MAT 2377, Probability and statistics for engineers

Assignment 4

Deadline : Before 3 pm on Friday, March 20

Submit the assignment in the drop box at 585 King Edward

UNless otherwise stated, solve the following exercises with a TI-30, TI-34, Casio FX-260 or Casio FX-300 calculator.

1. (based on 4.12) Compute the following, given the following sample data:

$$13 \;\; 49 \;\; 36 \;\; 3 \;\; 6 \;\; 38 \;\; 38 \;\; 30 \;\; 8 \;\; 40 \;\; 31 \;\; 5 \;\; 36:$$

(a) the sample mean
(b) The median
(c) The first and third quartiles (i.e. 25th and 75 percentile)

Solution:(a) There are 13 numbers here, so the sample mean is the sum of the numbers divided by 13.

$$(13+49+36+3+6+38+38+30+8+40+31+5+36)/13 = 25.61538.$$

(b) First we put the numbers in order:

$$3 \;\; 5 \;\; 6 \;\; 8 \;\; 13 \;\; 30 \;\; 31 \;\; 36 \;\; 36 \;\; 38 \;\; 38 \;\; 40 \;\; 49.$$

The middle number is then the median. More precisely, we take item number $(13+1)/2 = 7$ form the list, which is 31. (If the sample size were 13, then since $(14+1)/2 = 7.5$ is not an integer, we'd take the numeric average of item numbers 7 and 8.) (c) For the first quartile, we look at the numbers before the seventh (7 was the location of the median, calculated in part (b)), in our sorted list, in other words, 3 5 6 8 13 30. The median of this list is the first quartile. Here there are 6 numbers, and $(6+1)/2 = 3.5$ is not an integer, so the first quartile is the numeric average of item numbers 3 and 4 on the list, i.e. $(6+8)/2 = 7$. Similarly, the third quartile is $(38+38)/2 = 38$.

2. (based on 4.16) The heights of 1000 students are approximately normally distributed with a mean of 174.5 centimeters and a standard deviation of 6.9 centimeters. Suppose random samples of size 25 are drawn from this population. Determine:
(a) the mean and standard deviation of the sampling distribution of $\overline{X}$;
(b) the probability that the sample mean falls between 172.5 and 175.8 centimeters inclusive.
(c) If we repeated this sampling 200 times (each sample with 25 students), for how many samples would we expect the sample mean to be below 172.0 centimeters? (I.e. What is the mathematical expectation for this number?)

Solution: (a) The mean of $\overline{X}$ is just $\mu = 174.5$. The standard deviation is

$$\frac{\sigma}{\sqrt{n}} = \frac{6.9}{\sqrt{25}} = 1.38.$$

(b) To calculate the probability, we find the normalized (Z) values, since they will follow a standard normal distribution. Recall that

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}},$$

and the inequality $172.5 \leq \overline{X} \leq 175.8$ is equivalent to

$$\frac{172.5 - \mu}{\sigma/\sqrt{n}} \leq Z \leq \frac{175.8 - \mu}{\sigma/\sqrt{n}},$$

i.e.

$$\frac{172.5 - 174.5}{1.38} \leq Z \leq 175.8 - 174.51.38, \quad \text{or, more simply,} \quad -1.45 \leq Z \leq 0.94.$$

We use two-decimal-place accuracy because that's what our distribution table uses. To find the probability of this inequality, we must express it in terms of cumulative probabilities, as that's what that table gives.

$$P(-1.45 \leq Z \leq 0.94) = P(Z \leq 0.94) - P(Z < -1.45) = 0.8264 - 0.0735 = 0.7529.$$

In English, the probability is 75.29%.
(c) The probability that $\overline{X}$ is less than 172 is $P(Z \leq \frac{172-\mu}{\sigma/\sqrt{n}}) = P(Z < -1.81) = 0.0351$. If we repeat this 200 times and count the number of times it occurs, this is a binomial random variable with $p = 0.0351$ and $n = 200$, and the expected value is $np = 7.02$. (We'd expect about 7 samples to have a mean that low.)

