

Yerevan State University
Applied Statistics and Data Science
Applied Statistic with R
Midterm 1 Exam Test
29 October, 2019

Exam Time/Place: 18:30 - 20:30, MMF #241

Last Name: _____ First Name: _____

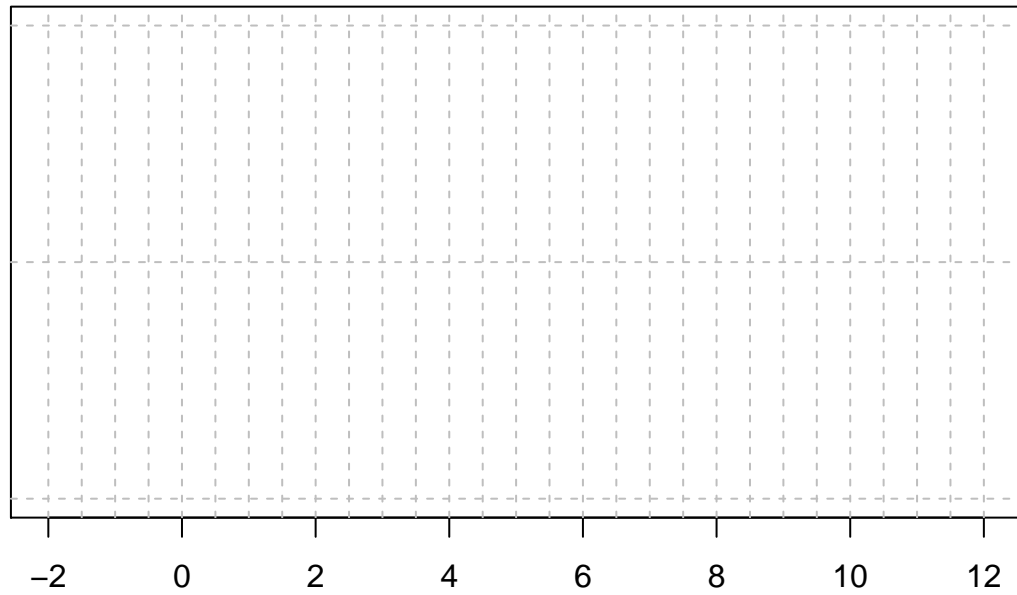
READ THESE INSTRUCTIONS CAREFULLY

- This test consists of 6 Show-Work Problems. The test booklet has 16 pages, including this cover page and empty pages for draft calculations.
- Each Show-Work Problem has its own grade. The overall test grade is 100.
- This is a closed-book test, and no notes, assignments, practice problems, books, formula sheets or other materials are allowed.
- The use of mobile phones or any other electronic devices are strongly prohibited. Please turn off your cell phones and place them out of reach. You can use only simple calculators.
- Sharing of stationery (pens, pencils, erasers, etc.) or calculators is not permitted.
- Talking to another student, looking at another student's paper, or communicating with other students in any way is strictly forbidden.
- Use the scratch pages of the test booklet to do your draft calculations. Please ask the instructor for extra scratch papers if necessary.
- If you run out of the space on the test pages, please use a scratch page to finish your work. Indicate in the test page that you will continue on the scratch page, and mark with the rectangle the portion on the scratch page that contains the solution. Any other work on the scratch page will not be graded.
- Good luck!

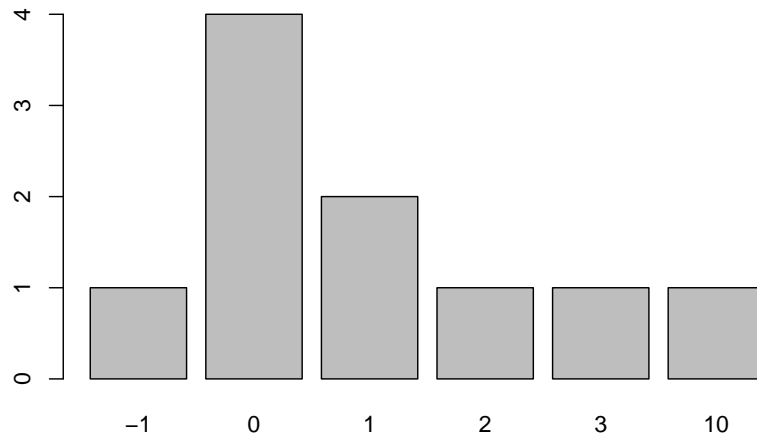
DO NOT OPEN THIS BOOKLET
UNTIL YOU HAVE BEEN TOLD TO DO SO

For the Problem 1

BoxPlot



Problem 1 (10 points) Here is a Barplot of the Dataset x :

