Department of Mathematics & Statistics
Math2560 (All Sections)
Homework Assignment #1 , September 05

---

1. How does marijuana affect willingness to work? Canadian researchers persuaded young adult men who used marijuana to live for 98 days in a "planned environment". The men earned money by weaving belts. They used their earnings to pay for meals and other consumption and could keep any money left over. One group smoked two potent marijuana cigarettes every evening. The other group smoked two weak marijuana cigarettes every evening. All subjects could buy more cigarettes but were given strong or weak cigarettes depending on the group they belonged to. Did the weak and strong groups differ in work output?

   (a) Outline a completely randomized design for an experiment in order to try and answer the question. Identify the factor and the response. List the different treatments.

   Solution This experiment has one factor : strength of the cigarettes. The response is work output. There are two levels of the factor : strong and weak cigarettes and therefore two treatments. An equal number of young men should be randomly assigned to each one of the treatments (strong or weak cigarettes) and the work output would be measured for each one of the young men.

   (b) The following table lists the names of the 20 subjects (or experimental units). Use Table B of random digits in your textbook, starting at line 131, to carry out the randomization that your design requires. Carefully explain how you assign different subjects to different treatments.

Abbott	Decker	Gutierrez	Lucero	Rosen
Afifi	Engel	Hwang	McNeill	Thompson
Brown	Fluharty	Iselin	Morse	Travers
Chen	Gerson	Kaplan	Quinones	Ullman

   Solution We label the experimental units $01, 02, \ldots, 20$ in this order following the first column, the second, third and fourth. We then slice the string of digits given in Table B, starting at line 131 into slices of two digits. As sson as encore any of the labels in these slices, we assign the person with that label to the first treatment until we have 10 experimental units for the first treatment. Then we assign them

to the second treatment.

(c) Using any computer software such as Minitab, generate 10 random numbers from the uniform distribution between 0 and 100. Writing the 10 numbers in a continuous string, use this mini-table of random digits to do the same randomization as in (b) above.

(d) If it was a well established fact that men and women react differently to different strength of marijuana cigarettes, how would you modify the completely randomized design in (a) above to take this fact into account?

Solution In this case, we use "blocking" that is, we run the experiment separately for males and females.

2. Below is a data showing the lengths (in millimeters) of 100 brown trout at a hatchery.

15.0	15.3	14.1	10.4	10.2	11.5	15.4	11.7	15.0	10.9
13.6	10.5	13.8	15.0	13.8	14.5	13.7	13.9	12.5	15.2
10.7	13.1	10.6	12.1	14.9	14.1	12.7	14.0	10.1	14.1
10.3	15.2	15.0	12.9	10.7	10.3	10.8	15.3	14.9	14.8
14.9	11.8	10.4	11.0	11.4	14.3	15.1	11.5	10.2	10.1
14.7	15.1	12.8	14.8	15.0	10.4	13.5	14.5	14.9	13.9
10.1	14.8	13.7	10.9	10.6	12.4	14.5	10.5	15.1	15.8
12.0	15.5	10.8	14.4	15.4	14.8	11.4	15.1	10.3	15.4
15.0	14.0	15.0	15.1	13.7	14.7	10.7	14.5	13.9	11.7
15.1	10.9	11.3	10.5	15.3	14.0	14.6	12.6	15.3	10.4

All the histograms requested in (a), (b) and (c) below should be constructed with classes closed on the left and open on the right.

(a) Construct a grouped frequency distribution using classes [10.0,10.5), [10.5,11.0) and so on and draw a histogram of the distribution.

Solution From Minitab we obtain the following stem and plot

```
12    10 111223334444
25    10 5556677788999
29    11 0344
34    11 55778
37    12 014
42    12 56789
```

```
 43   13 1
(10)  13 5677788999
 47   14 00011134
 39   14 555567788889999
 24   15 000000011111223333444
  2   15 58
```

and therefore the histogram with classes as required has classes with relative frequencies as given below

Class	Frequency	Relative Frequency
[10,10.5)	12	.12
[10.5,11)	13	.13
[11,11.5)	4	.04
[11.5,12)	5	.05
[12,12.5)	3	.03
[12.5,13)	5	.05
[13,13.5)	1	.01
[13.5,14)	10	.1
[14,14.5)	8	.08
[14.5,15)	15	.15
[15,15.5)	22	.22
[15.5,16)	2	.02

The distribution is bimodal with modes at 10-11 and 15-15.5.

