Assignment 01

Regression and Mathematics

EPsy 8264

This goal of this assignment is to review ideas from regression and mathematics that will be useful for the remainder of the course. Turn in a printed version of your responses to each of the questions on this assignment. Please adhere to the following guidelines for further formatting your assignment:

- All graphics should be set to an appropriate aspect ratio and sized so that they do not take up more room than necessary. They should also have an appropriate caption.
- Any typed mathematics (equations, matrices, vectors, etc.) should be appropriately typeset within the document.
- Syntax or computer output should not be included in your assignment unless it is specifically asked for.

This assignment is worth 20 points.

Use for Questions 1–2

Using expectation and summation rules, mathematically confirm the following:

1. $\mathbb{E}\left[\hat{Y}_i \times \epsilon_i\right] = 0$

2. $\sum\left(X_i - \bar{X}\right)^2 = \sum X^2 - \frac{\left(\sum X\right)^2}{n}$

Use for Questions 3–4

Suppose that the means and standard deviations of Y and X are the same. $\bar{Y} = \bar{X}$ and $S_Y = S_X$.

3. Mathematically show that $\hat{\beta}_{1(Y|X)} = \hat{\beta}_{1(X|Y)} = r_{XY}$; where $\hat{\beta}_{1(Y|X)}$ is the least-squares slope for the simple regression of Y on X, $\hat{\beta}_{1(X|Y)}$ is the least-squares slope for the simple regression of X on Y, and r_{XY} is the simple correlation between X and Y.

4. Also show that the intercepts for the two regressions are the same (i.e., $\hat{\beta}_{0(Y|X)} = \hat{\beta}_{0(X|Y)}$).

Use for Questions 5–6

Imagine that X is father's height and Y is son's height for a sample of father-son pairs. Suppose that $\bar{Y} = \bar{X}$ and $S_Y = S_X$, and that the regression of son's heights on father's heights is linear. Lastly, suppose that $0 < r_{XY} < 1$ (i.e., father's and son's heights are positively correlated, but not perfectly).

5. Mathematically show that the expected height of a son whose father is shorter than average is also less than average, but to a smaller extent; likewise a a son whose father is taller than average is also taller than average, but to a smaller extent. This idea of "regression to the mean" was the reason Galton chose the word "regression" to describe this methodology.

6. What is the expected height for a father whose son is shorter than average?

Use for Questions 7–8

Davis regressed subjects' reported weights on their actual weights and obtained the following coefficient-level output:

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic   p.value
##   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)   -0.948    0.858      -1.11 2.70e-  1
## 2 weight         1.01     0.0128     79.2  1.95e-142
```

7. Imagine the predictor-variable (`weight`) values in Davis' regression are transformed according to: $X' = X - 10$ and that Y (`repwt`) is regressed on X'. How does the coefficient-level output for the slope change? Explain. (*Hint:* Use rules of variances and covariances. Also feel free to check your response using the *davis-corrected.csv* data.)

8. Imagine the predictor-variable (`weight`) values in Davis' regression are transformed according to: $X' = 10(X - 1) = 10X - 10$. and that Y (`repwt`) is regressed on X'. How does the coefficient-level output for the slope change? Explain. (*Hint:* The SE is a square root of a variance.)

9. Imagine the outcome-variable (`repwt`) values in Davis' regression are transformed according to: $Y' = 5Y + 2$. and that Y' is regressed on X (`weight`). How does the coefficient-level output for the slope change? Explain.

10. In general, how are confidence intervals and hypothesis tests for the slope affected by linear transformations of X and Y.