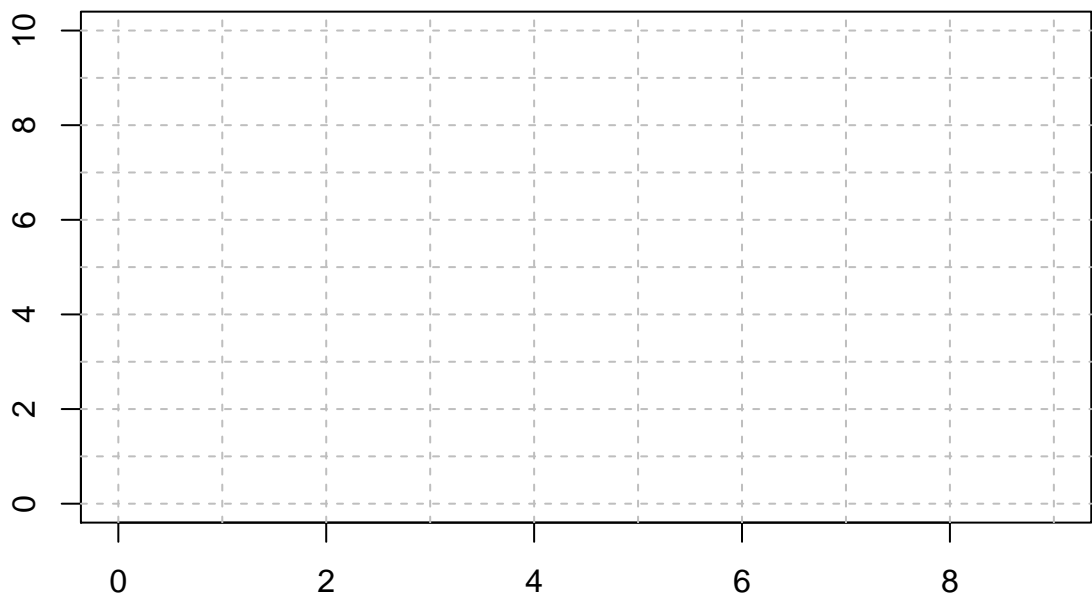


Using the empty graph template on the Page 2, construct the Boxplot of the Dataset x .

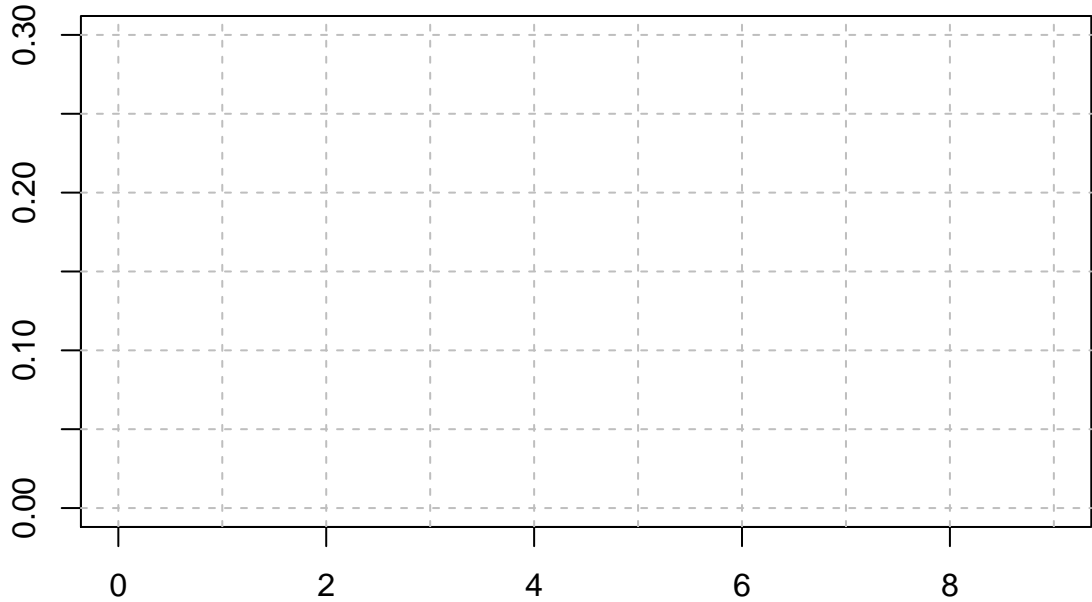
Show and explain below all your steps.

