Homework 5

Sahana Sarangi

29 April 2024

2022 Midterm Problem 6: Dr. Walburne is studying the change in cell phone usage over time. She conducts surveys every year for five years, asking participants to estimate the percentage of the time they spent on their phones that was used to make calls (as opposed to internet browsing, texting, taking photos, or any of the other things one can do on a phone). She graphs the year on the x-axis (starting from $x = 0$ in the first year) and the average percentage on the y-axis. She finds that the line of best fit for this data is $y = 23.6 - 0.2x$, with $R^2 = 0.6$. With what degree of confidence, if any, can Dr. Walburne conclude that the percentage of phone usage used for making calls is decreasing over time?

Solution: To find the degree of confidenced, we can perform an F-test. The formula to find the F-score of a line of best fit is

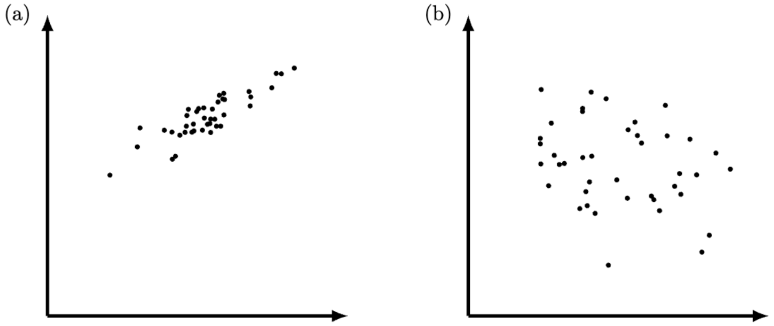
$$F^* = \frac{R^2}{1 - R^2}\,(N - 2)$$

where N is the number of samples used to find the line of best fit (which in this case is 5, as Walburne graphs the average percentage every year for five years). Substituting 0.6 for R^2 and 5 for N, the F-score is

$$F^* = \frac{0.6}{0.4}(3) = 4.5$$

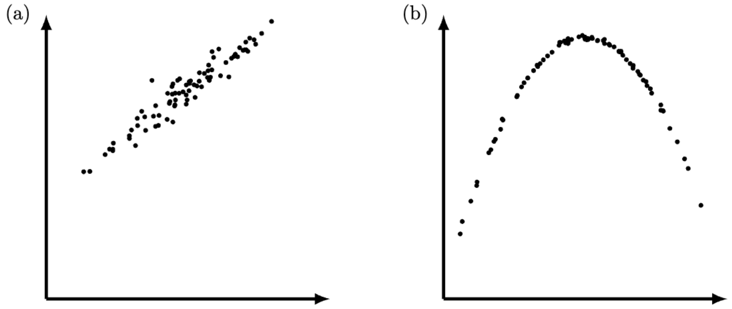
For a sample size of 5, an F-score of 4.5 does not even pass the 90% confidence threshold, which between 5.54 and 10.13. So Welburne $\boxed{\text{cannot conclude}}$ with any meaningful degree of confidence (90% or above) that the average percentage of phone usage used for making calls is decreasing over time.

2022 Midterm Problem 4: Two data sets are shown below, with each dot representing one point. Which one has a higher correlation between x and y? Briefly explain your reasoning.



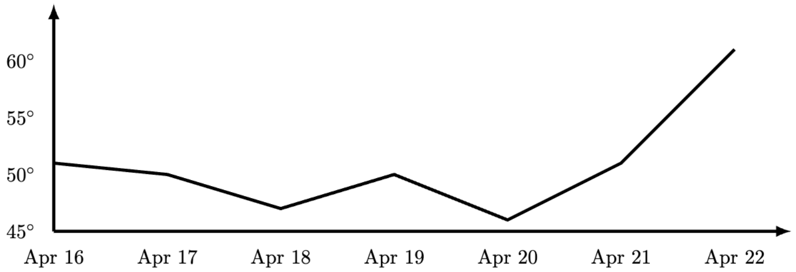
Solution: Correlation is measured on a scale from -1 to 1. In a dataset, if the x-coordinates increase while the y-coordinates increase, the correlation will be be positive. If the x-coordinates increase while the y-coordinates decreasse, the correlation will be negative. In dataset (a), the x-coordinates increase while the y-coordinates increase, unlike in (b) where the x-coordinates increase while the y-coordinates decrease. So the correlation for (a) is positive while the correlation for (b) in negative. Positive numbers are higher than negative numbers, so $\boxed{\text{(a) has a higher correlation.}}$

2023 Midterm Problem 4: Two data sets are shown below, with each dot representing one point. Which one has a stronger (either positive or negative) correlation between x and y? Briefly explain your reasoning.



Solution: To solve this problem, we can use the characteristics of having a positive versus negative correlation described in the 2022 Midterm Problem 4. Graph (a) shows a dataset where the y-coordinates increase as the x-coordinates increase. In graph (b), however, the y-coordinates both increase and decrease as the x-coordinates increase. This means the correlation for (b) is close to zero, but the correlation for (a) is positive. Having some correlation (either positive or negative) is stronger than having no correlation, so $\boxed{\text{graph (a) has a stronger correlation.}}$

2023 Midterm Problem 5: According to Weather Underground, the daily temperature highs (in degrees Fahrenheit) in the RC's zip code (98195) last week were the following:



The mean of the temperatures was 50.857° F; the sample standard deviation was 4.78° F. The correlation between the day of the month and the temperature was 0.49, and the line of best fit (using the day of the month as x and the temperature as y) was $y = 1.107x + 29.824$. The RMSE of this model is 3.938, and the R^2 is 0.24. Based on this information, can we conclude that the temperature was rising during that week? Briefly (in two or three sentences, or by showing work) explain your reasoning.

Solution: To see whether we can conclude that the temperature was rising, we can perform an F-test using the formula for an F-score used in the solution for the 2022 Midterm Problem 6. Substituting 0.24 for R^2 and 7 for N (because we have 7 datapoints) in the formula, we can say that this line of best fit has an F-score of

$$F^* = \frac{0.24}{0.76}(5) \approx 1.58$$

An F-score of 1.58 doesn't even satisfy the threshold for a 90% confidence level for a sample size of 7, which is between 4.06 and 6.61. This means that we $\boxed{\text{cannot conclude}}$ that the temperature was rising this last week.