Chapter 4
Randomized Blocks, Latin Squares, and Related Designs
Solutions

4.1. The ANOVA from a randomized complete block experiment output is shown below.

Source	*DF*	*SS*	*MS*	*F*	*P*
Treatment	4	1010.56	?	29.84	?
Block	?	?	64.765	?	?
Error	20	169.33	?		
Total	29	1503.71			

(a) Fill in the blanks. You may give bounds on the *P*-value.

Completed table is:

Source	*DF*	*SS*	*MS*	*F*	*P*
Treatment	4	1010.56	252.640	29.84	< 0.00001
Block	5	323.82	64.765		
Error	20	169.33	8.467		
Total	29	1503.71			

(b) How many blocks were used in this experiment?

Six blocks were used.

(c) What conclusions can you draw?

The treatment effect is significant; the means of the five treatments are not all equal.

4.2. Consider the single-factor completely randomized experiment shown in Problem 3.4. Suppose that this experiment had been conducted in a randomized complete block design, and that the sum of squares for blocks was 80.00. Modify the ANOVA for this experiment to show the correct analysis for the randomized complete block experiment.

The modified ANOVA is shown below:

Source	DF	SS	MS	F	P
Treatment	4	987.71	246.93	46.3583	< 0.00001
Block	5	80.00	16.00		
Error	20	106.53	5.33		
Total	29	1174.24			

4.3. A chemist wishes to test the effect of four chemical agents on the strength of a particular type of cloth. Because there might be variability from one bolt to another, the chemist decides to use a randomized block design, with the bolts of cloth considered as blocks. She selects five bolts and applies all four chemicals in random order to each bolt. The resulting tensile strengths follow. Analyze the data from this experiment (use $\alpha = 0.05$) and draw appropriate conclusions.

	Bolt				
Chemical	1	2	3	4	5
1	73	68	74	71	67
2	73	67	75	72	70
3	75	68	78	73	68
4	73	71	75	75	69

Design Expert Output

Response: Strength
ANOVA for Selected Factorial Model
Analysis of variance table [Partial sum of squares]

Source	**Sum of Squares**	**DF**	**Mean Square**	**F Value**	**Prob > F**	
Block	157.00	4	39.25			
Model	12.95	3	4.32	2.38	0.1211	not significant
A	*12.95*	*3*	*4.32*	*2.38*	*0.1211*	
Residual	21.80	12	1.82			
Cor Total	191.75	19				

The "Model F-value" of 2.38 implies the model is not significant relative to the noise. There is a 12.11 % chance that a "Model F-value" this large could occur due to noise.

Std. Dev.	1.35	R-Squared	0.3727
Mean	71.75	Adj R-Squared	0.2158
C.V.	1.88	Pred R-Squared	-0.7426
PRESS	60.56	Adeq Precision	10.558

Treatment Means (Adjusted, If Necessary)

	Estimated Mean	**Standard Error**
1-1	70.60	0.60
2-2	71.40	0.60
3-3	72.40	0.60
4-4	72.60	0.60

Treatment	**Mean Difference**	**DF**	**Standard Error**	**t for H_0 Coeff=0**	**Prob > |t|**
1 vs 2	-0.80	1	0.85	-0.94	0.3665
1 vs 3	-1.80	1	0.85	-2.11	0.0564
1 vs 4	-2.00	1	0.85	-2.35	0.0370
2 vs 3	-1.00	1	0.85	-1.17	0.2635
2 vs 4	-1.20	1	0.85	-1.41	0.1846
3 vs 4	-0.20	1	0.85	-0.23	0.8185

There is no difference among the chemical types at $\alpha = 0.05$ level.

4.4. Three different washing solutions are being compared to study their effectiveness in retarding bacteria growth in five-gallon milk containers. The analysis is done in a laboratory, and only three trials can be run on any day. Because days could represent a potential source of variability, the experimenter decides to use a randomized block design. Observations are taken for four days, and the data are shown here. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

	Days			
Solution	1	2	3	4
1	13	22	18	39
2	16	24	17	44
3	5	4	1	22

Design Expert Output

Response: Growth
ANOVA for Selected Factorial Model
Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	
Block	1106.92	3	368.97			
Model	703.50	2	351.75	40.72	0.0003	significant
A	*703.50*	*2*	*351.75*	*40.72*	*0.0003*	
Residual	51.83	6	8.64			
Cor Total	1862.25	11				

The Model F-value of 40.72 implies the model is significant. There is only a 0.03% chance that a "Model F-Value" this large could occur due to noise.

Std. Dev.	2.94	R-Squared	0.9314
Mean	18.75	Adj R-Squared	0.9085
C.V.	15.68	Pred R-Squared	0.7255
PRESS	207.33	Adeq Precision	19.687

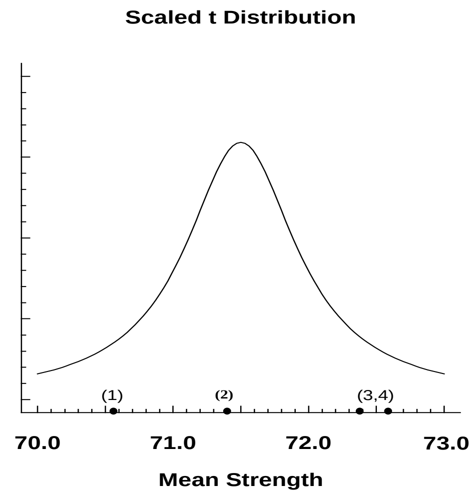
Treatment Means (Adjusted, If Necessary)

	Estimated Mean	Standard Error
1-1	23.00	1.47
2-2	25.25	1.47
3-3	8.00	1.47

Treatment	Mean Difference	DF	Standard Error	t for H_0 Coeff=0	Prob > |t|
1 vs 2	-2.25	1	2.08	-1.08	0.3206
1 vs 3	15.00	1	2.08	7.22	0.0004
2 vs 3	17.25	1	2.08	8.30	0.0002

There is a difference between the means of the three solutions. The Fisher LSD procedure indicates that solution 3 is significantly different than the other two.

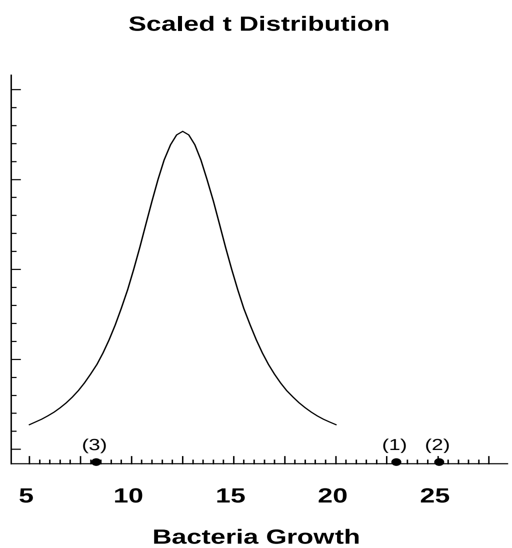
4.5. Plot the mean tensile strengths observed for each chemical type in Problem 4.3 and compare them to a scaled t distribution. What conclusions would you draw from the display?



$$S_{\bar{y}_{i.}} = \sqrt{\frac{MS_E}{b}} = \sqrt{\frac{1.82}{5}} = 0.603$$

There is no obvious difference between the means. This is the same conclusion given by the analysis of variance.

4.6. Plot the average bacteria counts for each solution in Problem 4.4 and compare them to an appropriately scaled t distribution. What conclusions can you draw?



$$S_{\bar{y}_{i.}} = \sqrt{\frac{MS_E}{b}} = \sqrt{\frac{8.64}{4}} = 1.47$$

There is no difference in mean bacteria growth between solutions 1 and 2. However, solution 3 produces significantly lower mean bacteria growth. This is the same conclusion reached from the Fisher LSD procedure in Problem 4.4.

4.7. Consider the hardness testing experiment described in Section 4.1. Suppose that the experiment was conducted as described and the following Rockwell C-scale data (coded by subtracting 40 units) obtained:

	Coupon			
Tip	1	2	3	4
1	9.3	9.4	9.6	10.0
2	9.4	9.3	9.8	9.9
3	9.2	9.4	9.5	9.7
4	9.7	9.6	10.0	10.2

(a) Analyize the data from this experiment.

There is a difference between the means of the four tips.

Design Expert Output

Response: Hardness

ANOVA for Selected Factorial Model

Analysis of variance table [Terms added sequentially (first to last)]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	
Bock	0.82	3	0.27			
Model	0.38	3	0.13	14.44	0.0009	significant
A	0.38	3	0.13	14.44	0.0009	
Residual	0.080	9	8.889E-003			
Cor Total	1.29	15				

The Model F-value of 14.44 implies the model is significant. There is only a 0.09% chance that a "Model F-Value" this large could occur due to noise.

Std. Dev.	0.094	R-Squared	0.8280
Mean	9.63	Adj R-Squared	0.7706
C.V.	0.98	Pred R-Squared	0.4563
PRESS	0.25	Adeq Precision	15.635

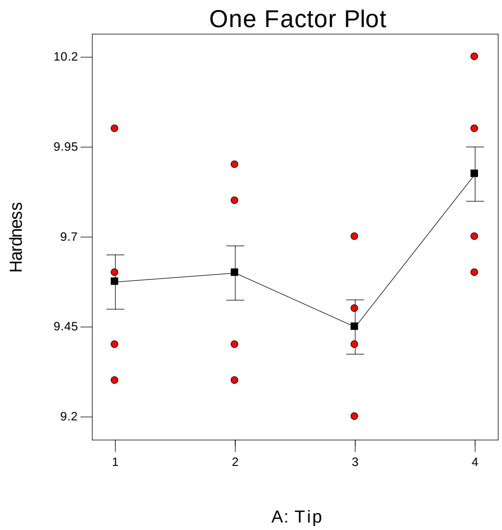
Treatment Means (Adjusted, If Necessary)

	Estimated Mean	Standard Error
1-1	9.57	0.047
2-2	9.60	0.047
3-3	9.45	0.047
4-4	9.88	0.047

Treatment	Mean Difference	DF	Standard Error	t for H0 Coeff=0	Prob > |t|
1 vs 2	-0.025	1	0.067	-0.38	0.7163
1 vs 3	0.13	1	0.067	1.87	0.0935
1 vs 4	-0.30	1	0.067	-4.50	0.0015
2 vs 3	0.15	1	0.067	2.25	0.0510
2 vs 4	-0.27	1	0.067	-4.12	0.0026
3 vs 4	-0.43	1	0.067	-6.37	0.0001

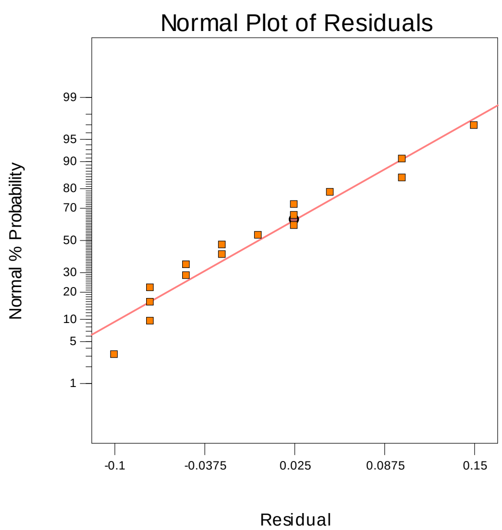
(b) Use the Fisher LSD method to make comparisons among the four tips to determine specifically which tips differ in mean hardness readings.

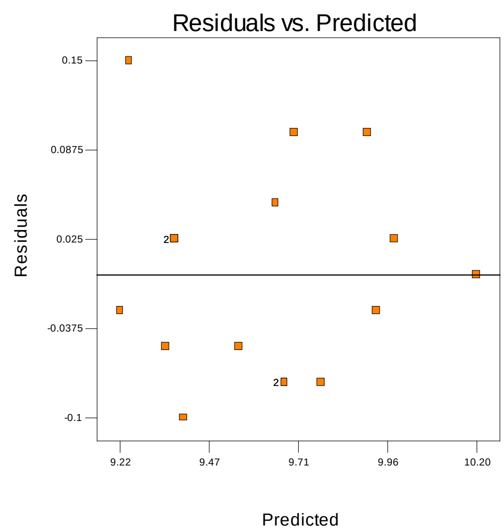
Based on the LSD bars in the Design Expert plot below, the mean of tip 4 differs from the means of tips 1, 2, and 3. The LSD method identifies a marginal difference between the means of tips 2 and 3.

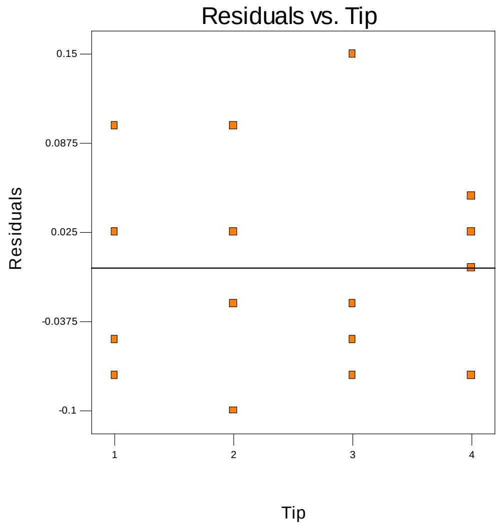


(c) Analyze the residuals from this experiment.

The residual plots below do not identify any violations to the assumptions.







4.8. A consumer products company relies on direct mail marketing pieces as a major component of its advertising campaigns. The company has three different designs for a new brochure and want to evaluate their effectiveness, as there are substantial differences in costs between the three designs. The company decides to test the three designs by mailing 5,000 samples of each to potential customers in four different regions of the country. Since there are known regional differences in the customer base, regions are considered as blocks. The number of responses to each mailing is shown below.

	Region			
Design	NE	NW	SE	SW
1	250	350	219	375
2	400	525	390	580
3	275	340	200	310

(a) Analyze the data from this experiment.

The residuals of the analsysis below identify concerns with the normality and equality of variance assumptions. As a result, a square root transformation was applied as shown in the second ANOVA table. The residuals of both analysis are presented for comparison in part (c) of this problem. The analysis concludes that there is a difference between the mean number of responses for the three designs.

Design Expert Output

Response: Number of responses
ANOVA for Selected Factorial Model
Analysis of variance table [Terms added sequentially (first to last)]

Source	**Sum of Squares**	**DF**	**Mean Square**	**F Value**	**Prob > F**	
Block	49035.67	3	16345.22			
Model	90755.17	2	45377.58	50.15	0.0002	significant
A	*90755.17*	*2*	*45377.58*	*50.15*	*0.0002*	
Residual	5428.83	6	904.81			
Cor Total	1.452E+005	11				

The Model F-value of 50.15 implies the model is significant. There is only a 0.02% chance that a "Model F-Value" this large could occur due to noise.

