Self-Assessment

Weeks 4 and 5: Multiple Regression; Squared Semi-Partial Correlations (ΔR^2)

Answers

1. Below is a regression analysis with four variables:

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

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.850(a)	.722	.704	160.898	.722	39.899	3	46	.000

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

ANOVA(c)

Model			Sum of Squares	df	Mean Square	F	Sig.	R Square Change
1	Subset Tests	percent of population in metropolitan area	1251474.526	1	1251474.526	48.341	.000(a)	.292
		percent of population in poverty	229834.852	1	229834.852	8.878	.005(a)	.054
		percent of population in single parent family	650399.008	1	650399.008	25.123	.000(a)	.152
	Regression		3098767.107	3	1032922.369	39.899	.000(b)	
	Residual		1190858.113	46	25888.220			
	Total		4289625.220	49				

a Tested against the full model.
b Predictors in the Full Model: (Constant), percent of population in single parent family, percent of population in metropolitan area, percent of population in poverty.
c 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)	-1197.538	180.487		-6.635	.000	-1560.840	-834.236
	percent metro	7.712	1.109	.565	6.953	.000	5.480	9.945
	percent poverty	18.283	6.136	.265	2.980	.005	5.932	30.634
	percent single parent	89.401	17.836	.446	5.012	.000	53.498	125.303

a Dependent Variable: violent crime rate per 100,000

The prediction equation for this regression model is

Predicted Y = b0 + b1 X1 + b2 X2 + b3 X3

Predicted Crime Rate = -1197.538 + 7.712 (% metro) + 18.283 (% poverty) + 89.401 (% Single-parent)

(a) What is the literal interpretation for values obtained for b0, b1, b2, and b3?

b0: the predicted violent crime rate per state is -1197.538 when percent of population in metro areas is 0.00, in poverty is 0.00, and in single-parent homes is 0.00.

b1: the predicted crime rate is expected to increase by 7.712 for each one percentage point increase in percent of population living in metro areas controlling for percent in poverty and single-parent homes.

b2: predicted crime rate is expected to increase by 18.28 for each one percentage point increase in the percent living in poverty controlling for the percent in metro areas and single-parent homes.

b3: predicted crime rate is expected to increase by 89.40 for each one percentage point increase in the percent living in single-parent homes controlling for the percent in metro areas and in poverty.

(b) What is the general interpretation for b1, b2, and b3?

As the percentage of the state's population living in poverty, metro areas, or single-parent homes increases, so too does the expected violent crime rate per capita.

(c) What percent of the crime rate variance can be predicted by knowing the percent of the population in metro areas, poverty, and in single-parent households?

R^2 = .72 and adjusted R^2 = .70, so about 70 to 72% of variance in violent crime rate can be predicted knowing a state's percent of population in poverty, metro areas, and single-parent homes.

(d) Is the overall model statistically significant; does the model predict more variance in crime rates than would be expected by chance? Explain how you arrived at your answer.

Yes, more variance is being predicted by the regression model than would be expected by chance since the model F ratio (39.899) is larger than expected by chance at the .05 and .01 level (p-value for F ratio is reported as .000), i.e., the model is statistically significant at the .05 and also .01 level. These numbers are reported in the ANOVA table above.

(e) Which of the predictors in this model are statistically significant at the .05 level? Explain how you arrived at your answer.

All are statistically significant at the .05 level since the reported p-values for each regression slope (b1, b2, and b3) are less than .05 with values of .000, .005, and .000 respectively (see "Sig." column in the Coefficients table). Another approach for testing Ho: $\beta_i = 0.00$ is to determine whether the value of 0.00 lies within each confidence interval. Since 0.00 is not within any of the reported confidence intervals for the slopes (see

Coefficients table), one may reject Ho for each slope tested and conclude that a slope of 0.00 (no relation) is not likely for any of the predictors.

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

One may be 95% confident that the population slope relating percent living in metro areas to violent crime rate, while controlling for percent in poverty and in single-parent homes, lies between 5.48 and 9.95.

