Lab 7

2024-10-22

Q1

The full SAS program to analyze the penicillin data is provided in the penicillin Lab7.sas file located in our course's shared folder in SAS Studio. Refer to the output (and modify the code where necessary) to complete the following exercises:

(a)

From the SAS output, find the full ANOVA table and provide an appropriate summary of results for analyzing the different processes on the yield of penicillin.

Inference for RCBD using PROC GLM

The GLM Procedure

Dependent Variable: yield

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	334.0000000	47.7142857	2.53	0.0754
Error	12	226.0000000	18.8333333		
Corrected Total	19	560.0000000			

R-Square	Coeff Var	Root MSE	yield Mean
0.596429	5.046208	4.339739	86.00000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
batch	4	264.0000000	66.0000000	3.50	0.0407
process	3	70.0000000	23.3333333	1.24	0.3387

Figure 1: Img

(b)

Perform all pairwise comparisons of treatment means using Tukey's HSD method. Write a summary of your findings

Differences of Least Squares Means

Effect	process	_process	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment	Adj P
process	A	B	-1.0000	2.7447	12	-0.36	0.7219	Tukey-Kramer	0.9827
process	A	C	-5.0000	2.7447	12	-1.82	0.0935	Tukey-Kramer	0.3105
process	A	D	-2.0000	2.7447	12	-0.73	0.4802	Tukey-Kramer	0.8838
process	B	C	-4.0000	2.7447	12	-1.46	0.1707	Tukey-Kramer	0.4905
process	B	D	-1.0000	2.7447	12	-0.36	0.7219	Tukey-Kramer	0.9827
process	C	D	3.0000	2.7447	12	1.09	0.2958	Tukey-Kramer	0.7002

Figure 2: Img

(c)

There are three orthogonal contrasts specified in the SAS code. Describe the analysis provided by these contrasts and determine which are statistically significant. Write a summary of your findings.

Inference for RCBD using PROC GLM

The GLM Procedure

Dependent Variable: yield

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
A-B	1	2.50000000	2.50000000	0.13	0.7219
C-(A+B)/2	1	67.50000000	67.50000000	3.58	0.0827
D-(A+B+C)/3	1	0.00000000	0.00000000	0.00	1.0000

Parameter	Estimate	Standard Error	t Value	Pr > |t|
A-B	-1.00000000	2.74469185	-0.36	0.7219
C-(A+B)/2	4.50000000	2.37697286	1.89	0.0827
D-(A+B+C)/3	0.00000000	2.24103151	0.00	1.0000

Figure 3: Img

A-B: The comparison between A and B is not statistically significant (P = 0.7219), indicating no meaningful difference between these two levels.

C - (A+B)/2: This contrast approaches statistical significance (P = 0.0827), suggesting that there may be a meaningful difference between C and the average of A and B, although it does not reach the conventional threshold for significance (0.05).

D - (A+B+C)/3: This contrast shows no significant difference (P = 1.0000), indicating that D is essentially equivalent to the average of A, B, and C in terms of yield.

(d)

Check the assumptions for the RCBD analysis using SAS output. Write a summary of your findings.

There are 4 assumptions in the RCBD: 1. Independence of residuals: assessed through the study design principles

2. Homogeneous (equal) residual variance: assessed through boxplots of residuals within each treatment, ratio of standard deviations, and equal variance statistical tests such as the Brown Forsythe
3. Normality of residuals: assessed through histograms, normal Q-Q plots, summary statistics (mean, median, skewness, excess kurtosis), and tests for normality of the residuals
4. Additive block and treatment effects: Using the penicillin example as illustration, the additivity assumption for the RCBD can be checked in SAS using both graphs and statistical tests • Residual vs. Fitted Graph (no replication required): The easiest method for diagnosing this new assumption is to plot the fitted values on the x-axis and the residuals on the y-axis and then examine this plot for any trends (trends indicate an interaction between blocks and treatments, whereas random scatter indicates additive effects).