3. (4.26) Two alloys A and B are being used to manufacture a certain steel product. An experiment needs to be designed to compare the two in terms of maximum load capacity in tons (the maximum weight that can be tolerated without breaking). It is known that the two standard deviations in load capacity are equal at 5 tons each. An experiment is conducted in which 30 specimens of each alloy (A and B) are tested and the results recorded as follows:

$$\bar{x}_A = 49.5, \quad \bar{x}_B = 45.5; \quad \bar{x}_A - \bar{x}_B = 4.$$

The manufacturers of alloy A are convinced that this evidence shows conclusively that $\mu_A > \mu_B$ and strongly supports the claim that their alloy is superior. Manufacturers of alloy B claim that the experiment could easily have given $\bar{x}_A - \bar{x}_B = 4$ even if the

two population means are equal. In other words, "the results are inconclusive!"
(a) Make an argument that manufacturers of alloy B are wrong. Do it by computing

$$P(\bar{X}_A - \bar{X}_B \geq 4 \mid \mu_A = \mu_B).$$

(b) Do you think these data strongly support alloy A?

Solution: (a) We use the formula for the difference of two means. Suppose that $\mu_A = \mu_B$. Then the formula for Z is

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{(\sigma_A^2/n_A) + (\sigma_B^2/n_B)}}$$

(and Z follows a standard normal distribution, approximately). Here we are told $n_A = n_B = 30$, $\sigma_A = \sigma_B = 5$ and $\mu_A = \mu_B$ (value unknown but irrelevant).

$$\begin{aligned} P(\bar{X}_A - \bar{X}_B \geq 4) &= P(Z \geq \frac{4 - (\mu_A - \mu_B)}{\sqrt{(\sigma_A^2/n_A) + (\sigma_B^2/n_B)}}) \\ &= P(Z \geq \frac{4}{\sqrt{(5^2)/30 + (5^2)/30}}) \\ &= P(Z \geq 3.10) = 1 - P(Z \leq 3.10) = 1 - 0.9990 = 0.001. \end{aligned}$$

(Again, we rounded the z value to two decimals because the table only takes two decimals.)
(b) It would be extremely unlikely to get the measurements we got, if the means were equal. So the data strongly supports alloy A being stronger.

4. The amount of cereal packed into a cereal box is supposed to be 500g. Either too high or too low a population mean is considered bad (either the company is wasting money by giving away free product, or customers will complain about being shortchanged), so the t-value is required to be in the middle 95%. A sample of 20 boxes is taken, and the sample mean is 480 and sample standard deviation 2.23.
(a) What is the t-value for our sample?
(b) What is the range of t-values that the company considers acceptable?
(c) Will the company find this t-value acceptable?

Solution: (a) It is

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{480 - 500}{2.23/\sqrt{20}} = -40.1088.$$