Std. Dev.	30.08	R-Squared	0.9436
Mean	351.17	Adj R-Squared	0.9247
C.V.	8.57	Pred R-Squared	0.7742
PRESS	21715.33	Adeq Precision	16.197

Treatment Means (Adjusted, If Necessary)

	Estimated Mean	**Standard Error**
1-1	298.50	15.04
2-2	473.75	15.04
3-3	281.25	15.04

Treatment	**Mean Difference**	**DF**	**Standard Error**	**t for H0 Coeff=0**	**Prob > |t|**
1 vs 2	-175.25	1	21.27	-8.24	0.0002
1 vs 3	17.25	1	21.27	0.81	0.4483
2 vs 3	192.50	1	21.27	9.05	0.0001

Design Expert Output for Transformed Data

Response: Number of responses Transform: Square root Constant: 0

ANOVA for Selected Factorial Model

Analysis of variance table [Terms added sequentially (first to last)]

Source	**Sum of Squares**	**DF**	**Mean Square**	**F Value**	**Prob > F**	
Block	35.89	3	11.96			
Model	60.73	2	30.37	60.47	0.0001	significant
A	*60.73*	*2*	*30.37*	*60.47*	*0.0001*	
Residual	3.01	6	0.50			
Cor Total	99.64	11				

The Model F-value of 60.47 implies the model is significant. There is only a 0.01% chance that a "Model F-Value" this large could occur due to noise.

Std. Dev.	0.71	R-Squared	0.9527
Mean	18.52	Adj R-Squared	0.9370
C.V.	3.83	Pred R-Squared	0.8109
PRESS	12.05	Adeq Precision	18.191

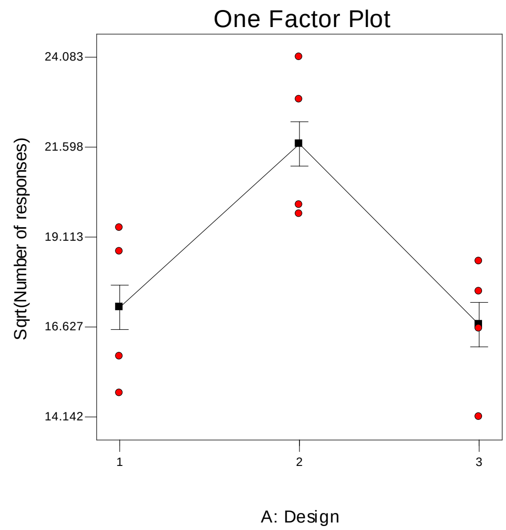
Treatment Means (Adjusted, If Necessary)

	Estimated Mean	**Standard Error**
1-1	17.17	0.35
2-2	21.69	0.35
3-3	16.69	0.35

Treatment	**Mean Difference**	**DF**	**Standard Error**	**t for H0 Coeff=0**	**Prob > |t|**
1 vs 2	-4.52	1	0.50	-9.01	0.0001
1 vs 3	0.48	1	0.50	0.95	0.3769
2 vs 3	4.99	1	0.50	9.96	< 0.0001

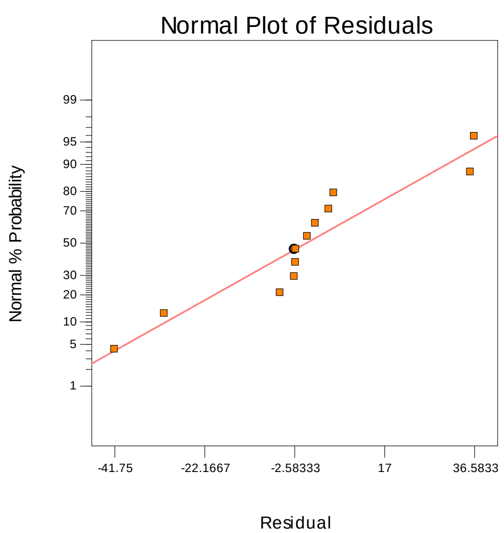
(b) Use the Fisher LSD method to make comparisons among the three designs to determine specifically which designs differ in mean response rate.

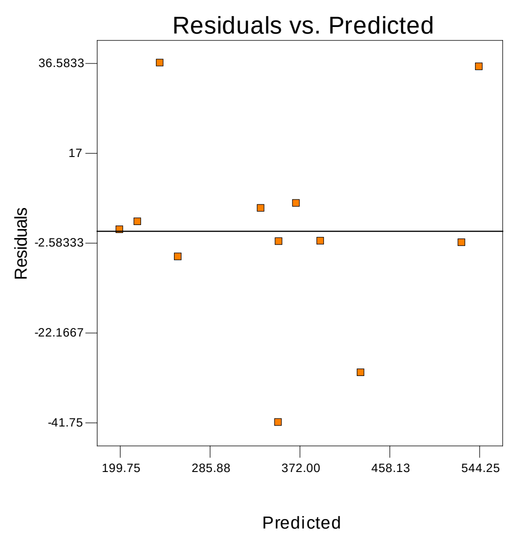
Based on the LSD bars in the Design Expert plot below, designs 1 and 3 do not differ; however, design 2 is different than designs 1 and 3.

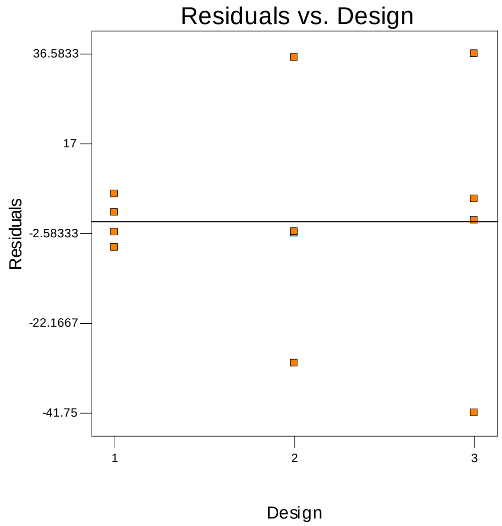


(c) Analyze the residuals from this experiment.

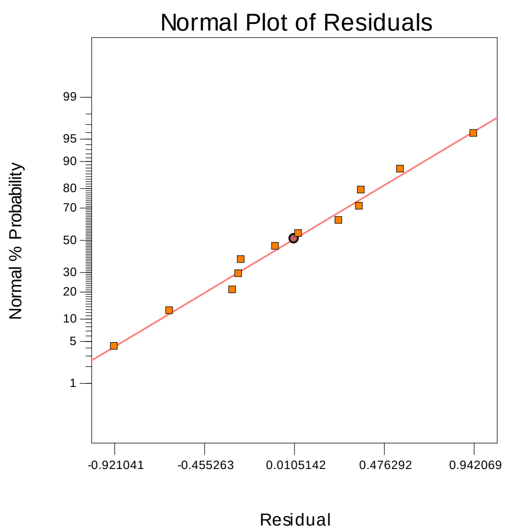
The first set of residual plots presented below represent the untransformed data. Concerns with normality as well as inequality of variance are presented. The second set of residual plots represent transformed data and do not identify significant violations of the assumptions. The residuals vs. design plot indicates a slight inequality of variance; however, not a strong violation and an improvement over the non-transformed data.

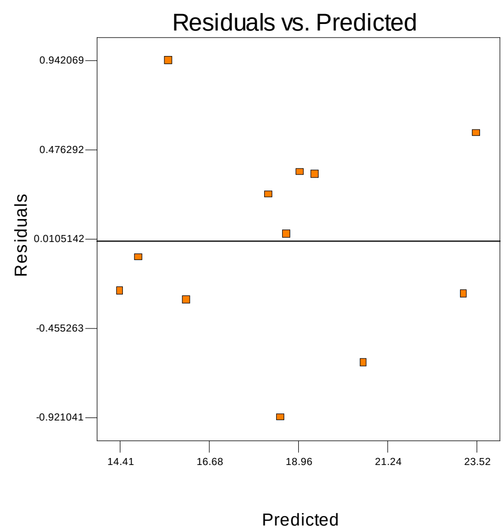


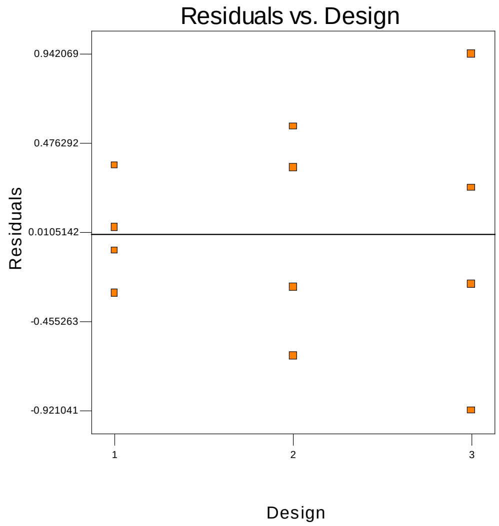




The following are the square root transformed data residual plots.







4.9. The effect of three different lubricating oils on fuel economy in diesel truck engines is being studied. Fuel economy is measured using brake-specific fuel consumption after the engine has been running for 15 minutes. Five different truck engines are available for the study, and the experimenters conduct the following randomized complete block design.

	Truck				
Oil	1	2	3	4	5
1	0.500	0.634	0.487	0.329	0.512
2	0.535	0.675	0.520	0.435	0.540
3	0.513	0.595	0.488	0.400	0.510

(a) Analyze the data from this experiment.

From the analysis below, there is a significant difference between lubricating oils with regards to fuel economy.

Design Expert Output

Response: Fuel consumption
ANOVA for Selected Factorial Model
Analysis of variance table [Terms added sequentially (first to last)]

Source	**Sum of Squares**	**DF**	**Mean Square**	**F Value**	**Prob > F**	
Block	0.092	4	0.023			
Model	6.706E-003	2	3.353E-003	6.35	0.0223	significant
A	*6.706E-003*	*2*	*3.353E-003*	*6.35*	*0.0223*	
Residual	4.222E-003	8	5.278E-004			
Cor Total	0.10	14				

The Model F-value of 6.35 implies the model is significant. There is only a 2.23% chance that a "Model F-Value" this large could occur due to noise.

Std. Dev.	0.023	R-Squared	0.6136
Mean	0.51	Adj R-Squared	0.5170
C.V.	4.49	Pred R-Squared	-0.3583
PRESS	0.015	Adeq Precision	18.814

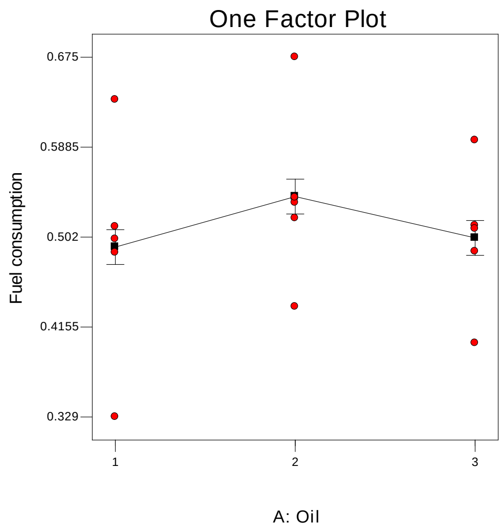
Treatment Means (Adjusted, If Necessary)

	Estimated Mean	Standard Error
1-1	0.49	0.010
2-2	0.54	0.010
3-3	0.50	0.010

Treatment	Mean Difference	DF	Standard Error	t for H_0 Coeff=0	Prob > |t|
1 vs 2	-0.049	1	0.015	-3.34	0.0102
1 vs 3	-8.800E-003	1	0.015	-0.61	0.5615
2 vs 3	0.040	1	0.015	2.74	0.0255

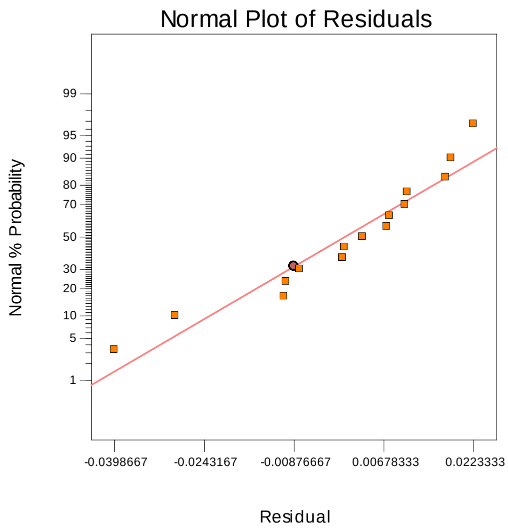
(b) Use the Fisher LSD method to make comparisons among the three lubricating oils to determine specifically which oils differ in break-specific fuel consumption.

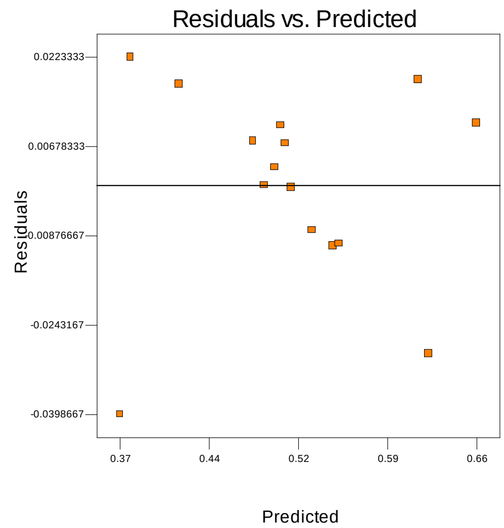
Based on the LSD bars in the Design Expert plot below, the means for break-specific fuel consumption for oils 1 and 3 do not differ; however, oil 2 is different than oils 1 and 3.

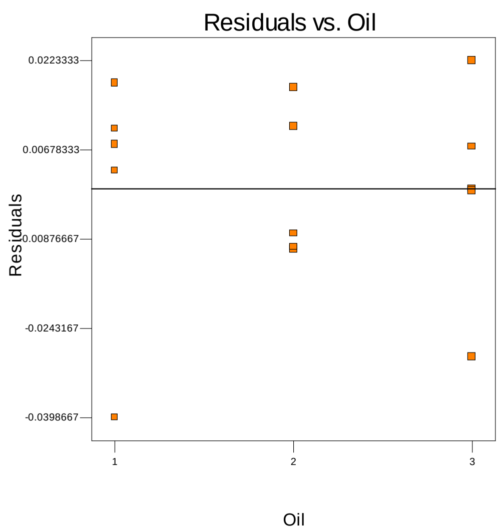


(c) Analyze the residuals from this experiment.

The residual plots below do not identify any violations to the assumptions.







4.10. An article in the *Fire Safety Journal* ("The Effect of Nozzle Design on the Stability and Performance of Turbulent Water Jets," Vol. 4, August 1981) describes an experiment in which a shape factor was determined for several different nozzle designs at six levels of jet efflux velocity. Interest focused on potential differences between nozzle designs, with velocity considered as a nuisance variable. The data are shown below:

	Jet Efflux Velocity (m/s)					
Nozzle Design	11.73	14.37	16.59	20.43	23.46	28.74
1	0.78	0.80	0.81	0.75	0.77	0.78
2	0.85	0.85	0.92	0.86	0.81	0.83
3	0.93	0.92	0.95	0.89	0.89	0.83
4	1.14	0.97	0.98	0.88	0.86	0.83
5	0.97	0.86	0.78	0.76	0.76	0.75

(a) Does nozzle design affect the shape factor? Compare nozzles with a scatter plot and with an analysis of variance, using $\alpha = 0.05$.