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

As noted above, if the value of 0.00 lies within the confidence interval one would fail to reject. If the value of 0.00 is not within the confidence interval, one may reject Ho and conclude that a relation exists between the predictor and dependent variable. The logic stems from the null which states that the relation between predictor and dependent variable is 0.00 (i.e., Ho: $\beta_i = 0.00$) so if 0.00 is not one of the values identified in the confidence interval, then the null of no relationship can be rejected.

(h) What is the predicted violent crime rate for the following states?

State	Observed Violent Crime Rate (per 100,000)	% in Metro Areas	% in Poverty	% in Single-parent Households
Alaska	761	41.80	9.10	14.30
California	1078	96.70	18.20	12.50
New Hampshire	138	59.40	9.90	9.20

Prediction Model:
Predicted Crime Rate = -1197.538 + 7.712 (% metro) + 18.283 (% poverty) + 89.401 (% Single-parent)

Alaska:

Predicted Crime Rate = -1197.538 + 7.712 (41.80) + 18.283 (9.10) + 89.401 (14.30)
569.633 = -1197.538 + 7.712 (41.80) + 18.283 (9.10) + 89.401 (14.30)

California:

Predicted Crime Rate = -1197.538 + 7.712 (96.70) + 18.283 (18.20) + 89.401 (12.50)
998.476 = -1197.538 + 7.712 (96.70) + 18.283 (18.20) + 89.401 (12.50)

New Hampshire:

Predicted Crime Rate = -1197.538 + 7.712 (59.40) + 18.283 (9.90) + 89.401 (9.20)
264.046 = -1197.538 + 7.712 (59.40) + 18.283 (9.90) + 89.401 (9.20)

(i) What is the residual for each of these states?

Residuals

Alaska:

Residual =	**Observed – Predicted**
191.367 =	**761 – 569.633**

California:

Residual =	**Observed – Predicted**
79.524 =	**1078 – 998.476**

New Hampshire:

Residual =	**Observed – Predicted**
-126.046 =	**138 – 264.046**

(j) What is the value of the squared semi-partial correlation (ΔR^2) for each of the three predictors?

Predictor	ΔR^2
% in metro areas	**.292**
% in poverty	**.054**
% in single-parent homes	**.152**

(k) What is the value of the inferential test statistic used to test the significance of ΔR^2 for each predictor, and is this value significant at the .05 level?

The partial F-ratio is used to test Ho: $\Delta R^2 = 0.00$. Partial F ratios may be found in the SPSS ANOVA table. The p-value for each partial F ratio is also reported in the ANOVA table and is found in the column labeled "Sig."

Predictor	F	p-value
% in metro areas	**48.34**	**.000**
% in poverty	**8.78**	**.005**
% in single-parent homes	**25.12**	**.000**

2. Below is a data file containing the following variables for cars taken between 1970 and 1982:

mpg:	miles per gallon
engine:	engine displacement in cubic inches
horse:	horsepower
weight:	vehicle weight in pounds
accel:	time to accelerate from 0 to 60 mph in seconds
year:	model year (70 = 1970, to 82 = 1982)
origin:	country of origin (1=American, 2=Europe, 3=Japan)
cylinder:	number of cylinders

SPSS Data: http://www.bwgriffin.com/gsu/courses/edur8132/selfassessments/Week04/cars_missing_deleted.sav

(Note: There are underscore marks between words in the SPSS data file name.)

Other Data Format: If you prefer a data file format other than SPSS, let me know.

For this problem our interest is in calculating and testing the partial contribution of two engine measures to MPG, horsepower and engine displacement. The regression model includes both engine measures and vehicle weight:

Predicted MPG = b0 + b1 (weight) + b2 (horse) + b3 (engine)

In this equation there are two measures of engine performance, horsepower (horse) and displacement (engine). Test the **combined** contribution of these two measures using a squared semi-partial correlation (ΔR^2).

(a) What is the value of the squared semi-partial correlation (ΔR^2) for the set of horsepower and displacement once vehicle weight is first entered into the regression model? (Stated differently, how much of an increase in R^2 results when both horsepower and displacement are included in the regression model after weight is first included?)

Answer

The ΔR^2 for both horsepower and displacement combined is .014.

