DATA 101, Midterm 2018

Problem 1. Which of these functions is **not** related to plotting?

(a) abline()
(b) histogram()
(c) par()
(d) legend()
(e) sample()

Answer: Both B and E are considered correct. B is correct, because the function for histogram in R is hist(), not histogram().

Problem 2. What would R say, if c("a",1,T) is entered to the console?
(a) [1] "a" "1" "TRUE"
(b) [1] "b" 1 T
(c) [1] "a" "1" "T"
(d) [1] "a" 1 T
(e) Something else.

Answer: A

Problem 3. Which of the following is a correct way to select 2 columns from the data frame traffic?

(a) traffic[1,2]
(b) traffic[1:2,]
(c) traffic[,c(1,2)]
(d) traffic[c(1,2)]
(e) traffic[2]

Answer: C

Problem 4. Suppose v <- c(-1,0,3,2,-10) is entered into the R console. What would R say if you enter v[v>0]?

(a) [1] FALSE FALSE TRUE TRUE FALSE
(b) [1] -1 0 3 2 -10
(c) [1] 3 4
(d) [1] 3 2
(e) Something else.

Answer: D

For Problems 5–10, assume that the following data frame has been entered into R:

```
t <- data.frame(x=c(1,2,3),y=c(3,2,1),z=c("a","b","c"))
```

In each problem, what would R say if you entered the given command in the console?

Problem 5. t[t$x==1,] what would R say?

```
  x y z
1 1 3 a
```

Problem 6. t[t$z=='a',] what would R say?

```
  x y z
1 1 3 a
```

Problem 7. t[,t$y==2] what would R say?

```
[1] 3 2 1
```

Problem 8. t[1,] what would R say?

```
  x y z
1 1 3 a
```

Problem 9. t[,2:1] what would R say?

y x
1 3 1
2 2 2
3 1 3

Problem 10. t[c(1,2),c(1,2)] what would R say?

x y
1 1 3
2 2 2

Statistics and plotting

Problem 11. The goal of a permutation test is to:

(a) Show how often the observed results could happen by random chance.
(b) Prove that the order in which elements are added to the dataset is irrelevant.
(c) Show that every permutation of the dataset gives the same results.
(d) Prove that the null hypothesis is true.

Answer: A

Problem 12.

Q1. The test has a mean score of 150 and a standard deviation of 20. If John's z-score is 1.50, what was his score on the test?
A) 180
B) 130
C) 30
D) 150
E) None of the above

Answer: A

Problem 13

Studies show that listening to music while studying can improve your memory.

To demonstrate this, a researcher experiments with two samples A and B of 36 college students each and gives both of them a standard memory test.

Sample A listens to some background music. Sample B works in silence.

For sample B the mean score obtained was 25 and standard deviation is 6. For sample A (i.e With music) the mean is 28 with the same standard deviation of 6.

After performing the Z-test, what can we conclude? (Z value for α = 0.05 would be 1.65).

A)Listening to music does not improve memory.
B) Listening to music significantly improves memory.
C) The information is insufficient for any conclusion.

Answer: B

What is the Z-value?

2.1213

Problem 14

Average time to travel by bus from College Ave to Livingston is 10 mins whereas average time to travel back from Livingston to College Ave is 15 mins. It seems that travelling back from Livingston takes D=5 minutes longer. Permutation test is performed 100,000 times and its results report that 4000 permutations show D>7 and 10000 permutations show D>6. What can we say about the p-value?

a) $p <= 0.06$
b) $p > 0.1$
c) $p <= 0.07$
d) $p < 0.5$

Answer: B

Problem 15

A poll shows that 55% of sample of population supports more gun control, while 45% oppose more gun control/ We would like to test alternative hypothesis that majority of population supports gun control. The p-value is calculated to be 0.10. Lets assume that our critical value is 0.05 (acceptable p-value) Which statement is correct?

A. We can conclude that more that 55% of entire population supports more gun control.
B. We can conclude that more than 10% of entire population does not support more gun control.
C. We can conclude that exactly 45% of entire population does not supports more gun control.
D. We cannot conclude that majority of population supports gun control.

Answer: D

Problem 16

On average Immigrants happiness index is 8.1. Natives happiness index is 7.5.

Are natives less happy than immigrants, indeed?

We run 10000 permutation tests and 9500 of these tests show mean(Immigrants_ Happiness) < mean(Natives Happiness) + 1
What can we say about p-value?

A) $p>0.05$
B) $p<0.05$
C) None of the above

Answer: A

Problem 17

Studies showed that vitamin C prevents sinus infections. Large study was performed and demonstrated that taking vitamin C lowers the risk of sinus infection by 20%. The p-value of was reported to be 10^-3.

The information about study showed that ten different types of vitamins were tested for improvements on ten different types of infections.

Can we accept the findings of the study that vitamin C helps in sinus infection?

Answer: No (because of Bonferroni correction)

For Problems 18–20, determine which type of plot would be best to use for visualizing/plotting the described data. Choose from the following answers:

A) Box plot
B) Bar plot
C) Scatter plot
D) Mosaic plot
E) Histogram

Problem 18. The distribution of midterm grades for this class. [which plot from A-E would you select]

Answer: A or E

Problem 19. Frequency of each grade in this class for each major [which plot from A-E would you select]

Answer: A, B, D or E

Problem 20. Prices of wines vs alcohol level [which plot from A-E would you select]

Answer: C