Design Expert Output

Response: Shape

ANOVA for Selected Factorial Model

Analysis of variance table [Partial sum of squares]

Source	**Sum of Squares**	**DF**	**Mean Square**	**F Value**	**Prob > F**	
Block	0.063	5	0.013			
Model	0.10	4	0.026	8.92	0.0003	significant
A	*0.10*	*4*	*0.026*	*8.92*	*0.0003*	
Residual	0.057	20	2.865E-003			
Cor Total	0.22	29				

The Model F-value of 8.92 implies the model is significant. There is only a 0.03% chance that a "Model F-Value" this large could occur due to noise.

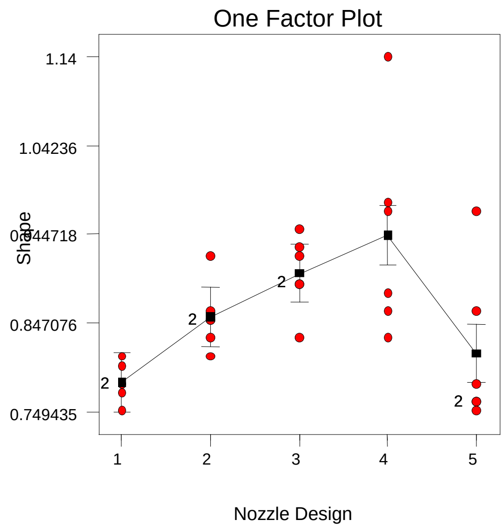
Std. Dev.	0.054	R-Squared	0.6407
Mean	0.86	Adj R-Squared	0.5688
C.V.	6.23	Pred R-Squared	0.1916
PRESS	0.13	Adeq Precision	9.438

Treatment Means (Adjusted, If Necessary)

	Estimated Mean	**Standard Error**
1-1	0.78	0.022
2-2	0.85	0.022
3-3	0.90	0.022
4-4	0.94	0.022
5-5	0.81	0.022

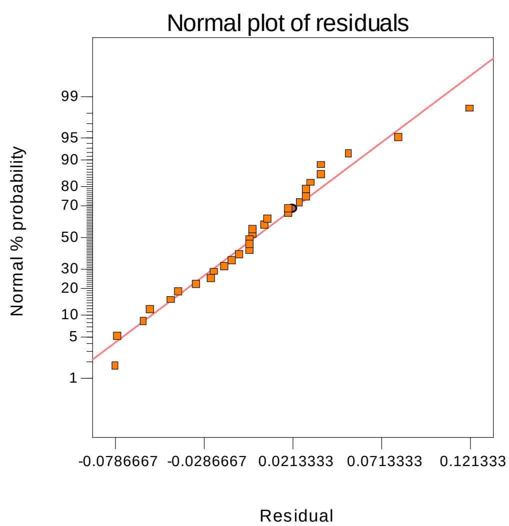
Treatment	**Mean Difference**	**DF**	**Standard Error**	**t for H0 Coeff=0**	**Prob > |t|**
1 vs 2	-0.072	1	0.031	-2.32	0.0311
1 vs 3	-0.12	1	0.031	-3.88	0.0009
1 vs 4	-0.16	1	0.031	-5.23	< 0.0001
1 vs 5	-0.032	1	0.031	-1.02	0.3177
2 vs 3	-0.048	1	0.031	-1.56	0.1335
2 vs 4	-0.090	1	0.031	-2.91	0.0086
2 vs 5	0.040	1	0.031	1.29	0.2103
3 vs 4	-0.042	1	0.031	-1.35	0.1926
3 vs 5	0.088	1	0.031	2.86	0.0097
4 vs 5	0.13	1	0.031	4.21	0.0004

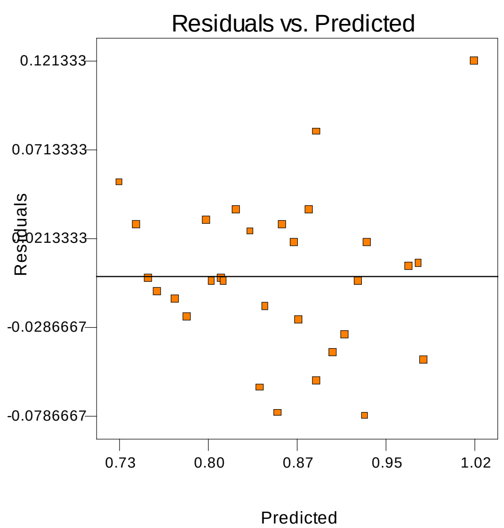
Nozzle design has a significant effect on shape factor.

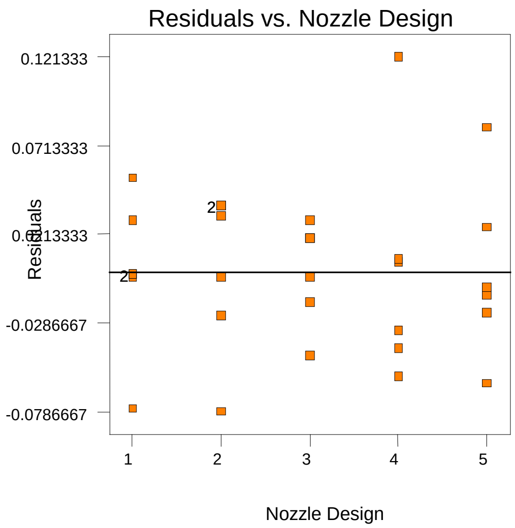


(b) Analyze the residual from this experiment.

The plots shown below do not give any indication of serious problems. Thre is some indication of a mild outlier on the normal probability plot and on the plot of residuals versus the predicted velocity.







(c) Which nozzle designs are different with respect to shape factor? Draw a graph of average shape factor for each nozzle type and compare this to a scaled *t* distribution. Compare the conclusions that you draw from this plot to those from Duncan's multiple range test.

$$S_{\bar{y}_{i.}} = \sqrt{\frac{MS_E}{b}} = \sqrt{\frac{0.002865}{6}} = 0.021852$$

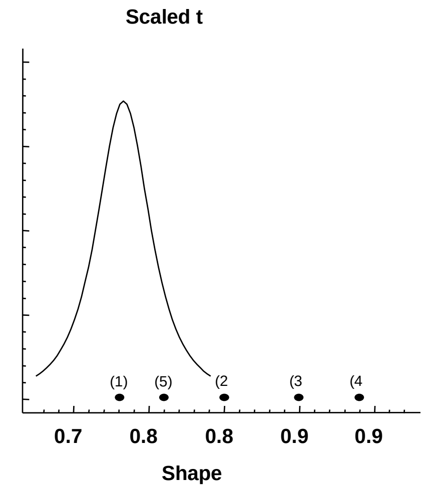
$$R_2 = r_{0.05}(2,20)\ S_{\bar{y}_{i.}} = (2.95)(0.021852) = 0.06446$$

$$R_3 = r_{0.05}(3,20)\ S_{\bar{y}_{i.}} = (3.10)(0.021852) = 0.06774$$

$$R_4 = r_{0.05}(4,20)\ S_{\bar{y}_{i.}} = (3.18)(0.021852) = 0.06949$$

$$R_5 = r_{0.05}(5,20)\ S_{\bar{y}_{i.}} = (3.25)(0.021852) = 0.07102$$

	Mean Difference		*R*	
1 vs 4	0.16167	>	0.07102	different
1 vs 3	0.12000	>	0.06949	different
1 vs 2	0.07167	>	0.06774	different
1 vs 5	0.03167	<	0.06446	
5 vs 4	0.13000	>	0.06949	different
5 vs 3	0.08833	>	0.06774	different
5 vs 2	0.04000	<	0.06446	
2 vs 4	0.09000	>	0.06774	different
2 vs 3	0.04833	<	0.06446	
3 vs 4	0.04167	<	0.06446	



4.11. An article in *Communications of the ACM* (Vol. 30, No. 5, 1987) studied different algorithms for estimating software development costs. Six algorithms were applied to several different software development projects and the percent error in estimating the development cost was observed. Some of the data from this experiment is show in the table below.

	Project					
Algorithm	1	2	3	4	5	6
1 (SLIM)	1244	21	82	2221	905	839
2 (COCOMO-A)	281	129	396	1306	336	910
3 (COCOMO-R)	220	84	458	543	300	794
4 (COCOMO-C)	225	83	425	552	291	826
5 (FUNCTION POINTS)	19	11	-34	121	15	103
6 (ESTIMALS)	-20	35	-53	170	104	199

(a) Do the algorithms differ in their mean cost estimation accuracy?

The ANOVA below identifies the algorithms are significantly different in their mean cost estimation error.

Design Expert Output

Response **Cost Error**
ANOVA for selected factorial model
Analysis of variance table [Classical sum of squares - Type II]

Source	**Sum of Squares**	**df**	**Mean Square**	**F Value**	**p-value Prob > F**	
Block	2.287E+006	5	4.575E+005			
Model	2.989E+006	5	5.978E+005	5.38	0.0017	significant
A-Algorithm	*2.989E+006*	*5*	*5.978E+005*	*5.38*	*0.0017*	
Residual	2.780E+006	25	1.112E+005			
Cor Total	8.056E+006	35				

The Model F-value of 5.38 implies the model is significant. There is only a 0.17% chance that a "Model F-Value" this large could occur due to noise.

Std. Dev.	333.44	R-Squared	0.5182
Mean	392.81	Adj R-Squared	0.4218
C.V. %	84.89	Pred R-Squared	0.0009
PRESS	5.764E+006	Adeq Precision	8.705

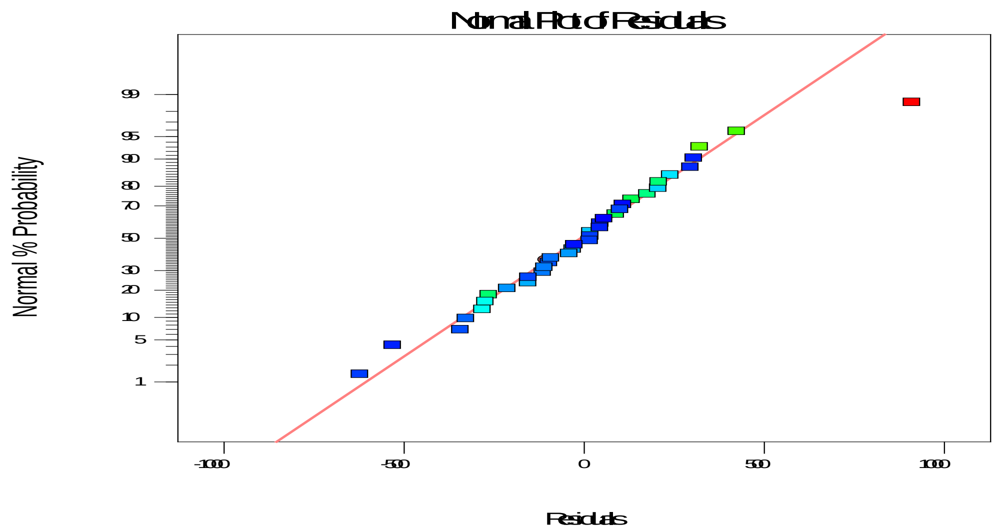
Treatment Means (Adjusted, If Necessary)

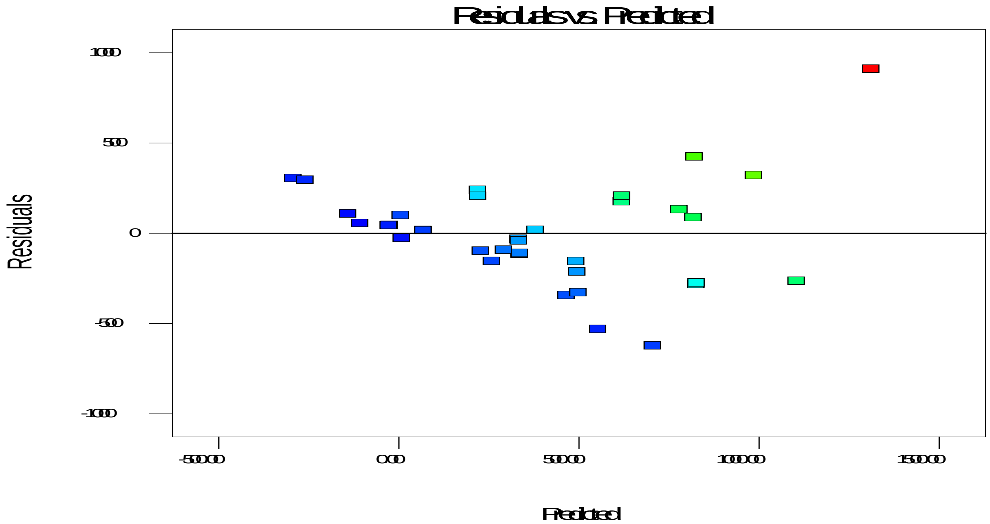
	Estimated Mean	Standard Error
1-SLM	885.33	136.13
2-COCOMO-A	559.67	136.13
3-COCOMO-R	399.83	136.13
4-COCOMO-C	400.33	136.13
5-FUNCTION POINTS	39.17	136.13
6-ESTIMALS	72.50	136.13

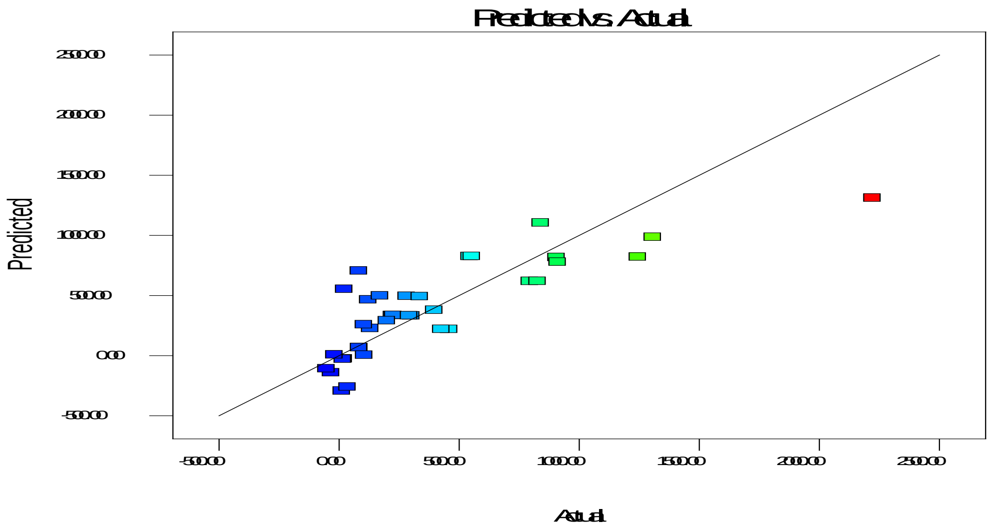
Treatment	Mean Difference	df	Standard Error	t for H0 Coeff=0	Prob > |t|
1 vs 2	325.67	1	192.51	1.69	0.1031
1 vs 3	485.50	1	192.51	2.52	0.0184
1 vs 4	485.00	1	192.51	2.52	0.0185
1 vs 5	846.17	1	192.51	4.40	0.0002
1 vs 6	812.83	1	192.51	4.22	0.0003
2 vs 3	159.83	1	192.51	0.83	0.4143
2 vs 4	159.33	1	192.51	0.83	0.4157
2 vs 5	520.50	1	192.51	2.70	0.0122
2 vs 6	487.17	1	192.51	2.53	0.0181
3 vs 4	-0.50	1	192.51	-2.597E-003	0.9979
3 vs 5	360.67	1	192.51	1.87	0.0727
3 vs 6	327.33	1	192.51	1.70	0.1015
4 vs 5	361.17	1	192.51	1.88	0.0724
4 vs 6	327.83	1	192.51	1.70	0.1010
5 vs 6	-33.33	1	192.51	-0.17	0.8639

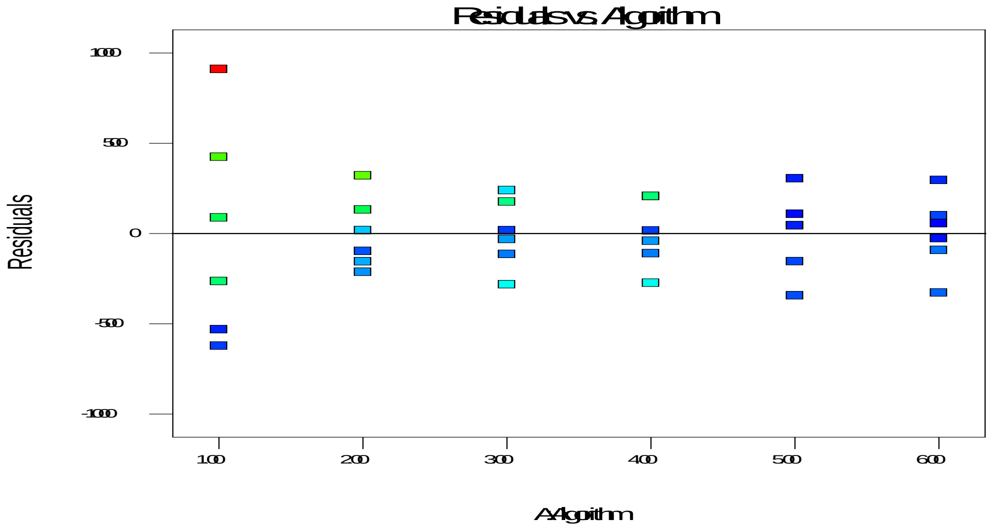
(b) Analyze the residuals from this experiment.

