STAT 500 Homework 7

Vu Thi-Hong-Ha; NetID: 851924086

October 15, 2020

1 Question 1

This is a randomized complete block experiment with patients as blocks. The block effects are fixed. The number of blocks: $n = 14$, the number of treatments: $J = 3$.

(a) ANOVA table:

Source of variation	D.f	SS	MS	F	p-value
Block	$n-1=13$	231.4923810	17.8071062	11.82	$< .0001$
Treatment	$J-1=2$	81.0876190	40.5438095	26.92	$< .0001$
Error	$(n-1)(J-1)=26$	39.1590476	1.5061172		
Corrected Total	$nJ-1=41$	351.7390476			

(b)

$H_0 : \mu_B = \mu_N = \mu_S$

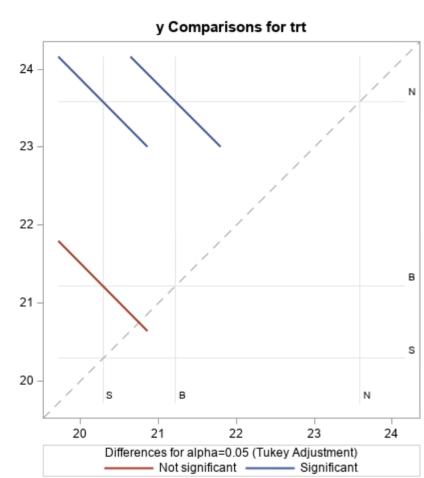
H_a : At least one μ_i is different, where $i = 1, 2, 3$

The test statistic $F = 26.92$ with p-value < 0.0001. At a significance level of 0.05, we reject the null hypothesis and conclude there is at least one mean of one treatment is different.

(c)

Least Squares Means for effect trt
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: y

i/j	1	2	3
1		<.0001	0.1318
2	<.0001		<.0001
3	0.1318	<.0001	



With an experiment-wise Type I error level of $\alpha = 0.05$, the mean difference between Neural (N) and Staggered (S) and the mean difference between Neural (N) and Back (B) are significantly different.

(d)

Dependent Variable: y

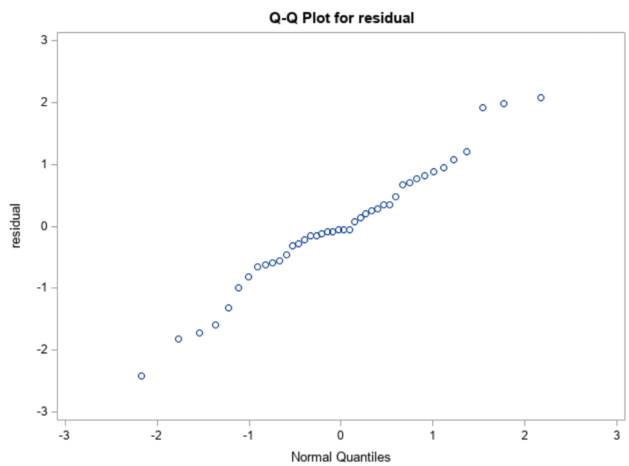
Parameter	Estimate	Standard Error	t Value	Pr > |t|
N-(B+S)/2	2.83571429	0.40170848	7.06	<.0001
B-S	0.92857143	0.46385300	2.00	0.0558

At a significance level of 0.05, we reject the null that mean of the N group is equal to the mean of the other two groups. We conclude that mean of the N group is different from the mean of the other two groups.

At a significance level of 0.05, we fail to reject the null that mean of the B group is equal to the mean of the S group. We conclude that mean of the B group is equal to the mean of the S group.

These conclusions agree with that in part (c).