Assessment:

Independence:

Homogenous variances:

Normality:

Block and Treatment effects:

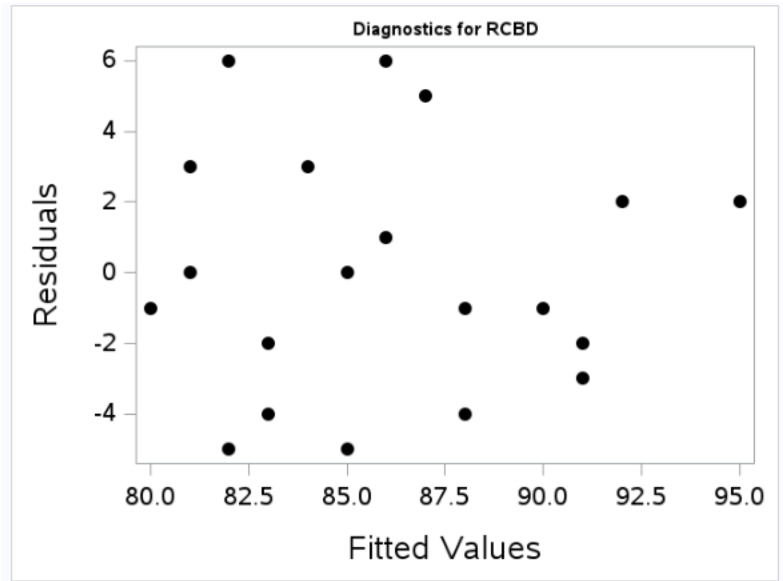

Figure 4: Img

Moments			
N	20	Sum Weights	20
Mean	0	Sum Observations	0
Std Deviation	3.44887472	Variance	11.8947368
Skewness	0.29080495	Kurtosis	-0.8382426
Uncorrected SS	226	Corrected SS	226
Coeff Variation	.	Std Error Mean	0.77119183

Basic Statistical Measures			
Location		Variability	
Mean	0.00000	Std Deviation	3.44887
Median	-0.50000	Variance	11.89474
Mode	-1.00000	Range	11.00000
		Interquartile Range	5.00000

Figure 5: Img

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.950472	Pr < W	0.3743
Kolmogorov-Smirnov	D	0.114073	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.03563	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.275682	Pr > A-Sq	>0.2500

