# PSTAT 10: Final Review Questions

December 03, 2023

---

**1. Which of the following is a valid variable name in `R`?**

  (a) `_awesomejob_`

  (b) `1st_place`

  (c) `perfect!`

  (d) `YOu_dId_iT`

  (e) None of the other answer choices

---

**2. What is returned by the following code?**

```
> typeof(matrix(4:9, nrow = 3, ncol = 2))
```

  (a) `"matrix"`

  (b) `"double"`

  (c) `"array"`

  (d) None of the other choices

  (e) `"integer"`

---

**3. Consider a matrix `mat`, defined as:**

```
> mat <- matrix(10:17, nrow = 3, ncol = 4, byrow = TRUE)
```

What code would we run to extract everything **except** the third column?

(a) `mat[ , -3]`

(b) `mat[ , !3]`

(c) `mat[-3, ]`

(d) `mat[ , ]`

(e) None of the other answer choices

---

**4. Given an array `arr_3d` and a function `myfunction()`, which of the following lines of code will, when run, apply the function `myfunction()` to the cells (both rows and columns) of `arr_3d`?**

(a) `apply(arr_3d, MARGIN = 1, FUN = myfunction)`

(b) `replicate(arr_3d, MARGIN = c(1,2), FUN = myfunction)`

(c) `apply(arr_3d, MARGIN = c(1,2), FUN = myfunction)`

(d) None of the other answer choices

(e) `apply(arr_3d, MARGIN = 2, FUN = myfunction)`

---

**5. What would be the output of running `Inf + Inf` in R?**

(a) `NaN`

(b) `Inf`

(c) `NA`

(d) `"Undefined"`

(e) None of the other answer choices

---

**6.** Let s be the string `"I need to study for the midterm exam by completing these midterm review questions!"` Which of the following snippets of code can, when run, replace all instances of the word `"midterm"` with the word `"final"` in s?

(a) `sub(pattern = "midterm", replacement = "final", s)`

(b) `gsub(s, "midterm", "final")`

(c) None of the other answer choices

(d) `gsub(pattern = "midterm", replacement = "final", s)`

(e) `gsub(pattern = midterm, replacement = final, s)`

---

**7.** Consider the following plot:

**Boxplot of y, Grouped by x**
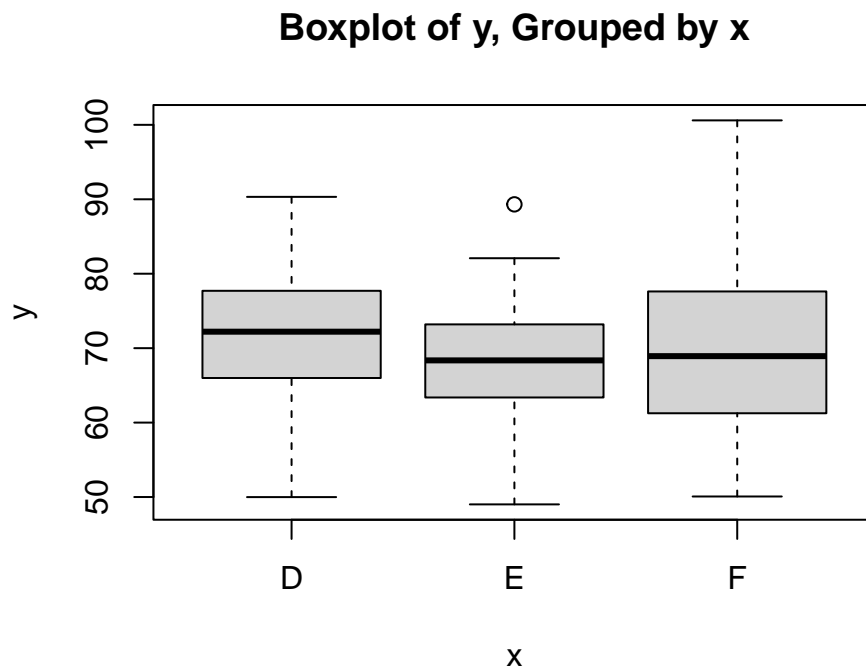


Which of the following lines of code will produce this plot? (Assume all relevant variables have been defined appropriately.)