(e)

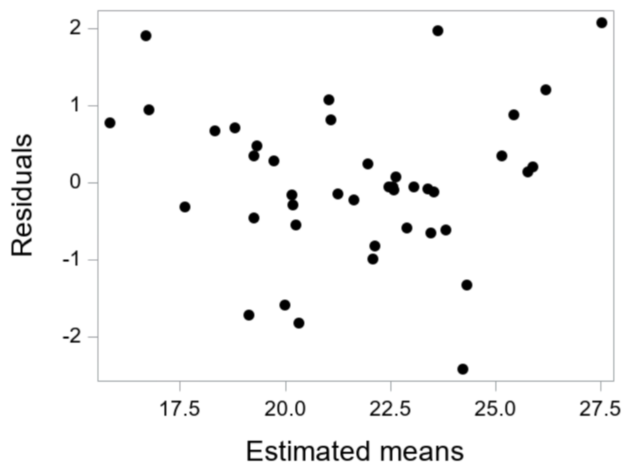


The Q-Q plot suggests that the data might be slightly light tailed, but the data points look quite close to a straight line, so the data might follow normal distribution. By Shapiro-Wilk test, we fail to reject the null that the data follows normal distribution at a significance level of 0.05.

Tests for Normality

Test	Statistic		p Value	
Shapiro-Wilk	W	0.976734	Pr < W	0.5388
Kolmogorov-Smirnov	D	0.086323	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.06534	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.399266	Pr > A-Sq	>0.2500

(f)



In the above plot, the residuals show a slight bend when plotted against the predicted values, indicating that we may have to add some non-linear component to our regression model. The data points do not form a "horizontal band" around the 0 line, suggesting that the variance of the error terms may not be equal. No data point is very separated from the others, so the residuals may not have any outliers.

(g)

$\hat{\sigma}^2_{RCBD} = MS_{error} = 1.5061172$

$\hat{\sigma}^2_{CRD} = \dfrac{(n-1)MS_{blocks} + n(J-1)MS_{error}}{nJ-1} = \dfrac{13 * 17.8071062 + 14 * 2 * 1.5061172}{41} = 6.674724$

Estimated efficiency: $\dfrac{(df_{RCBD}+1)(df_{CRD}+3)\hat{\sigma}^2_{CRD}}{(df_{RCBD}+3)(df_{CRD}+1)\hat{\sigma}^2_{RCBD}} = \dfrac{(26+1)(39+3)(6.674724)}{(26+3)(39+1)(1.5061172)} = 4.33241$ ($df_{CRD} = 39$ because there are $14 * 3 = 42$ units and 3 treatments).

Then, to have the same efficiency, $n_{CRD} = 4.33241 n_{RCBD}$, which means we need many more samples if doing a completely randomized experiment.

2 Question 2

In this experiment, there are two blocking factors: rows and columns. The number of levels for each blocking factor is 5, the number of treatments (or in this case, densities) is also 5. Denote $r = 5$.

(a) ANOVA table:

Source of variation	D.f	SS	MS	F	p-value
Row	$r-1=4$	9017.6	2254.4	2.23	0.1274
Column	$r-1=4$	4873.2	1218.3	1.20	0.3596
Spacing	$r-1=4$	6297.2	1574.3	1.55	0.2491
Error	$(r-1)(r-2)=12$	12158	1013.16667		
Corrected Total	$r^2-1=24$	32346			

(b)

$H_0 : \mu_2 = \mu_4 = \mu_6 = \mu_8 = \mu_{10}$

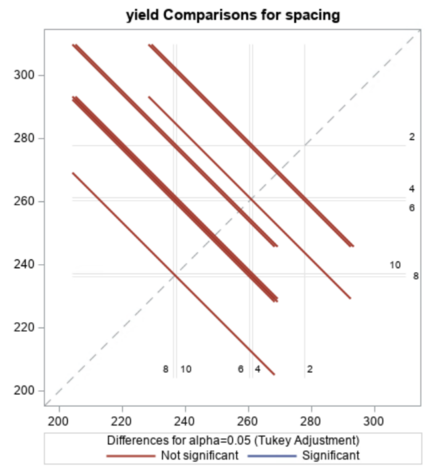
H_a : At least one mean of one density is different.

The test statistic $F = 1.55$ with p-value $= 0.2491$. At a significance level of 0.05, we fail to reject the null that the mean of each treatment is equal to each other. We conclude that there is no effect of treatments.

