Advanced Data Analysis HW1

Ao Liu, al3472

1.

Let η denote the median of a random variable X. Consider testing $H_0 : \eta = 0$ against $H_a : \eta \neq 0$ using $X_1, X_2, ..., X_{25}$, a random sample of size $n = 25$ from the distribution of X
(a) Let S denote the sign statistic. Determine the level of the test that rejects H_0 if $S \geq 16$.

Answer:
Let the sign test $S = \sum_{i=1}^{25} I(X_i \geq 0)$, then $S \sim Bin(25, \frac{1}{2})$ If under certain level the test will reject H_0 if $S \geq 16$, then according to symmetry, the test will also reject H_0 if $S \leq 9$.
So the level of the test is:

$$p(S \geq 16 or S \leq 9) = 0.2295$$

(b)
Determine the power of the teat in (a) if X has $N(0.5, 1)$ distribution?

Answer:
If $X \sim N(0.5, 1)$, then $P(X \geq 0) = 0.6915$, so $S \sim Bin(25, 0.6915)$. So the power of the test is:

$$1 - p(9 < S < 16 | H_\alpha) = 0.7842$$

2.

The data in in the talble below gives the pretest and posttest scores on the MLA listening test in Spanish for 20 high school teachers who attended an intensive course in Spanish.

subject	pretest	posttest	subject	pretest	posttest
1	30	20	11	30	32
2	28	30	12	29	22
3	31	32	13	31	34
4	26	30	14	29	32
5	20	16	15	34	32
6	30	25	16	20	27
7	34	31	17	26	28
8	15	18	18	25	29
9	28	33	19	31	32
10	20	25	20	29	32

Assume that the differences between these scores (pretest scores posttest) constitute a random sample from a distribution F with mean mu and variance σ^2

(a)
Use a t-test and $\alpha = 0.05$ to test $H_0 : \mu = 0$ against $H_\alpha : \mu = 0$. What is the p-value of the test? What assumption you need to make. Use a graphical technique to check this assumption.

Answer:

```
  pre <- c(30,28,31,26,20,30,34,15,28,20,30,29,31,29,34,20,26,25,31,29)
  post <- c(20,30,32,30,16,25,31,18,33,25,32,22,34,32,32,27,28,29,32,32)
  score <- pre-post
  t.test(score,mu=0)

```

and we get the following result:

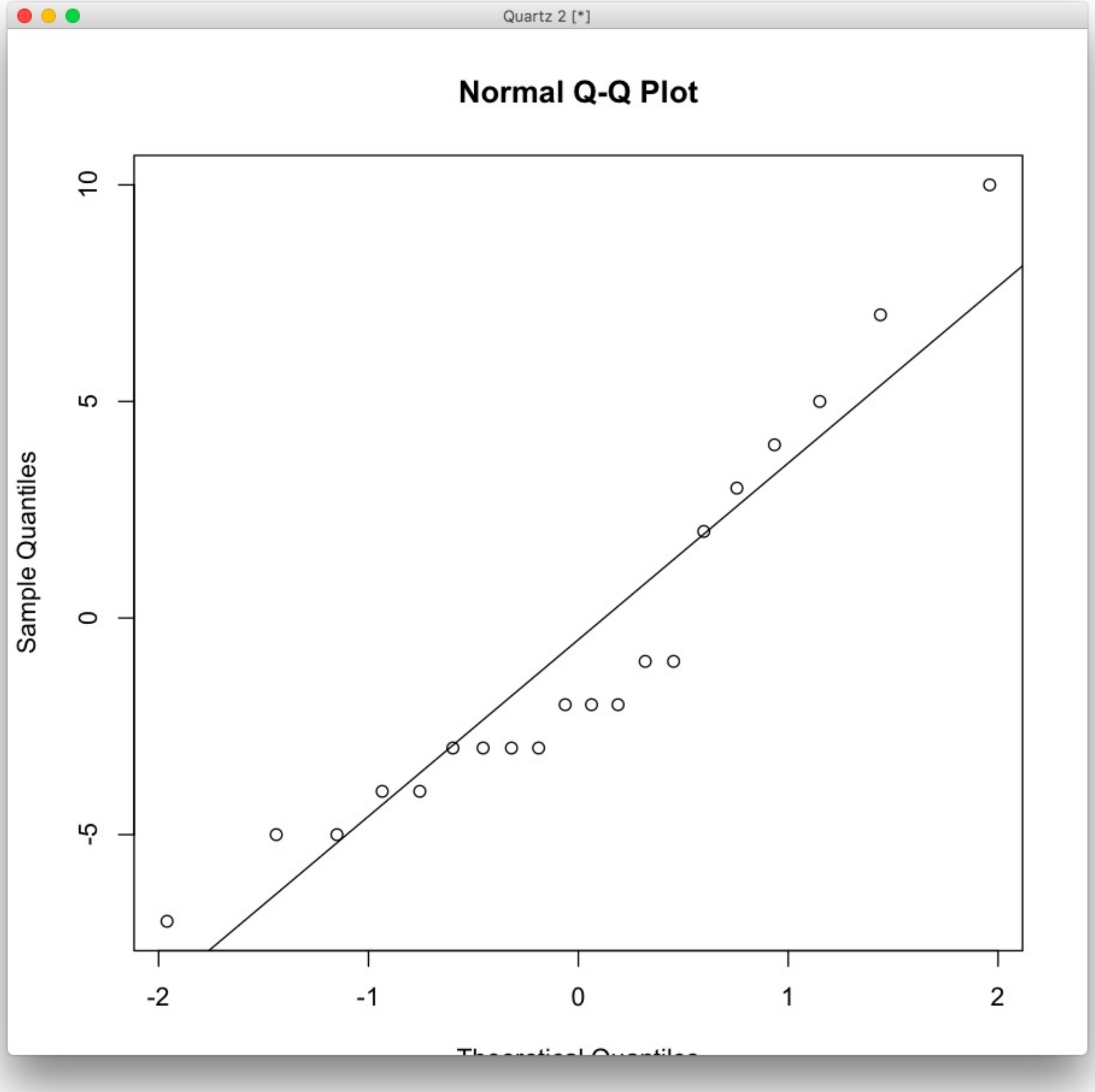
```
One Sample t-test

data:  score
t = -0.7054, df = 19, p-value = 0.4891
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -2.77699  1.37699
sample estimates:
mean of x
     -0.7
```