The residual plots below identify a single outlier that should be investigated.




Residuals vs. Predicted
Residuals
Predicted


Predicted vs. Actual
Predicted
Actual



(c) Which algorithm would you recommend for use in practice?

The FUNCTIONAL POINTS algorithm has the losest cost estimation error.

4.12. An article in *Nature Genetics* (2003, Vol. 34, pp. 85-90) "Treatment-Specific Changes in Gene Expression Discriminate in vivo Drug Response in Human Leukemia Cells" studied gene expressionas a function of different treatments for leukemia. Three treatment groups are: mercaptopurine (MP) only; low-dose methotrexate (LDMTX) and MP; and high-dose methotrexate (HDMTX) and MP. Each group contained ten subjects. The responses from a specific gene are shown in the table below:

	Project									
MP ONLY	334.5	31.6	701	41.2	61.2	69.6	67.5	66.6	120.7	881.9
MP + HDMTX	919.4	404.2	1024.8	54.1	62.8	671.6	882.1	354.2	321.9	91.1
MP + LDMTX	108.4	26.1	240.8	191.1	69.7	242.8	62.7	396.9	23.6	290.4

(a) Is there evidence to support the claim that the treatment means differ?

The ANOVA below identifies the treatment means are significantly different.

Design Expert Output

Response **Gene Expression**
ANOVA for selected factorial model
Analysis of variance table [Classical sum of squares - Type II]

Source	**Sum of Squares**	**df**	**Mean Square**	**F Value**	**p-value Prob > F**	
Block	9.206E+005	9	1.023E+005			
Model	5.384E+005	2	2.692E+005	3.68	0.0457	significant
A-Treatment	*5.384E+005*	*2*	*2.692E+005*	*3.68*	*0.0457*	
Residual	1.316E+006	18	73130.15			
Cor Total	2.775E+006	29				

The Model F-value of 3.68 implies the model is significant. There is only
a 4.57% chance that a "Model F-Value" this large could occur due to noise.

Std. Dev.	270.43	R-Squared	0.2903
Mean	293.82	Adj R-Squared	0.2114
C.V. %	92.04	Pred R-Squared	-0.9714
PRESS	3.657E+006	Adeq Precision	5.288

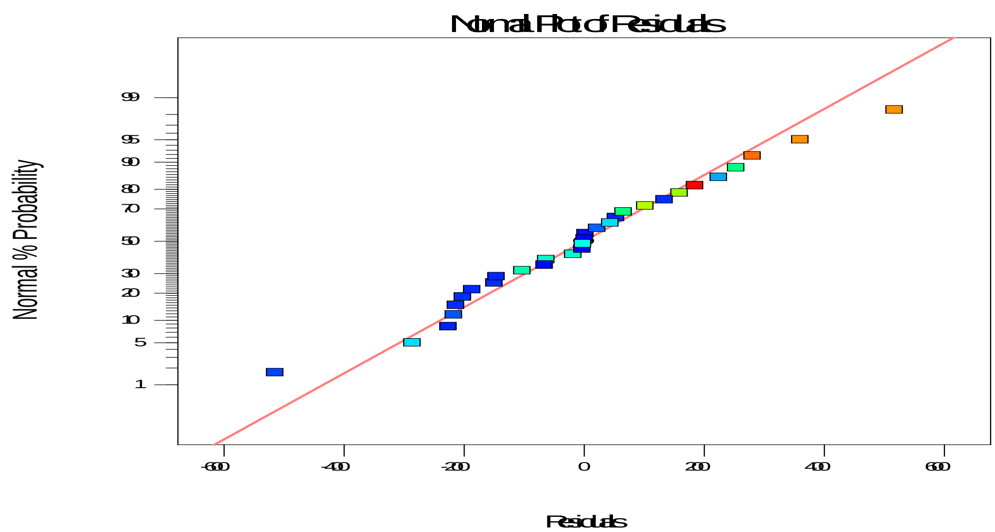
Treatment Means (Adjusted, If Necessary)

	Estimated Mean	Standard Error
1-MP Only	237.58	85.52
2-MP + HDMTX	478.62	85.52
3-MP + LDMTX	165.25	85.52

Treatment	Mean Difference	df	Standard Error	t for H0 Coeff=0	Prob > |t|
1 vs 2	-241.04	1	120.94	-1.99	0.0616
1 vs 3	72.33	1	120.94	0.60	0.5572
2 vs 3	313.37	1	120.94	2.59	0.0184

(b) Chec the normality assumption. Can we assume these samples are from normal populations?

The normal plot of residuals below identifies a slightly non-normal distribution.



(c) Take the logarithm of the raw data. Is there evidence to support the claim that the treatment means differ for the transformed data?

The ANOVA for the natural log transformed data identifies the treatment means as only moderately different with an *F* value of 0.07

Design Expert Output

Response **Gene Expression**
Transform:Natural Log Constant: 0
ANOVA for selected factorial model
Analysis of variance table [Classical sum of squares - Type II]

Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F
Block	14.75	9	1.64		
Model	6.30	2	3.15	3.09	0.0700
A-Treatment	*6.30*	*2*	*3.15*	*3.09*	*0.0700*
Residual	18.32	18	1.02		
Cor Total	39.37	29			

The Model F-value of 3.09 implies there is a 7.00% chance that a "Model F-Value" this large could occur due to noise.

Std. Dev.	1.01	R-Squared	0.2558
Mean	5.09	Adj R-Squared	0.1731
C.V. %	19.83	Pred R-Squared	-1.0672
PRESS	50.89	Adeq Precision	4.942

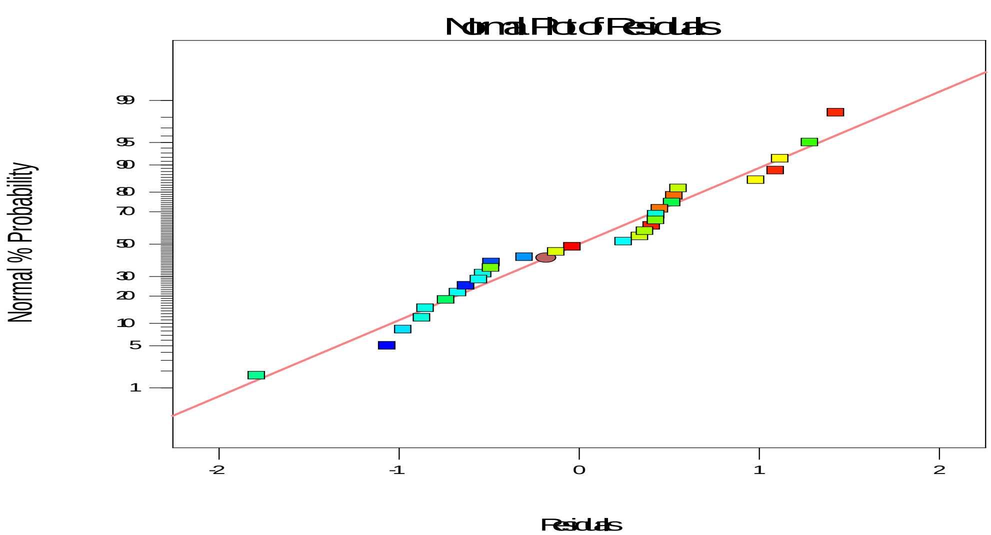
Treatment Means (Adjusted, If Necessary)

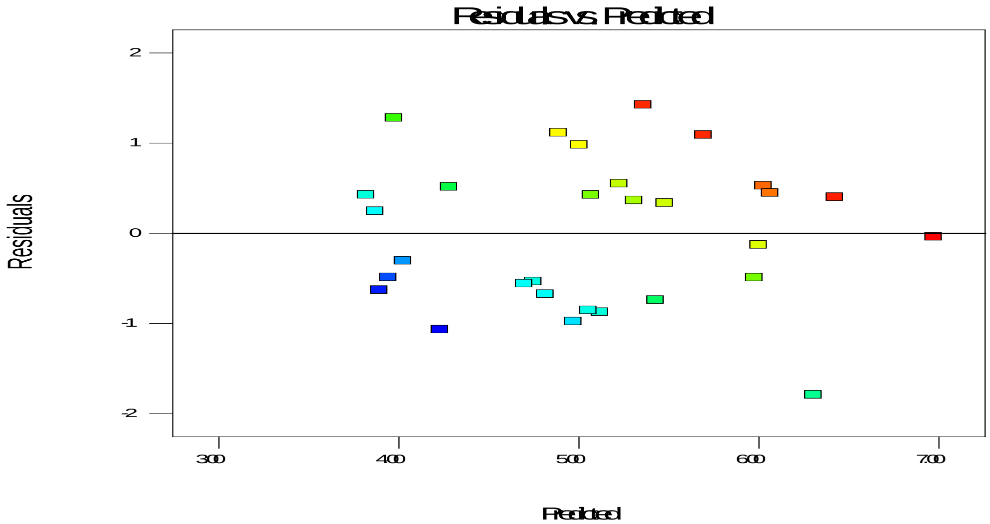
	Estimated Mean	Standard Error
1-MP Only	4.79	0.32
2-MP + HDMTX	5.73	0.32
3-MP + LDMTX	4.74	0.32

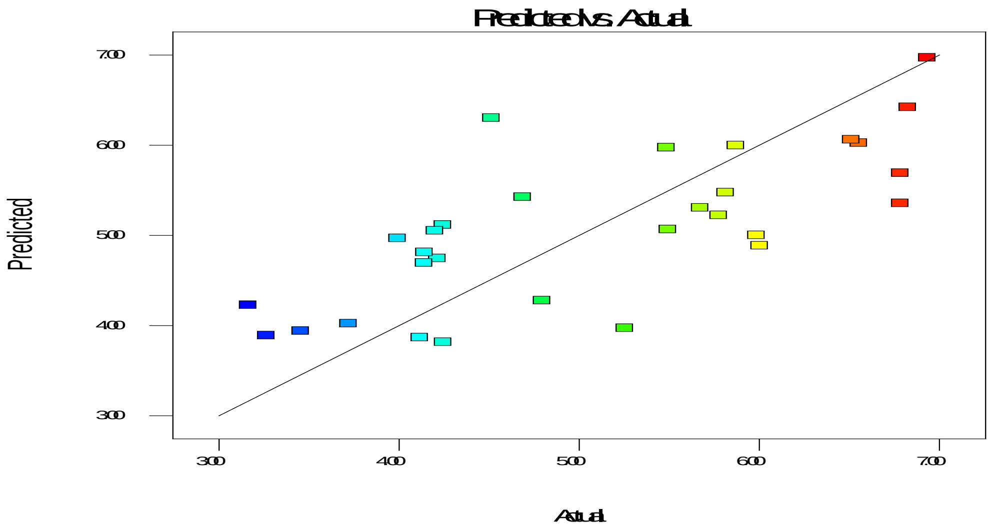
Treatment	Mean Difference	df	Standard Error	t for H0 Coeff=0	Prob > |t|
1 vs 2	-0.95	1	0.45	-2.10	0.0505
1 vs 3	0.050	1	0.45	0.11	0.9122
2 vs 3	1.00	1	0.45	2.21	0.0405

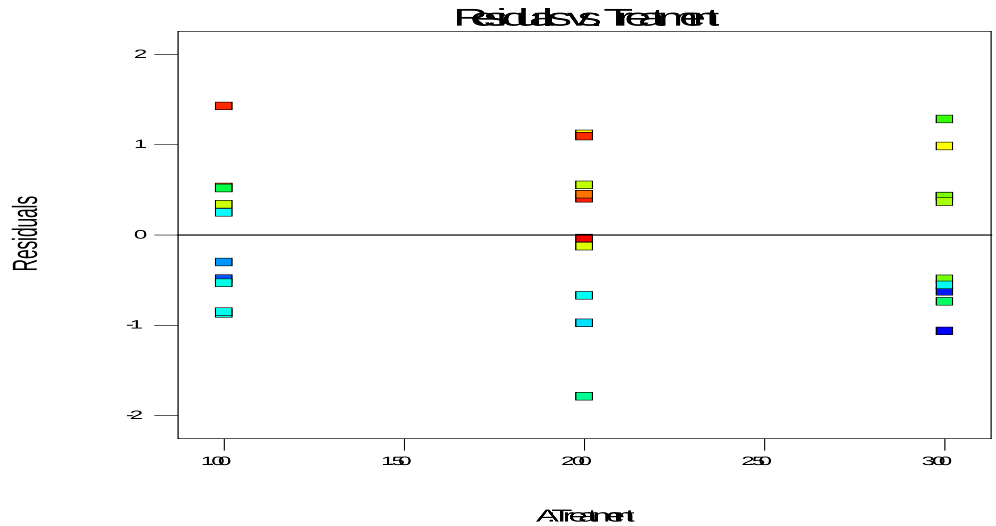
(d) Analyze the residuals from the transformed data and comment on model adequacy.

The residual plots below identify no concerns with the model adequacy.




