Quiz 4

STAT 109: Introductory Biostatistics

Quiz 4

Quiz 4 Practice Problems

These are practice problems covering:

- **Lecture 6:** conditional probability, multiplication rule, independence
- **Lecture 8:** practice problems 1–4 (COVID data; excludes Bayes rule / problems 5–6)
- **Lab 3 (R code by hand):** `rep()`, `sample()` with `replace = FALSE`, `for` loop with `if`, `vector()`, `mean()` / `sum()`, `which()`, logical `&` and `|`

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 in-class quiz in lab will be created from these practice problems.

Part A: Vocabulary and definitions (Lecture 6)

A1. Conditional probability

1. In words, what does $P(A \mid B)$ mean?
2. State the **formula** for $P(A \mid B)$ (when $P(B) > 0$).

A2. Multiplication rule

1. State the **multiplication rule** for $P(A \text{ and } B)$ in terms of $P(A \mid B)$ and $P(B)$.
2. State the **special case** when A and B are **independent**.

A3. Independence

1. In words, when are two events A and B **independent**?
2. Give **two** equivalent ways to check independence (formulas or conditions).

A4. Translate probability notation

Write a one-sentence interpretation for each:

1. $P(A \mid B) = 0.3$
2. $P(A \text{ and } B) = P(A)\,P(B)$

A5. Match probability notation with English

A person is randomly selected from a population. Define:

- A = "the person has the disease"
- B = "the person tests positive"

Match each probability expression on the left with the correct English interpretation on the right.

Expression	Interpretation
$P(A \text{ and } B)$	(a) The baseline chance of disease in the population before any test is done.
$P(A)$	(b) Among people who test positive, the chance they actually have the disease.
$P(A \mid B)$	(c) The proportion of people who would both have the disease and receive a positive test result.
$P(B \mid A)$	(d) Among people who have the disease, the chance the test correctly identifies them as positive.
$P(A \text{ or } B)$	(e) The chance that at least one of two things happens: the person has the disease, or tests positive (or both).

Write your matches (e.g., $P(A) \to$ (a)).

A6. Conditional probability with a die

A fair six-sided die is rolled once.

1. Find $P(\text{roll is } 2 \mid \text{roll is even})$.
2. Find $P(\text{roll is even} \mid \text{roll is } 2)$.

Part B: Lecture 6 problems

B1. Candy jar (without replacement)

A jar contains 10 candies: 6 are red and 4 are blue. Two candies are selected **without replacement**.

Let A = "first candy is red" and B = "second candy is red."

1. Find $P(A)$.
2. Find $P(B \mid A)$.
3. Use the multiplication rule to find $P(A \text{ and } B)$.

B2. Two marbles (without replacement)

A bag contains 8 marbles: 4 are blue and 4 are green. Two marbles are drawn **one at a time**; the first is not replaced before the second is drawn.

Find the probability the **first** marble is blue and the **second** marble is green.

B3. Tiles: color vs shape (independence)

A bag contains 6 tiles: 2 red, 2 blue, and 2 green. Each tile is either round or square — there are 3 round and 3 square in total (one of each color-shape combination). One tile is drawn at random.

Is the event "the tile is red" independent of the event "the tile is round"? Show your reasoning.

B4. Two coin tosses

Suppose a fair coin is tossed twice.

Find the probability of getting "heads" on the first toss **and** "tails" on the second toss.

B5. Two marbles (conditional probability)

A bag contains 8 marbles: 4 are blue and 4 are green. Two marbles are drawn **one at a time** without replacement.

Given that the first marble is blue, find the probability the second marble is green. (That is, find $P(\text{second is green} \mid \text{first is blue})$.)

B6. Two die rolls (independence)

Suppose a fair six-sided die is rolled twice.

Let A = "first roll is even" and B = "second roll is 6."

1. Are A and B independent? Explain briefly.
2. Compute $P(A \text{ and } B)$.

Part C: Lecture 8 problems (COVID data, no Bayes)

COVID Setup: A clinic tests a new rapid test for COVID on 1000 people. Of these, 20 have COVID and 980 do not. Among those with COVID, 17 test positive; among those without COVID, 78 test positive.

Let C = "has COVID" and T^+ = "tests positive."

C1. COVID probabilities

1. Find $P(C)$ and $P(\text{not } C)$ for a randomly selected person from the 1000.
2. Find $P(T^+ \mid C)$ (sensitivity).
3. Find $P(T^+ \mid \text{not } C)$ (false positive rate).

C2. COVID joint probabilities

1. Find $P(C \text{ and } T^+)$.
2. Find $P(\text{not } C \text{ and } T^+)$.

C3. COVID independence

1. If C and T^+ were **independent**, what would $P(C \text{ and } T^+)$ equal?

2. Are C and T^+ actually independent? Explain briefly using the definition.

C4. COVID overall positive rate

Using the COVID data, find $P(T^+)$ — the overall probability of a positive test.

C5. COVID interpretation

In one sentence, interpret what $P(T^+)$ means in the context of this COVID study.

Part D: R code by hand (Lab 3)

D1. `rep()` and balls in a bag

Write R code that creates a vector `balls` representing a bag with **5 red** and **2 white** balls. Use `"R"` and `"W"` as labels.

D2. Sampling without replacement and `for` loop

A bag has 3 red and 3 white balls. Write complete R code that simulates 10,000 draws of 2 balls (without replacement), records a success when the first is white and the second is red, and estimates the probability with `mean(successes)`.

D3. `which()` to find successes

Suppose `successes` is a numeric vector of length 10,000 from a simulation, where $1 =$ success and $0 =$ no success. Write one line of R code to find which replicate numbers had a success.

D4. Modify the `if` condition for "exactly one red"

The code below estimates $P(\text{white first and red second})$ when drawing 2 balls from a bag of 3 red and 3 white.

Your task: Cross out the part that needs to change and write the modification so the code instead estimates $P(\text{exactly one red})$ — i.e., either (R then W) or (W then R). Use the logical operator `|` for "or".

```r
balls <- rep(c("R", "W"), c(3, 3))
replicates <- 10000
successes <- vector("numeric", replicates)
set.seed(2018)
for (k in 1:replicates) {
  draw <- sample(balls, size = 2, replace = FALSE)
  if (draw[1] == "W" & draw[2] == "R") {
    successes[k] <- 1
  }
}
mean(successes)
```

D5. `sample()` with `prob`

Write R code that simulates 10,000 individuals where 70% like eggnog and 30% do not. Use `sample()` with the `prob` argument to specify these probabilities, and store the result in `eggnog`.

D6. Joint probability from simulated table

A population of 10,000 was simulated with: $P(\text{disease}) = 0.01$; among those with disease, $P(\text{positive}) = 0.95$; among those without disease, $P(\text{positive}) = 0.10$. The results are stored in vectors `disease` (0 = no, 1 = yes) and `test_result` (0 = negative, 1 = positive).

Write one line of R code to estimate $P(D \text{ and positive})$ — the proportion of individuals who have the disease **and** a positive test. Use `sum()` and `&`.
