Group Project

Disclaimer: any data given to you, the student, as part of this project is entirely fictional and does not represent any actual students or diseases.

Project Overview
This is the handout for the group project. **You will work in groups of 2-3 students.**
This handout has some definitions, and **5 sets of questions.**

Deadlines and grading information are covered on the last page.

Introduction
A disease called Big Blue Fever is spreading through Lexington. Medical experts are designing a new screening test to detect the disease. The screening test involves a measurement of the patient’s CATS levels, and the medical experts are trying to determine how to design the screening test to make it accurate at detecting Big Blue Fever.

- Ideally, the screening test should give a positive result if the patient has Big Blue Fever, and a negative result if the patient does not have the disease.
- That is, **the test should identify people who have the disease by giving them a positive test result**, and it should **identify people without the disease by giving them a negative test result.**
- However, the test **could give a positive test result to someone without the disease. The test misidentified that person as having the disease.** Similarly, the test **could give a negative test result to someone with the disease. Here, the test misidentified that person as not having the disease.**
- A true positive is when the patient tests positive, and indeed has the disease.
Question Set 1

Write complete sentences for your responses here.

   a) What would a true negative be in the context of this situation?
   b) What would a false positive be in the context of this situation?
   c) What would a false negative be in the context of this situation?

In this class, we have studied some conditional probabilities that can be used to measure how accurate these kinds of screening tests are. For the project, we will need these four measurements: sensitivity, specificity, false positive rate, and false negative rate.

d) Since sensitivity, etc. are probabilities, what is the largest number they can be?

e) Since sensitivity, etc. are probabilities, what is the smallest number they can be?

f) Which of the four conditional probabilities (sensitivity, specificity, false positive rate, or false negative rate) can tell us how often the test correctly identified someone with the disease by giving them a positive test result?

g) Which of the four conditional probabilities can tell us how often the test correctly identified someone without the disease by giving them a negative test result?

h) Which of the four conditional probabilities can tell us how often the test incorrectly identified someone with the disease by giving them a negative test result?

i) Which of the four conditional probabilities can tell us how often the test incorrectly identified someone without the disease by giving them a positive test result?
Where is the Data for My Group?
You will be given data in an email announcement from Canvas after you have formed your group, or after the instructor has put you in a group. If you cannot find this email, please go to your group’s page on Canvas. Here is a guide on finding your group page in Canvas:

https://guides.instructure.com/m/4212/l/55565-how-do-i-view-my-canvas-groups-as-a-student

Your group is in the MA111 class, and its name is a number (not a time and day).

Question Set 2
Give exact whole numbers for your answers here.

  a) How many patients in the study have a CATS level of 6 and have Big Blue Fever?
  b) How many patients have a CATS level of 4 and do not have Big Blue Fever?
  c) How many patients have a CATS level of 5 in total?
  d) How many patients have Big Blue Fever?
  e) How many patients have a CATS level of 4 or higher, and have Big Blue Fever?

What Should a Positive Test Result Be?
The test gives a whole number between 0 and 9 when measuring a patient’s CATS levels. The researchers need to decide how to use the CATS levels to diagnose whether a patient has Big Blue Fever or not by defining what CATS level readings are positive test results.

Their first idea is as follows:

A patient tests positive if their CATS level is 4 or higher.

A patient tests negative if their CATS level is 3 or less.

With this, the researchers can find the number of true positives, true negatives, false positives, and false negatives from their data. For example, the number of true positives is the number of patients in the study whose CATS level is 4 or higher, and indeed have Big Blue Fever.
Question Set 3

For parts (a-e), give whole numbers. For part (f), give fractions and decimals rounded to at least four places. For parts (g-k), give complete sentences.

a) How many true positives are there?
b) How many true negatives are there?
c) How many false positives are there?
d) How many false negatives are there?
e) Fill in the following table using the data from your Canvas group page, and compute the totals as well:

<table>
<thead>
<tr>
<th></th>
<th>No Big Blue Fever</th>
<th>Has Big Blue Fever</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tested Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tested Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

f) Compute the following conditional probabilities: sensitivity, specificity, false positive rate, and false negative rate. Give your answers as fractions, and as decimals rounded to at least 4 places. Make sure they are labelled with which probability they are.

