STAT 123 Midterm 1 (Practice Test)

Monday, February 1st, 2021

Duration: 10:00 am to 11:10 am + 10 minutes to upload

Submissions due by 11:20 am.

- The test is 70 minutes long + 10 minutes to upload.
- If you are not writing directly on the test paper (for example on a tablet or if you printed it out) then be sure to clearly label your solutions (indicating which question you are answering).
- You may type your answers or hand write your answers.
- Once you are done the test. Either take pictures of your handwritten work and convert them to a PDF or save your typed solutions as a PDF.
- You must have your test solutions uploaded to the Brightspace drop box by no later than 11:20pm unless you have a CAL time accommodation.

1. *[5 points]* A record collector wants to catalogue his collection. In a spreadsheet he inputs the artist, album name, year released, album length (in minutes), and the condition of the record (very good, good, some damage, poor quality).

   (a) What are the individuals being considered by the record collector?

   records

   (b) What are the variables being considered by the record collector?

   artist, album name, year released, album length, condition
   ↑ cat. ↑ cat. ↑ cat. ↑ num. ↑ cat.

   (c) Determine if each variable is categorical or numerical.

   (d) Is this an observational study or an experiment? Explain why.

   not treatments are imposed on the records.

2. *[3 points]* You want to buy a box of clementines from the grocery store but you don't want a rotten orange!

   (a) You sample 5 boxes from the top of the display and then open them up and observe the top layer of clementines. What type of sampling is being used?

   convenience because you "conveniently" chose the top of display.

   (b) Is the sampling design likely to result in bias? Why?

   Yes. There may be reasons for the store to put certain boxes on top display.

(c) Give an example of a way to sample the clementines differently. Be sure to indicate what kind of sampling you are describing.

many good answers

1. pick a box randomly.
2. pick some boxes on top, some at the bottom.

3. *[2 points]* Suppose you have a population consisting of 16 individuals. Using the table of random numbers given below, starting on the second row, choose a simple random sample of 4 individuals. Which individuals were chosen?

Part of a Table of Random Numbers

61424	20419	86546	00517
90222	27993	04952	66762
50349	71146	97668	86523
85676	10005	08216	25906
02429	19761	15370	43882
90519	61988	40164	15815
20631	88967	19660	89624
89990	78733	16447	27932

4. *[5 points]* In a class of 200 students a professor wants to predict the class average on the first homework assignment without marking every paper. She splits the students into groups of 50 (based on which tutorial they are registered in) and randomly selects 2 papers from each of the 4 tutorial sections. She marks these papers and it results in the following percentage scores:

85, 73, 56, 92, 81, 70, 70, 67

(a) What is the population of interest?

200 students in the professor's class.

(b) What is the parameter of interest?

class average on the first homework assignment.

(c) What is the statistic that should be used to estimate the paramater?

mean, median of the 8 papers. $\frac{X_1 + X_2 + \ldots + X_8}{8}$

(d) What is the observed value of the statistic?

Use calculator ... 74.25

(e) What type of sampling was used?

stratified : ① divide population into groups
② randomly select from each group.

5. *[4 points]* A researcher wants to estimate the proportion of the Canadian population who smoke cigarettes. She takes a simple random sample of 445 Canadians and determines that 164 of them smoke.

(a) Assuming the statistic used is the sample proportion, what is the observed value of the statistic?

(b) Estimate the margin of error for a 95% confidence interval.

(c) Compute the corresponding 95% confidence interval for the study results.

(d) Write a confidence statement about the true proportion of Canadians who smoke cigarettes.

6. *[7 points]* Consider the following R code:

```
4  apples = c("Gala","GrannySmith","McIntosh","Pink Lady")
5  quantity = c(40,60,20,20)
6  names(quantity) = apples
7  M = matrix(c(3:14), byrow = FALSE, ncol=4)
```

(a) What does quantity["GrannySmith"] return?

60

quantity [GrannySmith]

error : need quotation mark.

(b) What does apples[3] return?

"McIntosh"

(c) What does mean(quantity) return?

35

(d) What does mean(apples) return?

error : can't average character strings.

(e) If you print out the matrix M, what does it look like? 3 x 4 matrix

$$M = \begin{bmatrix} 3 & 6 & 9 & 12 \\ 4 & 7 & 10 & 13 \\ 5 & 8 & 11 & 14 \end{bmatrix}$$

filled by columns

(f) What value does rowSums(M) return?

30 34 38

(g) What does M[,1:2] return?

column 1 and 2

$$\begin{bmatrix} 3 & 6 \\ 4 & 7 \\ 5 & 8 \end{bmatrix}$$

[ ]

different in R