## Week 8 (Monday) Practice RAT

|  |  |
| --- | --- |
| 1 | Imagine that you use the number of Ikea stores per 10 million population to predict Nobel Laureates per 10 million population using a linear model. You then take the model residuals and plot a scatter plot of the residuals by the number of Ikea stores per 10 million. You find that the cloud of points looks like a cone, with the residuals tightly packed together at lower values and more spread out at higher values of Ikea stores per 10 million.  Based on this plot, which of the following conditions for linear regression would not seem to hold? |
|  |
| (a) | The Quantitative Variable Condition. |
| (b) | Straight Enough Condition. |
| (c) | Outlier Condition. |
| (d) | Does the Plot Thicken? Condition |

|  |  |
| --- | --- |
| 2 | A researcher has measured the shoe size and height of 100 clowns and standardized both variables using z-scores. She then fits a linear model of shoe size as a function of height.  Which of the following MUST be FALSE? |
|  |
| (a) | The researcher predicts that a clown whose height z-score is 1 will have a z-score of 1 for shoe size. |
| (b) | The researcher predicts that a clown whose height z-score is 1.2 will have a z-score of 1 for shoe size. |
| (c) | The researcher predicts that a clown whose height z-score is 1 will have a z-score of 1.2 for shoe size. |
| (d) | The researcher predicts that a clown whose height z-score is 1 will have a z-score of -1 for shoe size. |

|  |  |
| --- | --- |
| 3 | Which of the following is NOT an example of regression to the mean? |
|  |
| (a) | A mother whose height is 2 standard deviations above the mean is predicted to have a daughter whose height is 1.5 standard deviations above the mean. |
| (b) | A basketball player whose points in her first game of the season is 2 standard deviations above the mean is predicted to score a number of points in the second game that is 1.5 standard deviations above the mean. |
| (c) | A fair coin that has given a number of heads 2 standard deviations above the mean in 100 flips is predicted to produce a number of heads 1.5 standard deviations above the mean in the next 100 flips. |
| (d) | A student whose score was 2 standard deviations above the mean on the midterm exam is predicted to score 1.5 standard deviations above the mean on the final exam. |

|  |  |
| --- | --- |
| 4 | A researcher studying generational wealth collects data on the lifetime earnings (standardized and adjusted for inflation) of fathers (x-variable) and sons (y-variable) and finds a positive correlation between these two variables. The researcher identifies a man whose earnings were 2 standard deviations above the mean.  The earnings of his son are most likely to be: |
|  |
| (a) | The mean earnings of men in his age bracket. |
| (b) | Greater than the mean, but less than 2 standard deviations above the mean earnings of men is his age bracket. |
| (c) | 2 standard deviations above the mean earnings of men in his age bracket |
| (d) | A bit more than 2 standard deviations above the mean earnings of men in his age bracket. |

|  |  |
| --- | --- |
| 5 | Suppose that correlation between midterm exam scores and final exam scores is equal to 0.6. You then run a linear model which uses midterm exam scores to predict final exam scores and inspect the model outputs.  How much of the variation in final exam scores will the model account for? |
|  |
| (a) | 30% |
| (b) | 36% |
| (c) | 60% |
| (d) | 63% |

Answers:

(1) D. The question literally describes the **Does the Plot Thicken? Condition**.

(2) C. It is not possible for the *predicted* z-score of shoe size to be greater than the z-score of height.

(3) C. Unlike the other examples, the outcome of two sets of coin-flips are independent and unrelated to each other.

(4) B. This is another example of regression to the mean. Insofar as the wealth of parents and their children are positively correlated and the earnings of the father are two standard deviations above the mean for his age bracket, then we should expect the earnings of the son to be above the mean for his age bracket. However, we would not expect the son’s z-score to be above that of the father.

(5) B. The proportion of variation in final exam scores that the model accounts for is simply the square of the correlation coefficient (i.e. R-squared) which is 0.36 or 36%.