Spatial Ecology Workshop (Quiz 6 - Key)

Student: ____________________________ Grader: ____________________

1) **Define collinearity (+1 point)?**

Collinearity refers to the existence of a linear relationship between *two* (or more) explanatory variables. In other words, two or more variables vary together. This co-variation between any two variables can be either in sync (both variables get larger or smaller together) or out of sync (one variable gets smaller and other one gets larger).

Explain (in your own words) why collinearity causes statistical problems (+0.5 point)

Strong correlations between independent (explanatory) variables inhibit our ability to estimate their individual effects (e.g., regression coefficients) on the dependent variable. Collinearity renders ineffective the numerical methods used to solve regression equations. As a result, we cannot determine which independent variable controls the dependent variable.

List the two practical solutions to solve the problem of collinearity (+0.25 points each):

Collinearity implies redundancy in the set of independent (explanatory) variables. Two practical solutions entail: discarding some of the explanatory variables and combining the explanatory variables into combinations. These are termed synthetic variables.

2) **Define these terms (+0.25 Points each):**

(r12.3): **1^{st} order correlation of variables 1 and 2, partialing out the influence of variable 3**

(r12): **zero order correlation of variables 1 and 2. The same as *(r21)***

(r12.34): **2^{nd} order correlation of variables 1 and 2, partialing out the effects of variables 3 and 4**

(r21): **zero order correlation of variables 1 and 2. The same as *(r12)***

3) Explain what these results imply (what can we infer from them) (+0.25 Points each)?

(r12) = (r21): No surprise. Correlations are symmetrical (the same for a pair of variables)

(r12) = r (12.3): The zero order correlation and the 1st order correlation (partialing out the effect of variable 3) yield the same correlation coefficient (r). This result suggests that variable 3 has a negligible effect on variables 1 and 2

(r12) > > r (12.3): The zero order correlation yields a much larger correlation coefficient (r) than the 1st order correlation (partialing out the effect of variable 3). This result suggests that variable 3 has an effect on variables 1 and 2 and that the correlation between variables 1 and 2 is *spurious* (not real) and only the result of the influence of variable 3

(r12) < < r (12.3): The zero order correlation yields a much smaller correlation coefficient (r) than the 1st order correlation (partialing out the effect of variable 3). This result suggests that variable 3 has an effect on variables 1 and 2 by *suppressing* their covariation

4) List three environmental variables you would expect to vary collinearly as you travel from coastal Oahu to the subarctic North Pacific Gyre (off Alaska) (+ 0.25 points each):

As we move from a tropical to a sub-arctic environment (from low to high latitude), we would expect: water temperature declines, ocean productivity increases, and wind speed increases

5) Suppose that a rather cranky professor has just given an exam in his statistics course, and that for each student in the course we have data on the following three variables:

X = the amount of effort spent on studying for the exam beforehand

Y = the student's score on the exam

Z = a measure of the degree to which the professor inspires fear in the student

These are the correlations among the variables:		% Variable Explained (+0.25 points)
X versus Y: $r_{XY} = + 0.20$	$r^2_{XY} = 0.04$	X: 4%
X versus Z: $r_{XZ} = + 0.80$	$r^2_{XZ} = 0.64$	Y: 64%
Y versus Z: $r_{YZ} = - 0.40$	$r^2_{YZ} = 0.16$	Z: 16%

It is odd that the correlation between X and Y is so small ($r_{XY} = +0.20$). The greater the fear, the greater the effort students put into preparing for the exam ($r_{XZ} = +0.80$). On the other hand, the greater the fear, the less well students do on the exam ($r_{YZ} = -0.40$). Partial out the effect of Z on the XY correlation. Write the equation and show your calculations (+ 0.50 points):

$$r_{XY \cdot Z} = \frac{r_{XY} - [(r_{XZ}) \times (r_{YZ})]}{\text{sqrt}[1 - r^2_{XZ}] \times \text{sqrt}[1 - r^2_{YZ}]}$$

$$= \frac{0.20 - [(+0.80) * (-0.40)]}{\text{sqrt}(1 - 0.64) * \text{sqrt}(1 - 0.16)}$$

$$= \frac{0.20 + 0.32}{\text{sqrt}(0.36) * \text{sqrt}(0.84)}$$

$$= \frac{0.52}{0.60 * 0.92}$$

$$= \frac{0.52}{0.55} = 0.95$$

$r_{XY \cdot Z} =$ 0.95 (+0.50 point) $r^2_{XY \cdot Z} =$ 0.90 (+0.50 point)

Interpret the result of the correlation between the amount of time studying (X) and the scores in the exam (Y), after removing the effect of fear (Z) (+ 1 point):

This is a very strong positive correlation, as indicated by the large correlation coefficient (0.95) and the large amount of variance explained (90%). This result means that studying produces higher exam scores, once we factor out the "fear factor" of taking the exam.

Draw a path diagram of the relationships between the three variables (X,Y,Z), and show / label 8 features: control variable, initial variables, 3 correlations (rXY, rXZ, rYZ), partial rXY.Z correlation, residual variance of XY correlation (+0.25 points each)