**Quiz 7, Version 1**

**Figure 1. The black line on this graph is the best-fitting line from a linear regression analysis based on the three datapoints shown on the graph. Use this Figure to answer Questions 1-5.**



Q1. In Figure 1, which of these are correct values for the residuals?

1. The residual for point 1 (red) = -4

The residual for point 2 (yellow) = 0

The residual for point 3 (blue) = 4

1. The residual for point 1 (red) = 4

The residual for point 2 (yellow) = 0

The residual for point 3 (blue) = -4

1. The residual for point 1 (red) = 0

The residual for point 2 (yellow) = 0

The residual for point 3 (blue) = 0

1. The residual for point 1 (red) = 3

The residual for point 2 (yellow) = 6

The residual for point 3 (blue) = 8

Q2. In Figure 1, what is the value of the Y intercept?

1. 5
2. 0
3. -5
4. 9

Q3. In Figure 1, what is the value of the slope?

1. 1
2. 1.5
3. 0.67
4. 0

Q4. Using the regression analysis in Figure 1, we want to predict the Y value when X=18. Which of these statements is **correct**?

1. This is an example of extrapolation, and when X=18, Y=17.
2. This is an example of interpolation, and when X=18, Y=17.
3. This is an example of interpolation, and when X=18, Y=18.
4. This is an example of extrapolation, and when X=18, Y=18.

Q5. Based on this residual plot from the linear regression in Figure 1, were all of our assumptions for this regression analysis appropriate?



1. Yes, because our residuals appear to be normally distributed around zero.
2. No, because our residuals are not symmetrically distributed around zero. This means that our regression line was biased, because it consistently underpredicted or overpredicted our data.
3. No, because the amount of variation in our residuals is not homogeneous; there’s a funnel shape in our residual plot, where for low values of X, there is only small variation Y, and for high values of X, there is large variation in Y.
4. No, because our linear regression was not appropriate for our nonlinear data; there are runs of positive residuals and runs of negative residuals, instead of a random scatter of residuals above and below zero.

Q6. Why might the blue point in the following figure be problematic?



1. The blue point may be an outlier.
2. The blue point may be interpolated.
3. The blue point may be nonlinear.
4. The blue point may have a small residual.

**Quiz 7, Version 2**

**Figure 1. The black line on this graph is the best-fitting line from a linear regression analysis based on the datapoints shown on the graph. Use this Figure to answer Questions 1-5.**



Q1. In Figure 1, which of these are correct values for the residuals?

1. The residual for point 1 (red) = 4

The residual for point 2 (yellow) = -4

The residual for point 3 (blue) = 0

1. The residual for point 1 (red) = 4

The residual for point 2 (yellow) = 0

The residual for point 3 (blue) = -4

1. The residual for point 1 (red) = 0

The residual for point 2 (yellow) = 0

The residual for point 3 (blue) = 0

1. The residual for point 1 (red) = 0

The residual for point 2 (yellow) = 4

The residual for point 3 (blue) = 4

Q2. In Figure 1, what is the value of the Y intercept?

1. 15
2. 30
3. -1
4. -2

Q3. In Figure 1, what is the value of the slope?

1. -0.5
2. 2
3. 1
4. -1

Q4. Which of these could be the correlation coefficient (r) for the regression line in Figure 1?

1. -0.8
2. 1
3. -1
4. 0

Q5. Using the regression analysis in Figure 1, we want to predict the Y value when X=18. Which of these statements is **correct**?

1. This is an example of interpolation, and when X=18, Y=6.
2. This is an example of extrapolation, and when X=18, Y=6.
3. This is an example of interpolation, and when X=18, Y=18.
4. This is an example of extrapolation, and when X=18, Y=18.

Q6. The analysis in the following figure went viral on Twitter, even though it was substantially flawed. The authors of the paper concluded that coronavirus transmission rates (R on the Y axis) are lower in cities/countries with higher average temperatures. What is wrong with this analysis?

1. The point where R is greater than 4.5 is an outlier; that city/country has unusually high transmission, and without it, there would be no relationship.
2. The point where R is greater than 4.5 is non-significant; it is not different than the mean R for all cities/countries.
3. The point where R is greater than 4.5 has an unusual temperature; that city/country is much colder than every other country in the analysis.
4. The point where R is greater than 4.5 has an unusual temperature; that city/country is much warmer than every other country in the analysis.

