Name: Drashti Mehta

Lab Assignment 8

(40 points)

DIRECTIONS

- ***This assignment is due on next Friday after assignment.***
- ***In fill in box, please provide the hypotheses, R commands, solutions/output, and interpretation.***
- ***For hand work, attach excel or R output showing setup of the problem.***

A. Antihistamine drug A was produced as a response to allergic reactions. In a study of 14 children, the sample variance for response to the antihistamine for allergy was found to be 1.1 minutes. We wish to know if we may conclude that the population variance is not 3 minutes. What is the confidence interval for the variance at 95% by hand? In addition, provide the R commands to generate the critical values and the p-value (5 points).

R commands to generate the critical points:
> qchisq(0.025,13,lower.tail=TRUE) { for lower 2.5% quartile }
>qchisq(0.025,13,lower.tail=FALSE) { for upper 2.5% quartile }
test statistic , under Ho = (13*1.1/3)
= 4.767 { upto 3 decimal places }
R command for p value of test statistic:
> pchisq(4.767,13,lower.tail = TRUE) { gives lower side probability of test statistic }

B. Antihistamine drug B has a sample standard deviation of 1.2 minutes in a study of 18 children. Does antihistamine drug B have a larger variance than drug A? Please do the problem by hand. In addition, provide the R commands to generate the critical values and the p-value (5 points).

Sample 1 = sigma_a^2= 1.1
Sample 2 = sigma_b^2= 1.2
H0: sigma_a^2 >= sigma_b^2
H1: sigma_a^2 < sigma_b^2
F = sigma_b^2/sigma_a^2 = (1.2)^2/(1.1) = 1.31
1.31 < 2.50
R Command:
> qf(0.05,17,13,lower.tail=FALSE)
[1]2.498672
=> Not enough evidence to reject H0.
=> Not enough evidence to say drug B have larger variance than drug A

C. Mice weight gain response from different diets (10 points). Perform ANOVA by hand ONLY. Calculate the p-value and critical value using R. Use alpha=0.05. H0=u1=u2=u3.

Weight Gain in Mice

Group A (Control)	Group B (Junk Food)	Group C (Health Food)
10.8	12.7	9.8
11	13.9	8.6
9.7	11.8	8
10.1	13	7.5
11.2	11	9
9.8	10.9	10
10.5	13.6	8.1
9.5	10.9	7.8
10	11.5	7.9
10.2	12.8	9.1

R-code:

```
> 1-pf(1.117,2,27)
[1] 0.3419161
```

D. A lab wants to test whether there are differences in the average reading of the four pH meters they have. There are 24 soil samples with known pH in the study. These soil samples are randomly divided in four groups with each having 6 samples. For each sample the difference between the real pH and the measured pH is recorded. Also included are summary statistics for sub-question 1.

1) Perform an ANOVA by hand and in R to see if there is any evidence to indicate any difference among the mean pH. Use $\alpha\alpha = 0.05$. If there are differences, determine which devices differ from the other devices. (14 points)

device	1	2	3	4	5	6	Sample size	mean	Standard deviation
A	-0.307	-0.294	0.079	0.019	-0.136	-0.324	6	-0.1605	0.1767
B	-0.176	0.125	-0.013	0.082	0.091	0.459	6	0.0947	0.2091
C	0.137	-0.063	0.240	-0.050	0.318	0.154	6	0.1227	0.1532
D	-0.042	0.690	0.201	0.166	0.219	0.407	6	0.2735	0.2492

	A	B	C	D	E	F	G	H	I
1		x	s^2	n	$n*(x - xgrand)^2$	$(n - 1)*s^2$			
2	Group 1	-0.1605	0.0312	6	0.3546	0.1561			
3	Group 2	0.0947	0.0437	6	0.0009	0.2186			
4	Group 3	0.1227	0.0235	6	0.0096	0.1174			
5	Group 4	0.2735	0.0621	6	0.2187	0.3105			
6	xgrand	0.0826			Σ = SSTR	Σ = SSE			
7					0.5838	0.8026			
8									
9	Source	SS	df	MS	F	p-value			
10	Between	0.5838	3	0.1946	4.85	0.0108			
11	Error	0.8026	20	0.0401					
12	Total	1.39	23						
13									
14	q-critical value	3.958							
15									
16	Groups	Absolute Mean Difference	HSD	If Absolute Mean Difference > HSD, the groups are significantly different					
17	Group 1 - Group 2	0.2552	0.3237	Not Significantly Different					
18	Group 1 - Group 3	0.2832	0.3237	Not Significantly Different					
19	Group 1 - Group 4	0.4340	0.3237	Significantly Different					
20	Group 2 - Group 3	0.0280	0.3237	Not Significantly Different					
21	Group 2 - Group 4	0.1788	0.3237	Not Significantly Different					
22	Group 3 - Group 4	0.1508	0.3237	Not Significantly Different					
23									

	A	B	C	D	E	F
1		x	s^2	n	$n*(x - xgrand)^2$	$(n - 1)*s^2$
2	Group 1	-0.1605	=0.1767^2	6	=D2*(B2-B6)^2	=(D2-1)*C2
3	Group 2	0.0947	=0.2091^2	6	=D3*(B3-B6)^2	=(D3-1)*C3
4	Group 3	0.1227	=0.1532^2	6	=D4*(B4-B6)^2	=(D4-1)*C4
5	Group 4	0.2735	=0.2492^2	6	=D5*(B5-B6)^2	=(D5-1)*C5
6	xgrand	=AVERAGE(B2:B5)			Σ = SSTR	Σ = SSE
7					=SUM(E2:E5)	=SUM(F2:F5)
8						
9	Source	SS	df	MS	F	p-value
10	Between	=E7	=4-1	=B10/C10	=D10/D11	=FDIST(E10,C10,C11)
11	Error	=F7	=C12-C10	=B11/C11		
12	Total	=B10+B11	=D2+D3+D4+D5-1			
13						
14	q-critical value	3.958				
15						
16	Groups	Absolute Mean Difference	HSD	If Absolute Mean Difference > HSD, the groups are significantly different		
17	Group 1 - Group 2	=ABS(B2-B3)	=B14*SQRT((D11/2)*((1/D2)+(1/D3)))	=IF(B17>C17,"Significantly Different","Not Significantly Different")		
18	Group 1 - Group 3	=ABS(B2-B4)	=C17	=IF(B18>C18,"Significantly Different","Not Significantly Different")		
19	Group 1 - Group 4	=ABS(B2-B5)	=C18	=IF(B19>C19,"Significantly Different","Not Significantly Different")		
20	Group 2 - Group 3	=ABS(B3-B4)	=C19	=IF(B20>C20,"Significantly Different","Not Significantly Different")		
21	Group 2 - Group 4	=ABS(B3-B5)	=C20	=IF(B21>C21,"Significantly Different","Not Significantly Different")		
22	Group 3 - Group 4	=ABS(B4-B5)	=C21	=IF(B22>C22,"Significantly Different","Not Significantly Different")		
23						

The hypothesis being tested is:
Null Hypothesis, H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$
Alternative Hypothesis, H_a: At least one means is not equal
The test statistic, F = 4.85 The p-value is 0.0108.
Since the p-value (0.0108) is less than the significance level (0.05), we can reject the null hypothesis. Therefore, we can conclude that there is a difference among the mean pH. There is a significant difference between devices A and D.

The R-code is:

```
A <- c(-.307,-.294,.079,.019,-.136,-.324)
B <- c(-.176,.125,-.013,.082,.091,.459)
C <- c(.137,-.063,.24,-.05,.318,.154)
```

```
D <- c(-.042,.69,.201,.166,.219,.407)
CG <- data.frame(cbind(A,B,C,D))
SG <- stack(CG)
AR <- aov(values ~ ind, data = SG)
> summary(AR)
            Df  Sum  Sq  Mean  Sq  F  value
Pr(>F)    ind              3  0.5837  0.19458
4.847 0.0108 *
Residuals   20 0.8029 0.04014
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> TukeyHSD(AR)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = values ~ ind, data = SG)

$ind
        diff        lwr       upr     p adj
B-A 0.2551667 -0.06861103 0.5789444 0.1557875
C-A 0.2831667 -0.04061103 0.6069444 0.0999643
D-A 0.4340000  0.11022231 0.7577777 0.0063398
C-B 0.0280000 -0.29577769 0.3517777 0.9948495
D-B 0.1788333 -0.14494436 0.5026110 0.4303346
D-C 0.1508333 -0.17294436 0.4746110 0.5711518
```

