Logo

Description automatically generated

Week 2 Practice

Craig Perkins – [perkins.cr@northeastern.edu](mailto:perkins.cr@northeastern.edu)

Northeastern University

November 2022 – ALY 6150

Week 2 Practice

1. **Define incidence and prevalence. Calculate the incidence and prevalence for the month of July from the following chart assuming there were population of 505. Convert to per 100,000 scale rates.**

According to the CDC *prevalence* is the proportion of people who have a condition at a particular time period while *incidence* is the rate at which people develop a condition during a particular time period. <https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section2.html#:~:text=Prevalence%20and%20incidence%20are%20frequently,during%20a%20particular%20time%20period>.

Assuming that the blue bars represent the duration of an incident for the condition in the total population of 505 then:

There are 8 bars between July 1 -> August 1 so Prevalence is (8 / 505) \* 100,000 = 1584

There were 5 bars that start between July 1 and August 1 meaning the incidence in July is:

( 5 / 505 ) \* 100,000 = 990

1. **What is the Risk of the Outcome among those who received treatment (i.e., exposed)?**

P(Disease | Received Treatment) = P(Disease & Received Treatment) / P (Received Treatment)

= (165 / 535) / (250 / 535) = 66%

**What is the Risk of the Outcome among those who received the control (i.e., unexposed)?**

P(Disease | Control Group) = P(Disease & Control Group) / P (Control Group)

= (245 / 535) / (285 / 535) = 86%

**Calculate the Risk Ratio comparing exposed to unexposed.**

Risk Ratio = P(Disease | Received Treatment) / P(Disease | Control Group) = 0.66 / 0.86 = 77%

1. **What is a confounding variable? Please draw a diagram (DAG) demonstrating the relationship between an exposure (E), confounder (C), and disease (D).**

**Diagram

Description automatically generated**

1. **Calculate the Odds Ratio (OR) of cardiovascular disease (CVD) comparing obese to non obese using the table below (Note: the formula for odds ratios is different than that for Risk Ratios).**

Odds Ratio = (a / c) / (b / d) = (46 / 60) / (254 / 640) = 1.93

Young odds ratio: (10 / 35) / (90 / 465) = 1.48

Old odds ratio: (36 / 25) / (164 / 175) = 1.54

The odds ratios are fairly similar, but there’s a slight increase in the older group.

1. **Follow one of the heart disease prediction algorithms demonstrated on the following website. What were some of the strongest predictor variables? Explain the difference in the conceptualization of variables in prediction models compared to explanatory/causal inference models, using age as an example.**

The most important variables for prediction are whether or not there is a defect in the Thalassemia followed by whether or not the patient has asymptomatic chest pain. I was surprised at how much more the incidence for men is than women for heart disease based on the data presented. In the boosted tree model, age is the 5th most important feature for predicting heart disease. The features that the boosted tree model inferred as the most important fit with my understanding of heart disease and the risk factors. While age is an important feature when predicting heart disease, pre-existing conditions will make a patient much more susceptible to heart disease than age alone.

1. **Imagine you are working for a company developing a new medical device to detect cancer. Your team tests the new device on the current “gold standard” device, yielding the following results. Calculate the sensitivity and specificity by hand.**

Sensitivity = TP / (TP + FN) = 8 / (8 + 2) = 80%

Specificity = TN / (TN + FP) = 9490 / (9490 + 500) = 95%