**PH 538: Biostatistical Methods I**

**Practice Exam [October 25, 2016]**

**Dr. Fares Qeadan: fqeadan@salud.unm.edu**

|  |
| --- |
|  |

**Part I- No STATA Required:**

**Problem 1:**

In a particular community, the probability that an adult has both high blood pressure and high cholesterol is 11%. Given that the probability of having high blood pressure is 27% among adults and that of having high cholesterol is 32%, please answer the following questions:

1. What is the probability that a randomly selected adult from this community neither has high blood pressure nor high cholesterol?

***Solution:***

1. What’s the probability that a randomly selected adult from this community has high blood pressure if it’s known that he/she has high cholesterol?

***Solution:***

1. What’s the probability that a randomly selected adult from this community has high cholesterol if it’s known that he/she has high blood pressure?

***Solution*:**

1. What’s the probability that a randomly selected adult from this community has high cholesterol but not high blood pressure?

***Solution*:**

1. What’s the probability that a randomly selected adult from this community has high blood pressure but not high cholesterol?

***Solution*:**

1. What’s the probability that a randomly selected adult from this community doesn’t have high cholesterol?

***Solution*:**

1. What’s the probability that a randomly selected adult from this community either has high cholesterol or high blood pressure?

***Solution*:**

1. What’s the probability that a randomly selected adult from this community has only one of the two conditions?

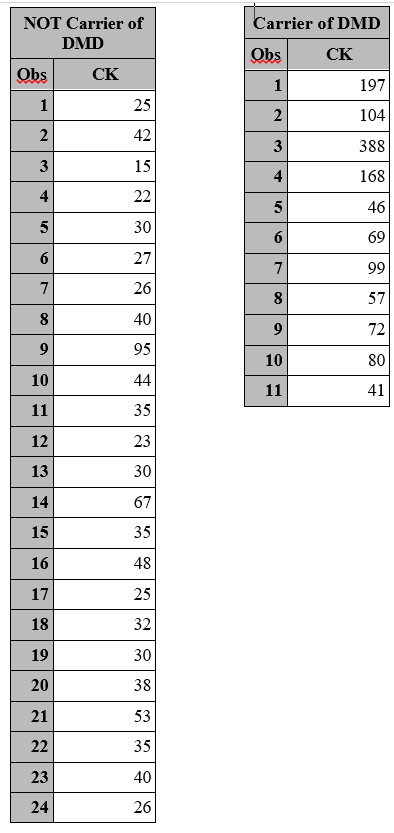
***Solution*:**

1. Is having high cholesterol independent from having high blood pressure?

***Solution*:**

**Problem 2:**

We consider a sample from the Duchenne Muscular Dystrophy (DMD) Dataset. DMD is a genetic disorder characterized by progressive muscle degeneration and weakness. This sample data is from M. Percy, and it corresponds to blood samples on 35 patients collected in a project to develop a screening program for female relatives of boys with DMD. The program's goal was to inform a woman of her chances of being a carrier based on serum markers. One of the serum markers to consider as a screening test is the Creatine Kinase (CK) such that a cut-off of 50 is considered as a reasonable diagnostic cut-point for CK. Specifically, any woman with CK greater or equal to 50 is Carrier of DMD, but with CK less than 50 is not Carrier of DMD. The CK levels of patients in the sample of 35 are shown below, stratified by disease status. Use this information to answer the following questions.



1. (3 points) Create a 2x2 table for this data:

**Answer:**

1. (3 points) Using the data above, what are the estimated sensitivity and specificity of the screening test?

**Answer:**

1. (3 points) In a particular population, the prevalence of DMD is about 0.00029. Using the estimated sensitivity and specificity in part (b), calculate the estimated probability that a woman diagnosed as a carrier of DMD doesn’t actually carry DMD?

**Answer:**

1. (3 points) What is the sample prevalence for carrier of DMD among women? Should we use this sample prevalence to compute the PPV and NPV as opposed to using the population prevalence given in part (c)? Explain

**Answer:**

1. (3 points) How would the estimated sensitivity and specificity change if the cut-point was lowered from 50 to 45? (You don’t need to calculate sensitivity and specificity, just describe what would happen).

**Answer:**

1. (3 points) Compute the likelihood ratio and interpret it?

**Answer:**

1. What is the probability of this test failing to detect diabetes?

**Answer:**

1. What is the probability of this test detecting diabetes when it’s not present?

**Answer:**

1. Calculate the PPV and the NPV (based on the given prevalence in part (c)) for this population.

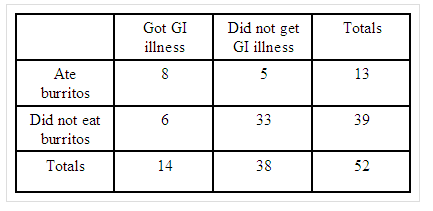
**Answer:**

1. Is it true that for very rare diseases it is nearly impossible to obtain high PPV values even for tests with good sensitivity and specificity? explain

**Answer:**

**Problem 3:**

The findings of a case control study conducted in Florida, USA are given in the 2×2 table below.



