STAT 500 Homework 6

Vu Thi-Hong-Ha; NetID: 851924086

October 2, 2020

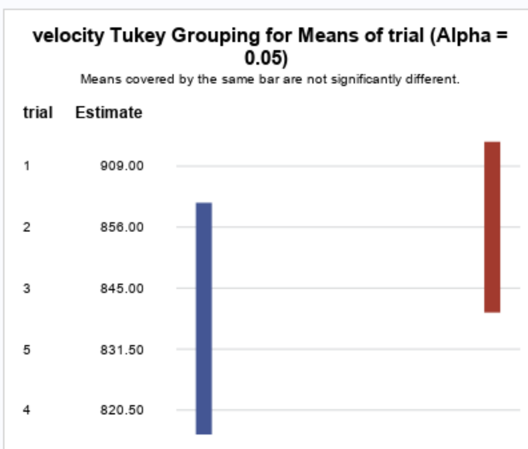
1 Question 1

(a)

Tukey's Studentized Range (HSD) Test for velocity

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	95
Error Mean Square	5510.632
Critical Value of Studentized Range	3.93274
Minimum Significant Difference	65.28



(i) Min distance required is 65.28.

(ii)

First Trial	Second Trial	Third Trial	Fifth Trial	Fourth Trial
909	856	845	831.5	820.5

(b) We have $r = 5$, so $m = \binom{5}{2} = 10$. Type I error rare for each test is $\alpha/m = 0.05/10 = 0.005$.

(c)

Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
1 vs 2345	70.7500000	18.5584071	3.81	0.0002	33.9069037	107.5930963
2 vs 345	23.6666667	19.1670404	1.23	0.2200	-14.3847196	61.7180529
3 vs 45	19.0000000	20.3297164	0.93	0.3524	-21.3595899	59.3595899
4 vs 5	-11.0000000	23.4747345	-0.47	0.6404	-57.6032401	35.6032401

Contrast 1:

- Contrast coefficients: $1, \frac{-1}{4}, \frac{-1}{4}, \frac{-1}{4}, \frac{-1}{4}$

- The estimate of the contrast: 70.75

- The standard error of the estimated contrast: 18.5584071

- p-value: 0.0002

- At a significance level of 0.05, we reject the null that the expected value of the contrast is zero.

Contrast 2:

- Contrast coefficients: $0, 1, \frac{-1}{3}, \frac{-1}{3}, \frac{-1}{3}$

- The estimate of the contrast: 23.6666667

- The standard error of the estimated contrast: 19.1670404

- p-value: 0.22

- At a significance level of 0.05, we fail reject the null that the expected value of the contrast is zero.

Contrast 3:

- Contrast coefficients: $0, 0, 1, \frac{-1}{2}, \frac{-1}{2}$

- The estimate of the contrast: 19

- The standard error of the estimated contrast: 20.3297164

- p-value: 0.3524

- At a significance level of 0.05, we fail reject the null that the expected value of the contrast is zero.

Contrast 4:

- Contrast coefficients: $0, 0, 0, 1, -1$

- The estimate of the contrast: -11

- The standard error of the estimated contrast: 23.4747345

- p-value: 0.6404

- At a significance level of 0.05, we fail reject the null that the expected value of the contrast is zero.

(d)

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
1 vs 2345	1	80089.00000	80089.00000	14.53	0.0002
2 vs 345	1	8401.66667	8401.66667	1.52	0.2200
3 vs 45	1	4813.33333	4813.33333	0.87	0.3524
4 vs 5	1	1210.00000	1210.00000	0.22	0.6404

Contrast 1:

- Sum of squares: 80089

- F statistic: 14.53

Contrast 2:

- Sum of squares: 8401.66667

- F statistic: 1.52

Contrast 3:

- Sum of squares: 4813.33333

- F statistic: 0.87

Contrast 4:

