Module 3E Questions

1. Which statement(s) is true about p-values?

    a. p-value is the probability that the null hypothesis is true or false

    b. p-value reflects the weight of evidence against the null hypothesis

    c. p-value measures the size of the effect

    d. if p value is less than or equal to the significance level, then the null hypothesis is not rejected.

2. Someone claims they make 90% of the shots they make on goal in soccer. If this is tested, what would be the null hypothesis – pick the best option **and justify it by writing out the full Ho and Ha.**

    a. You do not have enough information to make a null hypothesis because you don't know how they will do the test.

    b. $\bar{x} = 0.9$

    c. The person does not have a mean of making 90% of their soccer shots.

    d. The person does have a mean of making 90% of their soccer shots.

3. Fill in a "Confusion Matrix" with probability statements. There will be four cells that correspond to False Positive (alpha), True Positive, True Negative, False Negative (beta). Use the numbers from the following number to fill in a confusion matrix with counts:

You are testing a variant-calling pipeline on targeted sequencing data from **C57BL/6J mice**. You sequenced a small **panel of 1,000 genomic positions** that are known to sometimes carry mutations in certain mouse strains.

From previous studies, you know that in this panel:

- Exactly **100 positions** are truly **variant** (they carry a SNP in this strain).
- The remaining **900 positions** are truly **non-variant**.

Your pipeline analyzes the same 1,000 positions and:

- **Calls 120 positions as "variant."**
- Of these 120 calls, **90 are actually true variant positions** (they match the known truth set).

A. **Fill in the confusion matrix** for the variant caller, treating each of the 1,000 positions as either "variant" or "non-variant":
    - True Positives (TP)

- False Positives (FP)
- False Negatives (FN)
- True Negatives (TN)

B. You will then need to compute:
   - **Sensitivity (Recall)** = TP / (TP + FN)
   - **Specificity** = TN / (TN + FP)
   - **Precision** = TP / (TP + FP)