(a) `boxplot(x ~ y, xlabel = "x", ylabel = "y", main = "Boxplot of y, Grouped by x")`

(b) `boxplot(x ~ y, xlab = "x", ylab = "y", main = "Boxplot of y, Grouped by x")`

(c) None of the other answer choices

(d) `boxplot(y ~ x, xlab = "x", ylab = "y", title = "Boxplot of y, Grouped by x")`

(e) `boxplot(y ~ x, xlab = "x", ylab = "y", main = "Boxplot of y, Grouped by x")`

---

**8. Consider the following code:**

```
if(length(x) != length(y)) {
  if(x[1] == y[1]) {
    print("Krispy")
  } else if (x[2] == y[2]) {
    print("Kreme")
  } else {
    print("Donut")
  }
}
```

For which of the following values of `x` and `y` will the code above produce the result `"Donut"`?

(a) None of the other choices

(b) `x <- c(1, 3, 5), y <- c(1, 2, 5, 6)`

(c) `x <- 1:10, y <- 0:9`

(d) `x <- c(3, 6, 4), y <- c(1, 2, 3, 4)`

(e) `x <- 1:10, y <- 1:10`

---

**9. A bowl contains sweet and sour candies, of various flavors. The exact contents of the bowl are as follows:**

|       | Orange | Grape | Lemon |
|-------|--------|-------|-------|
| Sweet | 9      | 3     | 6     |
| Sour  | 1      | 7     | 4     |

A candy is selected, and its consistency is noted. Given that it is a sour candy, what is the probability it is grape flavored?

4

(a) 7/10

(b) 7/12

(c) 10/30

(d) 7/30

(e) None of the other choices

---

**10. Given two events $A$ and $B$ with $\mathbb{P}(A) = 0.4$, $\mathbb{P}(B) = 0.3$, and $\mathbb{P}(A \cap B) = 0.2$, are $A$ and $B$ independent?**

(a) There is not enough information to determine

(b) Yes

(c) No

---

**11. Suppose $X$ is a discrete random variable with the following probability mass function (p.m.f.):**

| x | $-1$ | 0 | 1 |
|---|------|-----|-----|
| $\mathbb{P}(\mathbf{X} = \mathbf{x})$ | 0.5 | 0.1 | 0.4 |

What is the value of $E[X]$, the expected value of $X$?

(a) 0.1

(b) None of the other choices

(c) 0.5

(d) 0

(e) 1

**12. Suppose we have a variable `CDF` that contains the cumulative distribution function of a random variable $X$. The contents of the variable `CDF` is shown below. Using `CDF`, which of the following will calculate $\mathbb{P}(X \leq 10.5)$?**

```
##    8    9   10   11   12   13
## 0.30 0.55 0.70 0.85 0.95 1.00
```

   (a) `CDF[10.5]`

   (b) `CDF[10]`

   (c) `CDF[2]`

   (d) `CDF[6] - CDF[3]`

   (e) `CDF[3]`

---

**13. Players in Major League Baseball (MLB) each have a 30% chance of hitting a home run with their lucky baseball bat. A sample of 15 MLB players is selected, with replacement. What is the probability that, in this sample, at least 5 players but fewer than 8 players hit a home run with their lucky baseball bat, i.e., $\mathbb{P}(5 \leq X < 8)$?**

   (a) `pbinom(7, size = 30, prob = 0.45) - pbinom(4, size = 30, prob = 0.45)`

   (b) `pbinom(7, size = 30, prob = 0.45) - pbinom(5, size = 30, prob = 0.45)`

   (c) `pbinom(8, size = 30, prob = 0.45) - pbinom(5, size = 30, prob = 0.45)`

   (d) `pbinom(8, size = 30, prob = 0.45) - pbinom(4, size = 30, prob = 0.45)`

   (e) `sum(dbinom(5:8, size = 30, prob = 0.45))`

---

**14. While waiting in line at the DMV office, you learn that the time it takes to reach the front of the line and receive help from an associate is uniformly distributed between 75 and 90 minutes. What is the probability a randomly selected person has to wait 80 minutes or more to to reach the front of the line and receive help from an associate?**

