Statistics with Recitation — Quiz 6

December 2, 2025

Name: ⎯⎯⎯⎯⎯⎯⎯⎯ **ID:** ⎯⎯⎯⎯⎯⎯⎯⎯

1. (**2 points**) Which of the following statements about `ggsave()` is FALSE?

    A. If no plot object is supplied, `ggsave()` saves the most recent plot drawn.

    B. We can save a plot as a PNG file by using `ggsave(file = .png)`.

    C. The `dpi` argument controls the quality of the saved image.

    D. By default, `ggsave()` saves the plot to the current working directory.

2. (**2 points**) If we want to add a vertical line to our graph, which function should we use?

    A. `geom_hline()`

    B. `geom_vline()`

    C. `geom_line()`

    D. `vertical_line()`

3. (**2 points**) Which of the following statements is FALSE?

    A. In `geom_hline()`, we use `intercept = ...` to set where the line is placed along the x-axis.

    B. `facet_wrap()` supports faceting by multiple variables.

    C. By default, the argument `scales = ...` in `facet_wrap()` uses `"fixed"`, meaning the x- and y-axis scales are the same across all subplots.

    D. `nrow` is an argument in `facet_wrap()` that controls the number of columns in the panel layout.

4. (**2 points**) Consider the following code:

```
1  data <- data.frame(
2    id = 1:5,
3    x  = c(3, 8, 2, 9, 5),
4    y  = c(10, 2, 10, 3, 10)
5  )
6
7  result <- data %>%
8    filter((x > 4 & y == 10) | (x >= 8 & y < 5)) %>%
9    select(id)
```

Which IDs are included in `result`?

A. 1, 2, 4

B. 2, 4, 5

C. 1, 2, 5

D. 2, 3, 4

5. (**2 points**) Which of the following statements is TRUE?

A. `&` only checks the first element of each vector; `&&` is vectorized.

B. The output of `cumsum(1:5 %in% list(1:5))` is 1 2 3 4 5.

C. `library(dplyr)` automatically loads all packages in the `tidyverse`.

D. The output of `5 < 6 & 3 > 1 | 7 < 3 & 4 < 2` is `TRUE`.

6. (**2 points**) What is the structure of the object `z` created by the following code?

```
1  x <- 1:5
2  y <- c(10,20,30,40,50)
3  z <- rbind(x, y)
```

A. A 2-by-5 matrix.

B. A 5-by-2 matrix.

C. A 10-by-1 matrix.

D. A 1-by-10 matrix.

7. (**2 points**) Which of the following correctly generates 100 random values from a Poisson distribution with standard deviation 4?

A. `rpois(100, lambda = 4)`

B. `rpois(100, rate = 4)`

C. `rpois(100, lambda = 16)`

D. `rpois(100, rate = 16)`

8. (**2 points**) What does the following code generate?

```
1  seq(as.Date("2024-01-01"),
2     as.Date("2024-01-10"),
3     by = "2 days")
```

A. A sequence of dates increasing by 2 days.

B. A sequence of length 2.

C. A character vector of dates.

D. An error because seq() cannot be used with Date objects.

9. (**2 points**) Suppose:

```
1  x <- c("AB12", "CD34", "EF56")
2  substring(x, 3, 4)
```

Which statement is TRUE?

    A. It returns `FALSE TURE FALSE`.

    B. It returns the first two characters of each string.

    C. It extracts the numeric parts `"12"`, `"34"`, `"56"`.

    D. It returns integers 12, 34, 56.

10. (**2 points**) Consider the following code:

```
1  df <- data.frame(
2    name = c("Alice", "Bob", "Carol", NA),
3    score = c(90, 85, NA, 70),
4    date = c("2025-01-01", "2025-02-15", "2025-03-10", "2025-04-20")
5  )
```

Which of the following statements is FALSE?

    A. The expression `dim(df)[1] == nrow(df)` always evaluates to TRUE for any data frame.

    B. Applying `na.omit(df)` will remove any row in which at least one column contains `NA`.

    C. Converting `df$date` using `as.Date(df$date)` works without specifying a format.

    D. The expression `grepl("0", df$score)` returns `FALSE FALSE FALSE FALSE`.

11. (**2 points**) Suppose:

```
1  counts <- table(c("A","A","B","B","A","B"),
2                  c("Yes","No","Yes","No","Yes","No"))
```

Which code correctly conducts a chi-squared test of independence?

    A. `chisq.test(prop.table(counts))`

    B. `chisq.test(counts)`

    C. `chisq.test(as.numeric(counts))`

    D. `chisq.test(table(counts))`

12. (**2 points**) What is the output of the following code?

```
1  grades <- data.frame(
2    class = c("A","A","B","B","B","A"),
3    pass  = c("Yes","No","Yes","Yes","No","Yes")
4  )
5
```

```
6  grades %>%
7    group_by(class) %>%
8    summarize(pass_rate = mean(pass == "Yes"))
```

A.

```
1  0.6667 0.6667
```

B.

```
1            [,1]
2  [1,]    0.6667
3  [2,]    0.6667
```

C.

```
1  class pass_rate
2  A         0.6667
3  B         0.6667
```

D.

```
1        pass_rate
2  A        0.6667
3  B        0.6667
```

13. (**2 points**) A company measures reaction time for each worker before and after a caffeine break. The results are stored in two equal-length vectors: `before` and `after`. Which R command correctly tests whether the caffeine break changed the average reaction time?

    A. `t.test(before, after)`

    B. `t.test(after, before)`

    C. `t.test(before, after, paired = TRUE)`

    D. `t.test(after ~ before)`

14. (**2 points**) Consider the following code:

```
1  damage_sunraku <- c(
2    1310, 1195, 1240, 1285, 1220, 1350, 1290, 1210,
3    1265, 1305, 1330, 1275, 1255, 1190, 1320
4  )
5
6  damage_psyger0 <- c(
7    1380, 1425, 1405, 1360, 1390, 1435, 1410, 1375,
8    1440, 1400, 1395, 1420
9  )
10
11  test_1 <- t.test(damage_psyger0, damage_sunraku)
12
```

```r
13  df <- data.frame(damage = c(damage_sunraku, damage_psyger0),
14                   player = c(replicate(15, "Sunraku"),
15                             replicate(12, "Psyger-0"))
16                   )
17
18  test_2 <- t.test(df$damage ~ df$player)
```

Which of the following statements about `test_1` and `test_2` is correct?

    A. The statistics calculated from these two tests are completely the same.

    B. The statistics calculated from these two tests are different because one uses the paired sample t-test, while the other uses the two independent sample t-test.

    C. The statistics calculated from these two tests are slightly different because one uses the two independent sample t-test, while the other uses Welch two sample t-test.

    D. Although `test_2` works, `test_1` does not because the lengths of `damage_psyger0` and `damage_sunraku` are different.

15. (**2 points**) What is the output of the following code?

```r
1  sales <- data.frame(
2    region = c("N","N","N","N","S","S","S","S"),
3    year   = c(2024,2024,2024,2025,2024,2025,2025,2025),
4    rep    = c("A","A","B","B","A","A","B","B"),
5    amount = c(100, 110, 120, 130, 90, 110, 150, 170),
6    cost   = c(40, 10, 60, 40, 20, 130, 230, 230)
7  )
8
9  stores <- sales %>%
10    group_by(region, year, rep) %>%
11    mutate(growth = amount - cost) %>%
12    summarize(mean_growth = mean(growth)) %>%
13    filter(mean_growth == max(mean_growth))
14
15  sum(stores$mean_growth)
```

    A. 90

    B. 160

    C. 210

    D. 220