(c)

Least Squares Means for effect spacing
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: yield

i/j	1	2	3	4	5
1		0.9211	0.9045	0.2942	0.3149
2	0.9211		1.0000	0.7233	0.7504
3	0.9045	1.0000		0.7504	0.7767
4	0.2942	0.7233	0.7504		1.0000
5	0.3149	0.7504	0.7767	1.0000	



With an experiment-wise Type I error level of $\alpha = 0.05$, the mean difference between each pair of treatments are not significantly different. We conclude that there is no difference between each pair of treatments.

(d)

ANOVA table for RCBD analysis after dropping rows

Source of variation	D.f	SS	MS	F	p-value
Column	$r-1=4$	4873.2	1218.3	0.92	0.4761
Spacing	$r-1=4$	6297.2	1574.3	1.19	0.3530
Error	$(r-1)(r-1)=16$	21175.6	1323.475		
Corrected Total	$r^2-1=24$	32346			

$MSE_{LS} = 1013.16667$

$MSE_{RCBD} = 1323.475 > 1013.16667$

Hence, Latin Square with rows is more precise.

Relative Efficiency $= MS_{LS}/MS_{RCBD} = 1013.16667/1323.475 = 0.77$

LS requires 0.77 as many observations as RCBD with column blocking does.

When dropping row blocking, the number of levels of blocks is $n = 5$, the number of treatments is $J = 5$.

$\hat{\sigma}^2_{RCBD} = MS_{error} = 1323.475$

$$\hat{\sigma}^2_{CRD} = \frac{(n-1)MS_{blocks} + n(J-1)MS_{error}}{nJ-1} = \frac{4*1218.3 + 5*4*1323.475}{24} = 1305.946$$

Hence, CRD is more precise than RCBD with column blocking, but still less precise than LS.

ANOVA table for RCBD analysis after dropping columns

Source of variation	D.f	SS	MS	F	p-value
Row	$r-1=4$	9017.6	2254.4	2.12	0.1259
Spacing	$r-1=4$	6297.2	1574.3	1.48	0.2550
Error	$(r-1)(r-1)=16$	17031.2	1064.45		
Corrected Total	$r^2-1=24$	32346			

$MSE_{LS} = 1013.16667$

$MSE_{RCBD} = 1064.45 > 1013.16667$

Hence, Latin Square with rows is more precise.

Relative Efficiency $= MS_{LS}/MS_{RCBD} = 1013.16667/1064.45 = 0.95$

LS requires 0.95 as many observations as RCBD with row blocking does.

$\hat{\sigma}^2_{RCBD} = MS_{error} = 1064.45$

$$\hat{\sigma}^2_{CRD} = \frac{(n-1)MS_{blocks} + n(J-1)MS_{error}}{nJ-1} = \frac{4*2254.4 + 5*4*1064.45}{24} = 1262.775$$

Hence, CRD is less precise than RCBD with row blocking.

In conclusion, we should keep using LS design for next year study.

3 Question 3

(a)

Treatment 1: sulfur dioxide pollution (2 levels: 0 or 1, i.e., treated with sulfur dioxide pollution or not; $a = 2$).

Treatment 2: nitrous oxides pollution (2 levels: 0 or 1, i.e., treated with nitrous oxides pollution or not; $b = 2$).

(b) This is a full factorial design because all possible combinations of factor levels are used (no pollution, treated with sulfur dioxide pollution, treated with nitrous oxides pollution, and treated with both pollution).

(c) Number of replications per treatment is $n = 4$.

Source of variation	D.f
Sulfur dioxide pollution	$a-1=1$
Nitrous oxides pollution	$b-1=1$
Interaction of Sulfur dioxide and Nitrous oxides pollution	$(a-1)(b-1)=1$
Error	$ab(n-1)=12$
Total	$abn-1=15$

(d)

Denote Sulfur dioxide pollution as factor 1 and Nitrous oxides pollution as factor 2. We have 2 level of factor 1 ($i = 1, 2$), 2 levels of factor 2 ($j = 1, 2$) and 4 units assigned to each of the combinations of the two factors ($k = 1, 2, 3, 4$).

(i) Is there any difference between treatments with air pollution and the control treatment (air without pollution)?