- Sum of squares: 1210

- F statistic: 0.22

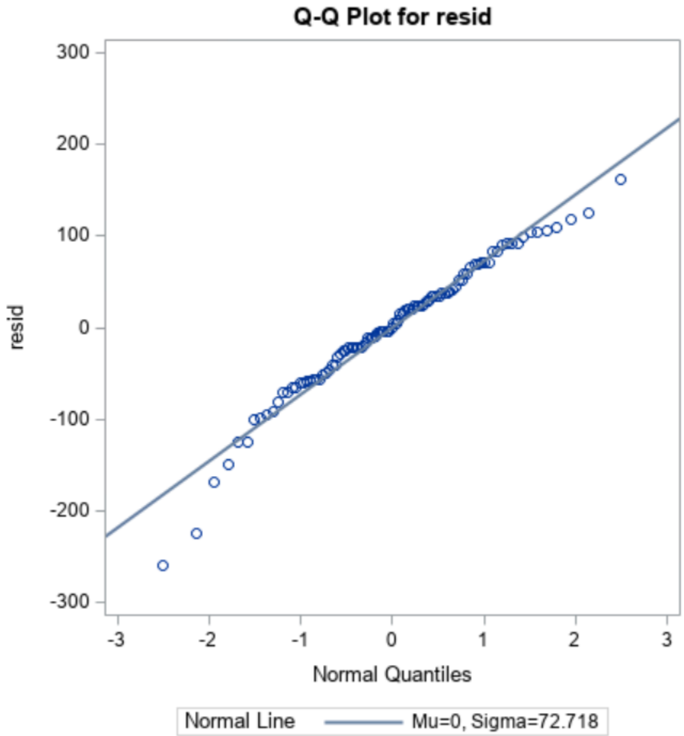
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	94514.0000	23628.5000	4.29	0.0031
Error	95	523510.0000	5510.6316		
Corrected Total	99	618024.0000			

We have that all of the mentioned contrasts are orthogonal to each other.

Sum of squares for this set of contrasts is $94514 = 80089 + 8401.66667 + 4813.33333 + 1210$, which are sums of squares of each contrast.

(e)

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.967789	Pr < W	0.0150
Kolmogorov-Smirnov	D	0.073745	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.077169	Pr > W-Sq	0.2291
Anderson-Darling	A-Sq	0.57775	Pr > A-Sq	0.1348



p-value of Shapiro-Wilk test is 0.015. At a significance level of 0.05, we reject the null that the measurements of the speed of light in air are normal. Based on the Q-Q plot, the residuals appear to be left-skewed.

(f)

Brown and Forsythe's Test for Homogeneity of velocity Variance ANOVA of Absolute Deviations from Group Medians					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
trial	4	17626.0	4406.5	1.67	0.1622
Error	95	249930	2630.8		

p-value is 0.1622. At a significance level of 0.05, we fail to reject the null that the variances of the five trials are homogeneous.

(g)

Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
15.0221	4	0.0047

Test statistic = 15.0221; p-value = 0.0047. At a significance level of 0.05, we reject the null that the five trials are sampled from the same distribution.

2 Question 2

(a)

$H_0: \dfrac{\mu_1+\mu_2+\mu_3+\mu_4+\mu_5+\mu_6}{6} = \dfrac{\mu_7+\mu_8+\mu_9}{3}$

$H_a: \dfrac{\mu_1+\mu_2+\mu_3+\mu_4+\mu_5+\mu_6}{6} \neq \dfrac{\mu_7+\mu_8+\mu_9}{3}$

Contrast: $\dfrac{\mu_1+\mu_2+\mu_3+\mu_4+\mu_5+\mu_6}{6} - \dfrac{\mu_7+\mu_8+\mu_9}{3}$

Point estimate of the contrast: $\hat{\gamma} = \dfrac{\bar{Y}_1+\bar{Y}_2+\bar{Y}_3+\bar{Y}_4+\bar{Y}_5+\bar{Y}_6}{6} - \dfrac{\bar{Y}_7+\bar{Y}_8+\bar{Y}_9}{3}$

$= \dfrac{7.347+7.369+7.428+7.487+7.563+7.568}{127+44+24+41+18+16} - \dfrac{8.214+8.272+8.297}{11+7+6} = -0.8668$

$MS_{error} = 126(0.4979^2)+43(0.4235^2)+23(0.3955^2)+40(0.3181^2)+17(0.3111^2)+15(0.4649^2)+10(0.2963^2)+6(0.3242^2)+5(0.5842^2) = 54.6956$

Then $S_{\hat{\gamma}} = \sqrt{54.6956 * (\frac{1}{36} * (\frac{1}{127}+\frac{1}{44}+\frac{1}{24}+\frac{1}{41}+\frac{1}{18}+\frac{1}{16}) + \frac{1}{9} * (\frac{1}{11}+\frac{1}{7}+\frac{1}{6}))} = 2.7598$

For a confidence coefficient of 95 percent and degrees of freedom in the numerator of $r-1 = 9-1 = 8$, and in the denominator of $N-r = 285$, we have: $\sqrt{(r-1)F_{r-1,N-r;1-\alpha/2}} = \sqrt{8 * F_{8,285;0.975}} = \sqrt{8 * 2.23657} = 4.23$, which is the critical value.