Describe the shape of the histogram.

(b) Construct a grouped frequency distribution using class midpoint 10.0, 10.5, 11.0 and so on and draw a histogram of the distribution.

Solution In this case, the classes and frequencies are as follows

Class	Frequency	Relative Frequency
[9.75,10.25)	5	.05
[10.25,10.75)	13	.13
[10.75,11.25)	6	.06
[11.25,11.75)	7	.07
[11.75,12.25)	5	.05
[12.25,12.75)	5	.05
[12.75,13.25)	3	.03
[13.25,13.75)	5	.05
[13.75,14.25)	11	.11
[14.25,14.75)	9	.09
[14.75,15.25)	23	.23
[15.25,15.75)	6	.06
[15.75,16.25)	1	.01

Describe the shape of the histogram.

Solution The histogram has two clear modes at 10.5 and 15.

(c) Construct a grouped frequency distribution with a class width of 2, and with 10 as the first class midpoint.

Solution

Class	Frequency	Relative Frequency
[9,11)	25	.25
[11,13)	17	.17
[13,15)	34	.34
[15,17)	26	.26

Describe the shape of the resulting histogram.

This histogram is rather flat with relatively small peaks for the first and third class.

(d) Discuss how the number of classes used and the choice of class boundaries affect the appearance of the resulting histogram.

Solution The difference between the histograms in (a) and (b) is the position of the center of the distribution. The histogram in (b) seems to capture the peaks (or modes) of the distribution more precisely since the region where most of the lower values occur are not split into two neighbouring regions as in (a). The histogram

in (c) is clearly not very informative: the classes are too wide.

3. The following data gives the amount of zinc (in mg) absorbed by 40 patients with rheumatoid arthritis.

8.0	12.9	13.0	8.9	10.1	17.3	11.1	10.9
6.2	8.1	8.8	10.4	15.7	13.6	19.3	9.9
8.5	11.1	10.7	8.8	10.7	6.8	7.4	4.8
11.8	13.0	9.5	8.1	6.9	11.5	11.2	13.6
4.9	18.8	15.7	10.8	10.7	11.5	16.1	9.9

(a) Draw a stem and leaf plot of this data set with tens of milligrams for the stem and with units of milligrams for the leaf. Round down the numbers to the lowest digit. Is this plot very informative?

Solution

0	4466678888888999
1	000000111111233335678 9

(b) Use any software such as Minitab to draw a stem and leaf plot of the data. Which unit did Minitab use to do the plot? Is the stem and leaf a good way to represent this data? Is the shape of distribution symmetric? If not, how would you describe the distribution?

Solution

```
Stem-and-leaf of C1        N  = 40
 Leaf Unit = 0.10

     2    4 89
     2    5
     5    6 289
     6    7 4
    13    8 0115889
    16    9 599
   (7)   10 1477789
    17   11 112558
```

```
11   12 9
10   13 0066
 6   14
 6   15 77
 4   16 1
 3   17 3
 2   18 8
 1   19 4
```

Minitab uses the tenth of a unit for the leaf and the plot gives good information. We can clearly see that the shape of the distribution is not symmetric but rather skewed to the right.

4. The output from the Minitab command "stem C1" where C1 is a given data vector is the following.

```
Stem-and-leaf of C1        N  = 15
Leaf Unit = 0.10

 1    2 2
 2    2 5
 3    3 4
 4    3 5
 6    4 11
(5)   4 66679
 4    5 44
 2    5 9
 1    6 0
```

(a) What is the meaning of the integers on the left most column?

Solution The integers of the lft most column give the cumulative frequency until we reach the mode. At the mode, (5) indicates the number of data point in the mode class. Below the mode the integers give the "inverse cumulatives frequencies, i.e. the cumulative frequencies staring from the highest values of the data.

(b) What is the sample mean for this data set? Do the computations by hand and show your work.

Solution The sample mean is 43.93.

(c) What is the sample standard deviation for this data set? Do the computations by hand and show your work.

Solution The sample standard devaition is 11.25.