$H_0 : \mu_{11} - \dfrac{\mu_{12} + \mu_{21} + \mu_{22}}{3} = 0$

Contrast: $\mu_{11} = \dfrac{\mu_{12} + \mu_{21} + \mu_{22}}{3}$

(ii) Averaged over sulfur dioxide pollution status (yes or no), is there any effect of nitrous oxides pollution?

$H_0 : \mu_{.1} - \mu_{.2} = 0$

Contrast: $\dfrac{\mu_{11} + \mu_{21}}{2} = \dfrac{\mu_{12} + \mu_{22}}{2}$

(iii) Does the effect of nitrous oxides pollution depend on the inclusion of sulfur dioxide pollution or not?

$H_0 : \mu_{11} - \mu_{12} - (\mu_{21} - \mu_{22}) = 0$

Contrast: $\mu_{11} - \mu_{12} = \mu_{21} - \mu_{22}$

(e)

(i) Non of the effects (because we did not average over any factors).

(ii) Main effect (because we averaged over the sulfur dioxide levels).

(iii) $\mu_{11} - \mu_{12}$ and $\mu_{21} - \mu_{22}$ are simple effects. Their contrast is an interaction effect.

(f) We have 4 treatments in total, so the number of orthogonal contrasts is 3.

(g) Yes. Because $\frac{1}{2} \times 1 + \frac{1}{2} \times (-1) + \frac{-1}{2} \times (-1) + \frac{-1}{2} \times 1 = 0$.

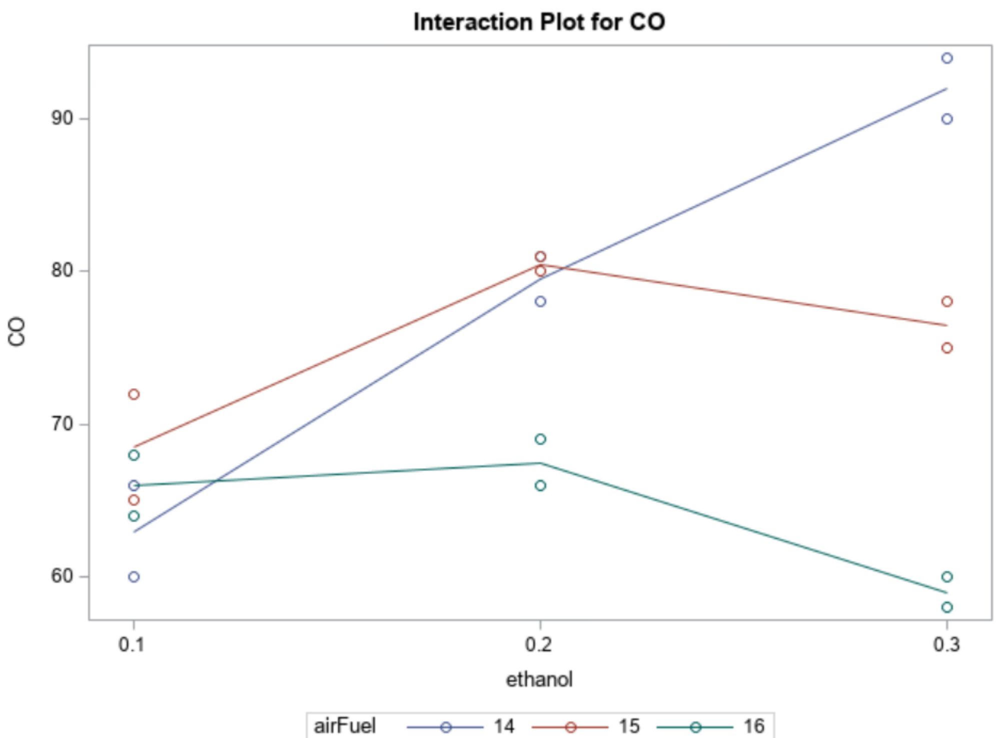
4 Question 4

Number of levels of factor 1 (ethanol): $a = 3$, number of levels of factor 2 (air-fuel ratio): $b = 3$, number of units per treatment: $n = 2$.