E. A random sample of 20 sicklebacks each were collected from three small streams and three lakes, we want to see if these populations differ in total length among 6 habitats. See file sicklebacks.txt uploaded on canvas. <u>In R only</u>, input the data and perform an ANOVA (alpha 0.05). If the means are not all equal, determine which means differ significantly from each other. (6 points)

<u>R-Code:</u>

The null and alternative hypothesis are
Ho: Populations means same among 6 habitats
H1: Populations means differ among 6 habitats

```
> data=read.table(file.choose("sicklebacks.txt"))

> head(data)

  V1   V2 V3

1 31 Lake  A
```

```
2 32 Lake  A

3 34 Lake  A

4 34 Lake  A

5 35 Lake  A

6 30 Lake  A

> data=data.frame("X"=data$V1,"Tr"=paste(data$V2,data$V3))

> anova=aov(X~Tr,data=data)

> summary(anova)

             Df Sum Sq Mean Sq

Tr           5   1586   317.1

Residuals   114   2206    19.4

            F value   Pr(>F)

Tr           16.38 3.69e-12 ***

Residuals

---

Signif. codes:

0 '***' 0.001 '**' 0.01 '*'

 0.05 '.' 0.1 ' ' 1
```

Here P-value=3.69e-12<alpha=0.05 then we reject the null hypothesis Ho
Conclusion: Populations means differ among 6 habitats

```
> T=TukeyHSD(anova)

> T

  Tukey multiple comparisons of means

    95% family-wise confidence level

Fit: aov(formula = X ~ Tr, data = data)
```

diff	lwr	upr	P adj
Lake B-Lake A -0.30	Lake B-Lake A -4.332779	Lake B-Lake A 3.732779	Lake B-Lake A 0.9999343
Lake C-Lake A 0.55	Lake C-Lake A -3.482779	Lake C-Lake A 4.582779	Lake C-Lake A 0.9987233
Stream A-Lake A 8.35	Stream A-Lake A 4.317221	Stream A-Lake A 12.382779	Stream A-Lake A 0.0000004
Stream B-Lake A 6.55	Stream B-Lake A 2.517221	Stream B-Lake A 10.582779	Stream B-Lake A 0.0001021
Stream C-Lake A 6.85	Stream C-Lake A 2.817221	Stream C-Lake A 10.882779	Stream C-Lake A 0.0000420
Lake C-Lake B 0.85	Lake C-Lake B -3.182779	Lake C-Lake B 4.882779	Lake C-Lake B 0.9900481
Stream A-Lake B 8.65	Stream A-Lake B 4.617221	Stream A-Lake B 12.682779	Stream A-Lake B 0.0000001
Stream B-Lake B 6.85	Stream B-Lake B 2.817221	Stream B-Lake B 10.882779	Stream B-Lake B 0.0000420
Stream C-Lake B 7.15	Stream C-Lake B 3.117221	Stream C-Lake B 11.182779	Stream C-Lake B 0.0000169
Stream A-Lake C 7.80	Stream A-Lake C 3.767221	Stream A-Lake C 11.832779	Stream A-Lake C 0.0000022
Stream B-Lake C 6.00	Stream B-Lake C 1.967221	Stream B-Lake C 10.032779	Stream B-Lake C 0.0004844
Stream C-Lake C 6.30	Stream C-Lake C 2.267221	Stream C-Lake C 10.332779	Stream C-Lake C 0.0002096
Stream B-Stream A -1.80	Stream B-Stream A -5.832779	Stream B-Stream A 2.232779	Stream B-Stream A 0.7878454
Stream C-Stream A -1.50	Stream C-Stream A -5.532779	Stream C-Stream A 2.532779	Stream C-Stream A 0.8890993
Stream C-Stream B 0.30	Stream C-Stream B -3.732779	Stream C-Stream B 4.332779	Stream C-Stream B 0.9999343