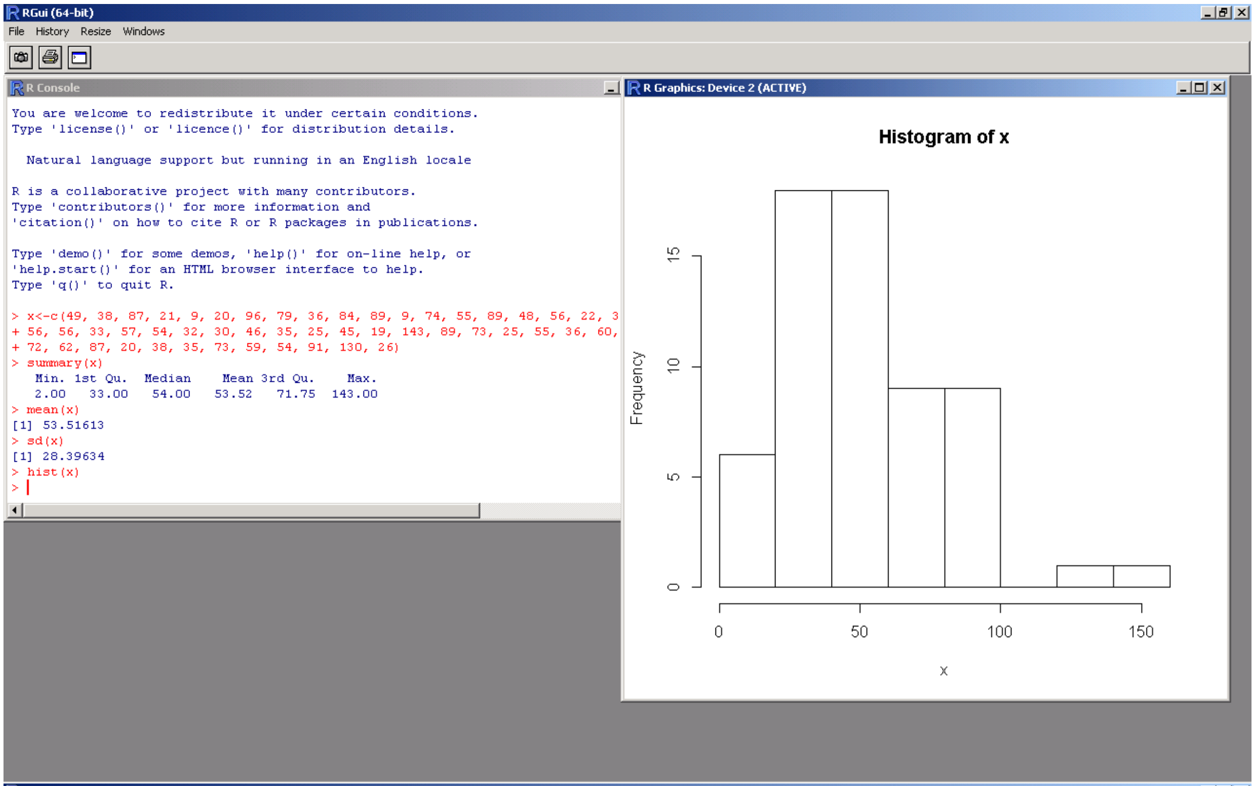
(b) Here $n = 20$, so we have $v = n - 1 = 19$ degrees of freedom. The α value is $1 - 0.95 = 0.05$, but we are ruling out both tails and hence will use $t_{\alpha/2} = t_{0.025}$. Reading along line row $v = 19$, we see that $t_{0.025} = 2.093$. The acceptable range will be from -2.093 to 2.093.
(c)The calculated t value is not in the acceptable range, so the company will not find it acceptable. They will be quite convinced that the population mean is lower than it is supposed to be.

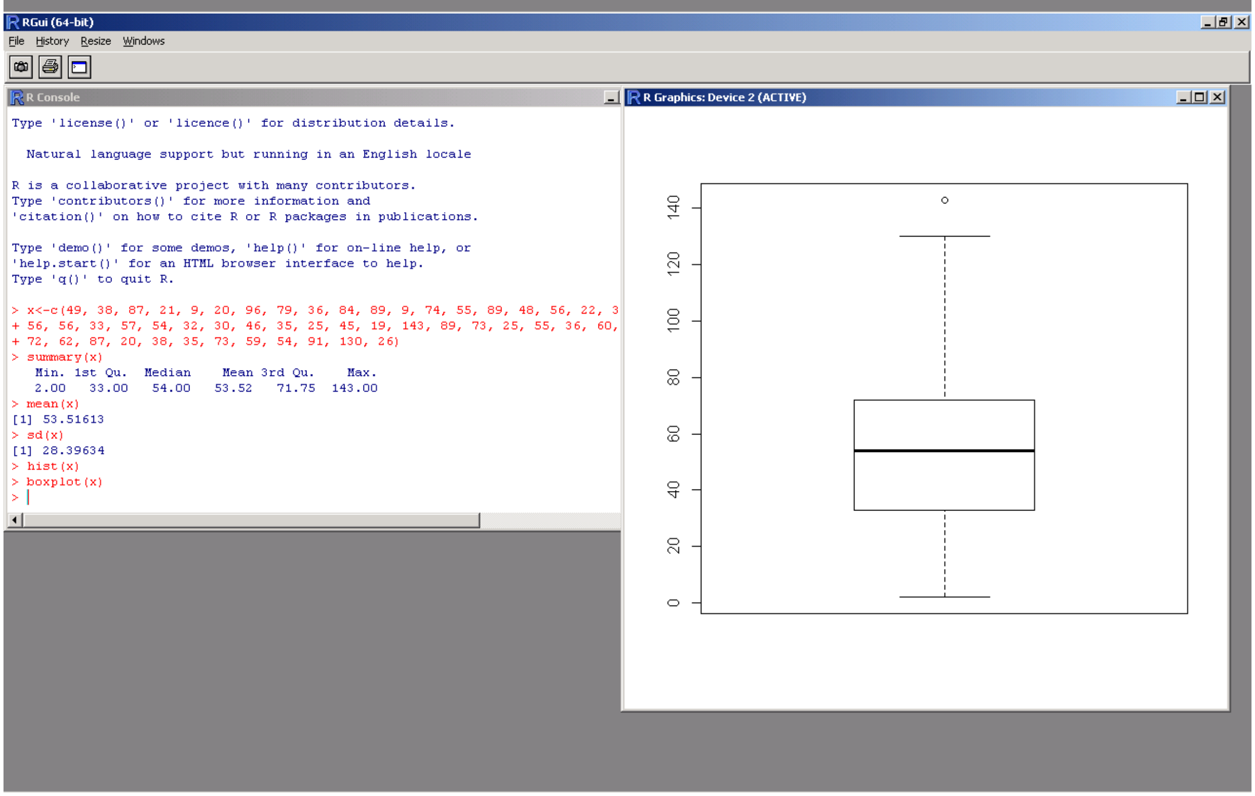
5. (R problem) (a) Input data into R using the command