Source of variation	D.f	SS	MS	F	p-value
Ethanol	$a - 1 = 2$	400	200	24.16	0.0002
Air/Fuel Ratio	$b - 1 = 2$	652	326	39.38	$< .0001$
Ethanol*Air/Fuel Ratio	$(a-1)(b-1) = 4$	678	169.5	20.48	0.0002
Error	$ab(n-1) = 9$	74.5	8.277778		
Corrected Total	$abn - 1 = 17$	1804.5			

At a significance level of 0.05, the ethanol factor, air/fuel ratio factor and the interaction between these two factors were all significant.

(b)



The interaction is significant as the lines in the plot are not parallel.

(c) + (d)

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Ethanol L3-L1	1	300.0000000	300.0000000	36.24	0.0002
Ethanol L2-L1/2-L3/2	1	100.0000000	100.0000000	12.08	0.0070
Air/Fuel L3-L1	1	588.0000000	588.0000000	71.03	<.0001
Air/Fuel L2-L1/2-L3/2	1	64.0000000	64.0000000	7.73	0.0214

For the two contrasts $L3 - L1$ and $L2 - (L1 + L3)/2$ of ethanol, they are orthogonal because $-1(-1/2) + 0(1) + 1(-1/2) = 0$.

$SS_{ethanol} = SS_{EthanolL3-L1} + SS_{EthanolL2-(L1+L3)/2} = 300 + 100 = 400$.

Both of the contrasts are statistically significant at a significance level of 0.05.

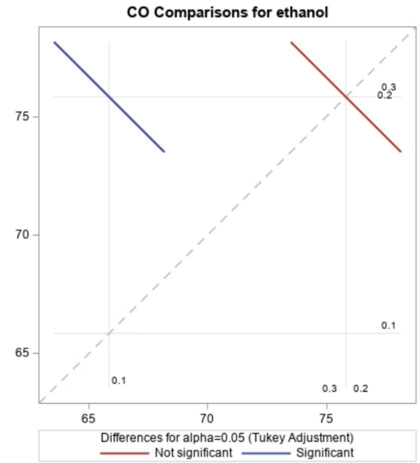
For the two contrasts $L3 - L1$ and $L2 - (L1 + L3)/2$ of air/fuel ratio, they are also orthogonal.

$SS_{airFuel} = SS_{airFuelL3-L1} + SS_{airFuelL2-(L1+L3)/2} = 588 + 64 = 652$.

Both of the contrasts are statistically significant at a significance level of 0.05.

(e)

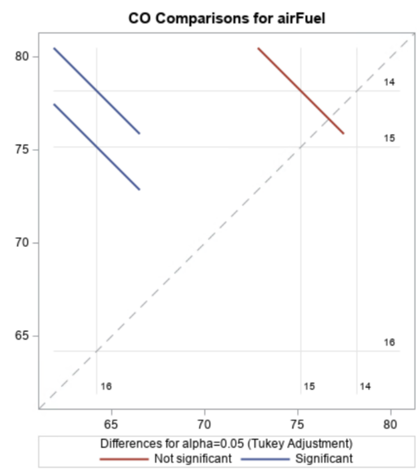
Least Squares Means for effect ethanol Pr > |t| for H0: LSMean(i)=LSMean(j) Dependent Variable: CO			
i/j	1	2	3
1		0.0005	0.0005
2	0.0005		1.0000
3	0.0005	1.0000	



At a significance level of 0.05, the marginal mean difference between ethanol level 1 and level 2 and ethanol level 1 and level 3 are statistically significant.

