STA5205 Spring, 2020
Design of Experiment

Problem1: le Riche and Csima (1964) evaluated four hypnotic drugs and a placebo to determine their effect on quality of sleep in elderly patients.

(a) What is the appropriate model for this data?
The model for a LSD is written as $yijk = \mu + ri + cj + \tau k + €ijk$
where ri represents the row blocking factor (patient), cj represents the column blocking factor (week), and τk represents the treatment factor (drug).

(b) Complete the ANOVA and determine if there are any significant differences among the treatments.
Some of the ANOVA tables are attached as below.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	3.02607200	0.25217267	5.28	0.0036
Error	12	0.57335200	0.04777933		
Corrected Total	24	3.59942400			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
patient	4	0.60734400	0.15183600	3.18	0.0536
week	4	0.36890400	0.09222600	1.93	0.1700
treat	4	2.04982400	0.51245600	10.73	0.0006

The F-test shows the fitted model is significant.
And the Type SS table above indicates there are significant differences among treatments by F-value=10.73 and p-value=0.0006 << alpha=0.05.

(c) Use an appropriate method to determine if there is a significant difference between the Placebo and the average of the other drugs, and if there are significant differences among the four drugs.
Use a contrast to test the significant difference between A and mean(BCDE) by

```
estimate 'A vs Others' treat -4 1 1 1 1;
```

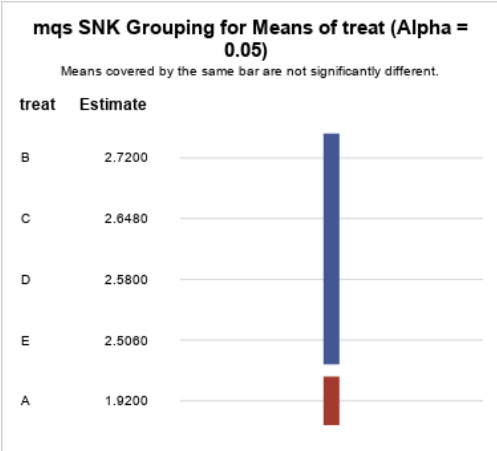
Parameter	Estimate	Standard Error	t Value	Pr > |t|
A vs Others	2.77400000	0.43716968	6.35	<.0001

The test result shown in the table above indicates there is a significant difference between the Placebo (A) and the average of the other drugs (BCDE).

Use means or lsmeans statement to see if there are significant differences among BCDE.

```
means treat/snk;
lsmeans treat / pdiff adjust = tukey;
```

The results of means and lsmeans are shown as below, respectively. The plot of means of BCDE on the left shows there are no significant differences among BCDE marked by blue square. The table on the right also shows BCDE are not significantly different with large p-values.

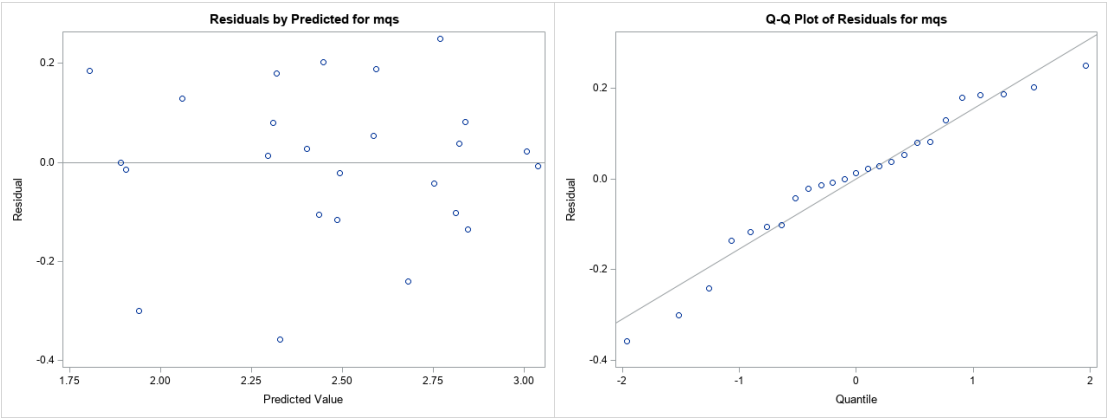


Least Squares Means for effect treat
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: mqs

i/j	1	2	3	4	5
1		0.0007	0.0015	0.0034	0.0083
2	0.0007		0.9835	0.8448	0.5537
3	0.0015	0.9835		0.9866	0.8382
4	0.0034	0.8448	0.9866		0.9818
5	0.0083	0.5537	0.8382	0.9818	

(d) Use residual plots to check the assumptions for the model you fit.

The plot of the residual vs the predicted value shows the variance seems to be constant. And the Q-Q plot also indicates the residuals basically follow the normality.



