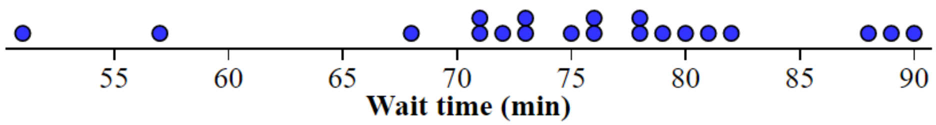
Mock FRQ #3

The Starnes family visited Yellowstone National Park in hopes of seeing the Old Faithful geyser erupt. When they pulled into the parking lot near Old Faithful, a large crowd of people was headed back to their cars from the geyser. Old Faithful had just finished erupting.

The Starnes family wondered how long they would have to wait until the next eruption. Fortunately, they had collected data on a random sample of 20 Old Faithful eruptions from the previous week, including the wait time until the next eruption (in minutes). Here are a dotplot and numerical summaries of the data.



Variable	N	Mean	StDev	Min	Q1	Med	Q3	Max
Wait time (min)	20	75.40	9.54	51.00	71.25	76.00	80.75	90.00

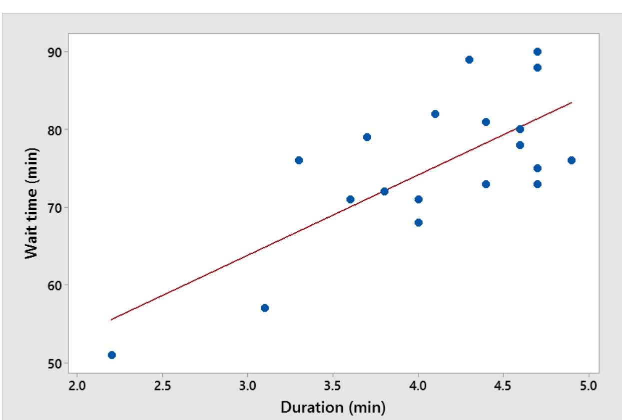
(a) Mr. Starnes decides there is time for the family to go and see the bubbling mud pots, which will take 70 minutes. Based on this sample of 20 eruptions, estimate the probability that the Starnes family will miss the start of the next eruption. That is, estimate *P*(wait time < 70).

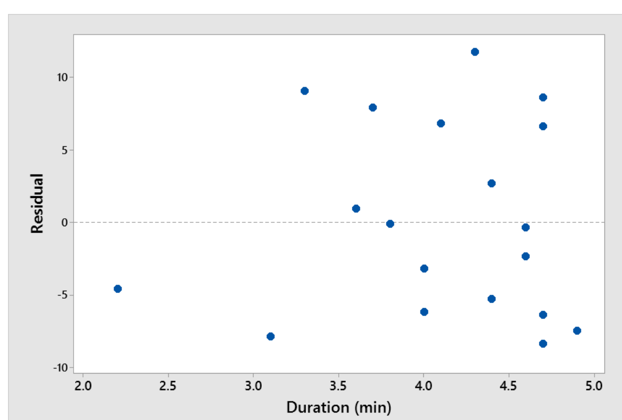
3/20 = 0.15

(b) The *z*-score for the wait time of 51 minutes is *z* = –2.56. Interpret this value.

This wait time was 2.56 standard deviations less than the mean wait time.

Would knowing the duration of the most recent Old Faithful eruption help the Starnes family predict the wait time until the next eruption? Using data on both of these variables from the random sample of 20 Old Faithful eruptions, Mr. Starnes performed a linear regression analysis. Here are a scatterplot with the least-squares regression line, a residual plot, and computer output from the analysis.





Predictor	Coef	SE Coef	T-value	P-value
Constant	33.0	9.3	3.5	0.0002

Duration (min)	10.3	2.2	4.4	0.000

S = 6.63 R-sq = 54.2% R-Sq(Adj) = 51.7%

(c) The Starnes family learns that the previous eruption lasted less than 3 minutes. Based on this fact and the positive association displayed in the scatterplot, do you think *P*(wait time < 70) will be less than, greater than, or about the same as your answer from part (a)? Justify your answer.

P(wait time < 70) should be greater than 0.15. In a positive association, smaller values of duration are typically paired with smaller values of wait time. Because we are told that the duration value is small (less than 3 minutes), it is more likely to get a small wait time (less than 70 minutes).

(d) Estimate and interpret the residual for the eruption at (2.2, 51).

From the residual plot, the eruption with a 2.2 minute duration has a residual of about –5. The actual wait time for this eruption was about 5 minutes less than predicted using the least-squares regression line with x = duration of previous eruption.

(e) Explain why the point in part (d) should be classified as an outlier when considering wait time alone, but not when considering both duration and wait time.

Based on part (b), an eruption with a 51 minute wait time should be classified as an outlier because it is more than 2 standard deviations from than the mean. However, the point (2.2, 51) is not an outlier on the scatterplot because it is less than 1 standard deviation from the least-squares regression line (5 < 6.63).

Raw Data:

Duration of last eruption (min)	Time until next eruption (min)
4.6	80
4.7	90
3.7	79
4.7	75
4.3	89
4.6	78
4	68
4.4	73
4.1	82
3.1	57
3.3	76
4.4	81
4.6	78
4.9	76
2.2	51
3.8	72
4.7	73
3.6	71
4	71
4.7	88