**Namita Mishra**

**Exercise 1: (1 point)**

The number of students in the MPH program by county is given below. If a student is selected at random from this dataset, find the probability that:

|  |  |  |
| --- | --- | --- |
| **MPH program** | | |
| **County** | **Number of students** | **Percentage** |
| Escambia | 106 | 21.2 |
| Walton | 151 | 30.2 |
| Santa Rosa | 150 | 30 |
| Okaloosa | 93 | 18.6 |
| Total | 500 | 100 |

1. The student is from Escambia county

106/500 = 0.212

1. The student is from Santa Rosa county

150/500 = 0.3

1. The student is from either Escambia or Santa Rosa county

106/500 + 150/500 = 0.512

1. The student is not from Santa Rosa county

(500-150)/500 = 350/500 = 0.7

1 point

(Show your work)

**Exercise 2:(1 point)**

A cross tabulation of the number of students in the MPH program by county and whether or not the students owned a computer when they started the program is given below. If the student is selected at random from this data set, determine the probability of selecting a student who is from Santa Rosa County **or** a student who had a computer at the start of the program. (Show your work) (HINT: see slide 10)



probability of selecting

1. student who is from Santa Rosa County(A) = 150/500 = 0.3
2. a student who had a computer (B) = 237/500 = 0.474
3. P (A and B) =

probability of selecting (a student who is from Santa Rosa County **or** a student who had a computer)

P (A or B) = P (A) +P (B) – P (A and B)

150/500 + 237/500 –64/500 = 323/500 = 0.646

1 point

**Exercise 3:(1 point)**

Out of 500 MPH students, we select at random a student and note whether that student ever smoked pipe. Then we select again, at random, from the entire group (sampling with replacement) and note again whether the newly selected student ever smoked pipe. A table with the values is given below.



1. What is the probability of selecting a student who never smoked pipe?

probability of selecting a student who smoked pipe = A

P(A)= 92/500

P (a student who never smoked pipe ) = 1- (92/500) =

500-92

----------- =408/500 = 0.816

500

1. What is the probability that neither student smoked pipe?

probability of selecting a student who never smoked pipe on 1 trial =

= 1- (92/500) =

500-92

----------- = 408/500 = 0.816

500

probability of selecting a student who never smoked pipe on 2 trial with replacement =

408/500 = 0.816

You still have not answered the question: what is the probability that neither student smoked pipe?

Multiplication Rule 1

Pr(A and B) = Pr(A) x Pr(B)

= Pr(No Smoked and No Smoked) = Pr(No Smoked) x Pr( No Smoked)

= Pr(No Smoked and No Smoked) = (408/500) x (408/500) = 0.816 x 0.816

= Pr(No Smoked and No Smoked) = 0.665856 = 0.67 or 66.6%

0.5 points

(Show your work)

**Exercise 4: (1 point)**

Investigators conducted a study to evaluate the use of a screening test of specific hormones in the blood to assess whether or not the fetus of a pregnant women is likely to have Down syndrome.  4810 pregnant women underwent the screening test and scored either positive or negative depending on the levels of hormones in the blood.There were 360 women with positive test; from them, 9 women had an affected fetus. There were 4800 unaffected fetus; 4449 of these had negative test results.

What was the sensitivity, specificity, positive predictive value and negative predictive value of the test? Interpret the results(Show your work including developing a 2 X 2 table).

|  |  |  |  |
| --- | --- | --- | --- |
|  | FETUS AFFECTED | FETUS NOT AFFECTED | TOTAL |
| screening test |  |  |  |
| positive | TP 9 a | FP 351 b | 360 |
| negative | FN 1 c | TN 4449 d | 4450 |
| TOTAL | 10 | 4800 | 4810 |
|  |  |  |  |

1. SENSITIVITY= true positive fraction = P(test +|disease);

TPF= TP/(TP+FN) where (TP + FN) is the number of patients with the disease.

= a/a+c = 9/(9 + 1) = 9/10 = .9 How would you interpret the sensitivity and specificity?

This test has 90% sensitivity, so if a pregnant woman is affected then there is a 90% chance the test will be positive.

1. SPECIFICITY= true negative fraction = P(test -|disease free)

TNF= TN/(TN + FP) where (TN + FP) is the number of patients that are disease free.

=d/d+b = 4449/(4449+351) = 4449/ 4800 = .93

1. Positive predictive value = a/a+b = 9/(9+351) = 9/360 = .025

It is the probability that are the subjects with positive screen truly have the fetus affected.

1. Negative predictive value = d/d+c = 4449/(4449 + 1) = 4449/4450 = .99

It is the probability that are the subjects with positive screen truly don’t have their fetus affected.

0.5 points

**Exercise 5: (6 points)**

1)

a) The first data set represents runs scored by 5 baseball players in a national tournament. We want to recode this data so that the players are rank ordered by their number of runs, with the player with the highest runs given a code of “1” and the player with the lowest score given a 5. Enter the following data in SPSS:

# of runs by players:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Number of runs by players | | | | |  |
| Players | 1 | 2 | 3 | 4 | 5 |  |
| Runs | 86 | 120 | 56 | 10 | 18 |  |

b) Recode the data so that the players are rank ordered by their number of runs, with the player with the highest runs given a code of "1" and the batsman with the lowest runs given a "5".



















6 points

**Exercise 6: (0.5 points) OPTIONAL**

Screencast 2 has a small error. Identify the error made while recoding the data and explain the consequences of this error.

The error while recoding the data the range of the second category to which score of 2 is given is wrong. Instead of 61 - 72 it should be 61 - 74. It would change the frequency of that range.

0.5 points

Grade: 9.5/10