Quiz 3

1. (True/False) If the Pearson’s correlation of two variables is 0, then there is no association between those two variables.
2. (Multiple choice) The Pearson’s correlation of the two variables plotted below is closest to which of the following?Chart, scatter chart

   Description automatically generated
3. -1
4. -0.5
5. 0
6. 0.5
7. 1
8. (True/False) The linearity assumption required for linear regression means that when the outcome is plotted against the predictor, the points must lie on a straight line.
9. (Multiple choice) The following is taken from a classic study on the association between lung cancer and smoking:

Twenty London hospitals were asked to co-operate by notifying of all patients admitted

to them with carcinoma of the lung. On receipt of the notification an [interviewer], engaged wholly on research, visited the hospital to interview the patient, using a set questionary. As well, however, as interviewing the notified patients with cancer, the [interviewers] were required to make similar inquiries of a group of "non-cancer control" patients. In this investigation the patients were closely questioned and asked (a) if they had smoked at any period of their lives; (b) the ages at which they had started and stopped; (c) the amount they were in the habit of smoking before the onset of illness which had brought them into hospital; (d) the main changes in their smoking history and the maximum they had ever been in the habit of smoking; (e) the varying proportions smoked in pipes and cigarettes; and (f) whether or not they inhaled.

Based on this excerpt, which of the following best describes the study design?

1. Cross-sectional
2. Randomized controlled trial
3. Prospective cohort
4. Case-control
5. (True/False) If X has a t-distribution with 5 degrees of freedom and Y has a standard Normal distribution, then the following inequality holds:

P(X > 2) > P(Y > 2)

1. (Short answer) Suppose I am trying to estimate the mean weight of fish living in Lake Washington. I manage to catch 100 fish and weigh them using a digital scale. After recording my measurements, I discover that my scale is miscalibrated: it adds half an ounce to every measurement it makes. (For example, a fish that truly weighs 8.0 ounces will be recorded as weighing 8.5 ounces.)

If I use my data to estimate the mean Lake Washington fish weight, will the miscalibration of my scale affect the accuracy of my estimate, the precision, or both? Briefly explain your reasoning.

1. (Short answer) I am studying the association between a child's height and their FEV (forced expiratory volume, a measure of lung function). I randomly sample 654 children, measure their height (in inches) and FEV (in liters per second), and run a linear regression of FEV on height. The estimated regression slope is 0.132.

In my analysis writeup, I write "We estimate that, on average, for every 1-inch increase in height, FEV increases by 0.132 liters/second."

Is this interpretation correct? Why or why not?

1. (Multiple choice) I run a linear regression of variable Y on variable X and get the following R output. Based on the hypothesis test for the slope, which of the following is a valid conclusion?

Text

Description automatically generated

1. I have significant evidence of no linear association between Y and X.
2. I do not have significant evidence of a linear association between Y and X.
3. I have significant evidence of a linear association between Y and X.
4. None of the above are valid conclusions.
5. (Multiple choice) Which of the following is true of the relationship between correlation and linear regression slope?
6. The correlation and the linear regression slope are always equal.
7. The correlation and the linear regression slope are never equal.
8. The correlation and the linear regression slope have the same sign.
9. The correlation is always smaller in magnitude than the linear regression slope.
10. (Short answer) The following plot shows the winning Olympic 100-meter dash times for men (blue) and women (red) dating back to 1900. The dashed lines are linear regression lines.

Based on this analysis, can we reasonably conclude that winning times for both men and women will approach 0 by the year 2884? Explain your rationale, and be specific about why this interpretation is correct or incorrect.

Chart, line chart

Description automatically generated