Name: _____
Date: _____

Quiz 11 – B

1. Consider the following equation: $0.25 \cdot age + 4 \cdot height = income$. A person's working age could vary between 18 and 65 years. A person's height can vary between 150 and 190 cm. Which predictor is more important, and by what factor?

2. The following table shows the correlations between three predictor variables {A, B, and C} and the response variable, D.

	B	C	D
A	0.20	0.82	0.43
B	-	0.15	0.38
C		-	0.25

   A. Which predictor variables should not be used together? Explain why clearly, using the correct statistical vocabulary.

   B. Name any models you might want to test. For example, A+B+C might be one model you want to test (or maybe not). Explain your thinking.

3. In the 1989 Batman movie (starring a very underrated Michael Keaton and an overrated Jack Nicholson), the Joker's plot involved putting various poisons into Gotham City's cosmetic supply. The Joker's plan was clever, though: it required certain combinations of cosmetics (such as hairspray and lipstick) to receive a toxic mixture that killed a person. Hairspray or lipstick alone wouldn't do it.

   Explain this nefarious scheme in terms of a statistical concept.

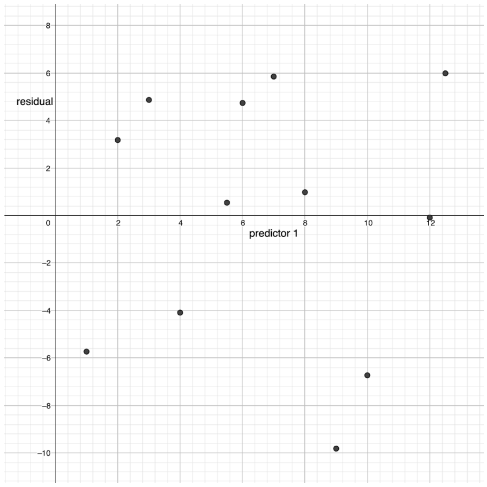
4. Name and explain the potential fallacy in the following statement: “On average, payday loans (for-profit businesses that give people small loans at high interest rates) put the typical family further into debt.”

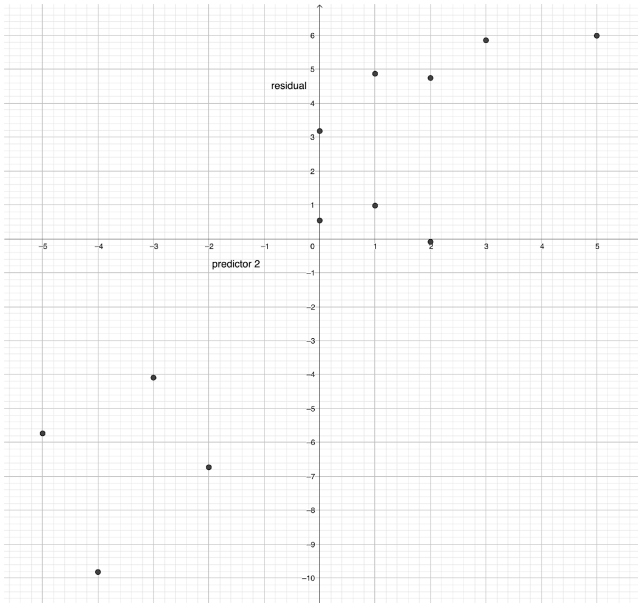
5. Consider the following statement: “Stress levels and income show a heteroskedastic relationship. At low-income levels, stress is typically high, but…”

   A. Complete the statement above.

   B. Explain why the complete statement might be true.

6. Look at the following two residual plots for the same model, $\beta_0 + \beta_1 \cdot predictor\ variable\ 1 = response\ variable$. Note that both show the <u>exact same residuals</u>, just in different orders.





What should you conclude? Why?

Quiz 11 – B – answer key

1. Age could add 11.75 points to the total, $(65 - 18) \cdot 0.25 = 11.75$, whereas height can add 160 points, $(190 - 150) \cdot 4 = 160$. Height is approximately 13.6 times more important, according to this equation.

2.
   A. Predictors A and C are highly correlated. They should not be used in the same model together; to do otherwise would be called a multicollinearity problem.

   B. Predictors A and B have a low correlation with each other and are both highly correlated with the response variable D. The best model is probably $\beta_0 + \beta_1 \cdot A + \beta_2 \cdot B = D$.

3. This in an interaction effect. Hairspray alone has no correlation with poisoning. Lipstick alone has no correlation with poisoning. But hairspray AND lipstick IS correlated with poisoning.

4. This is, potentially, a main effect fallacy. While the average effect may be bad, there may be important subgroups of clients that benefit from payday loans.

5.
   A. "… at high-income levels, stress varies quite widely."

   B. Some high-income jobs are incredibly stressful (e.g., emergency room doctor, trial lawyer), whereas other high-income jobs are less stressful (e.g., tax accountant, life insurance actuary). At low-income levels, the stress of the job is not the key source of stress; it's crushing financial problems.

6. The residuals for the model for predictor 1 look random, which is good. Predictor 1 should be kept in the model. Predictor 2 should also clearly be added to the model; the residuals after predictor 1 are not truly random as they can be largely explained by the values of predictor 2.