Residuals vs. Predicted
Residuals
2
1
0
-1
-2
300
400
500
600
700
Predicted


Predicted vs. Actual
Predicted
700
600
500
400
300
300
400
500
600
700
Actual



4.13. Consider the ratio control algorithm experiment described in Section 3.8. The experiment was actually conducted as a randomized block design, where six time periods were selected as the blocks, and all four ratio control algorithms were tested in each time period. The average cell voltage and the standard deviation of voltage (shown in parentheses) for each cell are as follows:

Ratio Control	Time Period					
Algorithms	1	2	3	4	5	6
1	4.93 (0.05)	4.86 (0.04)	4.75 (0.05)	4.95 (0.06)	4.79 (0.03)	4.88 (0.05)
2	4.85 (0.04)	4.91 (0.02)	4.79 (0.03)	4.85 (0.05)	4.75 (0.03)	4.85 (0.02)
3	4.83 (0.09)	4.88 (0.13)	4.90 (0.11)	4.75 (0.15)	4.82 (0.08)	4.90 (0.12)
4	4.89 (0.03)	4.77 (0.04)	4.94 (0.05)	4.86 (0.05)	4.79 (0.03)	4.76 (0.02)

(a) Analyze the average cell voltage data. (Use α = 0.05.) Does the choice of ratio control algorithm affect the cell voltage?

Design Expert Output

Response: Average
ANOVA for Selected Factorial Model
Analysis of variance table [Partial sum of squares]

Source	**Sum of Squares**	**DF**	**Mean Square**	**F Value**	**Prob > F**	
Block	0.017	5	3.487E-003			
Model	2.746E-003	3	9.153E-004	0.19	0.9014	not significant
A	*2.746E-003*	*3*	*9.153E-004*	*0.19*	*0.9014*	
Residual	0.072	15	4.812E-003			
Cor Total	0.092	23				

The "Model F-value" of 0.19 implies the model is not significant relative to the noise. There is a 90.14 % chance that a "Model F-value" this large could occur due to noise.

Std. Dev.	0.069	R-Squared	0.0366
Mean	4.84	Adj R-Squared	-0.1560
C.V.	1.43	Pred R-Squared	-1.4662
PRESS	0.18	Adeq Precision	2.688

Treatment Means (Adjusted, If Necessary)

	Estimated Mean	**Standard Error**

1-1	4.86	0.028
2-2	4.83	0.028
3-3	4.85	0.028
4-4	4.84	0.028

Treatment	Mean Difference	DF	Standard Error	t for H0 Coeff=0	Prob > |t|
1 vs 2	0.027	1	0.040	0.67	0.5156
1 vs 3	0.013	1	0.040	0.33	0.7438
1 vs 4	0.025	1	0.040	0.62	0.5419
2 vs 3	-0.013	1	0.040	-0.33	0.7438
2 vs 4	-1.667E-003	1	0.040	-0.042	0.9674
3 vs 4	0.012	1	0.040	0.29	0.7748

The ratio control algorithm does not affect the mean cell voltage.

(b) Perform an appropriate analysis of the standard deviation of voltage. (Recall that this is called "pot noise.") Does the choice of ratio control algorithm affect the pot noise?

Design Expert Output

Response: StDev Transform: Natural log Constant: 0.000

ANOVA for Selected Factorial Model

Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	
Block	0.94	5	0.19			
Model	6.17	3	2.06	33.26	< 0.0001	significant
A	*6.17*	*3*	*2.06*	*33.26*	*< 0.0001*	
Residual	0.93	15	0.062			
Cor Total	8.04	23				

The Model F-value of 33.26 implies the model is significant. There is only a 0.01% chance that a "Model F-Value" this large could occur due to noise.

Std. Dev.	0.25	R-Squared	0.8693
Mean	-3.04	Adj R-Squared	0.8432
C.V.	-8.18	Pred R-Squared	0.6654
PRESS	2.37	Adeq Precision	12.446

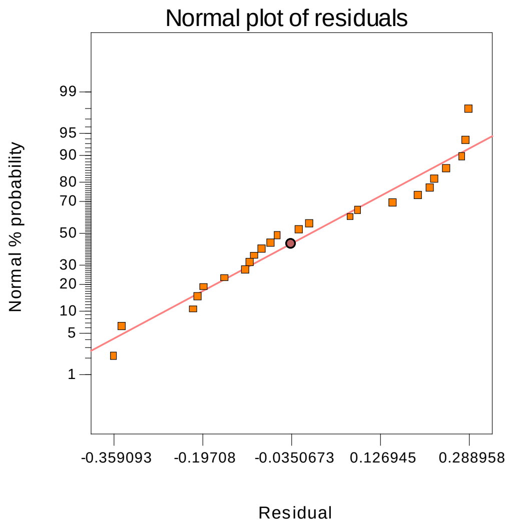
Treatment Means (Adjusted, If Necessary)

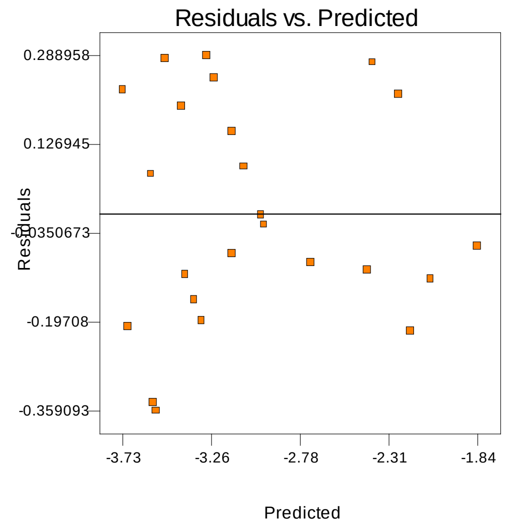
	Estimated Mean	Standard Error
1-1	-3.09	0.10
2-2	-3.51	0.10
3-3	-2.20	0.10
4-4	-3.36	0.10

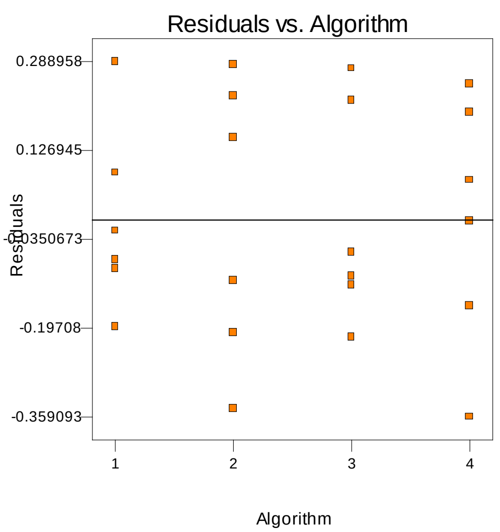
Treatment	Mean Difference	DF	Standard Error	t for H0 Coeff=0	Prob > |t|
1 vs 2	0.42	1	0.14	2.93	0.0103
1 vs 3	-0.89	1	0.14	-6.19	< 0.0001
1 vs 4	0.27	1	0.14	1.87	0.0813
2 vs 3	-1.31	1	0.14	-9.12	< 0.0001
2 vs 4	-0.15	1	0.14	-1.06	0.3042
3 vs 4	1.16	1	0.14	8.06	< 0.0001

A natural log transformation was applied to the pot noise data. The ratio control algorithm does affect the pot noise.

(c) Conduct any residual analyses that seem appropriate.







The normal probability plot shows slight deviations from normality; however, still acceptable.

(d) Which ratio control algorithm would you select if your objective is to reduce both the average cell voltage and the pot noise?

Since the ratio control algorithm has little effect on average cell voltage, select the algorithm that minimizes pot noise, that is algorithm #2.

4.14. An aluminum master alloy manufacturer produces grain refiners in ingot form. The company produces the product in four furnaces. Each furnace is known to have its own unique operating characteristics, so any experiment run in the foundry that involves more than one furnace will consider furnaces as a nuisance variable. The process engineers suspect that stirring rate impacts the grain size of the product. Each furnace can be run at four different stirring rates. A randomized block design is run for a particular refiner and the resulting grain size data is as follows.

	Furnace			
Stirring Rate	1	2	3	4
5	8	4	5	6
10	14	5	6	9
15	14	6	9	2
20	17	9	3	6

(a) Is there any evidence that stirring rate impacts grain size?

Design Expert Output

Response: Grain Size
ANOVA for Selected Factorial Model
Analysis of variance table [Partial sum of squares]

Source	**Sum of Squares**	**DF**	**Mean Square**	**F Value**	**Prob > F**	
Block	165.19	3	55.06			
Model	22.19	3	7.40	0.85	0.4995	not significant
A	*22.19*	*3*	*7.40*	*0.85*	*0.4995*	
Residual	78.06	9	8.67			
Cor Total	265.44	15				

The "Model F-value" of 0.85 implies the model is not significant relative to the noise. There is a 49.95 % chance that a "Model F-value" this large could occur due to noise.

Std. Dev.	2.95	R-Squared	0.2213
Mean	7.69	Adj R-Squared	-0.0382
C.V.	38.31	Pred R-Squared	-1.4610
PRESS	246.72	Adeq Precision	5.390

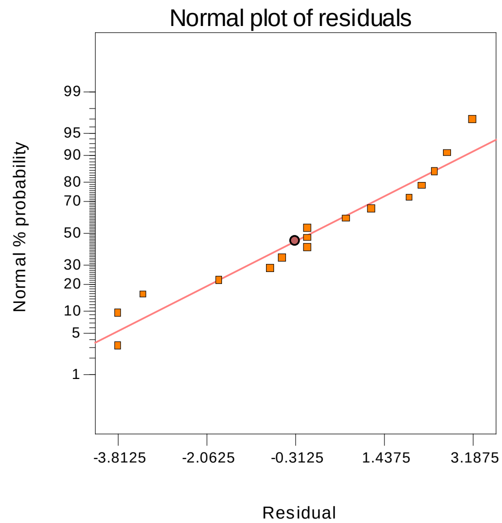
Treatment Means (Adjusted, If Necessary)

	Estimated Mean	**Standard Error**
1-5	5.75	1.47
2-10	8.50	1.47
3-15	7.75	1.47
4-20	8.75	1.47

Treatment	**Mean Difference**	**DF**	**Standard Error**	**t for H0 Coeff=0**	**Prob > |t|**
1 vs 2	-2.75	1	2.08	-1.32	0.2193
1 vs 3	-2.00	1	2.08	-0.96	0.3620
1 vs 4	-3.00	1	2.08	-1.44	0.1836
2 vs 3	0.75	1	2.08	0.36	0.7270
2 vs 4	-0.25	1	2.08	-0.12	0.9071
3 vs 4	-1.00	1	2.08	-0.48	0.6425

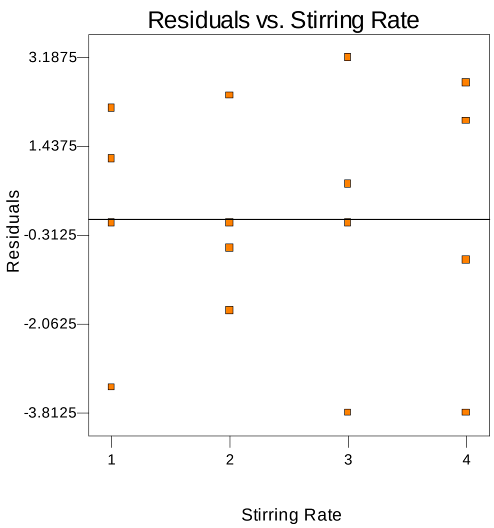
The analysis of variance shown above indicates that there is no difference in mean grain size due to the different stirring rates.

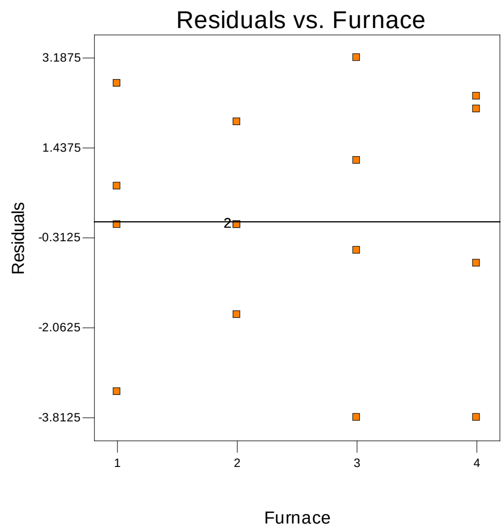
(b) Graph the residuals from this experiment on a normal probability plot. Interpret this plot.



The plot indicates that normality assumption is valid.

(c) Plot the residuals versus furnace and stirring rate. Does this plot convey any useful information?





The variance is consistent at different stirring rates. Not only does this validate the assumption of uniform variance, it also identifies that the different stirring rates do not affect variance.

(d) What should the process engineers recommend concerning the choice of stirring rate and furnace for this particular grain refiner if small grain size is desirable?

There really is no effect due to the stirring rate.

4.15. Analyze the data in Problem 4.4 using the general regression significance test.

$$\begin{array}{llllllllll}
\mu: & 12\hat{\mu} & +4\hat{\tau}_1 & +4\hat{\tau}_2 & +4\hat{\tau}_3 & +3\hat{\beta}_1 & +3\hat{\beta}_2 & +3\hat{\beta}_3 & +3\hat{\beta}_4 & =225 \\
\tau_1: & 4\hat{\mu} & +4\hat{\tau}_1 & & & +\hat{\beta}_1 & +\hat{\beta}_2 & +\hat{\beta}_3 & +\hat{\beta}_4 & =92 \\
\tau_2: & 4\hat{\mu} & & +4\hat{\tau}_2 & & +\hat{\beta}_1 & +\hat{\beta}_2 & +\hat{\beta}_3 & +\hat{\beta}_4 & =101 \\
\tau_3: & 4\hat{\mu} & & & +4\hat{\tau}_3 & +\hat{\beta}_1 & +\hat{\beta}_2 & +\hat{\beta}_3 & +\hat{\beta}_4 & =32 \\
\beta_1: & 3\hat{\mu} & +\hat{\tau}_1 & +\hat{\tau}_2 & +\hat{\tau}_3 & +3\hat{\beta}_1 & & & & =34 \\
\beta_1: & 3\hat{\mu} & +\hat{\tau}_1 & +\hat{\tau}_2 & +\hat{\tau}_3 & & +3\hat{\beta}_2 & & & =50 \\
\beta_1: & 3\hat{\mu} & +\hat{\tau}_1 & +\hat{\tau}_2 & +\hat{\tau}_3 & & & +3\hat{\beta}_3 & & =36 \\
\beta\ : & 3\hat{\mu} & +\hat{\tau}_1 & +\hat{\tau}_2 & +\hat{\tau}_3 & & & & +3\hat{\beta}_4 & =105
\end{array}$$

Applying the constraints $\sum \hat{\tau}_i = \sum \hat{\beta}_j = 0$, we obtain:

$$\hat{\mu} = \frac{225}{12},\ \hat{\tau}_1 = \frac{51}{12},\ \hat{\tau}_2 = \frac{78}{12},\ \hat{\tau}_3 = \frac{-129}{12},\ \hat{\beta}_1 = \frac{-89}{12},\ \hat{\beta}_2 = \frac{-25}{12},\ \hat{\beta}_3 = \frac{-81}{12},\ \hat{\beta}_4 = \frac{195}{12}$$

$$R(\mu,\tau,\beta) = \left(\frac{225}{12}\right)(225) + \left(\frac{51}{12}\right)(92) + \left(\frac{78}{12}\right)(101) + \left(\frac{-129}{12}\right)(32) + \left(\frac{-89}{12}\right)(34) + \left(\frac{-25}{12}\right)(50) + \left(\frac{-81}{12}\right)(36) + \left(\frac{195}{12}\right)(105) = 6029.17$$

$$\sum\sum y_{ij}^2 = 6081,\ SS_E = \sum\sum y_{ij}^2 - R(\mu,\tau,\beta) = 6081 - 6029.17 = 51.83$$

