Name: Drashti Mehta

Lab Assignment 9
(40 points)

DIRECTIONS
- *Be sure to show all of your work by providing your hypotheses, R commands, solutions/output, and interpretation.*
- *Please circle or highlight your final answers.*
- *This lab is due on Friday after assignment.*

By HAND means to calculate the test statistic by hand and find the critical value using the tables.   R means to use R functions and commands to calculate the test statistic, p-value, and critical value.

A. **The average time (in seconds) of mating for males and female grasshoppers were determined under four treatments. Assume the data are normally distributed. Test for equal variance using R only (show commands and conclusion). Then determine if the means are different by performing hypothesis test by hand and in R. Make a conclusion using alpha=0.05 with the test-statistic and p-value (15 points).**

Treatment A	Treatment B	Treatment C	Treatment D
9	30	3900	5
3	30	10800	9
10	480	28900	20
200	900	3600	180
1	2	200	15
2	1	120	20
21	5400	500	2
720	1500	600	17
1500	480	1980	30
60	3	160	8

H0 = All the mu_i are equal
Ha = Not all mu_i are equal

R code:
A:

```
>treatments<-
c(9,3,10,200,1,2,21,720,1500,60,30,30,480,900,2,1,5400,1500,480,3,3900,10800,28900,3600,2
00,120,500,600,1980,160,5,9,20,180,15,20,2,17,30,8)
>                                                                                    groups<-
c(rep("treatment_A",10),rep("treatment_B",10),rep("treatment_C",10),rep("treatment_D", 10))
> grasshoppers<-data.frame(treatments,groups)
> grasshoppers
     treatments    groups
1          9 treatment_A
2          3 treatment_A
3         10 treatment_A
4        200 treatment_A
5          1 treatment_A
6          2 treatment_A
7         21 treatment_A
8        720 treatment_A
9       1500 treatment_A
10        60 treatment_A
11        30 treatment_B
12        30 treatment_B
13       480 treatment_B
14       900 treatment_B
15         2 treatment_B
16         1 treatment_B
17      5400 treatment_B
18      1500 treatment_B
19       480 treatment_B
20         3 treatment_B
21      3900 treatment_C
22     10800 treatment_C
23     28900 treatment_C
24      3600 treatment_C
25       200 treatment_C
26       120 treatment_C
27       500 treatment_C
28       600 treatment_C
29      1980 treatment_C
30       160 treatment_C
31         5 treatment_D
32         9 treatment_D
33        20 treatment_D
34       180 treatment_D
```

```
35        15 treatment_D
36        20 treatment_D
37         2 treatment_D
38        17 treatment_D
39        30 treatment_D
40         8 treatment_D
```

> bartlett.test(treatments,groups)

    Bartlett test of homogeneity of variances

data: treatments and groups
Bartlett's K-squared = 120.84, df = 3, p-value < 2.2e-16

- We compare the p-value for the Bartlett's test, which is less than 0.05. This makes clear that the variances are different from each other.

> data = list(Treatment_A,Treatment_B,Treatment_C,Treatment_D)

> data

[[1]]

 [1]   9   3 10 200   1   2 21 720 1500 60

[[2]]

 [1] 30 30 480 900   2   1 5400 1500 480   3

[[3]]

 [1] 3900 10800 28900 3600 200 120 500 600 1980 160

[[4]]

 [1]   5   9 20 180 15 20   2 17 30   8 > kruskal.test(data)

Kruskal-Wallis rank sum test

data: data

Kruskal-Wallis chi-squared = 14.294, df = 3, p-value = 0.002532

> qchisq(0.950,3)

[1] 7.814728

- The p-value comes to be less than 0.05, adding to it the chi-swuared test statistic is greater than the critical value. Hence, we reject the null hypothesis.

By hand:

A. (Given → treatment table)

H_0 = mean ranks are equal → $\mu_1 = \mu_2 = \mu_3 = \mu_4$
H_a = mean ranks are not equal → not all μ_i are equal

	Treatment A		Treatment B		Treatment C		Treatment D	
1.	9	10.5	30	19	3900	37	5	8
2.	3	6.5	30	19	10800	39	9	10.5
3.	10	12	480	27.5	28900	40	20	15.5
4.	200	25.5	800	32	3600	36	180	24

	Treatment A		Treatment B		Treatment C		Treatment D	
5.	1	1.5	2	41	200	25.5	15	13
6.	2	4	1	1.5	120	22	20	15.5
7.	21	17	5400	38	500	29	2	4
8.	720	31	1500	335	600	30	17	14
9.	1500	35.5	480	27.5	1880	35	30	19
10.	60	21	3	6.5	160	23	8	9
mean	2526		882.6		5076		30.6	
median	155		255		1290		16	
sum		162.5		209.5		316.5		132.5
mean		16		20		31.65		13.25

