

Stat 324 Homework #1

Student Name Here

*Submit your homework to Canvas by the due date and time. Email your lecturer if you have extenuating circumstances and need to request an extension.

*If an exercise asks you to use R, include a copy of the code and output. Please edit your code and output to be only the relevant portions.

*If a problem does not specify how to compute the answer, you may use any appropriate method. I may ask you to use R or use manual calculations on your exams, so practice accordingly.

*You must include an explanation and/or intermediate calculations for an exercise to be complete.

*Be sure to submit the HWK1 Autograde Quiz which will give you 20 of your 40 accuracy points.

*50 points total: 40 points accuracy, and 10 points completion

Basics of Statistics and Summarizing Data Graphically (I)

Exercise 1. A number of individuals are interested in the proportion of citizens within a county who will vote to use tax money to upgrade a professional baseball stadium in the upcoming vote. Consider the following methods:

The **Baseball Team Owner** surveyed 8,000 people attending one of the baseball games held in the stadium. Seventy eight percent (78%) of respondents said they supported the use of tax money to upgrade the stadium.

The **Pollster** generated 1,000 random numbers between 1-52,661 (number of county voters in last election) and surveyed the 1,000 citizens who corresponded to those numbers on the voting roll. Forty three percent (43%) of respondents said they supported the use of tax money to upgrade the stadium.

- a. What is the population of interest? What is the parameter of interest? Will this parameter ever be calculated?

Type Answer Here.

- b. What were the sample sizes used and statistics calculated from those samples? Are these simple random samples from the population of interest?

Type Answer Here

- c. The baseball team owner claims that the survey done at the baseball stadium will better predict the voting outcome because the sample size was much larger. What is your response?

Type Answer Here

Exercise 2: After manufacture, computer disks are tested for errors. The table below summarizes the number of errors detected on each of the 100 disks produced in a day.

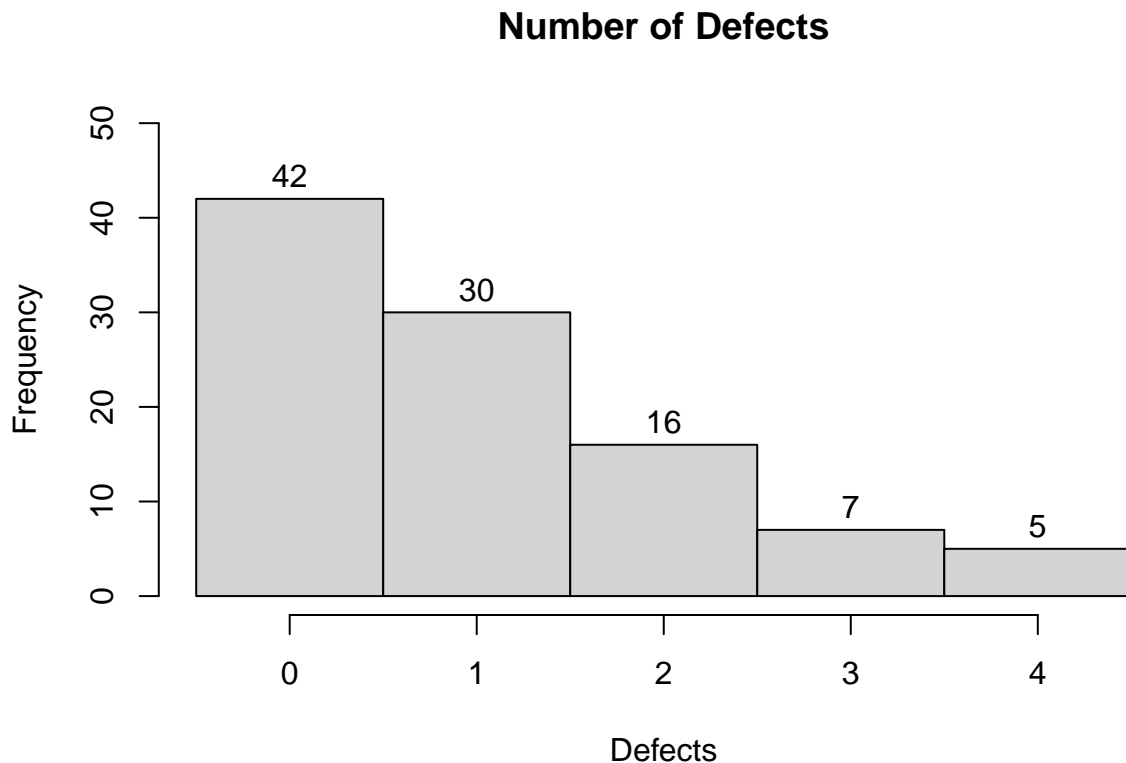
Number of Defects	Number of Disks
0	42
1	30
2	16
3	7
4	5

- a. Describe the type of data that is being recorded about the sample of 100 disks, being as specific as possible.

Type Answer Here

- b. Code for a frequency histogram showing the frequency for number of errors on the 100 disks is given below.
- bi. Knit the document and confirm that the histogram displays in the knitted file.

```
error.data=c(rep(0,42), rep(1,30), rep(2,16), rep(3,7), rep(4, 5))
hist(error.data, breaks=c(seq(from=-0.5, 4.5, by=1)),
     xlab="Defects", main="Number of Defects",
     labels=TRUE, ylim=c(0,50))
```



- bii. Describe what the `rep()` function does in this code chunk.

Type Answer Here

- biii. Describe how this `breaks` command affects the histogram's appearance in this code chunk. How would the histogram look different if the `breaks()` command was not included?

Type Answer Here

biv. Describe how setting `ylim=c(0,30)` instead of `ylim=c(0,50)` would change the histogram's appearance. Which value for `ylim` is preferable for clear communication of the data?

Type Answer Here