Model Restricted to $\tau_i = 0$:

$$\begin{array}{lllllll}
\mu: & 12\hat{\mu} & +3\hat{\beta}_1 & +3\hat{\beta}_2 & +3\hat{\beta}_3 & +3\hat{\beta}_4 & =225 \\
\beta_1: & 3\hat{\mu} & +3\hat{\beta}_1 & & & & =34 \\
\beta_2: & 3\hat{\mu} & & +3\hat{\beta}_2 & & & =50 \\
\beta_3: & 3\hat{\mu} & & & +3\hat{\beta}_3 & & =36 \\
\beta_4: & 3\hat{\mu} & & & & +3\hat{\beta}_4 & =105
\end{array}$$

Applying the constraint $\sum \hat{\beta}_j = 0$, we obtain:

$$\hat{\mu} = \frac{225}{12},\ \hat{\beta}_1 = -89/12,\ \hat{\beta}_2 = \frac{-25}{12},\ \hat{\beta}_3 = \frac{-81}{12},\ \hat{\beta}_4 = \frac{195}{12}.\ \text{Now:}$$

$$R(\mu,\beta) = \left(\frac{225}{12}\right)(225) + \left(\frac{-89}{12}\right)(34) + \left(\frac{-25}{12}\right)(50) + \left(\frac{-81}{12}\right)(36) + \left(\frac{195}{12}\right)(105) = 5325.67$$

$$R(\tau|\mu,\beta) = R(\mu,\tau,\beta) - R(\mu,\beta) = 6029.17 - 5325.67 = 703.50 = SS_{Treatments}$$

Model Restricted to $\beta_j = 0$:

$$\begin{array}{llllll}
\mu: & 12\hat{\mu} & +4\hat{\tau}_1 & +4\hat{\tau}_2 & +4\hat{\tau}_3 & =225 \\
\tau_1: & 4\hat{\mu} & +4\hat{\tau}_1 & & & =92 \\
\tau_2: & 4\hat{\mu} & & +4\hat{\tau}_2 & & =101 \\
\tau_3: & 4\hat{\mu} & & & +4\hat{\tau}_3 & =32
\end{array}$$

Applying the constraint $\sum \hat{\tau}_i = 0$, we obtain:

$$\hat{\mu} = \frac{225}{12},\ \hat{\tau}_1 = \frac{51}{12},\ \hat{\tau}_2 = \frac{78}{12},\ \hat{\tau}_3 = \frac{-129}{12}$$

$$R(\mu,\tau) = \left(\frac{225}{12}\right)(225) + \left(\frac{51}{12}\right)(92) + \left(\frac{78}{12}\right)(101) + \left(\frac{-129}{12}\right)(32) = 4922.25$$

$$R(\beta|\mu,\tau) = R(\mu,\tau,\beta) - R(\mu,\tau) = 6029.17 - 4922.25 = 1106.92 = SS_{Blocks}$$

4.16. Assuming that chemical types and bolts are fixed, estimate the model parameters τ_i and β_j in Problem 4.3.

Using Equations 4.18, applying the constraints, we obtain:

$$\hat{\mu} = \frac{35}{20},\ \hat{\tau}_1 = \frac{-23}{20},\ \hat{\tau}_2 = \frac{-7}{20},\ \hat{\tau}_3 = \frac{13}{20},\ \hat{\tau}_4 = \frac{17}{20},\ \hat{\beta}_1 = \frac{35}{20},\ \hat{\beta}_2 = \frac{-65}{20},\ \hat{\beta}_3 = \frac{75}{20},\ \hat{\beta}_4 = \frac{20}{20},\ \hat{\beta}_5 = \frac{-65}{20}$$

4.17. Draw an operating characteristic curve for the design in Problem 4.4. Does this test seem to be sensitive to small differences in treatment effects?

Assuming that solution type is a fixed factor, we use the OC curve in appendix V. Calculate

$$\Phi^2 = \frac{b\sum \tau_i^2}{a\sigma^2} = \frac{4\sum \tau_i^2}{3(8.64)}$$

using MS_E to estimate σ^2. We have:

$$\upsilon_1 = a - 1 = 2 \qquad \upsilon_2 = (a-1)(b-1) = (2)(3) = 6\,.$$

If $\sum \hat{\tau}^2_i = \sigma^2 = MS_E$, then:

$$\Phi = \sqrt{\frac{4}{3(1)}} = 1.15 \text{ and } \beta \cong 0.70$$

If $\sum \hat{\tau}_i = 2\sigma^2 = 2MS_E$, then:

$$\Phi = \sqrt{\frac{4(2)}{3(1)}} = 1.63 \text{ and } \beta \cong 0.55\text{, etc.}$$

This test is not very sensitive to small differences.

4.18. Suppose that the observation for chemical type 2 and bolt 3 is missing in Problem 4.3. Analyze the problem by estimating the missing value. Perform the exact analysis and compare the results.

$$y_{23} \text{ is missing.} \quad \hat{y}_{23} = \frac{ay'_{2.} + by'_{.3} - y'_{..}}{(a-1)(b-1)} = \frac{4(282)+5(227)-1360}{(3)(4)} = 75.25$$

Therefore, $y_{2.}$=357.25, $y_{.3}$=302.25, and $y_{..}$=1435.25

Source	*SS*	*DF*	*MS*	F_0
Chemicals	12.7844	3	4.2615	2.154
Bolts	158.8875	4		
Error	21.7625	11	1.9784	
Total	193.4344	18		

$F_{0.05,3,11}$=3.59, Chemicals are not significant. This is the same result as found in Problem 4.3.

4.19. Consider the hardness testing experiment in Problem 4.7. Suppose that the observation for tip 2 in coupon 3 is missing. Analyze the problem by estimating the missing value.

$$y_{23} \text{ is missing.} \quad \hat{y}_{23} = \frac{ay'_{2.} + by'_{.3} - y'_{..}}{(a-1)(b-1)} = \frac{4(28.6)+4(29.1)-144.2}{(3)(3)} = 9.62$$

Therefore, $y_{2.}$=38.22, $y_{.3}$=38.72, and $y_{..}$=153.82

Source	*SS*	*DF*	*MS*	F_0
Tip	0.40	3	0.133333	19.29
Coupon	0.80	3		
Error	0.0622	9	0.006914	
Total	1.2622	15		

$F_{0.05,3,9}$=3.86, Tips are significant. This is the same result as found in Problem 4.7.

4.20. ***Two missing values in a randomized block.*** Suppose that in Problem 4.3 the observations for chemical type 2 and bolt 3 and chemical type 4 and bolt 4 are missing.

(a) Analyze the design by iteratively estimating the missing values as described in Section 4.1.3.

$$\hat{y}_{23} = \frac{4y'_{2.} + 5y'_{.3} - y'_{..}}{12} \text{ and } \hat{y}_{44} = \frac{4y'_{4.} + 5y'_{.4} - y'_{..}}{12}$$

Data is coded *y*-70. As an initial guess, set y_{23}^0 equal to the average of the observations available for chemical 2. Thus, $y_{23}^0 = \frac{2}{4} = 0.5$. Then ,

$$\hat{y}_{44}^0 = \frac{4(8)+5(6)-25.5}{12} = 3.04$$

$$\hat{y}_{23}^1 = \frac{4(2)+5(17)-28.04}{12} = 5.41$$

$$\hat{y}_{44}^{1} = \frac{4(8)+5(6)-30.41}{12} = 2.63$$
$$\hat{y}_{44}^{2} = \frac{4(2)+5(17)-27.63}{12} = 5.44$$
$$\hat{y}_{44}^{2} = \frac{4(8)+5(6)-30.44}{12} = 2.63$$
$$\therefore \hat{y}_{23} = 5.44 \quad \hat{y}_{44} = 2.63$$

Design Expert Output

ANOVA for Selected Factorial Model
Analysis of variance table [Partial sum of squares]

Source	**Sum of Squares**	**DF**	**Mean Square**	**F Value**	**Prob > F**	
Block	156.83	4	39.21			
Model	9.59	3	3.20	2.08	0.1560	not significant
A	*9.59*	*3*	*3.20*	*2.08*	*0.1560*	
Residual	18.41	12	1.53			
Cor Total	184.83	19				

(b) Differentiate SS_E with respect to the two missing values, equate the results to zero, and solve for estimates of the missing values. Analyze the design using these two estimates of the missing values.

$$SS_E = \sum\sum y_{ij}^2 - \tfrac{1}{5}\sum y_{i.}^2 - \tfrac{1}{4}\sum y_{.j}^2 + \tfrac{1}{20}\sum y_{..}^2$$
$$SS_E = 0.6y_{23}^2 + 0.6y_{44}^2 - 6.8y_{23} - 3.7y_{44} + 0.1y_{23}y_{44} + R$$

From $\frac{\partial SS_E}{\partial y_{23}} = \frac{\partial SS_E}{\partial y_{44}} = 0$, we obtain:

$$\begin{aligned} 1.2\hat{y}_{23} + 0.1\hat{y}_{44} &= 6.8 \\ 0.1\hat{y}_{23} + 1.2\hat{y}_{44} &= 3.7 \end{aligned} \quad \Rightarrow \hat{y}_{23} = 5.45\text{ , } \hat{y}_{44} = 2.63$$

These quantities are almost identical to those found in part (a). The analysis of variance using these new data does not differ substantially from part (a).

(c) Derive general formulas for estimating two missing values when the observations are in *different* blocks.

$$SS_E = y_{iu}^2 + y_{kv}^2 - \frac{\left(y'_{i.} + y_{iu}\right)^2 + \left(y'_{k.} + y_{kv}\right)^2}{b} - \frac{\left(y'_{.u} + y_{iu}\right)^2 + \left(y'_{.v} + y_{kv}\right)^2}{a} + \frac{\left(y'_{..} + y_{iu} + y_{kv}\right)^2}{ab}$$

From $\frac{\partial SS_E}{\partial y_{23}} = \frac{\partial SS_E}{\partial y_{44}} = 0$, we obtain:

$$\hat{y}_{iu}\left[\frac{(a-1)(b-1)}{ab}\right] = \frac{ay'_{i.} + by'_{.j} - y'_{..}}{ab} - \frac{\hat{y}_{kv}}{ab}$$
$$\hat{y}_{kv}\left[\frac{(a-1)(b-1)}{ab}\right] = \frac{ay'_{k.} + by'_{.v} - y'_{..}}{ab} - \frac{\hat{y}_{iu}}{ab}$$

whose simultaneous solution is:

$$\hat{y}_{iu} = \frac{y'_{i.}\, a\left[1-(a-1)^2(b-1)^2-ab\right]+y'_{.u}\, b\left[1-(a-1)^2(b-1)^2-ab\right]-y'_{..}\left[1-ab(a-1)^2(b-1)^2\right]}{(a-1)(b-1)\left[1-(a-1)^2(b-1)^2\right]}+$$

$$\frac{ab\left[ay'_{k.}+by'_{.v}-y'_{..}\right]}{\left[1-(a-1)^2(b-1)^2\right]}$$

$$\hat{y}_{kv} = \frac{ay'_{i.}+by'_{.u}-y'_{..}-(b-1)(a-1)\left[ay'_{k.}+by'_{.v}-y'_{..}\right]}{\left[1-(a-1)^2(b-1)^2\right]}$$

(d) Derive general formulas for estimating two missing values when the observations are in the *same* block. Suppose that two observations y_{ij} and y_{kj} are missing, $i \neq k$ (same block j).

$$SS_E = y_{ij}^2 + y_{kj}^2 - \frac{\left(y'_{i.}+y_{ij}\right)^2+\left(y'_{k.}+y_{kj}\right)^2}{b} - \frac{\left(y'_{.j}+y_{ij}+y_{kj}\right)^2}{a} + \frac{\left(y'_{..}+y_{ij}+y_{kj}\right)^2}{ab}$$

From $\frac{\partial SS_E}{\partial y_{ij}} = \frac{\partial SS_E}{\partial y_{kj}} = 0$, we obtain

$$\hat{y}_{ij} = \frac{ay'_{i.}+by'_{.j}-y'_{..}}{(a-1)(b-1)} + \hat{y}_{kj}(a-1)(b-1)^2$$

$$\hat{y}_{kj} = \frac{ay'_{k.}+by'_{.j}-y'_{..}}{(a-1)(b-1)} + \hat{y}_{ij}(a-1)(b-1)^2$$

whose simultaneous solution is:

$$\hat{y}_{ij} = \frac{ay'_{i.}+by'_{.j}-y'_{..}}{(a-1)(b-1)} + \frac{(b-1)\left[ay'_{k.}+by'_{.j}-y'_{..}+(a-1)(b-1)^2\left(ay'_{i.}+by'_{.j}-y'_{..}\right)\right]}{\left[1-(a-1)^2(b-1)^4\right]}$$

$$\hat{y}_{kj} = \frac{ay'_{k.}+by'_{.j}-y'_{..}-(b-1)^2(a-1)\left[ay'_{i.}+by'_{.j}-y'_{..}\right]}{(a-1)(b-1)\left[1-(a-1)^2(b-1)^4\right]}$$

4.21. An industrial engineer is conducting an experiment on eye focus time. He is interested in the effect of the distance of the object from the eye on the focus time. Four different distances are of interest. He has five subjects available for the experiment. Because there may be differences among individuals, he decides to conduct the experiment in a randomized block design. The data obtained follow. Analyze the data from this experiment (use $\alpha = 0.05$) and draw appropriate conclusions.

	Subject				
Distance (ft)	1	2	3	4	5
4	10	6	6	6	6
6	7	6	6	1	6
8	5	3	3	2	5
10	6	4	4	2	3

Design Expert Output

Response: Focus Time
ANOVA for Selected Factorial Model
Analysis of variance table [Partial sum of squares]

Source	**Sum of Squares**	**DF**	**Mean Square**	**F Value**	**Prob > F**	
Block	36.30	4	9.07			
Model	32.95	3	10.98	8.61	0.0025	significant
A	*32.95*	*3*	*10.98*	*8.61*	*0.0025*	
Residual	15.30	12	1.27			
Cor Total	84.55	19				

The Model F-value of 8.61 implies the model is significant. There is only
a 0.25% chance that a "Model F-Value" this large could occur due to noise.

Std. Dev.	1.13	R-Squared	0.6829
Mean	4.85	Adj R-Squared	0.6036
C.V.	23.28	Pred R-Squared	0.1192
PRESS	42.50	Adeq Precision	10.432

Treatment Means (Adjusted, If Necessary)

	Estimated Mean	**Standard Error**
1-4	6.80	0.50
2-6	5.20	0.50
3-8	3.60	0.50
4-10	3.80	0.50

Treatment	**Mean Difference**	**DF**	**Standard Error**	**t for H0 Coeff=0**	**Prob > |t|**
1 vs 2	1.60	1	0.71	2.24	0.0448
1 vs 3	3.20	1	0.71	4.48	0.0008
1 vs 4	3.00	1	0.71	4.20	0.0012
2 vs 3	1.60	1	0.71	2.24	0.0448
2 vs 4	1.40	1	0.71	1.96	0.0736
3 vs 4	-0.20	1	0.71	-0.28	0.7842

Distance has a statistically significant effect on mean focus time.

