

Session 5: Post Class Test Solutions

1.
 - a. Correlation, covariance and R-squared all measure co-movement, and all three will show higher values, in absolute terms, as the relationship gets stronger.
 - b. The correlation is bounded between minus one and plus one, and is unaffected by the units in which your data is denominated. The covariance always has the same sign as the correlation, with positive (negative) covariance indicating that the two variables move together (in opposite directions), but it is unbounded and is in the same units as the underlying data.
 - c. The R-squared is just the correlation squared. As a consequence, it is just a measure of the strength of the relationship between two variables, no matter which direction (together or in opposite directions). It is bounded between zero and one.
2.
 - a. The scatter plot (and the best-fit line) indicate that as X increases, Y decreases, i.e., they move in opposite directions.
 - b. The best-fit line is the one that minimizes the squared distances of the points from the line. It has an intercept (the value of the Y variable, when the X variable is zero) and a slope (which is negative, since they move in opposite directions).
 - c. You use the regression to get predicted values for Y, for any given X. You then compute the difference between these predicted values for Y and the actual values to get the residuals. If the regression is perfect (every point falls on the line, and the R-squared is 100%), the residuals will all be zero.
3.
 - a. You can measure the correlations across each pair of the independent variables to see if they are different from zero. If they are (positive or negative), you have multicollinearity and you can try to minimize or eliminate the problem by dropping one of the correlated independent variables or replacing it with a different (uncorrelated variable). If you cannot get rid of multi-collinearity, you can still use the multiple regression for predictions, but the coefficients you get on individual independent variables can have the wrong sign and be unexplainable.
 - b. To check to see if each independent variable is adding statistically to the predictive power of your regression, you should look at the t statistics (or p values) of your regression. If the t statistic is low (well below 2), the p value will be high, indicating that there is a high probability that your coefficient is really zero. You can try a different independent variable that yields statistical significance or take the variable out entirely.
4.
 - a. Once you use a regression to get predicted values for your Y variable, you can compute the difference between your predicted and actual values, i.e., the residual for each Y value in your regression. Those residuals can then be plotted against the level of the predicted Y values to see if there is a pattern. In other words, are the errors greater or smaller for higher predicted values for Y? In a good regression, the residuals should show no such pattern and they should be close to normally distributed.
 - b. If you see a pattern in the residuals, i.e., they get larger or smaller as your predicted Y values increase, and/or they are non-normal, you can try weighted or generalized least squared regressions.
5.
 - a. larger sample gives you the freedom to bring in more independent variables into your regression. For very small samples, you may be stuck with just a single independent variable. In addition, the effects of outliers will be far greater in small than large samples, pushing R-squared up in some cases and down in others

- b. In a regression, sample size plays a role in explaining the difference between R-squared and adjusted R-squared, with the difference decreasing as sample size increases. In addition, it plays a role in whether you find statistical significant in your results. With large sample sizes, even small differences in explained value can be statistically significant. Thus, you could see strong significance for the independent variables co-existing with low R-squared in very large samples..