Test statistic: $\dfrac{0.8668}{2.7598} = 0.3141$

Since $0.3141 < 4.23$, we fail to reject the null that the two averages are equal at a significance level of 0.05.

(b) A 95% confidence interval is $-0.8668 \pm 2.7598 * 4.23 = (-12.54075; 10.80715)$

(c) In place of $\sqrt{(r-1)F_{r-1,N-r;1-\alpha/2}}$, we use $t_{N-r,1-\alpha/2} = t_{285,0.975} = 1.96832$. A 95% confidence interval is $-0.8668 \pm 2.7598 * 1.96832 = (-6.29897; 4.56537)$.

The CI constructed using t-test is narrower than that of Scheffe'.

(d) We need to use Scheffe' because we have already taken a look at the data and made some type pf comparison in order to choose to compare the first 6 species with the rest.

3 Question 3

(a)

The experiment is a randomized complete block experiment because units within each block are similar; within each block, the treatments are randomly assigned to the units so that one unit is for one treatment. Also, the number of units in each block is the same as the number of treatments.

- Experiment units: land plots.
- Treatments: densities. Number of treatment $J = 6$.
- Blocks: fields. Number of blocks $n = 5$.

Source of variation	Degrees of freedom
Blocks	$n-1=4$
Treatment	$J-1=5$
Error	$(n-1)(J-1)=20$
Total	$nJ-1=29$

(b)

The experiment is a completely randomized experiment because the assignments of boards to cutting speeds was done at random.

- Experiment units: circuit boards. Number of units $N = 15$
- Treatments: cutting speeds. Number of treatments $r = 3$
- Blocks: no.

Source of variation	Degrees of freedom
Model	$r-1=2$
Error	$N-r=12$
Total	$N-1=14$

(c)

The experiment is an incomplete block design experiment because the set of treatments appeared together in the same block an equal number of times.

- Experiment units: plants over each week.
- Treatments: music programs. Number of treatments $J = 5$
- Blocks: plants. Number of blocks $n = 8$.

Source of variation	Degrees of freedom
Blocks	$n-1=7$
Treatment	$J-1=4$
Error	$(n-1)(J-1)=28$
Total	$nJ-1=39$

4 Question 4

(a)

- Blocks: patients. Number of blocks: $n = 10$.
- Treatments: drugs. Number of treatments: $J = 2$.

(b)

$H_0 : \mu_{placebo} = \mu_{drug}$

$H_a : \mu_{placebo} - \mu_{drug} < 0$

Using paired t-test:

Patient	Difference (placebo - drug)
1	$d_1 = Y_{11} - Y_{12} = -0.7$
2	$d_2 = Y_{21} - Y22 = 0$
3	$d_3 = Y_{31} - Y32 = -3.7$
4	$d_4 = Y_{41} - Y42 = -0.8$
5	$d_5 = Y_{51} - Y52 = -2.0$
6	$d_6 = Y_{61} - Y62 = 1.6$
7	$d_7 = Y_{71} - Y72 = -3.4$
8	$d_8 = Y_{81} - Y82 = 0.2$
9	$d_9 = Y_{91} - Y92 = 1.2$
10	$d_{10} = Y_{10,1} - Y_{10,2} = 0.1$

Sample mean: $\bar{d} = -0.75$

Sample variance: $S_{\bar{d}}^2 = 3.2006$

Standard deviation: $S_{\bar{d}} = 1.789$

t-statistic: $t = \dfrac{\bar{d} - 0}{S_{\bar{d}}/\sqrt{n}} = -1.3257$ on $n - 1 = 9$ df.

One-sided p-value: $P(t_{n-1} < t) = 0.1088$.

(c) At a significance level of 0.05, we fail to reject the null that the mean difference between two treatments is 0.