4.22. The effect of five different ingredients (*A*, *B*, *C*, *D*, *E*) on reaction time of a chemical process is being studied. Each batch of new material is only large enough to permit five runs to be made. Furthermore, each run requires approximately 1 1/2 hours, so only five runs can be made in one day. The experimenter decides to run the experiment as a Latin square so that day and batch effects can be systematically controlled. She obtains the data that follow. Analyze the data from this experiment (use α = 0.05) and draw conclusions.

	Day				
Batch	1	2	3	4	5
1	*A*=8	*B*=7	*D*=1	*C*=7	*E*=3
2	*C*=11	*E*=2	*A*=7	*D*=3	*B*=8
3	*B*=4	*A*=9	*C*=10	*E*=1	*D*=5
4	*D*=6	*C*=8	*E*=6	*B*=6	*A*=10
5	*E*=4	*D*=2	*B*=3	*A*=8	*C*=8

The *Minitab* output below identifies the ingredients as having a significant effect on reaction time.

Minitab Output

General Linear Model

```
Factor     Type Levels Values
Batch    random      5 1 2 3 4 5
Day      random      5 1 2 3 4 5
Catalyst  fixed      5 A B C D E

Analysis of Variance for Time, using Adjusted SS for Tests

Source     DF     Seq SS     Adj SS     Adj MS       F      P
Catalyst    4    141.440    141.440     35.360   11.31  0.000
Batch       4     15.440     15.440      3.860    1.23  0.348
Day         4     12.240     12.240      3.060    0.98  0.455
Error      12     37.520     37.520      3.127
Total      24    206.640
```

4.23. An industrial engineer is investigating the effect of four assembly methods (*A, B, C, D*) on the assembly time for a color television component. Four operators are selected for the study. Furthermore, the engineer knows that each assembly method produces such fatigue that the time required for the last assembly may be greater than the time required for the first, regardless of the method. That is, a trend develops in the required assembly time. To account for this source of variability, the engineer uses the Latin square design shown below. Analyze the data from this experiment (α = 0.05) draw appropriate conclusions.

Order of	Operator			
Assembly	1	2	3	4
1	*C*=10	*D*=14	*A*=7	*B*=8
2	*B*=7	*C*=18	*D*=11	*A*=8
3	*A*=5	*B*=10	*C*=11	*D*=9
4	*D*=10	*A*=10	*B*=12	*C*=14

The Minitab output below identifies assembly method as having a significant effect on assembly time.

Minitab Output

General Linear Model

```
Factor     Type Levels Values
Order    random      4 1 2 3 4
Operator random      4 1 2 3 4
Method    fixed      4 A B C D


Analysis of Variance for Time, using Adjusted SS for Tests

Source     DF     Seq SS     Adj SS     Adj MS       F      P
Method      3     72.500     72.500     24.167   13.81  0.004
Order       3     18.500     18.500      6.167    3.52  0.089
Operator    3     51.500     51.500     17.167    9.81  0.010
Error       6     10.500     10.500      1.750
Total      15    153.000
```

4.24. Consider the randomized complete block design in Problem 4.4. Assume that the days are random. Estimate the block variance component.

The block variance component is:

$$\hat{\sigma}_\beta^2 = \frac{\left[MS_{\text{Blocks}} - MS_E\right]}{a} = \frac{\left[368.97 - 8.64\right]}{3} = 120.11$$

4.25. Consider the randomized complete block design in Problem 4.7. Assume that the coupons are random. Estimate the block variance component.

The block variance component is:

$$\hat{\sigma}_\beta^2 = \frac{\left[MS_{\text{Blocks}} - MS_E\right]}{a} = \frac{\left[0.27 - 0.008889\right]}{4} = 0.06528$$

4.26. Consider the randomized complete block design in Problem 4.9. Assume that the trucks are random. Estimate the block variance component.

The block variance component is:

$$\hat{\sigma}_\beta^2 = \frac{\left[MS_{\text{Blocks}} - MS_E\right]}{a} = \frac{\left[0.023 - 0.0005278\right]}{3} = 0.007491$$

4.27. Consider the randomized complete block design in Problem 4.11. Assume that the software projects that were used as blocks are random. Estimate the block variance component.

The block variance component is:

$$\hat{\sigma}_\beta^2 = \frac{\left[MS_{\text{Blocks}} - MS_E\right]}{a} = \frac{\left[457500 - 111200\right]}{6} = 57716.67$$

4.28. Consider the gene expression experiment in Problem 4.12. Assume that the subjects used in this experiment are random. Estimate the block variance component

The block variance component is:

$$\hat{\sigma}_\beta^2 = \frac{\left[MS_{\text{Blocks}} - MS_E\right]}{a} = \frac{\left[102300 - 73130.15\right]}{3} = 9723.28$$

4.29. Suppose that in Problem 4.22 the observation from batch 3 on day 4 is missing. Estimate the missing value and perform the analysis using this value.

y_{354} is missing. $\hat{y}_{354} = \dfrac{p\left[y'_{i..} + y'_{.j.} + y'_{..k}\right] - 2y'_{...}}{(p-2)(p-1)} = \dfrac{5\left[28 + 15 + 24\right] - 2(146)}{(3)(4)} = 3.58$

Minitab Output

General Linear Model

```
Factor     Type Levels Values
Batch    random      5 1 2 3 4 5
Day      random      5 1 2 3 4 5
Catalyst  fixed      5 A B C D E

Analysis of Variance for Time, using Adjusted SS for Tests

Source     DF     Seq SS     Adj SS     Adj MS        F      P
Catalyst    4    128.676    128.676     32.169    11.25  0.000
Batch       4     16.092     16.092      4.023     1.41  0.290
Day         4      8.764      8.764      2.191     0.77  0.567
Error      12     34.317     34.317      2.860
Total      24    187.849
```

4.30. Consider a p x p Latin square with rows (α_i), columns (β_k), and treatments (τ_j) fixed. Obtain least squares estimates of the model parameters α_i, β_k, τ_j.

$$\mu : p^2\hat{\mu} + p\sum_{i=1}^{p}\hat{\alpha}_i + p\sum_{j=1}^{p}\hat{\tau}_j + p\sum_{k=1}^{p}\hat{\beta}_k = y_{...}$$

$$\alpha_i : p\hat{\mu} + p\hat{\alpha}_i + p\sum_{j=1}^{p}\hat{\tau}_j + p\sum_{k=1}^{p}\hat{\beta}_k = y_{i..} \quad , \; i = 1,2,...,p$$

$$\tau_j : p\hat{\mu} + p\sum_{i=1}^{p}\hat{\alpha}_i + p\hat{\tau}_j + p\sum_{k=1}^{p}\hat{\beta}_k = y_{.j.} \, , \; j = 1,2,...,p$$

$$\beta_k : p\hat{\mu} + p\sum_{i=1}^{p}\hat{\alpha}_i + p\sum_{j=1}^{p}\hat{\tau}_j + p\hat{\beta}_k = y_{..k} \, , \; k = 1,2,...,p$$

There are 3p+1 equations in 3p+1 unknowns. The rank of the system is 3p-2. Three side conditions are necessary. The usual conditions imposed are: $\sum_{i=1}^{p}\hat{\alpha}_i = \sum_{j=1}^{p}\hat{\tau}_j = \sum_{k=1}^{p}\hat{\beta}_k = 0$. The solution is then:

$$\hat{\mu} = \frac{y_{...}}{p^2} = \bar{y}_{...}$$
$$\hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{...}, i = 1, 2, ..., p$$
$$\hat{\tau}_j = \bar{y}_{.j.} - \bar{y}_{...}, \; j = 1, 2, ..., p$$
$$\hat{\beta}_k = \bar{y}_{i..} - \bar{y}_{...}, k = 1, 2, ..., p$$

4.31. Derive the missing value formula (Equation 4.27) for the Latin square design.

$$SS_E = \sum\sum\sum y_{ijk}^2 - \sum\frac{y_{i..}^2}{p} - \sum\frac{y_{.j.}^2}{p} - \sum\frac{y_{..k}^2}{p} + 2\left(\frac{y_{...}^2}{p^2}\right)$$

Let y_{ijk} be missing. Then

$$SS_E = y_{ijk}^2 - \frac{\left(y'_{i..} + y_{ijk}\right)^2}{p} - \frac{\left(y'_{.j.} + y_{ijk}\right)^2}{p} - \frac{\left(y'_{..k} + y_{ijk}\right)^2}{p} + \frac{2\left(y'_{...} + y_{ijk}\right)^2}{p^2} + R$$

where R is all terms without y_{ijk}.. From $\frac{\partial SS_E}{\partial y_{ijk}} = 0$, we obtain:

$$y_{ijk}\,\frac{(p-1)(p-2)}{p^2} = \frac{p\left(y'_{i..} + y'_{.j.} + y'_{..k}\right) - 2y'_{...}}{p^2}, \text{ or } y_{ijk} = \frac{p\left(y'_{i..} + y'_{.j.} + y'_{..k}\right) - 2y'_{...}}{(p-1)(p-2)}$$

4.32. ***Designs involving several Latin squares***. [See Cochran and Cox (1957), John (1971).] The *p x p* Latin square contains only *p* observations for each treatment. To obtain more replications the experimenter may use several squares, say *n*. It is immaterial whether the squares used are the same are different. The appropriate model is

$$y_{ijkh} = \mu + \rho_h + \alpha_{i(h)} + \tau_j + \beta_{k(h)} + (\tau\rho)_{jh} + \varepsilon_{ijkh} \qquad \begin{cases} i = 1,2,\ldots,p \\ j = 1,2,\ldots,p \\ k = 1,2,\ldots,p \\ h = 1,2,\ldots,n \end{cases}$$

where y_{ijkh} is the observation on treatment *j* in row *i* and column *k* of the *h*th square. Note that $\alpha_{i(h)}$ and $\beta_{k(h)}$ are row and column effects in the *h*th square, and ρ_h is the effect of the *h*th square, and $(\tau\rho)_{jh}$ is the interaction between treatments and squares.

(a) Set up the normal equations for this model, and solve for estimates of the model parameters. Assume that appropriate side conditions on the parameters are $\sum_h \hat{\rho}_h = 0$, $\sum_i \hat{\alpha}_{i(h)} = 0$, and $\sum_k \hat{\beta}_{k(h)} = 0$ for each *h*, $\sum_j \hat{\tau}_j = 0$, $\sum_j (\hat{\tau\rho})_{jh} = 0$ for each *h*, and $\sum_h (\hat{\tau\rho})_{jh} = 0$ for each *j*.

$$\begin{aligned}
\hat{\mu} &= \bar{y}_{....} \\
\hat{\rho}_h &= \bar{y}_{...h} - \bar{y}_{....} \\
\hat{\tau}_j &= \bar{y}_{.j..} - \bar{y}_{....} \\
\hat{\alpha}_{i(h)} &= \bar{y}_{i..h} - \bar{y}_{...h} \\
\hat{\beta}_{k(h)} &= \bar{y}_{..kh} - \bar{y}_{...h} \\
\left(\hat{\tau\rho}\right)_{jh} &= \bar{y}_{.j.h} - \bar{y}_{.j..} - \bar{y}_{...h} + \bar{y}_{....}
\end{aligned}$$

(b) Write down the analysis of variance table for this design.

Source	SS	DF
Treatments	$\sum \frac{y_{.j..}^2}{np} - \frac{y_{....}^2}{np^2}$	p-1
Squares	$\sum \frac{y_{...h}^2}{p^2} - \frac{y_{....}^2}{np^2}$	n-1
Treatment x Squares	$\sum \frac{y_{.j.h}^2}{p} - \frac{y_{....}^2}{np^2} - SS_{Treatments} - SS_{Squares}$	(p-1)(n-1)
Rows	$\sum \frac{y_{i..h}^2}{p} - \frac{y_{...h}^2}{np}$	n(p-1)
Columns	$\sum \frac{y_{..kh}^2}{p} - \frac{y_{...h}^2}{np}$	n(p-1)
Error	subtraction	n(p-1)(p-2)
Total	$\sum\sum\sum\sum y_{ijkh}^2 - \frac{y_{....}^2}{np^2}$	np^2-1

4.33. Discuss how the operating characteristics curves in the Appendix may be used with the Latin square design.

For the fixed effects model use:

$$\Phi^2 = \frac{\sum p\tau_j^2}{p\sigma^2} = \sum \frac{\tau_j^2}{\sigma^2}\ ,\ \upsilon_1 = p-1 \quad \upsilon_2 = (p-2)(p-1)$$

For the random effects model use:

$$\lambda = \sqrt{1 + \frac{p\sigma_\tau^2}{\sigma^2}}\ ,\ \upsilon_1 = p-1 \quad \upsilon_2 = (p-2)(p-1)$$

4.34. Suppose that in Problem 4.22 the data taken on day 5 were incorrectly analyzed and had to be discarded. Develop an appropriate analysis for the remaining data.

Two methods of analysis exist: (1) Use the general regression significance test, or (2) recognize that the design is a Youden square. The data can be analyzed as a balanced incomplete block design with $a = b = 5$, $r = k = 4$ and $\lambda = 3$. Using either approach will yield the same analysis of variance.

Minitab Output

General Linear Model

```
Factor     Type Levels Values
Catalyst  fixed      5 A B C D E
Batch    random      5 1 2 3 4 5
Day      random      4 1 2 3 4
```

```
Analysis of Variance for Time, using Adjusted SS for Tests

Source     DF     Seq SS     Adj SS     Adj MS       F      P
Catalyst    4    119.800    120.167     30.042    7.48  0.008
Batch       4     11.667     11.667      2.917    0.73  0.598
Day         3      6.950      6.950      2.317    0.58  0.646
Error       8     32.133     32.133      4.017
Total      19    170.550
```

4.35. The yield of a chemical process was measured using five batches of raw material, five acid concentrations, five standing times, (A, B, C, D, E) and five catalyst concentrations (α, β, γ, δ, ε). The Graeco-Latin square that follows was used. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

	Acid Concentration				
Batch	1	2	3	4	5
1	$A\alpha$=26	$B\beta$=16	$C\gamma$=19	$D\delta$=16	$E\varepsilon$=13
2	$B\gamma$=18	$C\delta$=21	$D\varepsilon$=18	$E\alpha$=11	$A\beta$=21
3	$C\varepsilon$=20	$D\alpha$=12	$E\beta$=16	$A\gamma$=25	$B\delta$=13
4	$D\beta$=15	$E\gamma$=15	$A\delta$=22	$B\varepsilon$=14	$C\alpha$=17
5	$E\delta$=10	$A\varepsilon$=24	$B\alpha$=17	$C\beta$=17	$D\gamma$=14

The Minitab output below identifies standing time as having a significant effect on yield.