For parts (g)-(j), be sure to cite one of your probabilities from part (f) and explain why its size justifies your answer.

g) Based on the probabilities you computed, is the test good at correctly identifying people with the disease?
h) Is the test bad about misidentifying people as having the disease?
i) Is the test good at identifying people as not having the disease?
j) Is the test bad about misidentifying people as not having the disease?
k) Overall, based on your answers to parts (g)-(j), is the test good at both identifying people with and without the disease? If not, is it better at identifying people with the disease or people without the disease?
Redesigning the Test: Changing What a Positive Result Is

The researchers did not like some of the conditional probabilities from Question Set 3, and thought they needed to make the test better. They decided to change what positive and negative test results are:

A patient tests **positive** if their **CATS level is 8 or higher**.

A patient tests **negative** if their **CATS level is 7 or less**.

For example, the number of true positives is the number of patients in the study whose CATS level is 8 or higher, and indeed have Big Blue Fever.

**Question Set 4**

For parts (a-e), give whole numbers. For part (g), give fractions and decimals rounded to at least four places. For parts (e), and (h-n), give complete sentences.

a) How many true positives are there?

b) How many true negatives are there?

c) How many false positives are there?

d) How many false negatives are there?

e) Fill in the following table using the data from your Canvas group page:

<table>
<thead>
<tr>
<th></th>
<th>No Big Blue Fever</th>
<th>Has Big Blue Fever</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tested Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tested Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

f) Compare this table to the one from Question Set 3. Which entries got larger? Which entries got smaller? Which entries stayed the same? Please give complete sentences.

g) Compute the following conditional probabilities: sensitivity, specificity, false positive rate, and false negative rate. Give your answers as **fractions, and as decimals rounded to at least 4 places**. Make sure they are labelled with which probability they are.

h) Compare your answers to part (f) from Question Set 3. Which probabilities got bigger? Which ones got smaller?
For parts (i)-(l), be sure to cite one of your probabilities from part (g) and explain why its size justifies your answer.

i) Based on the probabilities you computed, is the test good at correctly identifying people with the disease?
j) Is the test bad about misidentifying people as having the disease?
k) Is the test good at identifying people as not having the disease?
l) Is the test bad about misidentifying people as not having the disease?
m) Compared to the results from Questions Set 3, did the test get better at identifying patients with Big Blue Fever? Cite at least one of the probabilities you computed from question sets 3 and 4 and how comparing them led you to your answer.
n) Compared to the results from Questions Set 3, did the test get better at identifying patients as not having Big Blue Fever? Cite at least one of the probabilities you computed from question sets 3 and 4 and how comparing them led you to your answer.

What Assumptions Are Being Made?

Question Set 5

a) Notice that the researchers define a positive test result based on whether the patient’s CATS level is too high (above a threshold). Based on this, how do the researchers believe CATS levels and Big Blue Fever are related?
b) Based on the table from Canvas, do you believe this assumption is true or appropriate?
c) Based on the table from Canvas, give an alternate relationship between CATS levels and the presence of Big Blue Fever.
What to turn in

First, you will need to form groups of 2-3 people.

**You will need to send an email to the instructor with the first and last names of every member of the group.**

You may request help forming a group using the tool on Canvas (check the announcements in Canvas). Please sign up for a time slot when you are best available for group work. The instructor will form groups based on common availability.

You must send this email or sign up for a time slot before **October 25th, 11:59pm**.

Your group will need to write (type in MS word) a report that includes your responses to all five question sets. Each group turns in one report, and it should have every group member's name on it. **You must submit your assignment on Canvas before November 20th, 11:59pm.** See instructions below:

https://community.canvaslms.com/docs/DOC-3128

**Do not upload from a smartphone!**

After uploading your submission, you may want to check if it was successful. After submitting, on the right side of the screen, there should be a green box with a green checkmark and the words “Turned In!” Inside this box is a link to your file. You can download it to see if the submission was uploaded correctly.

**Email submissions will not be accepted.**
Rubric
This project is out of 100 points, and is worth 10% of your final grade.

Group members
(15 points) Email the names of every member of your group (one email per group).

On-time upload
(15 points) The group’s report must be submitted to Canvas before November 20th, 11:59pm.

Responses to questions
- Question set 1: 9 points total
- Question set 2: 5 points total
- Question set 3: 22 points total
- Question set 4: 28 points total
- Question set 5: 6 points total