→ Test statistics:

$$H = \left[\frac{12}{M(N+1)} \times \frac{\Sigma T_c^2}{nC} \right] - \left[3 \times (n+1) \right]$$

$$= \left[\frac{12}{40(40+1)} \times \Sigma \frac{(162.5)^2}{10} + \frac{(208.5)^2}{10} + \frac{(316.5)^2}{10} \right.$$
$$\left. + \frac{(132.5)^2}{10} \right] - \left[3 \times (40+1) \right]$$

$$= 14.27341$$

B. **The blood potassium levels (K+) are measured in three sets of freshwater clams, Pond, DI water, and Sodium Solution. Assume the data are normally distributed. Test for equal variance using R only (show commands and conclusion). Then determine if the means are different by performing hypothesis test by hand and in R. Make a conclusion using alpha=0.05 with the test-statistic and p-value (15 points).**

Pond	DI	Sodium
0.518	0.318	0.393
0.523	0.342	0.415
0.495	0.301	0.351
0.502	0.39	0.39
0.525	0.327	0.385
0.49	0.32	0.397

R – code:

```
>potassium_levels<-
c(0.518,0.523,0.495,0.502,0.525,0.49,0.318,0.342,0.301,0.39,0.327,0.32,0.393,0.415,0.351,0.39
,0.385,0.397)
> groups<- c(rep("pond",6),rep("DI",6),rep("sodium",6))
> potassium <- data.frame(potassium_levels,groups)
> potassium
  potassium_levels groups
```

	potassium_levels	groups
1	0.518	pond
2	0.523	pond
3	0.495	pond
4	0.502	pond
5	0.525	pond
6	0.490	pond
7	0.318	DI
8	0.342	DI
9	0.301	DI
10	0.390	DI
11	0.327	DI
12	0.320	DI
13	0.393	sodium
14	0.415	sodium
15	0.351	sodium
16	0.390	sodium

17   0.385 sodium
18   0.397 sodium

> bartlett.test(potassium_levels,groups)

Bartlett test of homogeneity of variances

data: potassium_levels and groups
Bartlett's K-squared = 2.3104, df = 2, p-value = 0.315

- The p-value of the Bartlett's test is greater than 0.05 which shows that the variances are not significantly different from each other.
- This data is approximated normally and as the variances are not much different, we will use the ANOVA as the test for hypothesis testing.

R-code:
> qf(0.05,2,15,lower.tail = F)
[1] 3.68232
> pf(8526.7*10^-5,2,15,lower.tail = F)
[1] 0.9187089

- F value is greater than critical value
- We reject the null hypothesis

potassium_levels<-c(0.518,0.523,0.495,0.502,0.525,0.49,0.318,0.342,0.301,0.39,0.327,0.32,+
0.393,0.415,0.351,0.39,0.385,0.397)
> groups<-c(rep("pond",6),rep("DI",6),rep("sodium",6))
> potassium<-data.frame(potassium_levels,groups)
> results=aov(potassium_levels~groups,data = potassium)
> results
Call:
   aov(formula = potassium_levels ~ groups, data = potassium)

Terms:
                groups Residuals
Sum of Squares 0.09695544 0.00813833
Deg. of Freedom   2   15

Residual standard error: 0.02329282

Estimated effects may be unbalanced

> summary(results)

       Df Sum Sq Mean Sq F value   Pr(>F)

groups   2 0.09696 0.04848   89.35 4.65e-09 ***

Residuals   15 0.00814 0.00054

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4576.709

Estimated effects may be unbalanced

- P value is less than 0.05
- We reject the null hypothesis

By hand:

$H_0 = \mu_1 = \mu_2 = \mu_3$

$H_a = $ not all μ (all 3) are equal.

Variety (values)

Pond	0.518, 0.523, 0.495, 0.502, 0.525, 0.49
DI	0.318, 0.342, 0.301, 0.39, 0.327, 0.32
sodium	0.393, 0.415, 0.351, 0.39, 0.385, 0.397.

	sample size	means	variance
Pond	6	0.509	2.28×10^{-4}
DI.	6	0.333	9.56×10^{-4}
sodium	6	0.389	4.413×10^{-4}
(mean of means)		$\overline{0.41033}$	

$$SSA = \sum_{j=1}^{k} n_j \left(\overline{x_j} - \overline{x} \right)^2$$

$$= 6(0.509 - 0.410)^2 + 6(0.333 - 0.410)^2 +$$
$$6(0.389 - 0.410)^2$$

$$= 0.058806 + 0.035574 + 0.002646.$$

$$= 0.097026.$$

C. Transformation (10 pts)

Log-transform the data from question A using R. Test for normality and equal variances for the log-transformed data. Then determine if the means are different by performing hypothesis testing in R ONLY. Make a conclusion using alpha=0.05 with the test-statistic and p-value.

