

Stat 324: Lecture 01

Descriptive statistics

Moo K. Chung
mchung@stat.wisc.edu

January 18, 2005

1. A *population* is a collection of object under investigation. A *sample* is a subset of the population. The *frequency* of any particular value is the number of times that value occurs in the data set. The *relative frequency* of a value is the proportion of times the value occurs.
2. Let us consider following multivariate data. It can be downloaded from
<http://www.stat.wisc.edu/~mchung/teaching/data/312score.data>

```
score <- read.table("d:/stat324/312score.data", header=T)
```

```
> score
```

	mid1	mid2	hw	final
1	93	100	97	96
2	96	97	98	94
3	96	100	98	86
...				
48	24	87	90	22
49	56	80	28	33

```
> plot(score$mid1, score$mid2)
```

```
> hist(score$mid1)
```

To find out more about `hist` command, use

```
> help(hist). It will display a new window with detailed information about hist.
```

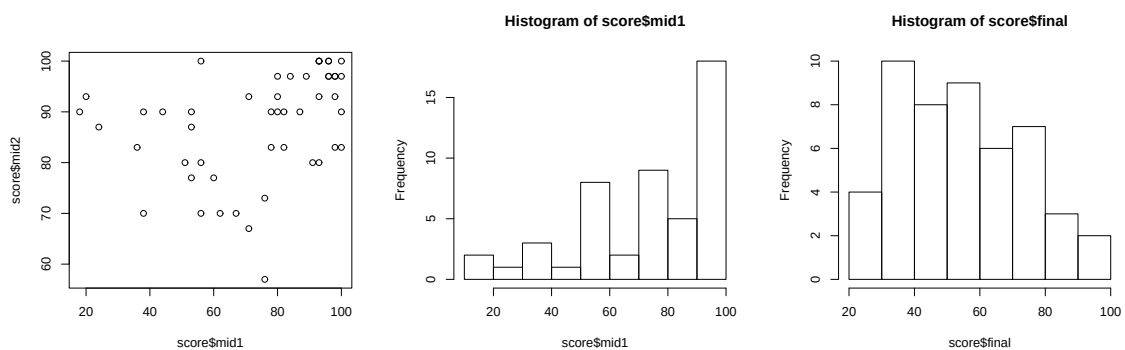


Figure 1: Left: scatter plot of midterm1 vs. midterm2, middle: histogram of the first midterm, right: histogram of the final exam score

3. The *sample mean*. is a summary measure of location.

```
> mean(score$mid1)
[1] 74.10204
> mean(score$mid2)
[1] 87.57143
> mean(score$final)
[1] 54.42857
```

4. The *sample variance* is a summary measure of variability.

```
> var(score$mid1)
[1] 552.6352
> sd(score$mid1)
[1] 23.50819
> sd(score$mid2)
[1] 10.56330
> sd(score$final)
[1] 19.04818
```

5. The *sample median* is obtained by ordering from data from smallest to largest and taking the middle value.

```
> median(score$final)
[1] 54
> boxplot(score$mid1, score$mid2, score$hw, score$final)
```

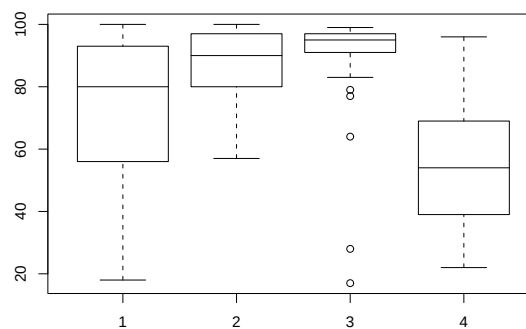


Figure 2: Boxplot of midterm, homework and final exam scores.

6. *Importing example and exercise data.* In order to load Example 1.5 data into R, follow the following command lines. Data is the percentage of binge drinkers on 140 campuses across the United States.

```
> library(Devore6)
> data(xmp01.05)
> attach(xmp01.05)
> xmp01.05
  bingePct
1         4
2        11
.
.
.
.
140       68
```

To load Exercise 1.14 data, use command `>data(ex01.14)`.

Assigned Problems Exercise 1.24. Exercise 1.56. Use R to generate figures. Click R figures with the right mouse button to save and print them.