

Roll No.: _____

Name: _____

BM5033 Statistical Inference Methods in Bioengineering

Exam 1 Solution

1. Does Google Sheets use $(n - 1)$ or n in the denominator for variance calculation? Only circle the correct answer from the choices given below. **(5 marks)**

(a) $n - 1$ ✓

(b) n

2. A clinical trial evaluating a novel microfluidic urine test strip for diabetes screening recorded urinary glucose levels across 120 patients. The microfluidic strip operates colorimetrically, and upon exposure to urine, the test pad changes color, and the output is given by the following labels: *negative*, *trace*, *moderate*, *severe*.

The results from one trial using this device all 120 participants, are summarized in the table below:

Labels	negative	trace	moderate	severe
Number of Patients	18	42	36	24

Identify and calculate the *appropriate* central tendency and measure of dispersion in this dataset. Give justification for your choice. If you calculate all measures of central tendency and all measures of dispersion covered in the class, it will result in 0 marks. **(20 marks)**

Given the ordinal nature of the data, the correct measures of the central tendency and dispersion are median and IQR, respectively **(6)**

For this example, the median value is between 60th and 61st observations in ordered data. This implies the central tendency lies between trace and moderate.

Acceptable answers **(7)**

- between trace and moderate
- trace
- moderate

The IQR is the range between the 30th and 90th observations; that is IQR = trace to moderate. **(7)**

3. Recall the following question from the problem set

"A medical technology startup launches a novel gene-therapy drug for a rare, chronic pediatric autoimmune disease. Before the drug was available, the disease was incurable and fatal, with children surviving an average of 2 years post-diagnosis. The new drug does not cure the disease or prevent new cases from occurring, but it stops disease progression, extending average survival post-diagnosis to 8 years. Assume that the population size and the Incidence Rate of the disease (10 new cases per 100, 000 person-years) remain completely constant over a 15-year evaluation period. Immediately after the drug is widely adopted, the local health department notices that the prevalence (the proportion of the population with the disease) increases dramatically over the decade. A public official claims "*This new drug is failing. It is causing a massive outbreak of the disease!*". Is this statement correct?"

I plugged in this question to Google Gemini and it gave the following answer

No, the public official's statement is incorrect.

The increase in prevalence is a result of treatment success, not a disease outbreak. The rate of new cases remains unchanged at 10 per 100,000 person-years. An "outbreak" requires an increase in incidence, which did not occur. Because patients live longer without being cured, they accumulate in the population over time. The rising prevalence reflects extended patient survival, not a failure of the drug.

Is the answer given by the Google Gemini tool correct? If you think the answer and argument are correct, just write that *it is correct*. No justification is required. Giving justification in this scenario will result in zero marks.

If it is not correct, justify why. **(20 marks)**

It is correct.