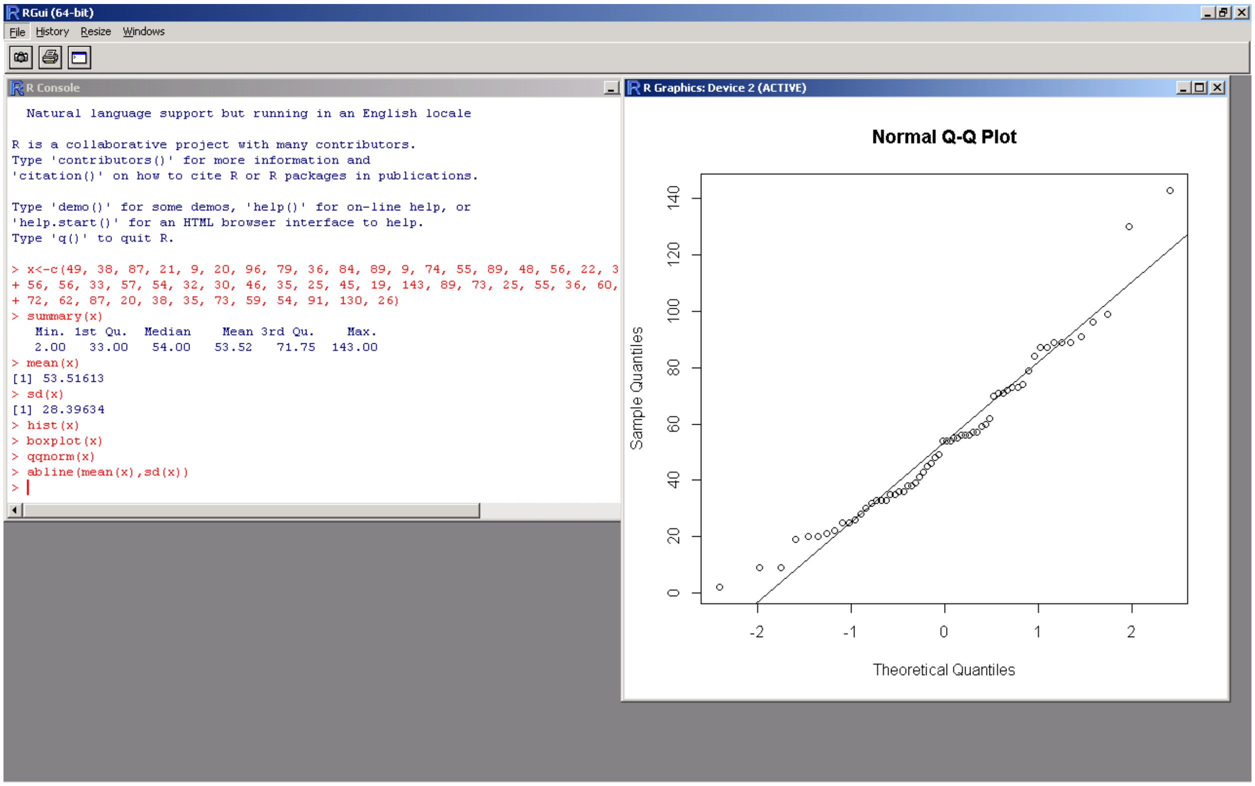
```
x<-c(49, 38, 87, 21, 9, 20, 96, 79, 36, 84, 89, 9, 74, 55, 89, 48, 56, 22,
39, 99, 71, 41, 28, 70, 54, 56, 56, 33, 57, 54, 32, 30, 46, 35, 25, 45, 19,
143, 89, 73, 25, 55, 36, 60, 71, 43, 33, 57, 33, 2, 72, 62, 87, 20, 38, 35,
73, 59, 54, 91, 130, 26)
```

