Applied Statistics - Homework 6

Sai Nikhil
NUID: 001564864

T. Sai Nikhil

(1)

13.6.10

(a)

```
> data.bulimic <- read.table(file="daily_calories_bulimic.txt", header = TRUE)
> data.healthy <- read.table(file="daily_calories_healthy.txt", header = TRUE)
> data.bulimic
   bulimic
1     15.9
2     16.0
3     16.5
4     17.0
5     17.6
6     18.1
7     18.4
8     18.9
9     18.9
10    19.6
11    21.5
12    21.6
13    22.9
14    23.6
15    24.1
16    24.5
17    25.1
18    25.2
19    25.6
20    28.0
21    28.7
22    29.2
23    30.9
> data.healthy
   healthy
1     20.7
2     22.4
3     23.1
4     23.8
5     24.5
6     25.3
7     25.7
8     30.6
9     30.6
10    33.2
11    33.7
12    36.6
13    37.1
14    37.4
15    40.8
>
> wilcox.test(data.bulimic$bulimic, data.healthy$healthy, alternative="two.sided")

   Wilcoxon rank sum test with continuity correction

data:  data.bulimic$bulimic and data.healthy$healthy
W = 61.5, p-value = 0.0009651
alternative hypothesis: true location shift is not equal to 0
```

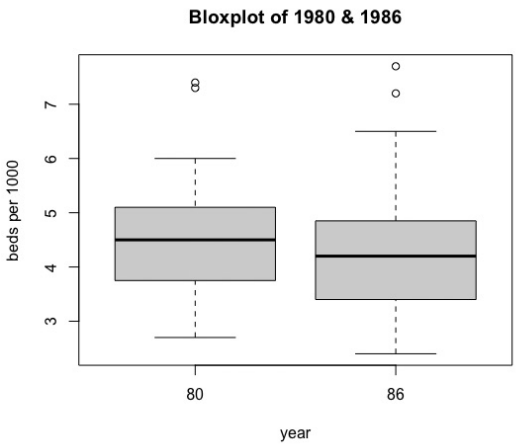
$< 0.05 \Rightarrow$ reject H_0

$\therefore$ median bulimic $\neq$ median healthy

(b) From the sample it looks like the bulimic patients take lower caloric intake than healthy adolescents. This is counterintuitive from the definition of bulimia.

13.6.13

```
> data <- read.table(file="bed2.txt", header=TRUE)
> boxplot(data$bed~data$year,
+         main="Bloxplot of 1980 & 1986",
+         xlab="year",
+         ylab="beds per 1000")
```



a

```
> data2 <- read.table("bed.txt", header=TRUE );
> x <- data2$bed80
> y <- data2$bed86
> wilcox.test (x, y, paired=TRUE, alternative="two.sided")

   Wilcoxon signed rank test with continuity correction

data:  x and y
V = 1196.5, p-value = 5.683e-07
alternative hypothesis: true location shift is not equal to 0
```

c

Same Conclusion

```
>
>
> wilcox.test (x, y, paired=FALSE, alternative="two.sided")

   Wilcoxon rank sum test with continuity correction

data:  x and y
W = 1576, p-value = 0.06544
alternative hypothesis: true location shift is not equal to 0
```

b

```
>
> t.test(data2$bed80, data2$bed86, paired=T)

   Paired t-test

data:  data2$bed80 and data2$bed86
t = 6.8721, df = 50, p-value = 9.513e-09
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.2289696 0.4180892
sample estimates:
mean of the differences
              0.3235294
```

e) $\therefore$ median of beds in 1980 $\neq$ median of beds in 1986

$\Rightarrow$ We reject H_0

d) In b) we accept H_0, where as in c) we reject it. So, Conclusion is different.

Assuming data in 1980 & 1986 is independent, we can say Rank-Sum is appropriate compared to Signed-Rank test.

13.6.16

a) The sample doesn't appear to be normally distributed. Hence, t-test is not appropriate.

b)

```
> data <- read.table(file="insure.txt", header = TRUE)
> uninsured_women <- subset(data, data$group == 0)
> insured_women <- subset(data, data$group == 1)
> wilcox.test(uninsured_women$stage, insured_women$stage)

   Wilcoxon rank sum test with continuity correction

data:  uninsured_women$stage and insured_women$stage
W = 28758, p-value = 1.496e-05
alternative hypothesis: true location shift is not equal to 0
```

$< \alpha = 0.05$

$\Rightarrow$ median of insured $\neq$ median of uninsured

c) From the data, it can be seen that median stage of uninsured data is less than median stage of insured data.

(2)

A: $m = 10$

B: $n = 9$

a) Table A.7

$$\Rightarrow p(W_0 = 36 \mid m = 10, n = 9) = 0.0000$$

$$p(W_0 = 38 \mid m = 10, n = 9) = 0.0001$$

b) Smallest p-value using permutation test

$$= \frac{1}{\binom{n+m}{n}}$$

$$= \frac{1}{\frac{19!}{9! \times 10!}}$$

$$= \frac{1}{92378}$$

$$\simeq 0.0000108$$