## Week 8 (Thursday) Practice RAT

8 mins (IRAT)

|  |  |
| --- | --- |
| 1 | Thor Heyerdahl wants to know about how the height of Mo’ai—the large stone carvings of human heads found on Easter Island—relates to their base width (both measured in meters). He samples 100 Mo’ai and finds that height has a correlation of 0.8 with the base width. He fits a line of best fit of height as a function of width, and finds that the intercept = 3 and the slope = 2.5.  How much of the variation in the height of Mo’ai is NOT accounted for by base width? |
|  |
| (a) | 20% |
| (b) | 36% |
| (c) | 64% |
| (d) | 80% |

|  |  |
| --- | --- |
| 2 | Consider one of the 100 Mo’ai with a base width of 1.5 meters. Thor’s regression equation predicts that the Mo’ai should be 3 + 2.5 \* 1.5 = 6.75 meters in height. The residual of this prediction is -1.25 meter.  Is this prediction an overestimation or an underestimation? |
|  |
| (a) | It is an overestimation. |
| (b) | It is an underestimation. |
| (c) | It is neither an overestimation nor an underestimation. |
| (d) | We do not have enough information to answer this question. |

|  |  |
| --- | --- |
| 3 | Consider the plot below with residuals against predicted values from a linear regression.  /Users/stevenoliver/Dropbox/ync/teaching/modules/YCC 1122 - QR/2020/Lesson Preps/Week 8 Thursday/Original Material/Week 9 Lesson 1/RATS/heteroscedasticity2.pdf  Which condition of linear regression seems violated? |
|  |
| (a) | Straight Enough Condition. |
| (b) | Outlier Condition. |
| (c) | Does the Plot Thicken? Condition. |
| (d) | All of the options. |

|  |  |
| --- | --- |
| 4 | Consider the scatter plots A, B, C, and D below which each plot y versus x for a different dataset. Consider linear regressions with y as outcome and x as predictor.  /Users/stevenoliver/Dropbox/ync/teaching/modules/YCC 1122 - QR/2020/Lesson Preps/Week 8 Thursday/Original Material/Week 9 Lesson 1/RATS/correlation2.pdf  Which dataset would yield the **lowest** *R*2? |
|  |
| (a) | Dataset A |
| (b) | Dataset B |
| (c) | Dataset C |
| (d) | Dataset D |

|  |  |
| --- | --- |
| 5 | Again, consider the scatter plots A, B, C, and D below which each plot y versus x for a different dataset. Consider linear regressions with y as outcome and x as predictor.  /Users/stevenoliver/Dropbox/ync/teaching/modules/YCC 1122 - QR/2020/Lesson Preps/Week 8 Thursday/Original Material/Week 9 Lesson 1/RATS/correlation2.pdf  Which dataset would yield the **highest** *R*2? |
|  |
| (a) | Dataset A |
| (b) | Dataset B |
| (c) | Dataset C |
| (d) | Dataset D |

Answers

1. (b) The amount of variation of y (height) NOT accounted for by x (base width) is simply the one minus the square of the correlation coefficient (r=0.8). Since the R-squared is 0.64, the amount of variation of y NOT accounted for is 0.36 or 36%.
2. (a) Based on the residual, we know that the Mo’ai is 5.5 meters in height. As such, the prediction overestimates the true value.
3. (a) The plot of residuals appears to bend, thus suggesting it violates the Straight Enough Condition. At the same time, the thickness of the plot is fairly uniform and there are no clear outliers.
4. (c) is correct. A linear model can explain virtually none of the variation in the data, yielding the lowest R-squared.
5. (d) given the tightly packed points that seem to follow a line from the bottom left to the top right corners, it would appear that (d) would produce the highest R-squared.