H0 = $\mu1 = \mu2 = \mu3 = \mu4$
Ha = not all the μi are equal

> log(Treatment_A)
 [1] 2.1972246 1.0986123 2.3025851 5.2983174 0.0000000 0.6931472 3.0445224 6.5792512
7.3132204
[10] 4.0943446
> log(Treatment_B)
 [1]   3.4011974   3.4011974   6.1737861   6.8023948   0.6931472   0.0000000   8.5941542
7.3132204,6.1737861
[10] 1.0986123
> log(Treatment_C)
 [1]   8.268732   9.287301 10.271597   8.188689   5.298317   4.787492   6.214608   6.396930
7.590852, 5.075174
> log(Treatment_D)
 [1] 1.6094379 2.1972246 2.9957323 5.1929569 2.7080502 2.9957323 0.6931472 2.8332133
3.4011974
[10] 2.0794415
>
treatments<c(2.1972246,1.0986123,2.3025851,5.2983174,0.0000000,0.6931472,3.0445224,6.5792512,7.3132204,4.0943446,3.4011974,3.4011974,6.1737861,6.8023948,0.6931472,0.0000000,8.5941542,7.3132204,6.1737861,1.0986123,8.268732,9.287301,10.271597,8.188689,5.298317,4.787492,6.214608,6.396930,7.590852,5.075174,1.6094379,2.1972246,2.9957323,5.1929569,2.7080502,2.9957323,0.6931472,2.8332133,3.4011974,2.0794415)
> groups <-
c(rep("treatment_A",10), rep("treatment_B",10), rep("treatment_C",10), rep("treatment_D",10))
>   grasshoppers<-data.frame(treatments,groups)
> grasshoppers   treatments   groups
1   2.1972246 treatment_A
2   1.0986123 treatment_A
3   2.3025851 treatment_A

4	5.2983174	treatment_A
5	0.0000000	treatment_A
6	0.6931472	treatment_A
7	3.0445224	treatment_A
8	6.5792512	treatment_A
9	7.3132204	treatment_A
10	4.0943446	treatment_A
11	3.4011974	treatment_B
12	3.4011974	treatment_B
13	6.1737861	treatment_B
14	6.8023948	treatment_B
15	0.6931472	treatment_B
16	0.0000000	treatment_B
17	8.5941542	treatment_B
18	7.3132204	treatment_B
19	6.1737861	treatment_B
20	1.0986123	treatment_B
21	8.2687320	treatment_C
22	9.2873010	treatment_C
23	10.2715970	treatment_C
24	8.1886890	treatment_C
25	5.2983170	treatment_C
26	4.7874920	treatment_C
27	6.2146080	treatment_C
28	6.3969300	treatment_C
29	7.5908520	treatment_C
30	5.0751740	treatment_C
31	1.6094379	treatment_D
32	2.1972246	treatment_D
33	2.9957323	treatment_D
34	5.1929569	treatment_D
35	2.7080502	treatment_D
36	2.9957323	treatment_D
37	0.6931472	treatment_D
38	2.8332133	treatment_D
39	3.4011974	treatment_D

40   2.0794415 treatment_D
> bartlett.test(treatments,groups)
        Bartlett test of homogeneity of variances
data: treatments and groups
Bartlett's K-squared = 7.483, df = 3, p-value = 0.058

- The p-value from Bartlett's test is higher than 0.05
- Hence, we reject the null hypothesis

Unlike already mentioned in question A and B, here we need to find if the data is normal or not, for which we use Shapiro-Wilk test.

> shapiro.test(treatments)

Shapiro-Wilk normality test

data: treatments
W = 0.35008, p-value = 4.21e-12

> treatment_A<-
c(2.1972246,1.0986123,2.3025851,5.2983174,0.0000000,0.6931472,3.0445224,6.5792512,7.31 32204,4.0943446)
> treatment_B<-
c(3.4011974,3.4011974,6.1737861,6.8023948,0.6931472,0.0000000,8.5941542,7.3132204,6.17 37861,1.0986123)
> treatment_C<-
c(8.268732,9.287301,10.271597,8.188689,5.298317,4.787492,6.214608,6.396930,7.590852,5.0 75174)
> treatment_D<-
c(1.6094379,2.1972246,2.9957323,5.1929569,2.7080502,2.9957323,0.6931472,2.8332133,3.40 11974,2.0794415)
> Treatments<-c(treatment_A,+ treatment_B,+ treatment_C,+ treatment_D)

> groups<-
c(rep("treatment_A",10),rep("treatment_B",10),rep("treatment_C",10),rep("treatment_D",10))
>   grasshoppers<-data.frame(Treatments,groups)
> results = aov(Treatments~groups,data = grasshoppers)
> results

Call: aov(formula = Treatments ~ groups, data = grasshoppers)

Terms:

          groups Residuals
Sum of Squares 117.7650 184.7404
Deg. of Freedom   3   36
Residual standard error: 2.265321
Estimated effects may be unbalanced
> summary(results)
        Df Sum Sq Mean Sq F value   Pr(>F)
groups   3 117.8   39.25   7.65 0.000443 ***
Residuals 36 184.7   5.13

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

By hand: