Name: Drashti Mehta

Lab Assignment 7

(40 points)

DIRECTIONS

- ***An uploaded copy is due by next Friday.***

For the following questions, determine whether to use a parametric or nonparametric test (look at assumptions or test for normality). Perform the hypothesis testing by hand and list the test statistics, critical values, R commands and the conclusions. Make conclusions using both the p-value and test statistic. ATTACH YOUR HAND WORK TO THE ASSIGNMENT OR YOU WILL NOT BE GIVEN CREDIT. Where needed perform tests at α=0.05.

A) The zika virus has been linked to severe neurological diseases and can cause microcephaly, reducing the brain size of a baby. Ultrasounds have measured the brain size (cm) of 3-week-old babies in the womb before and after infection of zika. Assume the data is normally distributed.

Before zika: 12.9, 13.5, 12.8, 15.6, 17.2, 19.2, 12.6, 15.3, 14.4, 11.3
After zika: 12.7, 11.6, 10, 15.2, 16.8, 20, 12, 15.5, 15, 10.9

Is there a decrease in brain size after zika infection? (9 points)

A → Hypothesis:

H_0 = difference in means ≤ 0, H_a = ~~μ~~ $\mu_1 > \mu_2$

→ $n_1 = n_2 = n = 10$

$d_1 = 0.2, d_2 = 1.9, d_3 = 2.8, d_4 = 0.4, d_5 = 0.4,$

$d_6 = -0.8, d_7 = 0.6, d_8 = -0.2, d_9 = -0.6, d_{10} = 0.4$

mean of differences ($\bar{d}$) = 0.51 ; sd = 1.094

→ Test statistics = $\dfrac{0.51}{1.094 \times \sqrt{10}}$ = 1.474

→ df = 9; critical value : t(9, 0.05) = 1.833

R code:

```
bz <- c(12.9, 13.5, 12.8, 15.6, 17.2, 19.2, 12.6, 15.3, 14.4, 11.3)
az <- c(12.7, 11.6, 10, 15.2, 16.8, 20, 12, 15.5, 15, 10.9)
```

> qt(0.95,9) = 1.833
> t.test(bz,az,alternative = "greater", var.equal =TRUE, paired = TRUE)
> Results:
t = 1.4744, df = 9, p-value = 0.08724
alternative hypothesis: true difference in means is greater than 0. 95 percent confidence interval: (-0.1240969, Inf)

The p-value was greater than 0.05, therefore the averages of the two groups were concluded to be similar. We do not have enough evidence to reject the null hypothesis. Also, our t-test statistic was less than the critical t-value for 9 degrees of freedom.

B) You are asked to test whether pollution levels in a nearby pond decrease after stopping industrial waste dumping. You cannot assume a Gaussian distribution and the data is not normally distributed. Here are the values of pollution at each site:

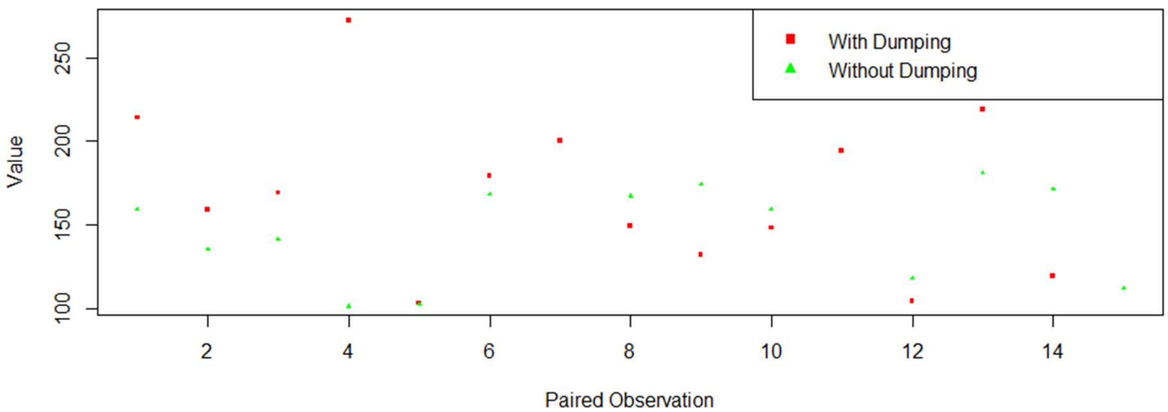
During dumping: 214, 159, 169, 272, 103, 179, 200, 149, 132, 148, 194, 104, 219, 119, 234
After dumping stops: 159, 135, 141, 101, 102, 168, 62, 167, 174, 159, 66, 118, 181, 171, 112

Does the level of pollution in the pond decrease after waste dumping stops?
Plot both the with dumping and without dumping values in different colors and symbols on the same plot. Add a legend. Label both axis. (13 points)

B. Hypothesis:
- H_0 = population (1) ≤ population (2)
  H_a = population (1) > population (2)
- $A_1 - B_1 = 55$, $A_2 - B_2 = 24$, $A_3 - B_3 = 28$, $A_4 - B_4 = 171$,
  $A_5 - B_5 = 1$, $A_6 - B_6 = 11$, $A_7 - B_7 = 138$, $A_8 - B_8 = -18$,
  $A_9 - B_9 = -42$, $A_{10} - B_{10} = -11$, $A_{11} - B_{11} = 128$,
  $A_{12} - B_{12} = -14$, $A_{13} - B_{13} = 38$, $A_{14} - B_{14} = -52$,
  $A_{15} - B_{15} = 122$.
- $T+ = 1 + 2.5 + 6 + 7 + 8 + 11 + 12 + 13 + 14 + 15 = 89.5$
  $T- = 2.5 + 4 + 5 + 9 + 10 = 30.5$, $n = 15$
- T- is smaller value and 30.5 > 30, the critical value for n=15, for one tailed test.
- We have evidence to reject null hypothesis & thus conclude that level of pollution decreases after waste dumping is stopped.

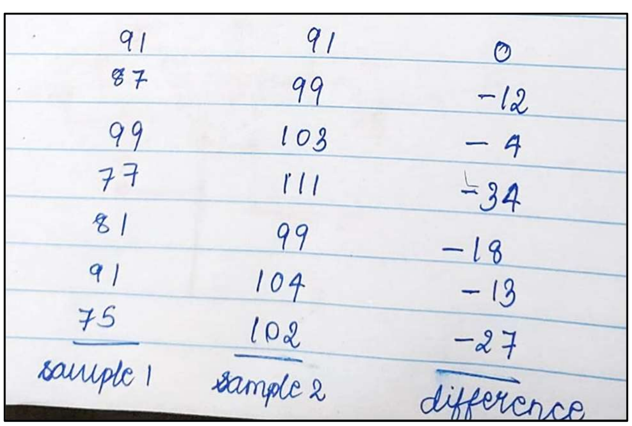
In R:

```
> g1<-c(214, 159, 169, 272, 103, 179, 200, 149, 132, 148, 194, 104, 219, 119, 234)
>g2<-c(159, 135, 141, 101, 102, 168, 62, 167, 174, 159, 66, 118, 181, 171, 112)
> wilcox.test(g1,g2, paired=TRUE, alternative="greater")
> Result:
V = 89.5, p-value = 0.04974 alternative hypothesis: true location shift is greater than 0
>plot(g1, pch = 15, col = "red", cex = 0.5, ylab = "Value", xlab = "Paire d Observation")
>points(g2,pch=17,col="green",cex=0.5)
>legend("topright",legend=c ("With Dumping" ,"Without Dumping"), col = c("red","green"), pch=c(15,17))
```



C) 6 subjects were given a drug to dull pain. Their reaction time to a shock stimulus was measured (in ms). Assume the distribution is not normal.

Prior to treatment: 91, 87, 99, 77, 81, 91, 75
After Treatment: 91, 99, 103, 111, 99, 104, 102
Did the drug increase the resistance to pain?



sample 1	sample 2	difference
91	91	0
87	99	-12
99	103	-4
77	111	-34
81	99	-18
91	104	-13
75	102	-27

sample 1: $\bar{x}_1$ mean = 85.857, sd = 9.63
sample2: $\bar{x}_2$ mean = 101.286, sd = 6.075
difference: mean = -15.429, sd = 12.053.
$n = 7$.
sample means are $\bar{x}_1$ & $\bar{x}_2$ for sample 1&2 resp

→ Hypothesis:
H_0: population 1 ≥ population 2.
H_a: population 1 < population 2

- Absolute values of test: 12, 4, 34, 18, 13, 27.
- Ranks (in order): 2, 1, 6, 4, 3, 5.
- sum of ranks: 21, $n = 6$.
- $T- = 21$, $T+ = 0$.
- We take smaller value $(T+)$ 0 < critical value for $n = 6$ (2). Therefore we have enough evidence to reject null hypothesis.

In R:
> g1<-c(91, 87, 99, 77, 81, 91, 75); >g2<-c(91, 99, 103, 111, 99, 104, 102)
> wilcox.test(g1,g2, paired=TRUE, alternative="greater")
Results:
V = 0, p-value = 0.9895 alternative hypothesis: true location shift is greater than 0
We can conclude that the drug did increase resistance to pain.

D) You are asked to compare the average wing size (in cm) of two groups of *Drosophila melanogaster*. The first group consists of *Drosophila* from brazil; the second group is North America *Drosophila*. The data are given below, (assume the variances of the groups are equal and the distribution is normal):

Brazilian = 18, 22, 21, 17, 20, 17, 23, 20, 22 ,21
North America = 16, 17, 14, 21, 19, 18, 13, 15, 17, 21
Is the wing size equal in both groups?

D. Hypothesis :

H_0 : difference in means = 0

H_a : difference in means ≠ 0

$\bar{x}_1$: 20.1 $\bar{x}_2$: 17.1

sd_1 : 2.1318 sd_2 : 2.7264

n_1 : 10 n_2 : 10

→ Rejection region :

$\alpha = 0.05$

$d_f = (n_1 + n_2 - 2) = 18$

critical value at $\alpha = 0.05$ & $d_f = 18$ is

$t_c = \underline{\underline{2.1009}}$.

$R = t : |t| > 2.1009$.

→ Test statistics :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{sp(1/n_1 + 1/n_2)}}$$

→ $$sp^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2} = \underline{\underline{5.98891}}.$$

$t = \underline{\underline{2.741}}$.

- $t = 2.741 > t_c = 2.101$, ∴ H_0 is rejected
- p-value = 0.0134, p value < $\alpha/2 = 0.025$
- Hence we reject H_0

In R:

```
> a<-c(18, 22, 21, 17, 20, 17, 23, 20, 22, 21)
> b<-c(16, 17, 14, 21, 19, 18, 13, 15, 17, 21),
>qt(0.95,18) = 1.734064
> t.test(a,b,var.equal=TRUE)
```

Results:

t = 2.7412, df = 18, p-value = 0.01342 alternative hypothesis: true difference in means is not equal to 0. 95 percent confidence interval: (0.7006872, 5.2993128)

Our p-value was less than 0.05, therefore the average wing size of the two groups are not equal.