Self-Assessment

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

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?

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

(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?

(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.

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

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

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

(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

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

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

(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?

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?)

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

(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.

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.