The p-value of the test is 0.4891

In order to make the t-test result more plausible, we have to make an assumption that the underlying dietribution is not extremely skewed. To check our assumption, we make a qq plot of the given data:

```
  qqnorm(y)
  qqline(y, col = 1)
```

The following is a way to insert pics:



As we can see from the figure, our assumption is met.

(b)

Obtain a 95% **confidence interval for the mean in (a)**

Answer:

From the result that we get from (a), the 95% confidence interval is:

$$(-2.77699, 1.37699)$$

(c)

If the median of F is η,use the sign test and α=0.05 to test $H_0 : \eta = 0$ against $H_\alpha : \eta \neq 0$. What is the p-value of this test?

Answer:

The values of ”score” are:

```
[1] 10 -2 -1 -4  4  5  3 -3 -5 -5 -2  7 -3 -3  2 -7 -2 -4 -1 -3
```

It turns out there are 6 positive socres in the sample.

So we do the following sign test:

```
    Exact binomial test

    data:  6 and 20
    number of successes = 6, number of trials = 20, p-value = 0.1153
    alternative hypothesis: true probability of success is not equal to 0.5
    95 percent confidence interval:
     0.1189316 0.5427892
    sample estimates:
    probability of success
                       0.3


```

The p-value of this test is 0.1153

(d)

Obtain a 95% confidence interval for η and compare use answer the answer in (b)

Answer:

According to the result in (c), the 95% confidence interval for η is:

$$(0.1189316, 0.5427892)$$

Compared with the confidence interval that we get in the t test above, this one is smaller, more accurate.

3.

Twelve one week old infants were randomly assigned into two groups of six infant each. One group participated in an experimental active-exercise to learn to walk and the other was used as a control group. The following are the ages at which these infants first walked alone

Active-exercise group	No-exercise group
9.00	11.50
9.50	12.00
9.75	9.00
10.00	11.50
13.00	13.25
9.50	13.00

Call the no-exercise group Y sample and the active-exercise group X sample. Compare the two groups using two tests (one parametric and one nonparametric) (take $= 0.05$). State all the assumptions that you make in carrying out these tests.

Answer:

(1) Non Parametric Test:
(Wilcoxon) Mann-Whitney two sample procedure:
We assume that the populations have the same shape and differ only in location.

```
  active <- c(9.00,9.50,9.75,10.00,13.00,9.50)
  no <- c(11.50,12,9,11.50,13.25,13)
  wilcox.test(active,no,correct = FALSE)
```

```
  Wilcoxon rank sum test with continuity correction

data:  active and no
W = 9, p-value = 0.1705
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(active, no) : cannot compute exact p-value with ties
```

Since p-value is 0.1705, we cannot reject the Null Hypothesis.

(2) Parametric Test: We use ANOVA F-Test to check whether the two population have the same mean.
Assume every population of interest has unknown population mean and variance.

```
level <- c(rep("active", 6), rep("no", 6))
age <- c(9, 9.5, 9.75, 10, 13, 9.5, 11.5, 12, 9, 11.5, 13.25, 13)
data <- data.frame(level, age)
summary(aov(age ~ level))
```

```
Df Sum Sq Mean Sq F value Pr(>F)
level         1   7.521    7.521    3.415 0.0943 .
Residuals    10  22.021    2.202
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1
[Finished in 0.497s]
```

So $SSB = 7.521$, $SSE = 22.021$, $MSB = 7.521$, $MSE = 2.2021$, $F = 3.415$
If we take $\alpha = 0.05$, we have

```
>qf(.95, df1=1, df2=10)
 [1] 4.964603
```

$F(1-0.05, 1, 10) = 4.965$
Since $3.415 > 4.965$, we cannot reject H_0

Also, p-value $= 0.0943 > 0.05$, we cannot reject H_0