SPSS Commands

To obtain squared semi-partial correlations, I used the SPSS test command. I changed the SPSS syntax (command language) from /METHOD = ENTER to /METHOD = TEST then I grouped the two engine performance measures and isoluated the weight variable, i.e. the original line

/METHOD=ENTER horse engine weight .

was changed to this

/METHOD=test (horse engine) (weight) .

The full SPSS command to obtain results appears below.

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT mpg
 /METHOD=test ( horse engine) ( weight) .
```

SPSS Results

ANOVA(c)

Model			Sum of Squares	df	Mean Square	F	Sig.	R Square Change
1	Subset Tests	Horsepower, Engine Displacement (cu. inches)	340.701	2	170.350	9.446	.000(a)	.014
		Vehicle Weight (lbs.)	992.001	1	992.001	55.005	.000(a)	.042
	Regression		16630.316	3	5543.439	307.375	.000(b)	
	Residual		6979.459	387	18.035			
	Total		23609.775	390				

a Tested against the full model.

b Predictors in the Full Model: (Constant), Vehicle Weight (lbs.), Horsepower, Engine Displacement (cu. inches).

c Dependent Variable: Miles per Gallon

(b) What would be the null hypothesis, both written and symbolic, for the set contribution — ΔR^2 — of both horsepower and displacement?

Symbolic:

Ho: ΔR^2 (Horsepower, Displacement) = 0.00.

Written:

Combined, horsepower and engine displacement add no predictive benefit to vehicle MPG once vehicle weight is taken into controlled.

(c) Is the combined contribution for the set of horsepower and displacement statistically significant? Present the F ratio, degrees of freedom, and p-value for this combined test. Explain if Ho is rejected.

F = 9.45
DF = 2
P = .000

Since p < .05, reject Ho and conclude that horsepower and engine displacement contribute to model fit (prediction of MPG variance) over and above that provided by vehicle weight.

3. Using cars data presented above in Question 2, run a regression model with the previously identified variables as noted in the equation below.

Predicted MPG = b0 + b1 (weight) + b2 (horse) + b3 (engine)

In this exercise, there is no need to calculate and present the set contribution of horsepower and displacement as was done in Question 2. Instead, we are now interested in learning the individual, partial contribution of each of the three predictors to MPG.

For this analysis, set alpha = .01 (which means the confidence intervals should be 99%).

Present results in APA style.

Table 1

Descriptive Statistics and Correlations among MPG, Displacement, Horsepower, and Weighty

Variable	Correlations			
	1	2	3	4
1. MPG	---			
2. Displacement	-.81*	---		
3. Horsepower	-.78*	.90*	---	
4. Weight	-.83*	.93*	.86*	---
Mean	23.48	194.13	104.24	2973.10
SD	7.78	104.63	38.28	845.83

Note: n = 391

* p < .01

Table 2

Regression of MPG on Engine Displacement, Horsepower, and Weight

Variable	b	se	ΔR^2	99%CI	t
Horsepower	-.041	.013	.008	-.075, -.008	-3.20*
Displacement	-.006	.007	.001	-.023, .011	-0.89
Weight	-.005	.001	.042	-.007, -.003	-7.42*
Intercept	44.82	1.22		41.67, 47.97	36.81*

Note: R^2 = .70, adj. R^2 = .70, F = 307.38*, df = 3,387, MSE = 18.04, n = 391. The symbol ΔR^2 represents the squared semi-partial correlation.

*p < .01.

Bivariate correlations show that all predictors are strongly, negatively, and significantly correlated with MPG. Regression results show, however, that only horsepower and weight are statistically significant predictors of MPG once all predictors are used to model MPG simultaneously. Both horsepower and weight are negatively associated with MPG—as either horsepower or weight increases, vehicle MPG declines. Engine displacement is not related to MPG once horsepower and weight are considered, thus engine displacement does not help predict MPG.

(Note: Given the very high correlations among the predictors (all .86 or greater), this regression model likely suffers from something called multicollinearity and therefore cannot provide an adequate test of which variables contribute to predicting MPG. We may cover multicollinearity later in the semester.)