   (a) `punif(80, min = 75, max = 90)`

(b) `punif(80, min = 75, max = 90, lower.tail = F)`

(c) `1 - punif(79, min = 75, max = 90)`

(d) `punif(79, min = 75, max = 90, lower.tail = F)`

(e) `1 - punif(80, min = 75, max = 90, lower.tail = F)`

---

**15. A professional swimmer knows that the length of time it takes each member of their swimming team to complete a 1-mile lap in the pool, X, is normally distributed with a mean of 40 minutes and a variance of 9 minutes. In this context, what time separates the top 20 percent of the fastest swimmers to complete the 1-mile lap in the pool?**

(a) `qnorm(0.2, mean = 40, sd = 9)`

(b) `qnorm(0.8, mean = 40, sd = 3)`

(c) `pnorm(20, mean = 40, sd = 3)`

(d) `qnorm(0.2, mean = 40, sd = 9, lower.tail = F)`

(e) `qnorm(0.8, mean = 40, sd = 3, lower.tail = F)`

---

**16. Which of the following is NOT part of the relational model?**

(a) Manipulative

(b) Integrity

(c) Structural

(d) Security

(e) All of the above are part of the relational model

---

**17.**

**I. Existence of multiple foreign keys in a relation is NOT possible.**

**II. Existence of multiple primary keys in a relation is NOT possible.**

   (a)  Both I. and II. are false

   (b)  I. is correct, but II. is incorrect.

   (c)  I. is incorrect, but II. is correct.

   (d)  Both I. and II. are true

   (e)  There is not enough information to determine.

---

**For 18. and 19., use the following information:**

A database in a gaming universe contains only two relations: **PLAYERS** and **GUILDS**. Each player is assigned a unique `PLAYER_ID`, and each guild is given a unique `GUILD_ID`. `PLAYER_NAME` and `PLAYER_LOCATION` represent the player's name and location respectively. `JOIN_DATE` signifies the year a player first joins a guild. Assume that new tuples may be added to the relations at any time.

Table 3: **GUILDS**

| PLAYER_ID | JOIN_DATE | GUILD_ID |
|-----------|-----------|----------|
| P1        | 2020      | G1       |
| P2        | 2010      | G2       |
| P3        | 2007      | G3       |
| P1        | 2010      | G4       |

Table 4: **PLAYERS**

| PLAYER_ID | PLAYER_NAME | PLAYER_LOCATION |
|-----------|-------------|-----------------|
| P1        | FRANCIS     | State           |
| P2        | SHAUN       | Garden          |
| P3        | LUCA        | Storke          |
| P7        | FOX         | Hollister       |

**18.** The primary key of relation `GUILDS` is GUILD_ID. The primary key of relation `PLAYERS` is PLAYER_ID. The tuple `<P5, 2005, G3>` is added to the relation `GUILDS`. Which of the following is true?

(a) Referential integrity is violated.

(b) Entity integrity is violated.

(c) Both Referential integrity and Entity integrity are violated.

(d) Neither Referential integrity nor Entity integrity are violated.

(e) There is not enough information to determine.

---

**19.** Assume the database in the gaming universe is called `GAMING` and is assigned to a variable called `gaming_db`. Write an RSQLite expression to add a new player named **CYBORG** who lives at Mission to the relation `PLAYERS`. Assume CYBORG has been assigned the player ID 'P8'.

(a) `dbWriteQuery(gaming_db, 'INSERT VALUES ("P8","CYBORG","Mission") INTO PLAYERS')`

(b) `dbSendQuery(gaming_db, 'INSERT VALUES ("P8","CYBORG","Mission") INTO PLAYERS')`

(c) `dbWriteQuery(gaming_db, 'INSERT INTO PLAYERS VALUES ("P8","CYBORG","Mission")')`

(d) `dbSendQuery(gaming_db, 'INSERT INTO PLAYERS VALUES ("P8","CYBORG","Mission")')`

(e) `dbSendQuery(gaming_db, 'INSERT PLAYERS VALUES ("P8","CYBORG","Mission")')`

---

**20.** Kate is eager to identify employees in her company whose names end with the letter 'e'. Within her database, there is a relation named `EMPLOYEES`, and inside this relation exists a field called `Name`, that encompasses both the first and last names of the employees. Which of the following SQL clauses is the correct choice for extracting the names of employees in her company that end with the letter 'e'?