For the Problem 2

Frequency Histogram



Density Histogram



Problem 2 (15 points) We are given the Dataset x (generated in **R**)

```
## [1] 0.05 3.07 2.51 5.25 0.77 3.62 7.76 2.43 4.29 2.24 1.42 3.72 4.39 1.38  
## [15] 2.26 5.44 0.24 8.56 5.82 0.98
```

- On the empty graph templates on the Page 4, construct the Frequency and Density Histograms for x , using Bins $[0, 2)$, $[2, 3)$, $[3, 7)$ and $[7, 9]$;
- Find the $\text{range}(x)$;
- Calculate the sum of deviations $\sum_{k=1}^n (x_k - \bar{x})$ for this Dataset (here n is the size of x).

Show and explain below all your steps.

Scratch Paper

Problem 3 (15 points) We want to draw the Q-Q Plot for the Dataset

$$x : 2, 3, 1, 1, 3, 4$$

versus the Distribution given by the PDF

$$f(x) = \frac{5}{4x^2}, \quad x \in [1, 5] \quad \text{and} \quad f(x) = 0, \quad x \notin [1, 5].$$

Using the 20%, 50% and 80% quantiles, draw this Q-Q Plot.

Scratch Paper

Problem 4 (20 points) Insurance company wants to model the number of car accidents in Yerevan in a day. The company chooses as an appropriate model the Poisson Distribution $Pois(\lambda)$. The company collects information about the number of car accidents in Yerevan for several days, and get the following data:

10, 14, 21, 7, 9, 14, 18, 22, 21, 19.

The company's statistician is not strong in Statistics (he was not attending my lectures ☺), and he knows and can use only two Estimators for the unknown Parameter λ :

$$\hat{\lambda} = \frac{X_1 + 2X_{10}}{3} \quad \text{and} \quad \tilde{\lambda} = \frac{X_1 + X_2 + \dots + X_{10}}{10}.$$

What is the **best Estimate** for λ that the company's statistician can get using only these estimators? Justify your answer.

Note: Without justification you will get only 1 point!

Show and explain all your steps.

Scratch Paper

Problem 5 (20 points) We have 2 Datasets:

$$x : 1, 2, 3, \quad \text{and} \quad y : -1, 2, \alpha.$$

- a. Find all values of α such that the Correlation Coefficient $\text{cor}(x, y)$ is maximal;
- b. Find all values of α such that the Datasets x and y are Uncorrelated.

Show and explain all your steps.

Scratch Paper

Problem 6 (20 points)

- a. Assume we have a sequence of IID r.v.s $X_1, X_2, \dots \sim \mathcal{N}(-1, 2^2)$. Calculate the limit in the Probability sense:

$$\lim_{n \rightarrow \infty} \frac{n}{X_1^2 + X_2^2 + \dots + X_n^2}.$$

- b. Let X_n and X be defined on the same Probability Space with

$$X_n \sim \text{Unif} \left[1 - \frac{1}{n^2}, 1 + \frac{1}{n^2} \right] \quad \text{and} \quad X \equiv 1.$$

Using the definition, show that $X_n \xrightarrow{\mathbb{P}} X$.

Show and explain all your steps.

Scratch Paper

Scratch Paper

Scratch Paper