Quiz 5

STAT 109: Introductory Biostatistics

Quiz 5

Quiz 5 Practice Problems

These are practice problems covering:

- **Lecture 7:** Bayes' rule, law of total probability
- **Lecture 8:** diagnostic-test style problems (prevalence, sensitivity, specificity, $P(C \mid T^+)$)
- **Lecture 9:** factorials, ${}_nP_r$, ${}_nC_r$, counting (permutations and combinations)
- **Lecture 10:** binomial distribution (four conditions, identifying n, x, p)
- **Lab 4 (R code by hand):** binomial in R (`dbinom`, `pbinom`, `rbinom`)

Instructions

- Work **without notes** first; then check your answers.
- For **R questions**, write **valid R code** exactly as you would type it (by hand).
- The online quiz in lab will be created from these practice problems.

Part A: Lecture 7 — Bayes' rule

A1. State Bayes' rule

State **Bayes' rule** in formula form. (You may use A and B for the two events.)

A2. Law of total probability (two events)

State the **law of total probability** for the case where the sample space is partitioned into two events A and "not A." That is, give the formula for $P(B)$ in terms of conditional probabilities and $P(A)$, $P(\text{not } A)$.

Part B: Lecture 8 — Diagnostic test (like Problem 5)

B1. First diagnostic scenario

Suppose the **prevalence** of a disease in a population is $P(D) = 0.03$ (3%). A test has **sensitivity** $P(T^+ \mid D) = 0.90$ and **specificity** $P(T^- \mid \text{not } D) = 0.95$ (where T^+ = tests positive, T^- = tests negative).

1. Find $P(T^+ \mid \text{not } D)$ (the false positive rate).
2. Find $P(D \text{ and } T^+)$ using the multiplication rule.
3. Find $P(\text{not } D \text{ and } T^+)$.
4. Find $P(T^+)$ — the overall probability of a positive test.
5. Find $P(D \mid T^+)$ — the probability a person has the disease given they test positive. (Use the definition of conditional probability or Bayes' rule.)

B2. Second diagnostic scenario

Suppose the prevalence of a disease is $P(D) = 0.01$ (1%). A test has sensitivity $P(T^+ \mid D) = 0.98$ and specificity $P(T^- \mid \text{not } D) = 0.90$.

1. Find the false positive rate $P(T^+ \mid \text{not } D)$.
2. Find $P(T^+)$ using the law of total probability.
3. Find $P(D \mid T^+)$.

Part C: Lecture 9 — Factorials, permutations, combinations

C1. Factorials

1. Compute 4!.
2. What is 0! (by convention)?

C2. Compute ${}_nP_r$ and ${}_nC_r$

1. Compute ${}_5P_2$.
2. Compute ${}_5C_2$.

C3. Order matters vs does not matter

1. How many ways can you choose 3 people from 7 to form a **committee** (order doesn't matter)? Write the expression (e.g., ${}_7C_3$ or $\binom{7}{3}$) and compute it.
2. How many ways can you choose 3 people from 7 to fill **three distinct roles** (president, secretary, treasurer)? Write the expression and compute it.

C4. Counting problem

A team of 4 must be chosen from 10 players. Order does not matter — we only care which 4 players are on the team.

1. Write the expression for the number of ways to choose 4 players from 10 (use ${}_nC_r$ or $\binom{n}{r}$).
2. Compute the numerical value.

Part D: Lecture 10 — Binomial distribution

D1. Four conditions for binomial

State the **four conditions** that must hold for a random variable to have a **binomial** distribution.

D2. Identify n, x, and p in context

A fair coin is flipped 12 times. Let X = number of heads.

1. In this binomial setting, what is n? What does it represent?

2. What does x represent when we write $P(X = x)$?
3. What isthe definition of p? What does it represent in this context?

D3. Another binomial context

Twenty percent of seeds from a certain plant fail to germinate. You plant 8 seeds. Let X = number of seeds that fail to germinate.

1. Identify n and p in this binomial setting.
2. Write the expression (using binomial notation or ${}_nC_x$) for $P(X = 3)$. you do not need to simplify it.

Part E: Lab 4 — R code (binomial)

E1. `dbinom`

Write one line of R code to compute $P(X = 5)$ when X is Binomial$(n = 10, p = 0.4)$. Store the result in `prob_five`.

E2. `pbinom`

Write one line of R code to compute $P(X \leq 3)$ when X is Binomial$(n = 10, p = 0.4)$. Store the result in `prob_at_most_3`.

E2b. `pbinom` with `lower.tail = FALSE`

Write one line of R code to compute $P(X > 2)$ when X is Binomial$(n = 10, p = 0.4)$. Use `pbinom` with `lower.tail = FALSE`. Store the result in `prob_more_than_2`.

E3. `rbinom`

Write one line of R code to simulate 1000 values from a Binomial$(n = 8, p = 0.2)$ distribution. Store the result in `sims`.

E4. Estimate probability from simulation

Using the vector `sims` from E3, write one line of R code to estimate $P(X \leq 2)$ (the proportion of simulated values that are 2 or less). Store the result in `p_est`.

E5. Meaning of `aes()` in ggplot2

What does `aes()` do in a ggplot2 plot? Select the single best answer:

a. It creates the axes (x-axis and y-axis) of the plot.

b. It maps variables in the data to visual properties (e.g., which column is x, which is y, color, fill).

c. It chooses which geometry to use (e.g., bars, points, lines).

d. It defines the data frame (i.e. "spreadsheet") that the plot uses.

E6. ggplot2 bar plot syntax (fill in the blanks)

Fill in the blanks using the options so the code creates a data frame with one column called `x` that contains the possible numbers of heads in 10 tosses and another column called `prob` that contains $P(X = x)$ with $n = 10$ and $p = 0.4$ and then a bar plot with ggplot2 using `geom_col()`. M

Options: `0:10` | `dbinom(0:10, size = 10, prob = 0.4)` | `x` | `prob`

```
df <- data.frame(x = _______, prob = _______)
ggplot(df, aes(x = _______, y = _______)) + geom_col()
```