AMS 380.01: Problem Set 3

Due on 02/23

Wei Zhu

Harris Temuri

Problem 1

Dr.Oz has a crazy hypothesis. He hypothesized that the average weight of adult US men and US women are equal. A group of smart Stony Brook University students do not agree with this hypothesis, and they got the SBU IRB (https://www.stonybrook.edu/commcms/research-compliance/) approval to draw two independent random samples of 9 adult men and 9 adult women. The measured weight (in kg) is as the following:

ID	group	weight
1	Woman	38.9
2	Woman	61.2
3	Woman	73.3
4	Woman	21.8
5	Woman	63.4
6	Woman	64.6
7	Woman	48.4
8	Woman	48.8
9	Woman	48.5
10	Man	67.8
11	Man	60.0
12	Man	63.4
13	Man	76.0
14	Man	89.4
15	Man	73.3
16	Man	67.3
17	Man	61.3
18	Man	62.4

Solution

(a) Please visualize the two groups using the Box plot in R.

```
ggplot(data = data_1, aes(x = group, y = weight)) + geom_boxplot()
```

(b) Please test the normality and equal variance assumptions using R.

P-value of Shapiro test is greater than 0.05 for both man and woman, so data is normally distributed.

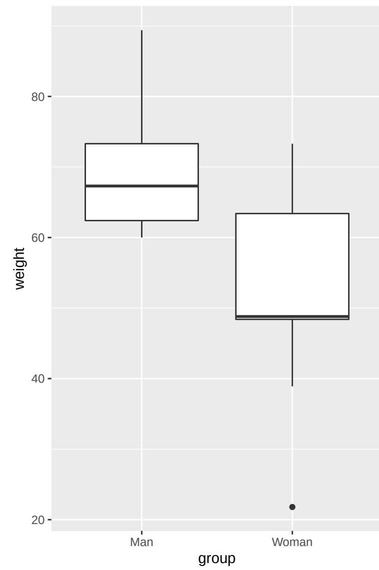

Figure 1: Box Plot of the Data_1

```
# Normality Test
# for male
shapiro.test(data_1$weight[data_1$group=="Man"])
# for female
shapiro.test(data_1$weight[data_1$group=="Woman"])
```

Variance is equal (p-value above 0.05)

```
# Equal variance assumptions test
var.test(weight ~ group, data = data_1)
```

(c) Please perform the correct test comparing whether the two means are equal or not using R. Please report the p-value of your test. Please also construct the 95% confidence interval for the mean difference using R.

```
# 1.3 2 sample t-test to test for equal means
t.test(data_1$weight[data_1$group=="Woman"], data_1$weight[data_1$group=="Man"],
    var.equal = TRUE)
```

The p-value of the t-test is 0.01327 which is less than 0.05.

The 95% confidence interval is $[-29.748019, -4.029759]$

(d) Please perform the correct test comparing whether the two means are equal or not by hand. Please make decision for you var.test(weight $\sim$ group, data = data_1) r hypothesis test at the significance level of $\alpha = 0.05$. Please also construct the 95% confidence interval for the mean difference by hand.

T-test:

$$t = \frac{m_A - m_B}{\sqrt{\frac{S^2}{n_A} + \frac{S^2}{n_B}}}$$

$$S^2 = \frac{\sum(x - m_A)^2 + \sum(s - m_B)^2}{n_A + n_B - 2}$$

$$S^2 = \frac{1946.06 + 703.1888889}{16} = 165.5780556$$

$$t = \frac{68.99 - 52.1}{\sqrt{165.5780556/9 + 165.5780556/9}}$$

$$t = 2.784235553, p \approx 0.015$$

Confidence Interval:

$$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$68.99 - 52.1 \pm t_{0.05/2} \sqrt{165.578} \sqrt{\frac{1}{9} + \frac{1}{9}}$$

$$16.89 \pm (2.120)(12.8677)(0.471404521)$$

$$16.89 \pm 12.86$$

$$[4.03, 29.75]$$

Problem 2

While most of the time we wish to reduce our weight, there are times when we wish to gain some weight. The MiraGro Co. has developed just such a weight gain formula and they tested it on 10 mice. The weight of the mice before and after the treatments are as follows (in unknown unit):

ID	group	weight
1	before	200.1
2	before	190.9
3	before	192.7
4	before	213.0
5	before	241.4
6	before	196.9
7	before	172.2
8	before	185.5
9	before	205.2
10	before	193.7
1	after	392.9
2	after	393.2
3	after	345.1
4	after	393.0
5	after	434.0
6	after	427.9
7	after	422.0
8	after	383.9
9	after	392.3
10	after	352.2

Solution

(a) Please visualize the two groups using the Box plot and the Profile plot in R.

```
# Box Plot
ggplot(data = data_2, aes(x = group, y = weight)) + geom_boxplot()
# Profile Plot
before <- subset(data_2, group=="before", weight, drop = TRUE)
after <- subset(data_2, group=="after", weight, drop = TRUE)
pd <- paired(before, after)
plot(pd, type="profile")
```

Problem 2 continued on next page. . .