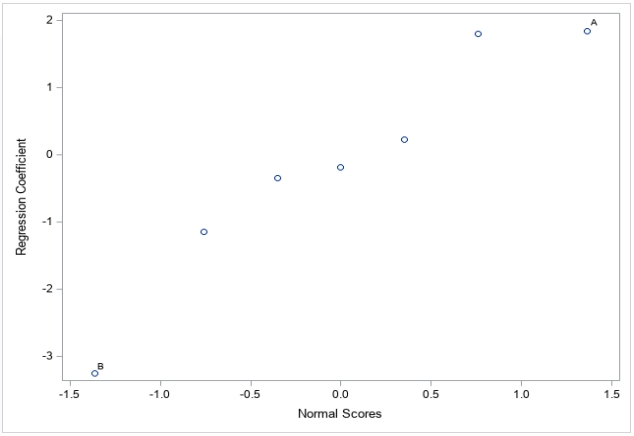
Problem2: Melo et al. (2007) used a 2^(4-1) factional factorial design with generator D = ABC to study the factors that influence the production of levan by aerobic fermentation using the yeast Zymomonas mobilis.

(a) What is the defining relation and complete alias structure for this design?
Given D=ABC, so the defining relation is I=ABCD. And the complete alias structure should be: I+ABCD, A+BCD, B+ACD, C+ABD, D+ABC, AB+CD, AC+BD, AD+BC.

(b) What is the resolution of this design?
Since I=ABCD is the shortest word with length 4, the resolution should be 4.

(c) The results (in g/L) for the 8 experiments (in standard order) were: 4.70, 14.67, 1.71, 3.73, 9.47,7.61, 0.15, 4.78. From this data, calculate the 7 effects and make a normal probability plot to determine what is significant.
All estimates of the 7 effects (A, B, C, D, AB, AC, AD) are shown as below. And the normal probability plot indicates the main effects A and B have larger effects than others, but none of them are significant from the estimate table as below.

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	5.852500000	.	.	.
A	1.845000000	.	.	.
B	-3.260000000	.	.	.
C	-0.350000000	.	.	.
D	1.805000000	.	.	.
A*B	-0.182500000	.	.	.
A*C	-1.152500000	.	.	.
A*D	0.222500000	.	.	.



(d) Delete the smallest three effects and fit the model again. Are the four effects left in the model significant?
Delete the smallest three effects (C, AB, AD) and refit the model (A, B, D, AC). The estimates of four effects in the model are significant with small p-values indicated by the table below.

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	5.852500000	0.26160562	22.37	0.0002
A	1.845000000	0.26160562	7.05	0.0059
B	-3.260000000	0.26160562	-12.46	0.0011
D	1.805000000	0.26160562	6.90	0.0062
A*C	-1.152500000	0.26160562	-4.41	0.0217

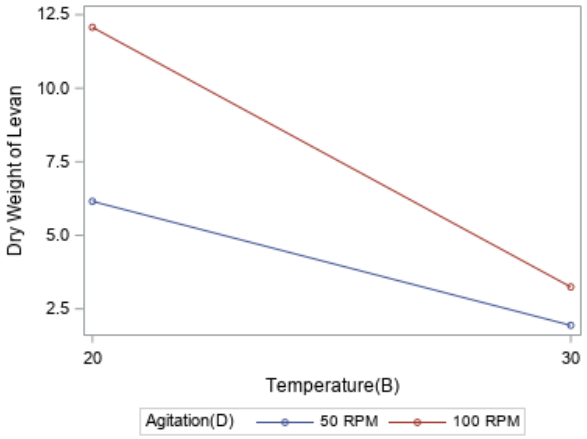
(e) Based on the effect heredity principle, what do you think the significant string of aliased two factor interactions could represent ?
From the result of (d), we can see the interaction AC is significant and confounded with BD. Given C is not significant, but both B and D are significant. Therefore, based on the effect

heredity principle, we may intend to include the significant interaction BD in the model, instead of AC.

(f) Can the design be collapsed to a full factorial by ignoring insignificant factors?
Yes. Due to the hidden replications in the factorial design, it means there are three factors (A, B, D) in the design with 2^(3)=8 runs which are sufficient for a full factorial design to do the F test and ANOVA analysis.

(g) Based on what you said in e) write an interpretation of the significant effects and interaction. Based on your model determine the factor levels that will produce the maximum dry weight of levan.

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	5.852500000	0.26160562	22.37	0.0002
A	1.845000000	0.26160562	7.05	0.0059
B	-3.260000000	0.26160562	-12.46	0.0011
D	1.805000000	0.26160562	6.90	0.0062
B*D	-1.152500000	0.26160562	-4.41	0.0217



Overall, the significant effects included in the model are A, B, D, and BD. The positive estimate of A means one unit increase in A (sucrose initial concentration) results in the dry weight of levan increases by 1.845. we intend to choose the higher A (250), instead of the lower one (150). Similarly, the higher D (agitation) may lead to the higher dry weight of levan, while the higher B (temperature) may cause the lower dry weight of leven. The interaction plot of BD tells that keeping a higher level of D (agitation) at 100 rpm and a lower level of B (temperature) at 20 ^{O}C can get a higher dry weight of levan. For reaching the maximum dry weight of levan, we may have this combination of levels: A (+), B (-), D (+).