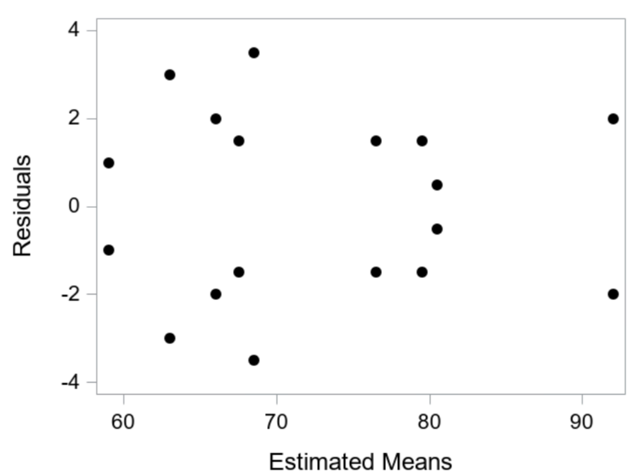
(f)

Least Squares Means for effect airFuel Pr > |t| for H0: LSMean(i)=LSMean(j) Dependent Variable: CO			
i/j	1	2	3
1		0.2219	<.0001
2	0.2219		0.0003
3	<.0001	0.0003	



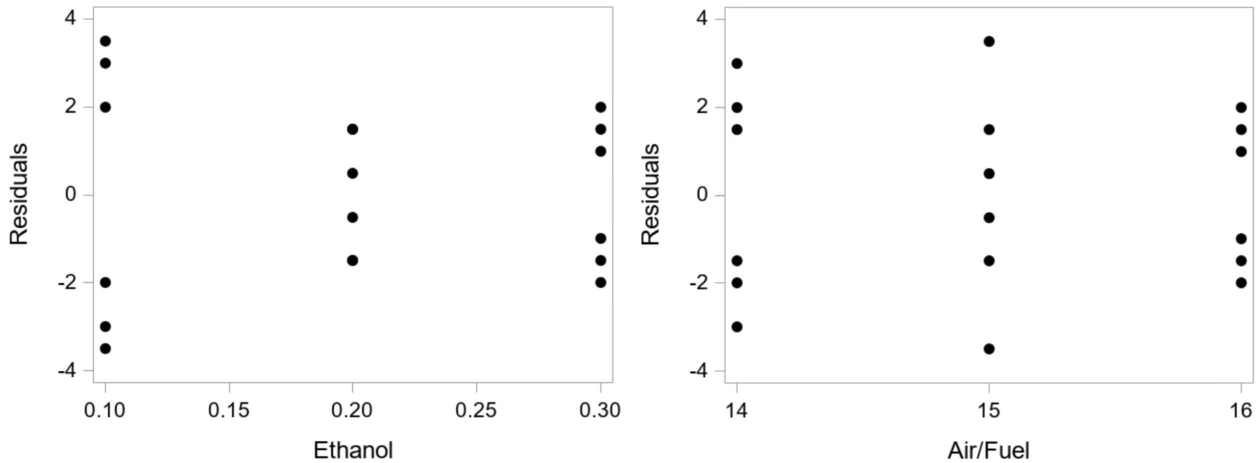
At a significance level of 0.05, the marginal mean difference between air/fuel ratio level 1 and level 2 and ethanol level 2 and level 3 are statistically significant.

(g)



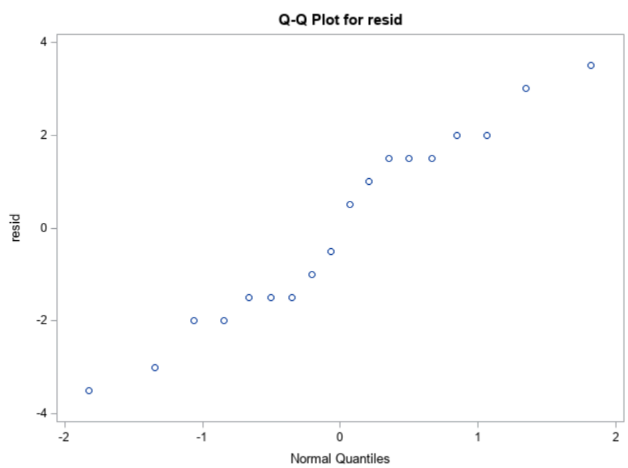
In the above plot, the data points seem to form a "horizontal band" around the 0 line, suggesting that the variance of the error terms may be equal. No data point is very separated from the others, so the residuals may not have any outliers.

(h)



It seems like the residuals under different treatments have similar mean because their distributions overlap for the most part.

(i)



The residual points quite align to the normal line, indicating that the data might follow a normal distribution.