(b) Use the summary, mean, and sd functions to determine the quartiles, mean and standard deviation of this sample. Write down the results.
(c) Use hist to produce a histogram of x, and boxplot to produce a box plot of x. Print out the result.
(d) Use qqnorm to produce a QQ plot of x. Print out the result. Does the plot suggest that x is normal? Why or why not?

Solution: See the attached pages. The QQ plot of any distribution will go from the bottom left corner to the top right corner of the box, and will always be increasing. The QQ plot of a normal distribution will be a straight line, and the QQ plot of an approximately normal distribution will have only slight bobbles away from a straight line. The command `abline(mean(x),sd(x))` draws an appropriate reference line. We see that the both ends of the QQ plot are well to the left of the line. This means that our data seems to have a larger number of extreme values at both ends than a sampling from a normal distribution would, i.e. the true population distribution (from which x is a sample) has fatter tails than a normal distribution would.


RGui (64-bit)
File History Resize Windows
R Console
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
> x<-c(49, 38, 87, 21, 9, 20, 96, 79, 36, 84, 89, 9, 74, 55, 89, 48, 56, 22, 3
+ 56, 56, 33, 57, 54, 32, 30, 46, 35, 25, 45, 19, 143, 89, 73, 25, 55, 36, 60,
+ 72, 62, 87, 20, 38, 35, 73, 59, 54, 91, 130, 26)
> summary(x)
Min. 1st Qu. Median Mean 3rd Qu. Max.
2.00 33.00 54.00 53.52 71.75 143.00
> mean(x)
[1] 53.51613
> sd(x)
[1] 28.39634
> hist(x)
>
R Graphics: Device 2 (ACTIVE)
Histogram of x
Frequency
0
5
10
15
0
50
100
150
x


RGui (64-bit)
File History Resize Windows
R Console
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
> x<-c(49, 38, 87, 21, 9, 20, 96, 79, 36, 84, 89, 9, 74, 55, 89, 48, 56, 22, 3
+ 56, 56, 33, 57, 54, 32, 30, 46, 35, 25, 45, 19, 143, 89, 73, 25, 55, 36, 60,
+ 72, 62, 87, 20, 38, 35, 73, 59, 54, 91, 130, 26)
> summary(x)
Min. 1st Qu. Median Mean 3rd Qu. Max.
2.00 33.00 54.00 53.52 71.75 143.00
> mean(x)
[1] 53.51613
> sd(x)
[1] 28.39634
> hist(x)
> boxplot(x)
>
R Graphics: Device 2 (ACTIVE)
0
20
40
60
80
100
120
140


RGui (64-bit)
File History Resize Windows
R Console
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
> x<-c(49, 38, 87, 21, 9, 20, 96, 79, 36, 84, 89, 9, 74, 55, 89, 48, 56, 22, 3
+ 56, 56, 33, 57, 54, 32, 30, 46, 35, 25, 45, 19, 143, 89, 73, 25, 55, 36, 60,
+ 72, 62, 87, 20, 38, 35, 73, 59, 54, 91, 130, 26)
> summary(x)
Min. 1st Qu. Median Mean 3rd Qu. Max.
2.00 33.00 54.00 53.52 71.75 143.00
> mean(x)
[1] 53.51613
> sd(x)
[1] 28.39634
> hist(x)
> boxplot(x)
> qqnorm(x)
> abline(mean(x),sd(x))
>
R Graphics: Device 2 (ACTIVE)
Normal Q-Q Plot
Sample Quantiles
0
20
40
60
80
100
120
140
-2
-1
0
1
2
Theoretical Quantiles