(d)

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	W	0.925806	Pr < W	0.4079
Kolmogorov-Smirnov	D	0.188852	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.056048	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.346907	Pr > A-Sq	>0.2500

p-value is 0.4079. At a significance level of 0.05, we fail to reject the null that the differences in hours of sleep follow normal distribution.

(e)

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	t	-1.32571	Pr > |t|	0.2176
Sign	M	-0.5	Pr >= |M|	1.0000
Signed Rank	S	-8.5	Pr >= |S|	0.3594

H_0 : The median difference is equal to zero.

H_a : The median difference is smaller than zero.

Patient	$\mid$ Difference (placebo - drug) $\mid$	Rank
1	$\|-0.7\|$	4
2	$\|0\|$	1
3	$\|-3.7\|$	10
4	$\|-0.8\|$	5
5	$\|-2.0\|$	8
6	$\|1.6\|$	7
7	$\|-3.4\|$	9
8	$\|0.2\|$	3
9	$\|1.2\|$	6
10	$\|0.1\|$	2

The test statistics in SAS is S = -8.5 = $T - n(n+1)/4 = T - 27.5$. Then $T = 19$.

One-sided p-value = 0.3594/2 = 0.1797.

At a significance level of 0.05, we fail to reject the null that the median difference is equal to zero.

(f)

H_0 : The median difference is equal to 0.

H_a : The median difference is smaller than 0.

Patient	$\mid$ Difference (placebo - drug) $\mid$	Sign
1	$\|-0.7\|$	-
2	$\|0\|$	NA
3	$\|-3.7\|$	-
4	$\|-0.8\|$	-
5	$\|-2.0\|$	-
6	$\|1.6\|$	+
7	$\|-3.4\|$	-
8	$\|0.2\|$	+
9	$\|1.2\|$	+
10	$\|0.1\|$	+

Sample size is 10, but there is one difference = 0, so here n = 9. Also, S = 4 (the number of positive sign).

One-sided p-value = $(\binom{9}{0} + \binom{9}{1} + \binom{9}{2} + \binom{9}{3} + \binom{9}{4})*0.5^9 = 0.5$.

At a significance level of 0.05, we fail to reject the null that the median difference is equal to zero.

(g)

```
proc power;
   onesamplemeans    /* one sample */
     power=0.95       /* specify power */
     stddev=1.789       /* specify sigma */
     mean=0.75          /* specify mean under Ha */
     ntotal=.;        /* and solve for # obs */
run;
```

Hours of sleep under placebo and drug

The POWER Procedure
One-Sample t Test for Mean

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Mean	0.75
Standard Deviation	1.789
Nominal Power	0.95
Number of Sides	2
Null Mean	0
Alpha	0.05

Computed N Total	
Actual Power	N Total
0.950	76

The number of patients needed is 76.

5 Question 5

Let the preference for compazine indicate positive, and the preference for THC indicate negative. Excluding the patients expressing no preference, here $n = 25$.

p-value $= 2*(\binom{25}{0} + \binom{25}{1} + \binom{25}{2} + \binom{25}{3} + \binom{25}{4} + \binom{25}{5}) * 0.5^{25} = 0.0041$

At a significance level of 0.05, we reject the null that the two drugs are equally effective in preventing side effects.

6 Question 6

Number of blocks $n = 5$; number of treatment $J = 3$. $\bar{Y}_{..} = 0.844$

Sums of squares:

$$J\sum_{i=1}^{n=5}(\bar{Y}_{i.} - \bar{Y}_{..})^2 = 3*((0.5166667-0.844)^2 + (0.6066667-0.844)^2 + (0.67-0.844)^2 + (1.1566667-0.844)^2 + (1.27-0.844)^2) = 1.41896$$

$$n\sum_{j=1}^{J=3}(\bar{Y}_{.j} - \bar{Y}_{..})^2 = 5*((1.11-0.844)^2 + (0.992-0.844)^2 + (0.43-0.844)^2) = 1.32028$$

$$\sum_{i=1}^{n=5}\sum_{j=1}^{J=3}(Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2 = 0.01932$$

$$\sum_{i=1}^{n=5}\sum_{j=1}^{J=3}(Y_{ij} - \bar{Y}_{..})^2 = 2.75856$$

Source of variation	D.f	SS	MS	F	p-value
Block	$n-1=4$	1.41896	0.35474	146.89	$< .0001$
Diet	$J-1=2$	1.32028	0.66014	273.35	$< .0001$
Error	$(n-1)(J-1)=8$	0.01932	0.002415		
Total	$nJ-1=14$	2.75856			

(b)

Test the null hypothesis of no diet effects: $F = \frac{MS_{diet}}{MS_{error}} = \frac{0.66014}{0.002415} = 273.3499$

$F_{2,8;0.95} = 4.45897$

We reject the null hypothesis that diets have no effects at a significance level of 0.05.

