Homework

Nicholas Alexeev

8/23/2021

1. For each of the following hypothetical populations, give a plausible sample of size 4:

a. All distances that might result when you throw a football.

{5 yards,3 yards,4 yards,1 yard}

b. Page lengths of books published 5 years from now.

{112,221,812,351}

c. All possible earthquake-strength measurements (Richter scale) that might be recorded in California during the next year.

{3.2,4.2,3.8,2.1}

d. All possible yields (in grams) from a certain chemical reaction carried out in a laboratory.

{10g,20g,15g,8g}

2.

a. Give two different examples of concrete populations and two different examples of hypothetical populations.

Concrete Populations:

- length of salmon swimming up the copper river
- mass of each female salmon's fish egg sack. ### Hypothetical Population
- How many fish each fisherman will catch on each day.
- Revenue on each day for Hem's charters in the next season.

b. For one each of your concrete and your hypothetical populations, give an example of a probability question and an example of an inferential statistics question.

Length of Salmon swimming up the Copper River

Probability Question: What is the probability that a given salmon has a length longer then two inches. Inferential Question: Does species of salmon correlate with size? ### Mass of Each Female Salmon's Fish Egg Sack. Probability Question: What is the probability that the mass of the egg sack is greater then 50g.

Inferential Question: Does time in the fishing season correlate with mass of sack. ### How many fish each fisherman will catch in the next season. Probability Question: What is the probability that a fisherman will catch their limit on a given day. Inferential Question: Do low temperatures correlate with increased catches.

Revenue on each day for Hem's charters in the next season.

Probability Question: What is the probability that any given day is profitable. Inferential Question: Do high water levels correlate well with decreased income.

3. Many universities and colleges have instituted supplemental instruction (SI) programs, in which a student facilitator meets regularly with a small group of students enrolled in the course to promote discussion of course material and enhance subject mastery. Suppose that students in a large statistics course (what else?) are randomly divided into a control group that will not participate in SI and a treatment group that will participate. At the end of the term, each student's total score in the course is determined.

a. Are the scores from the SI group a sample from an existing population? If so, what is it? If not, what is the relevant conceptual population?

scores from the SI group are a sample from the existing population which is the entire class taking student body.

b. What do you think is the advantage of randomly dividing the students into the two groups rather than letting each student choose which group to join?

People are randomly assigned inorder to prevent biases from impacting the results. For example more active students may be more likely to join a SI group and also have higher scores, creating a bais.

c. Why didn't the investigators put all students in the treatment group? Note: The article "Supplemental Instruction: An Effective Component of Student Affairs Programming" (J. of College Student Devel., 1997: 577–586) discusses the analysis of data from several SI programs.

Investigators did not put all of their students in the treatment group so that they knew what the status quo scores were. If all students were in the SI group there would be no way to tell if their scores improved because there would be nothing to compare the SI scores to.

4. Install R and RStudio on your computer. (In your solution to this problem, it's okay to say something like "Done" (if you did).)

```r
print("hello R")
```

```
## [1] "hello R"
```

5. Use R to construct a stem-and-leaf plot of this data set: { 66,66,69,74,74,75,75,76,76,76,76,78,79,79,81,81,82,83,83,84,86,87,87,92,98 } Be sure to include your R code and the resulting plot. Also, what is a fairly typical value, based on the stem-and-leaf plot?

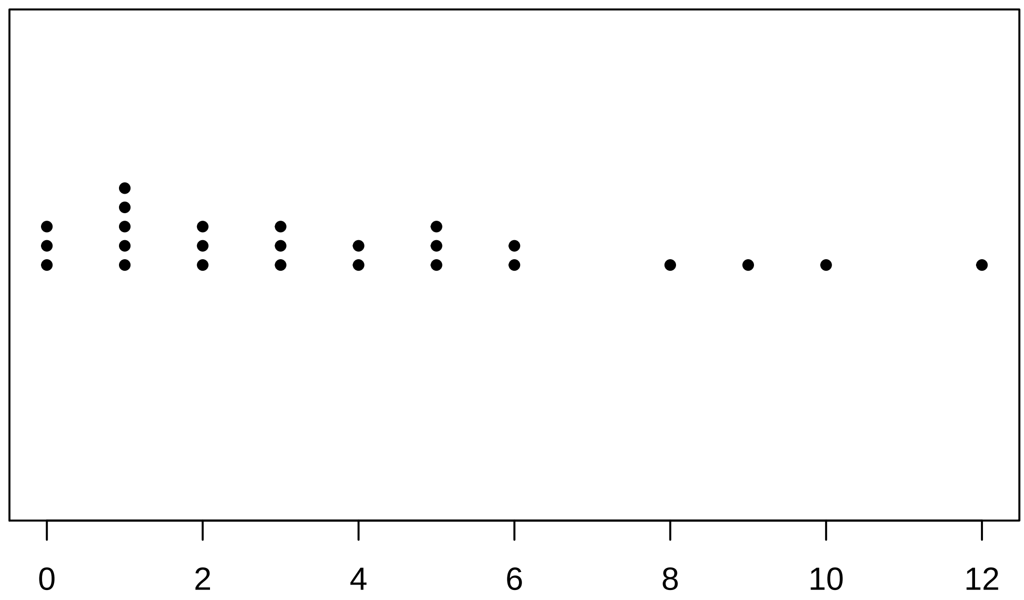
```
data <-c(66,66,69,74,74,75,75,76,76,76,76,78,79,79,81,81,82,83,83,84,86,87,87,92,98);
stem(data)
```

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   6 | 669
##   7 | 44556666899
##   8 | 112334677
##   9 | 28
```

A fairly typical value is 76 based on the plot

6. Use R to construct a dot plot of this data set: { 2,3,3,0,12,0,1,4,6,2,0,6,5,9,8, } Be sure to include your R code and the resulting plot. Describe the general appearance of this data set, based on the plot.

```
data <- c(2,3,3,0,12,0,1,4,6,2,0,6,5,9,8,1,1,1,2,3,10,4,5,5,1);
stripchart(data,method="stack",offset=0.5,pch=20);
```



The data is skewed to the left, it is unimodal with the center at 1.