homes in a state to violent crime rate lies between 85.81 and 174.41.

(g) How can the 95% confidence interval for b1 be used to test Ho: b1 = 0.00?

The default null for a regression slope is Ho: β_1 = 0.00. This means no relationship in the population.

The confidence interval provides a range of possible values estimated for the population slope.

Given the two explanations above, one may therefore use the confidence interval to test hypotheses about slopes with the following logic:

(a) If the value of 0.00 lies within the confidence interval, then a population slope of 0.00 is viewed as possible so fail to reject Ho: β_1 = 0.00.

(b) If the value of 0.00 does not lie within the confidence interval, then a population slope of 0.00 is viewed as not probable so one may reject Ho: β_1 = 0.00.

(h) What is the predicted violent crime rate for North Dakota (8.4% of households are single parents)?

Predicted Crime Rate per 100,000 = -878.86 + 130.11 (Percent in Single-parent Households)
Predicted Crime Rate per 100,000 = -878.86 + 130.11 (8.4)
Predicted Crime Rate per 100,000 = -878.86 + 1092.924
Predicted Crime Rate per 100,000 = 214.064

(i) What is the predicted violent crime rate for Louisiana (14.9% of households are single parents)?

Predicted Crime Rate per 100,000 = -878.86 + 130.11 (14.9)
Predicted Crime Rate per 100,000 = -878.86 + 1938.639
Predicted Crime Rate per 100,000 = 1059.779

(j) What would be the residual for North Dakota (observed violent crime rate for ND = 82)?

Observed violent crime rate for ND = 82
Predicted violent crime rate for ND = 214.064

Residual = observed – predicted
Residual = 82 – 214.064
Residual = -132.064

(k) What would be the residual for Louisiana (observed violent crime rate for LA = 1062)?

Observed violent crime rate for LA = 1062
Predicted violent crime rate for ND = 1059.779

Residual = observed – predicted
Residual = 1062 – 1059.779
Residual = 2.221

2. Using the blood pressure data from the Week 1 self-assessment, perform a regression analysis in which the IV = body weight (pounds) and DV = systolic blood pressure.

SPSS Data: http://www.bwgriffin.com/gsu/courses/edur8132/selfassessments/Week01/Week01Q5Data.sav
Excel Data: http://www.bwgriffin.com/gsu/courses/edur8132/selfassessments/Week01/Week01Q5Data.xlsx

Present APA styled results for this analysis.

Table 1
Descriptive Statistics and Correlations Between Systolic Blood Pressure and Body Weight

Variable	Correlations	
	Systolic BP	Body Weight (lbs)
Systolic BP	---	
Body Weight (lbs)	.49*	---
Mean	151.21	217.50
SD	12.92	7.94

Note: **n = 28**
*** p < .05**

Table 2
Regression of Systolic Blood Pressure on Body Weight

Variable	b	se	95%CI	t
Body Weight	0.79	0.28	0.22, 1.36	2.83*
Intercept	-20.38	60.79	-145.34, 104.58	-0.34

Note: **R^2 = .235, adj. R^2 = .205, F = 7.98*, df = 1,26, MSE = 132.71, n = 28.**
***p < .05.**

Regression results show a statistically significant, positive relation between body weight and systolic blood pressure. The greater the body weight, the higher will be systolic blood pressure.

3. What does mean squared error (MSE, or mean squared residual, MSR) measure?

It is a measure of residual variance; variance in DV after subtracting predicted values.

4. What does standard error of estimate (SEE, or standard error of residual, SER) measure?

It is a measure of residual standard deviation; the SD in DV after subtracting predicted values.

5. If the mean squared error (MSE, or mean squared residual, MSR) is 25, what would be the standard error of estimate (SEE, or standard error of residual, SER)?

MSE = SEE^2 , or
SEE = $\sqrt{MSE}$

So MSE = 25, then SEE = 5.

6. If the original variance of a dependent variable (DV) is 50, and the MSE is 35, what is the value of adjusted R^2?

Adjusted R^2 is a percentage measure showing the amount reduction in the DV variance when comparing VAR(DV) to MSE.

Adjusted R^2 = $\frac{VAR(DV)-MSE}{VAR(DV)} = \frac{50-35}{50} = \frac{15}{50} = .30$

7. There is a commonly used data file in statistics that contains a number of automobile measurements including miles per gallon (MPG) and vehicle weight in pounds. The data file contains over 400 records, but for this problem I selected 10 observations. I ran a regression analysis with

IV = vehicle weight in pounds
DV = miles per gallon (MPG)

Below is the SPSS coefficient table for this analysis.

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	52.519	6.836		7.683	.000	36.756	68.282
	Vehicle Weight (lbs.)	-.011	.003	-.811	-3.919	.004	-.017	-.004

a Dependent Variable: Miles per Gallon

The prediction equation is as follows:

Predicted Y = b0 + b1 X

Predicted MPG = 52.52 + (-0.011 x vehicle weight)

Below is at table that shows the observed MPG and the predicted MPG. The predicted MPG values were obtained using the equation above. The values for vehicle weight have been removed from the table.

Predicted MPG Table

mpg	weight	predicted mpg
15	---	11.897
17	---	14.581
19	---	23.546
22	---	26.032
25	---	29.134
28	---	27.352
29	---	31.983
32	---	32.324
36	---	32.445
43	---	30.685

Using the information provided in the Predicted MPG Table, calculate and report two model fit statistics, R and R^2.

Model R is the Pearson correlation between the observed DV and the predicted DV.

In this example the DV = MPG, so model R would be the correlation between MPG and predicted MPG,

R = .812

Model R^2 is the value of R squared.

R^2 = .812^2 = .659