Figure 6: Img

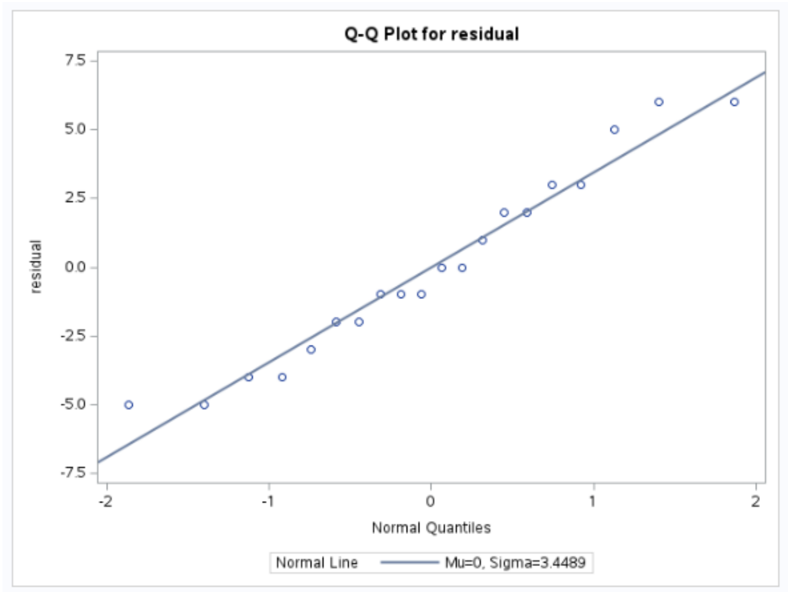

Figure 7: Img

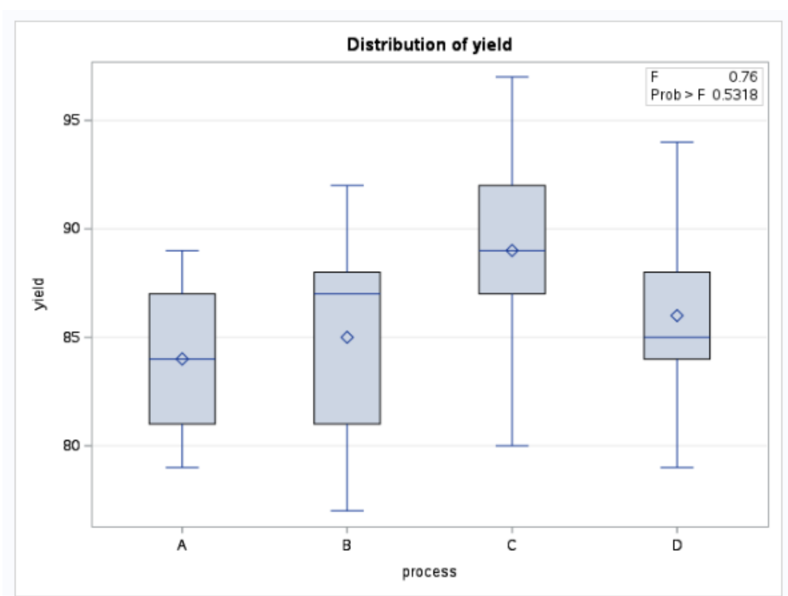

Figure 8: Img

(e)

Did blocking help? Compare the RCBD design with a design dropping the block effect and summarize your findings. Include a discussion of efficiency

Inference for RCBD using PROC GLM

The GLM Procedure

Dependent Variable: yield

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	334.0000000	47.7142857	2.53	0.0754
Error	12	226.0000000	18.8333333		
Corrected Total	19	560.0000000			

R-Square	Coeff Var	Root MSE	yield Mean
0.596429	5.046208	4.339739	86.00000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
batch	4	264.0000000	66.0000000	3.50	0.0407
process	3	70.0000000	23.3333333	1.24	0.3387

Figure 9: Img

Results after dropping the blocks

The GLM Procedure

Dependent Variable: yield

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	70.0000000	23.3333333	0.76	0.5318
Error	16	490.0000000	30.6250000		
Corrected Total	19	560.0000000			

R-Square	Coeff Var	Root MSE	yield Mean
0.125000	6.434867	5.533986	86.00000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
process	3	70.00000000	23.33333333	0.76	0.5318

Figure 10: Img

Blocking (using batch) significantly improves the efficiency of the model, as demonstrated by a higher R-squared (59.84% vs. 12.5%) and a lower Root MSE (4.33 vs. 5.53). The block effect (batch) is statistically significant (P = 0.0407), which indicates that it captures important variability in the data. The process factor is not significant in either model, but blocking leads to a more precise and efficient estimation of the treatment effects. Therefore, the RCBD design with blocks is more effective than the model without blocks.

Efficiency Gain from Blocking: The model with blocking is clearly more efficient. The lower Root MSE and higher R-squared show that the model with blocks provides a better fit and explains more of the variability in yield. Blocking reduces unexplained error variance, leading to a more precise estimate of treatment effects.

Significance of Blocking: The batch (block) effect is significant, meaning that blocking helps account for variability that would otherwise inflate the error term in the unblocked model. This makes the RCBD design more effective at detecting true differences between the treatments.

Process Effect: In both models, the process effect is not statistically significant. However, the RCBD model still performs better overall due to the inclusion of blocks, which capture important variation not explained by the process alone.

Q2

The full SAS program to analyze the brome data is provided in the brome Lab7.sas file located in our course's shared folder in SAS Studio. Refer to the output (and modify the code where necessary) to complete the following exercises:

(a)

Find the full ANOVA table and provide an appropriate summary of results for analyzing the different management plans on the yield of brome.

