Quiz 9

1. (Short answer) Suppose we obtain information regarding observed follow-up time (Y) and event indicators () for individuals 1 through 6. We record this information in the table below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Observation ID | 1 | 2 | 3 | 4 | 5 | 6 |
| Y | 15 | 8 | 7 | 3 | 20 | 12 |
|  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 1 | 0 | 1 | 1 | 0 |

Who is in the risk set at time t = 12, and what is the size of the risk set at time t = 12?

1. (Short answer) Explain in your own words why the size of the risk set necessarily shrinks over time when we have right-censored data (do not just copy text from slides).
2. (Multiple response) Suppose we collect data on time to death (from birth) in months for children. If , then… (select all that apply)

* Approximately 91% of the population will die within the first 5 years of life.
* Approximately 91% of the population will not die within the first 5 years of life.
* Approximately 9% of the population will die within the first 5 years of life.
* Approximately 9% of the population will not die within the first 5 years of life.

1. (Short answer) Identify which of the following three “survival curves” could not possibly be a survival curve, and explain why it could not be a survival curve.

Chart, line chart

Description automatically generated

1. (True/False) Suppose we collect time-to-event data for 200 individuals, randomly assigned to a treatment and placebo group. We lose no one to follow-up, and there is no administrative censoring. We observe an event for every individual in our dataset. Is the following statement true or false?

We can fit a linear regression to our data to obtain an unbiased estimate of the mean difference in time-to-event between treatment and placebo groups.

1. (True/False) Researchers enroll individuals in a randomized clinical trial to determine whether a new drug improves quality of sleep for individuals with chronic sleeping problems. Quality of sleep is measured in number of times per night that an individual wakes up. The researchers hypothesize that individuals taking the new drug will have lower mean number of times per night that they wake up compared to individuals taking a placebo (and therefore, will have better quality sleep). Participants are randomly assigned to either the new drug or the placebo, and randomization is successful.

Taylor gains access to the clinical trial data, and decides to run a secondary analysis on the data. In particular, she is interested in the association between anxiety (binary: yes/no) and number of times per night that an individual wakes up. She hypothesizes that individuals with anxiety will have higher mean number of times per night that they wake up compared to individuals without anxiety. She plans to fit a simple linear regression analysis with times per night woken up as the outcome and anxiety as the predictor of interest to answer her scientific question.

Is the following statement true or false?

Taylor does not need to worry about confounding because her data came from a randomized clinical trial.

1. (Multiple choice) We collect information on survival time in days for patients with advanced lung cancer from the North Central Cancer Treatment Group. Some patients are censored, so their actual time of death is unknown. I tell you that the last observation (the individual with the longest observed time) is censored. Which of the following two Kaplan-Meier curves could possibly represent the overall survival curve for the data we collected?

Chart, line chart, scatter chart

Description automatically generated

1. Kaplan-Meier survival curve 1
2. Kaplan-Meier survival curve 2
3. (Short answer) Suppose I am studying people with severe bone fractures, and I am interested in knowing whether or not the time it takes for the fracture to heal varies depending on whether the individual with the fracture had surgery. The more severe a fracture is, the more likely that surgery will be recommended. We hypothesize that individuals who had surgery will have longer healing times than those who did not have surgery, given that more severe fractures take longer to heal. We enroll people in our study at the time they have a severe bone fracture, and follow them over time to record when their fracture healed completely, measured at a follow-up clinic appointment. We lose a few people in our study to follow-up, and we end our study after 20 weeks, where a few individuals are subject to administrative censoring.

Determine whether or not the independent censoring assumption holds for the following scenario, and justify your answer:

Participants exit the study because they can no longer afford to travel to the clinic for a follow-up appointment.

1. (True/False) Suppose I run a simple linear regression model with continuous outcome Y and binary predictor of interest X (0,1). The slope coefficient for X in my regression model represents the estimated difference in mean outcome Y between groups with X = 0 and X = 1. I compute a 95% confidence interval for my estimated difference. Is the following statement true or false?
2. (True/False) Suppose I run a simple linear regression model with continuous outcome Y and binary predictor of interest X (0,1). The slope coefficient for X in my regression model represents the estimated difference in mean outcome Y between groups with X = 0 and X = 1. I compute a 95% confidence interval for my estimated difference. Is the following statement true or false?

The *true* difference in mean outcome Y will fall within my 95% confidence interval exactly 95% of the time.