(a) `WHERE Name like '%e_'`

(b) `WHERE Name like '%e'`

(c) `WHERE Name like 'e%'`

(d) `WHERE Name like '%e%'`

(e) `WHERE Name like '_e%'`

---

Table 5: **SCORES**

| Student_ID | Exam_ID | Score |
|------------|---------|-------|
| 101 | 0 | 88 |
| 101 | 1 | 84 |
| 102 | 0 | 96 |
| 102 | 1 | 83 |
| 103 | 0 | 95 |
| 103 | 1 | 90 |
| 104 | 0 | 85 |
| 104 | 1 | 81 |

**21. Suppose we have the following relation called `SCORES` in our database that stores exam scores for students. The relation has three fields, *Student_ID*, *Exam_ID*, and *Score*. Which of the below queries, would, *FOR EACH* `Student_ID`, find the *highest* exam score and filter for only those students having a max score higher than 90? (Assume `dbGetQuery()` will be used for each query below.)**

(a) "SELECT student_id, MAX(score) AS max_score FROM SCORES WHERE max_score > 90 GROUP BY student_id"

(b) "SELECT student_id, MAX(score) AS max_score FROM SCORES GROUP BY student_id WHERE max_score > 90"

(c) "SELECT student_id,MAX(score) AS max_score FROM SCORES HAVING max_score > 90"

(d) "SELECT student_id,MAX(score) AS max_score FROM SCORES GROUP BY student_id"

(e) "SELECT student_id, MAX(score) AS max_score FROM SCORES GROUP BY student_id HAVING max_score > 90"

---

**22.** **Which of the following statements is true about the difference between the HAVING and WHERE clauses in SQL?**

(a) WHERE is used before grouping/aggregating, whereas HAVING is used after grouping/aggregating.

(b) HAVING is used before grouping/aggregating, whereas WHERE is used after grouping/aggregating.

(c) There is no difference; the two clauses may be used interchangeably.

(d) Every SQL query **must** include a WHERE clause, whereas it is allowed for a HAVING clause to be absent in a SQL query.

(e) None of the other answer choices

Table 6: **CUSTOMER**

| CUST_NO | Name | Address |
|---------|------|---------|
| C1 | ALEX | State |
| C2 | BOB | Hollister |
| C3 | CAROL | Ocean |
| C6 | JUAN | Phelps |

Table 7: **SALES_ORDER**

| ORDER_NO | DATE | CUST_NO |
|----------|------|---------|
| 1 | 11/11/19 | C1 |
| 2 | 7/9/19 | C3 |
| 9 | 8/16/19 | C6 |
| 10 | 10/12/19 | C6 |

**23. Suppose we want to join the following two relations from the `TinyClothes` database. Which query would correctly *INNER JOIN* the relation `CUSTOMER` to `SALES_ORDER`? Assume we select all attributes. (Also assume `dbGetQuery()` will be used for each query below)**

(a) `SELECT * FROM CUSTOMER JOIN SALES_ORDER ON CUSTOMER.CUST_NO = SALES_ORDER.ORDER_NO`

(b) `SELECT * FROM CUSTOMER INNER JOIN SALES_ORDER ON CUST_NO = CUST_NO`

(c) `SELECT * FROM CUSTOMER JOIN SALES_ORDER ON CUSTOMER.CUST_NO = SALES_ORDER.CUST_NO`

(d) `SELECT * FROM CUSTOMER INNER JOIN SALES_ORDER ON CUST_NO`

(e) `SELECT * FROM CUSTOMER INNER JOIN SALES_ORDER CUSTOMER.CUST_NO = SALES_ORDER.CUST_NO`

---

**24. Assume that `fall_df` is a data frame we created in `R`. We are interested in adding it as a relation to our database, `mydb`, with the name `FALL`. Which of the following expressions will result in adding the relation `FALL` to our database?**

(a) None of the other answer choices.

(b) `dbAddTable(mydb, "fall_df", FALL, overwrite = T)`

(c) `dbAddTable(mydb, "FALL", fall_df, overwrite = T)`

(d) `dbWriteTable(mydb, "FALL", fall_df, overwrite = T)`

(e) `dbWriteTable(mydb, "fall_df", FALL, overwrite = T)`

---

**25. Assume you are working in RSQLite with the database `mydb`. Which of the following commands closes the connection?**

(a) `disconnect(mydb)`

(b) `exit(mydb)`

(c) `dbDisconnect()`

(d) `Dbdisconnect(mydb)`

(e) None of the other answer choices.

---