(c)

$HSD = \frac{1}{\sqrt{2}} q_{r,df_{error},1-\alpha} \sqrt{MS_{error}(\frac{1}{n}+\frac{1}{n})} = \frac{1}{\sqrt{2}} 4.04 \sqrt{0.002415(\frac{1}{5}+\frac{1}{5})} = 0.08878813$, where $q_{3,8,0.95} = 4.04$

Rank the mean of diets in increasing order:

Diet 3	Diet 2	Diet 1
0.43	0.992	1.11

Based on these means and the HSD value, at an experiment-wise Type I error level of 0.05, we reject the null that mean difference between each pair of diets is zero.

(d)

Choose contrast 2 as γ_2: $1 * \alpha_2 - 0.5 * (\alpha_1 + \alpha_3)$, which is the contrast between mean of the second diet with mean of the first and third diet.

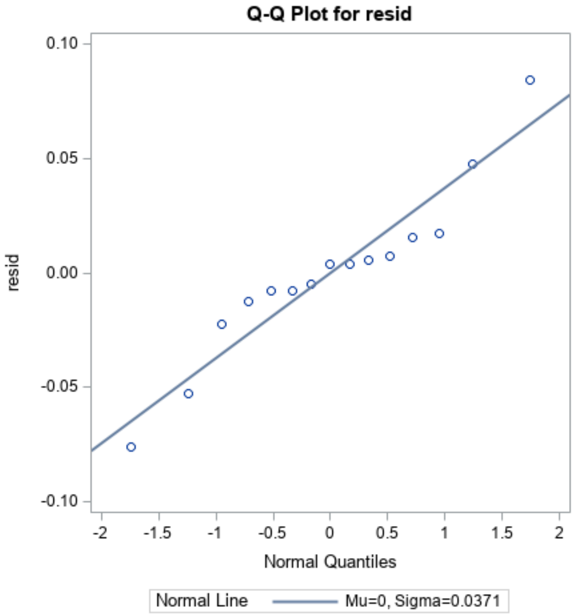
γ_1 and γ_2 are orthogonal as $(-1)(-0.5) + 0(1) + 1(-0.5) = 0$.

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
1 vs 3	1	1.15600000	1.15600000	478.67	<.0001
2 vs 13	1	0.16428000	0.16428000	68.02	<.0001

For γ_1, sum of squares is 1.156, F-statistics is 478.67, p-value is < 0.0001. At a significance level of 0.05, we reject the null that the contrast mean is zero.

For γ_2, sum of squares is 0.16428, F-statistics is 68.02, p-value is < 0.0001. At a significance level of 0.05, we reject the null that the contrast mean is zero.

(e)



Based on the q-q plot, the data might not be normal and might have heavy tail.

By Shapiro-Wilk test with p-value = 0.3135, at a significance level of 0.05, we fail to reject the null that the data follow normal distribution.

(f)

$\hat{\sigma}^2_{RCBD} = MS_{error} = 0.002415$

$\hat{\sigma}^2_{CRD} = \dfrac{(n-1)MS_{blocks} + n(J-1)MS_{error}}{nJ-1} = \dfrac{4 * 0.35474 + 5 * 2 * 0.002415}{14} = 0.1031$

Estimated efficiency: $\dfrac{(df_{RCBD}+1)(df_{CRD}+3)\hat{\sigma}^2_{CRD}}{(df_{RCBD}+3)(df_{CRD}+1)\hat{\sigma}^2_{RCBD}} = \dfrac{(8+1)(12+3)(0.1031)}{(8+3)(12+1)(0.002415)} = 40.30318$ ($df_{CRD} = 12$ because there are 15 units and 3 treatments).

Then, to have the same efficiency, $n_{CRD} = 40.30318 n_{RCBD}$, which means we need many more samples if doing a completely randomized experiment.