Hong Soon Bin 2009312066

Assignment #2

1. (5 pts) The data below are the number of points scored in 30 games by the Portland Trailblazers.

Scores: 90,95,89,71,73,96,87,95 107,89,96,80,97,95 102,97,93 101,82,83,74,91,83,98,95 111,99 120,93,84

a. Estimate the sample mean score. What does the quantity estimate?
b. Is the estimate in part(a) likely to equal the population parameter? Why or why not?
c. Calculate the standard error for your sample estimate.
d. What does the quantity in part(c) measure?
e. Calculate a 95% confidence interval for the population mean.
f. Provide an interpretation for the interval you calculated in part (e).

a.

```
> mean(Score)
[1] 92.2
```

b.

```
> t.test(Score)

        One Sample t-test

data:  Score
t = 46.647, df = 29, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 88.15754 96.24246
sample estimates:
mean of x
     92.2
```

Since p-value is lesser than 0.05, we could reject H0(=sample mean is equal to population mean).

c.

```
> SE <- sd(Score)/sqrt(length(Score))
> se
Error: object 'se' not found
> SE
[1] 1.976529
```

d. The standard error is the standard deviation of the sampling distribution of statistic, most commonly of the mean.

e.

```
> t.test(Score)

        One Sample t-test

data:  Score
t = 46.647, df = 29, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 88.15754 96.24246
sample estimates:
mean of x
     92.2
```

(88.15754, 96.24246)

2. (5 pts) Using the following data, test the null hypothesis that male and females have the same mean cholesterol concentrations. Include descriptive statistics, hypothesis testing (e.g., t-test) and 95% confidence intervals.

Male: 220.1, 218.6, 229.6, 228.8,222.0,224.1, 226.5
Female: 223.4,221.5,230.2,224.3,223.8,230.8

```
        Two Sample t-test

data:  Male and Female
t = -0.96415, df = 11, p-value = 0.3557
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -13.115635   5.125159
sample estimates:
mean of x mean of y
 221.6714  225.6667

> mean(Male)
[1] 221.6714
> sd(Male)
[1] 9.446995
> mean(Female)
[1] 225.6667
> sd(Female)
[1] 3.866609
```

3. (5 pts) A clinical trail was carried out to test whether a new treatment has an effect on the rate of recovery of patients. The null hypothesis "H_0: the treatment has no effect" was rejected with a P-value of 0.04. The researchers used a significance level of 5%. State whether the following conclusions is correct. If not, explain why.

a. The treatment has only a small effect.
b. The treatment has some effect.
c. The probability of committing a Type I error is 0.04.
d. The probability of committing a Type II error is 0.04.
e. The null hypothesis would not have been rejected if the significance level was α=0.01.

a. False- We do not have any idea of the size of effect of the treatment from this test. We only have information whether this treatment has effect.
b. True- We can reject the null hypothesis in a significance level of 5%, so this treatment has some effect.
c. False- This is a false sentence because we have a danger of reject in probability of size of significance level.
d. False- This is not correct. We cannot derive the Type 2 error from the test.
e. True- The null hypothesis could not reject in a significance level of 1% cause the p-value is 0.04

4. (5 pts) The data below are volumes of red blood cells from two individuals. Test the hypothesis (using the Mann-Whitney test) that the red blood cells of person B are 1.5 times the volume of person A.

person A: 248, 236, 269, 254, 249, 251, 260, 245, 239, 255
person B: 380, 391, 377, 392, 398, 374

```
> pa=c(248,236,269,254,249,251,260,245,239,255)
> pb=c(380,391,377,392,398,374)
> pa2 <- 1.5*pa
> wilcox.test(1.5*pa,pb,alter="two.sided")

        Wilcoxon rank sum test

data:  1.5 * pa and pb
W = 16, p-value = 0.1471
alternative hypothesis: true location shift is not equal to 0
```

5. (5 pts) What is the difference between the standard error of mean and the standard deviation? Provide example data that illustrates their difference.

SE shows how close the sampling mean to population mean.
SD is the information about the dispersion of the sample data.