1. (3 points) What is the empirical probability of getting GI illness?

**Answer:**

1. (3 points) What is the marginal distribution of GI illness?

**Answer:**

1. (3 points) What is the conditional probability of getting GI illness given that one Ate burritos?

**Answer:**

1. (3 points) Are getting GI illness and eating burritos independent events? Why or why not?

**Answer:**

1. (3 points) What is the conditional distribution of those who got GI illness?

**Answer:**

**Problem 4**:

In NM (based on 2014), 17.4% of the adult population have an education level of less than High School (L) , 26.9% of High School or GED (H), 33% of some post-high school (S), and 22.8% of college or more (C). The rate of obesity by educational level in NM during 2014 is known to be as follows: 35.8% of those with less than High school are obese, 28.1% of those with high school or GED are obese, 29.5% of those with some post-high school are obese, and 22.2% of those with college+ are obese.

1. What is the rate of obesity in NM during 2014?

**Answer:**

1. Given that an adult individual from NM is obese, what is the chance that this individual went to college?

**Answer:**

**Part II- STATA Required:**

**Problem 5:**

Consider the Very Low Birth Weight Infants Dataset (<http://www.mathalpha.com/PH-538/BirthWeightfall16.dta> [dataset 3 on the course webpage]). Data on 671 infants with very low (<1600 grams) birth weight from 1981-87 were collected at Duke University Medical Center by Dr. Michael O'Shea, now of Bowman Gray Medical Center. The data set has 8 variables as follows:

|  |  |
| --- | --- |
| **Variable** |  |
| lowph | lowest pH in first 4 days of life |
| pltct | platelet count |
| race | white  black  native American  oriental |
| bwt | birth weight in grams |
| gest | gestational age in weeks |
| delivery | abdominal  vaginal |
| sex | female  male |
| dead | 0 or 1 |

1. (5 points) Describe the race variable numerically and graphically?

**Answer:**

1. (5 points) Describe the birth weight variable in grams numerically and graphically?

**Answer:**

1. (5 points) Using gestational age, create a new variable called ***birth*** which classify births into broad categories as:

|  |  |
| --- | --- |
| **Gestational Age in Weeks** | **Classification** |
| <28 weeks | extremely preterm |
| 28 to <32 weeks | very preterm |
| 32 to <37 weeks | moderate to late preterm |
| 37 to <42 weeks | full term |

Provide the PMF of the newly generated variable?

**Answer:**

1. (5 points) Is death higher among extremely preterm births than very preterm ones? Explain

**Answer:**

1. (5 points) What is the sample proportion of stillbirth among infants with very low (<1600 grams) birth weight from 1981-87 at Duke University Medical Center?

**Answer:**

1. (5 points) Suppose you are conducting study about stillbirth and you obtain a random sample of 10 infants with very low (<1600 grams) birth weight. What is the average number of stillbirths you would expect to see in the sample?

**Answer:**

1. (5 points) What is the probability that at least 6 of the sampled infants are dead?

**Answer:**

1. (5 points) Platelets are cells in your blood that help your blood to clot when it needs to. The normal number of platelets is between 150 and 400 million per millilitre (ml) of blood. Please construct a histogram for this variable while fitting the theoretical normal curve?

**Answer:**

1. (5 points) What is the sample mean platelet count among women who gave birth to infants with very low (<1600 grams) birth weight from 1981-87 at Duke University Medical Center?

**Answer:**

1. (5 points) Plot the distribution of stillbirth by delivery type? Among which type of delivery, the proportion of stillbirth is the highest?

**Answer:**

1. (5 points) Plot the distribution of platelet count by race? Women with which race have the highest average platelet count?

**Answer:**

**Problem 6:**

About 59% of the people in New Mexico have had a tetanus shot. If 8 individuals are selected at random from NM, using STATA:

1. What is the probability that exactly 4 of them have had a tetanus shot?

**Answer:**

1. What is the probability that less than 5 of them have had a tetanus shot?

**Answer:**

1. What is the probability that at most 3 of them have had a tetanus shot?

**Answer:**

1. What is the probability that at least 7 of them have had a tetanus shot?

**Answer:**

1. What is the probability that between 2 and 6 of them have had a tetanus shot?

**Answer:**

1. What is the average number of subjects who have had a tetanus shot are you expecting to observe in this sample?

**Answer:**