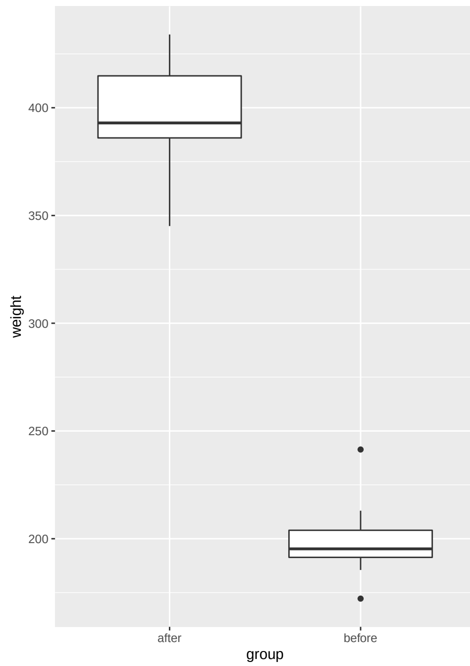

Figure 2: Box Plot of the Data_2

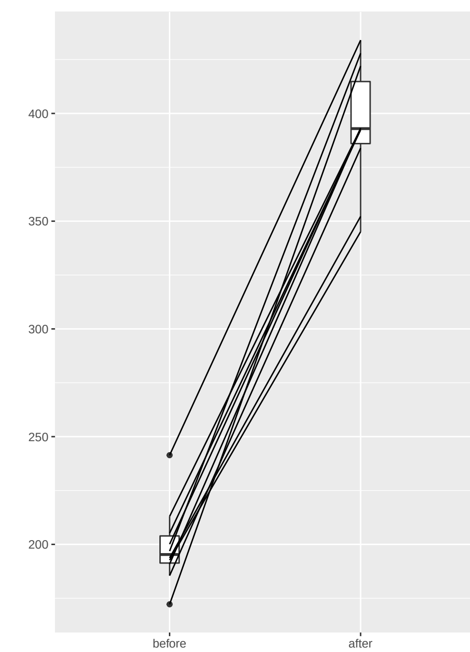

Figure 3: Profile Plot of the Data_2

Problem 2 continued on next page...

(b) Please test the normality assumption using R.

P-value of Shapiro test is greater than 0.05 so data is normally distributed.

```
    # Normality Test
shapiro.test(data_2$weight[data_2$group=="after"] - data_2$weight[data_2$group=="before
    "])
```

(c) Please perform the correct test examining whether the mean difference is zero or not using R. Please report the p-value of your test. Please also construct the 95% confidence interval for the mean difference using R.

```
    # Paired t-test to test for equal means
    t.test(before, after, paired=TRUE)
```

The p-value of the paired t-test is 6.2e-09 which is less than 0.05.

The 95% confidence interval is $[-215.5581, -173.4219]$

(d) Please perform the correct test examining whether the mean difference is zero or not by hand. Please make decision for your hypothesis test at the significance level of $\alpha = 0.05$. Please also construct the 95% confidence interval for the mean difference by hand.

Paired t-test:

$$t = \frac{m}{s/\sqrt{n}}$$

$$t = \frac{393.65 - 199.16}{29.45111/\sqrt{10}}$$

$$t = 20.883130793, p << 0.05$$

Confidence Interval:

$$\bar{x}_1 - \bar{x}_2 \pm t_{n-1,\alpha/2}(s_d/\sqrt{n})$$

$$194.49 \pm (2.262)(29.45111/\sqrt{9})$$

$$194.49 \pm 22.20613694$$

$$[172.2839, 216.69613694]$$

Problem 3

A true story goes as follows. Many years ago, when we were still using the typewriter to record data and write research reports, a group of lung cancer researchers hired a wonderful typist to enter the lung cancer patient data. The lady was a heavy smoker. In the midst of entering the patient data; however, she quit smoking once and for all. The following is a data set of smokers among lung cancer patients and healthy individuals. I personally found the data to be very unrealistic, however; it is only an example, and most likely a fake example.

Group A: lung cancer patients: n = 500, 490 smokers;

Group B: healthy individuals: n = 500, 400 smokers.

Solution

(a) Please perform the correct test examining whether the proportion of smokers are equal between Group A and Group B, using R. Please report the p-value of your test. Please also construct the 95% confidence interval for the mean difference in proportions using R.

```
prop.test(x = c(490,400), n = c(500,500))
```

The p-value of the paired t-test is 2.2e-16 which is less than 0.05.

The 95% confidence interval is $[0.1408536, 0.2191464]$

(b) Please perform the correct test examining whether the proportion of smokers are equal between Group A and Group B, by hand. Please make decision for your hypothesis test at the significance level of $\alpha = 0.05$. Please also construct the 95% confidence interval for the mean difference in proportions by hand.

Problem 3 continued on next page...

Paired t-test:

$$z = \frac{p_A - p_B}{\sqrt{pq/n_A + pq/n_B}}$$

$$z = \frac{0.98 - 0.80}{\sqrt{(0.89)(1-0.89)/500 + (0.89)(1-0.89)/500}}$$

$$z = \frac{0.18}{\sqrt{0.0003916}}$$

$$z = 9.096014909, p - value << 0.05$$

Confidence Interval:

$$p_A - p_B \pm z_{\alpha/2}(SE)$$

$$SE = \sqrt{\frac{p_A(s_d)}{n_A} + \frac{p_B(s_d)}{n_B}}$$

$$SE = \sqrt{\frac{0.98(0.02)}{500} + \frac{0.80(0.20)}{500}}$$

$$SE = \sqrt{0.0000392 + 0.00032}$$

$$0.18 \pm 1.96(0.018952572)$$

$$[0.142852959, 0.217147042]$$