Analysis of Latin Square

The GLM Procedure

Dependent Variable: yield

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	2387.749175	265.305464	109.11	<.0001
Error	6	14.589400	2.431567		
Corrected Total	15	2402.338575			

R-Square	Coeff Var	Root MSE	yield Mean
0.993927	3.142342	1.559348	49.62375

Source	DF	Type I SS	Mean Square	F Value	Pr > F
freeway	3	3.015725	1.005242	0.41	0.7497
stream	3	108.960825	36.320275	14.94	0.0034
trt	3	2275.772625	758.590875	311.98	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
freeway	3	3.015725	1.005242	0.41	0.7497
stream	3	108.960825	36.320275	14.94	0.0034
trt	3	2275.772625	758.590875	311.98	<.0001

Figure 11: Img

(b)

Is there any difference between the management plan of "in situ" versus the other plans? Provide appropriate output from SAS and summarize your findings.

Analysis of Latin Square

The GLM Procedure

Dependent Variable: yield

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
insitu - rest	1	13.63200833	13.63200833	5.61	0.0557

Parameter	Estimate	Standard Error	t Value	Pr > |t|
insitu - rest	2.13166667	0.90029008	2.37	0.0557

Figure 12: Img

(c)

Did the consideration of the stream block help? Compare the LS design with a design dropping the stream effect and summarize your findings. Include a discussion of efficiency.

RCBD Analysis after dropping stream

The GLM Procedure

Dependent Variable: yield

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	2278.788350	379.798058	27.67	<.0001
Error	9	123.550225	13.727803		
Corrected Total	15	2402.338575			

R-Square	Coeff Var	Root MSE	yield Mean
0.948571	7.466395	3.705105	49.62375

Source	DF	Type I SS	Mean Square	F Value	Pr > F
freeway	3	3.015725	1.005242	0.07	0.9729
trt	3	2275.772625	758.590875	55.26	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
freeway	3	3.015725	1.005242	0.07	0.9729
trt	3	2275.772625	758.590875	55.26	<.0001

Figure 13: Img

Analysis of Latin Square

The GLM Procedure
Least Squares Means

trt	yield LSMEAN	Standard Error	Pr > |t|
harvest	62.5250000	0.7796741	<.0001
insitu	51.2225000	0.7796741	<.0001
uncut	30.2325000	0.7796741	<.0001
windrow	54.5150000	0.7796741	<.0001

Figure 14: Img

RCBD Analysis after dropping stream

The GLM Procedure
Least Squares Means

trt	yield LSMEAN	Standard Error	Pr > |t|
harvest	62.5250000	1.8525525	<.0001
insitu	51.2225000	1.8525525	<.0001
uncut	30.2325000	1.8525525	<.0001
windrow	54.5150000	1.8525525	<.0001

Figure 15: Img

The “stream” factor is significant (P = 0.0041) in the LS design, meaning it has a notable impact on yield. Meanwhile, the treatment (trt) remains highly significant in both designs. The freeway factor is not significant in either design.

The RCBD model shows the following sums of squares for yield: Freeway: 3.015725 Trt: 2275.772825

In the LS design: Freeway: 3.015725 Stream: 108.080825 Trt: 2275.772825

The inclusion of the “stream” factor adds significant explanatory power to the model, as shown by the Type III sum of squares for stream (108.08). This shows that the stream factor accounts for a meaningful portion of the variability in the data, which the RCBD design without stream misses.

The inclusion of the “stream” block improves the efficiency of the design. This is evident from the improved R-squared, lower Root MSE, and lower coefficient of variation in the Latin Square design. In statistical models, efficiency typically refers to minimizing the error variance, and the LS design achieves this with a lower Root MSE (1.5590 vs. 3.7051).

The stream factor’s significance (P = 0.0041) suggests that its inclusion helps to control for variability that would otherwise be attributed to error, leading to a more efficient use of experimental data.

By considering the stream as a block, the model becomes more effective at isolating the treatment effect (trt), as evidenced by the much higher F value for treatment in the LS design (311.98 vs. 55.68).