Minitab Output

General Linear Model

```
Factor     Type Levels Values
Time      fixed      5 A B C D E
Catalyst random      5 a b c d e
Batch    random      5 1 2 3 4 5
Acid     random      5 1 2 3 4 5

Analysis of Variance for Yield, using Adjusted SS for Tests

Source     DF     Seq SS     Adj SS     Adj MS       F      P
Time        4    342.800    342.800     85.700   14.65  0.001
Catalyst    4     12.000     12.000      3.000    0.51  0.729
Batch       4     10.000     10.000      2.500    0.43  0.785
Acid        4     24.400     24.400      6.100    1.04  0.443
Error       8     46.800     46.800      5.850
Total      24    436.000
```

4.36. Suppose that in Problem 4.23 the engineer suspects that the workplaces used by the four operators may represent an additional source of variation. A fourth factor, workplace (α, β, γ, δ) may be introduced and another experiment conducted, yielding the Graeco-Latin square that follows. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

Order of	Operator			
Assembly	1	2	3	4
1	$C\beta$=11	$B\gamma$=10	$D\delta$=14	$A\alpha$=8
2	$B\alpha$=8	$C\delta$=12	$A\gamma$=10	$D\beta$=12
3	$A\delta$=9	$D\alpha$=11	$B\beta$=7	$C\gamma$=15
4	$D\gamma$=9	$A\beta$=8	$C\alpha$=18	$B\delta$=6

Minitab Output

General Linear Model

```
Factor     Type Levels Values
Method    fixed      4 A B C D
Order    random      4 1 2 3 4
Operator random      4 1 2 3 4
Workplac random      4 a b c d

Analysis of Variance for Time, using Adjusted SS for Tests

Source     DF     Seq SS     Adj SS     Adj MS       F      P
Method      3     95.500     95.500     31.833    3.47  0.167
Order       3      0.500      0.500      0.167    0.02  0.996
Operator    3     19.000     19.000      6.333    0.69  0.616
Workplac    3      7.500      7.500      2.500    0.27  0.843
Error       3     27.500     27.500      9.167
Total      15    150.000
```

Method and workplace do not have a significant effect on assembly time. However, there are only three degrees of freedom for error, so the test is not very sensitive.

4.37. Construct a 5 x 5 hypersquare for studying the effects of five factors. Exhibit the analysis of variance table for this design.

Three 5 x 5 orthogonal Latin Squares are:

ABCDE	$\alpha\beta\gamma\delta\varepsilon$	12345
BCDEA	$\gamma\delta\varepsilon\alpha\beta$	45123
CDEAB	$\varepsilon\alpha\beta\gamma\delta$	23451
DEABC	$\beta\gamma\delta\varepsilon\alpha$	51234
EABCD	$\delta\varepsilon\alpha\beta\gamma$	34512

Let rows = factor 1, columns = factor 2, Latin letters = factor 3, Greek letters = factor 4 and numbers = factor 5. The analysis of variance table is:

Source	SS	DF
Rows	$\frac{1}{5}\sum_{i=1}^{5} y_{i....}^2 - \frac{y_{.....}^2}{25}$	4
Columns	$\frac{1}{5}\sum_{m=1}^{5} y_{....m}^2 - \frac{y_{.....}^2}{25}$	4
Latin Letters	$\frac{1}{5}\sum_{j=1}^{5} y_{.j...}^2 - \frac{y_{.....}^2}{25}$	4
Greek Letters	$\frac{1}{5}\sum_{k=1}^{5} y_{..k..}^2 - \frac{y_{.....}^2}{25}$	4
Numbers	$\frac{1}{5}\sum_{l=1}^{5} y_{...l.}^2 - \frac{y_{.....}^2}{25}$	4
Error	SS_E by subtraction	4
Total	$\sum_{i=1}^{5}\sum_{j=1}^{5}\sum_{k=1}^{5}\sum_{l=1}^{5}\sum_{m=1}^{5} y_{ijklm}^2 - \frac{y_{.....}^2}{25}$	24

4.38. Consider the data in Problems 4.23 and 4.36. Suppressing the Greek letters in 4.36, analyze the data using the method developed in Problem 4.32.

Square 1 - Operator

Batch	1	2	3	4	Row Total
1	*C*=10	*D*=14	*A*=7	*B*=8	(39)
2	*B*=7	*C*=18	*D*=11	*A*=8	(44)
3	*A*=5	*B*=10	*C*=11	*D*=9	(35)
4	*D*=10	*A*=10	*B*=12	*C*=14	(46)
	(32)	(52)	(41)	(36)	164=$y_{...1}$

Square 2 - Operator

Batch	1	2	3	4	Row Total
1	*C*=11	*B*=10	*D*=14	*A*=8	(43)
2	*B*=8	*C*=12	*A*=10	*D*=12	(42)
3	*A*=9	*D*=11	*B*=7	*C*=15	(42)
4	*D*=9	*A*=8	*C*=18	*B*=6	(41)
	(37)	(41)	(49)	(41)	168=$y_{...2}$

Assembly Methods	Totals
A	$y_{.1..}$=65
B	$y_{.2..}$=68
C	$y_{.3..}$=109
D	$y_{.4..}$=90

Source	*SS*	*DF*	*MS*	F_0
Assembly Methods	159.25	3	53.08	14.00*
Squares	0.50	1	0.50	
A x *S*	8.75	3	2.92	0.77
Assembly Order (Rows)	19.00	6	3.17	
Operators (columns)	70.50	6	11.75	
Error	45.50	12	3.79	
Total	303.50	31		

Significant at 1%.

4.39. Consider the randomized block design with one missing value in Problem 4.19. Analyze this data by using the exact analysis of the missing value problem discussed in Section 4.1.4. Compare your results to the approximate analysis of these data given from Problem 4.19.

To simplify the calculations, the data in Problems 4.19 was transformed by multiplying by 10 and substracting 95.

$$\begin{array}{llllllllllll}
\mu: & 15\hat{\mu} & +4\hat{\tau}_1 & +3\hat{\tau}_2 & +4\hat{\tau}_3 & +4\hat{\tau}_4 & +4\hat{\beta}_1 & +4\hat{\beta}_2 & +3\hat{\beta}_3 & +4\hat{\beta}_4 & =17 \\
\tau_1: & 4\hat{\mu} & +4\hat{\tau}_1 & & & & +\hat{\beta}_1 & +\hat{\beta}_2 & +\hat{\beta}_3 & +\hat{\beta}_4 & =3 \\
\tau_2: & 3\hat{\mu} & & +3\hat{\tau}_2 & & & +\hat{\beta}_1 & +\hat{\beta}_2 & & +\hat{\beta}_4 & =1 \\
\tau_3: & 4\hat{\mu} & & & +4\hat{\tau}_3 & & +\hat{\beta}_1 & +\hat{\beta}_2 & +\hat{\beta}_3 & +\hat{\beta}_4 & =-2 \\
\tau_4: & 4\hat{\mu} & & & & +4\hat{\tau}_4 & +\hat{\beta}_1 & +\hat{\beta}_2 & +\hat{\beta}_3 & +\hat{\beta}_4 & =15 \\
\beta_1: & 4\hat{\mu} & +\hat{\tau}_1 & +\hat{\tau}_2 & +\hat{\tau}_3 & +\hat{\tau}_4 & +4\hat{\beta}_1 & & & & =-4 \\
\beta_2: & 4\hat{\mu} & +\hat{\tau}_1 & +\hat{\tau}_2 & +\hat{\tau}_3 & +\hat{\tau}_4 & & +4\hat{\beta}_2 & & & =-3 \\
\beta_3: & 3\hat{\mu} & +\hat{\tau}_1 & & +\hat{\tau}_3 & +\hat{\tau}_4 & & & +3\hat{\beta}_3 & & =6 \\
\beta_4: & 4\hat{\mu} & +\hat{\tau}_1 & +\hat{\tau}_2 & +\hat{\tau}_3 & +\hat{\tau}_4 & & & & +4\hat{\beta}_4 & =18
\end{array}$$

Applying the constraints $\sum \hat{\tau}_i = \sum \hat{\beta}_j = 0$, we obtain:

$$\hat{\mu} = \frac{41}{36},\ \hat{\tau}_1 = \frac{-14}{36},\ \hat{\tau}_2 = \frac{-24}{36},\ \hat{\tau}_3 = \frac{-59}{36},\ \hat{\tau}_4 = \frac{94}{36},\ \hat{\beta}_1 = \frac{-77}{36},\ \hat{\beta}_2 = \frac{-68}{36},\ \hat{\beta}_3 = \frac{24}{36},\ \hat{\beta}_4 = \frac{121}{36}$$

$$R(\mu,\tau,\beta) = \hat{\mu}y_{..} + \sum_{i=1}^{4}\hat{\tau}_i y_{i.} + \sum_{j=1}^{4}\hat{\beta}_j y_{.j} = 138.78$$

With 7 degrees of freedom.

$$\sum\sum y_{ij}^2 = 145.00,\ SS_E = \sum\sum y_{ij}^2 - R(\mu,\tau,\beta) = 145.00 - 138.78 = 6.22$$

which is identical to SS_E obtained in the approximate analysis. In general, the SS_E in the exact and approximate analyses will be the same.

To test H_o: $\tau_i = 0$ the reduced model is $y_{ij} = \mu + \beta_j + \varepsilon_{ij}$. The normal equations used are:

$$\begin{array}{lllllll}
\mu: & 15\hat{\mu} & +4\hat{\beta}_1 & +4\hat{\beta}_2 & +3\hat{\beta}_3 & +4\hat{\beta}_4 & =17 \\
\beta_1: & 4\hat{\mu} & +4\hat{\beta}_1 & & & & =-4 \\
\beta_2: & 4\hat{\mu} & & +4\hat{\beta}_2 & & & =-3 \\
\beta_3: & 3\hat{\mu} & & & +3\hat{\beta}_3 & & =6 \\
\beta_4: & 4\hat{\mu} & & & & +4\hat{\beta}_4 & =18
\end{array}$$

Applying the constraint $\sum \hat{\beta}_j = 0$, we obtain:

$$\hat{\mu} = \frac{19}{16},\ \hat{\beta}_1 = \frac{-35}{16},\ \hat{\beta}_2 = \frac{-31}{16},\ \hat{\beta}_3 = \frac{13}{16},\ \hat{\beta}_4 = \frac{53}{16}.\ \text{Now}\ R(\mu,\beta) = \hat{\mu}y_{..} + \sum_{j=1}^{4}\hat{\beta}_j y_{.j} = 99.25$$

with 4 degrees of freedom.

$$R(\tau|\mu,\beta) = R(\mu,\tau,\beta) - R(\mu,\beta) = 138.78 - 99.25 = 39.53 = SS_{Treatments}$$

with 7-4=3 degrees of freedom. $R(\tau|\mu,\beta)$ is used to test H_o: $\tau_i = 0$.

The sum of squares for blocks is found from the reduced model $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$. The normal equations used are:

Model Restricted to $\beta_j = 0$:

$$\begin{array}{llllllll}
\mu: & 15\hat{\mu} & +4\hat{\tau}_1 & +3\hat{\tau}_2 & +4\hat{\tau}_3 & +4\hat{\tau}_4 & =17 \\
\tau_1: & 4\hat{\mu} & +4\hat{\tau}_1 & & & & =3 \\
\tau_2: & 3\hat{\mu} & & +3\hat{\tau}_2 & & & =1 \\
\tau_3: & 4\hat{\mu} & & & +4\hat{\tau}_3 & & =-2 \\
\tau_4: & 4\hat{\mu} & & & & +4\hat{\tau}_4 & =15
\end{array}$$

Applying the constraint $\sum \hat{\tau}_i = 0$, we obtain:

$$\hat{\mu} = \frac{13}{12},\ \hat{\tau}_1 = \frac{-4}{12},\ \hat{\tau}_2 = \frac{-9}{12},\ \hat{\tau}_3 = \frac{-19}{12},\ \hat{\tau}_4 = \frac{32}{12}$$

$$R(\mu,\tau) = \hat{\mu}y_{..} + \sum_{i=1}^{4} \hat{\tau}_i y_{i.} = 59.83$$

with 4 degrees of freedom.

$$R(\beta|\mu,\tau) = R(\mu,\tau,\beta) - R(\mu,\tau) = 138.78 - 59.83 = 78.95 = SS_{Blocks}$$

with 7-4=3 degrees of freedom.

Source	*DF*	*SS*(exact)	*SS*(approximate)
Tips	3	39.53	39.98
Blocks	3	78.95	79.53
Error	8	6.22	6.22
Total	14	125.74	125.73

Note that for the exact analysis, $SS_T \neq SS_{Tips} + SS_{Blocks} + SS_E$.

4.40. An engineer is studying the mileage performance characteristics of five types of gasoline additives. In the road test he wishes to use cars as blocks; however, because of a time constraint, he must use an incomplete block design. He runs the balanced design with the five blocks that follow. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

	Car				
Additive	1	2	3	4	5
1		17	14	13	12
2	14	14		13	10
3	12		13	12	9
4	13	11	11	12	
5	11	12	10		8

There are several computer software packages that can analyze the incomplete block designs discussed in this chapter. The *Minitab* General Linear Model procedure is a widely available package with this capability. The output from this routine follows. The adjusted sums of squares are the appropriate sums of squares to use for testing the difference between the means of the gasoline additives. The gasoline additives have a significant effect on the mileage.

Minitab Output

General Linear Model

```
Factor     Type Levels Values
Additive  fixed      5 1 2 3 4 5
Car      random      5 1 2 3 4 5

Analysis of Variance for Mileage, using Adjusted SS for Tests

Source     DF     Seq SS     Adj SS     Adj MS       F      P
Additive    4    31.7000    35.7333     8.9333    9.81  0.001
Car         4    35.2333    35.2333     8.8083    9.67  0.001
Error      11    10.0167    10.0167     0.9106
Total      19    76.9500
```

4.41. Construct a set of orthogonal contrasts for the data in Problem 4.40. Compute the sum of squares for each contrast.

One possible set of orthogonal contrasts is:

$$H_0 : \mu_4 + \mu_5 = \mu_1 + \mu_2 \qquad (1)$$
$$H_0 : \mu_1 = \mu_2 \qquad (2)$$
$$H_0 : \mu_4 = \mu_5 \qquad (3)$$
$$H_0 : 4\mu_3 = \mu_4 + \mu_5 + \mu_1 + \mu_2 \qquad (4)$$

The sums of squares and *F*-tests are:

Brand ->	1	2	3	4	5			
Q_i	33/4	11/4	-3/4	-14/4	-27/4	$\sum c_i Q_i$	*SS*	F_0
(1)	-1	-1	0	1	1	-85/4	30.10	33.06
(2)	1	-1	0	0	0	22/4	4.03	4.426
(3)	0	0	0	-1	1	-13/4	1.41	1.55
(4)	-1	-1	4	-1	-1	-15/4	0.19	0.21

Contrasts (1) and (2) are significant at the 1% and 5% levels, respectively.

4.42. Seven different hardwood concentrations are being studied to determine their effect on the strength of the paper produced. However the pilot plant can only produce three runs each day. As days may differ, the analyst uses the balanced incomplete block design that follows. Analyze this experiment (use $\alpha = 0.05$) and draw conclusions.

Hardwood	Days						
Concentration (%)	1	2	3	4	5	6	7
2	114				120		117
4	126	120				119	
6		137	117				134
8	141		129	149			
10		145		150	143		
12			120		118	123	
14				136		130	127

There are several computer software packages that can analyze the incomplete block designs discussed in this chapter. The Minitab General Linear Model procedure is a widely available package with this capability. The output from this routine for Problem 4.35 follows. The adjusted sums of squares are the appropriate sums of squares to use for testing the difference between the means of the hardwood concentrations.

Minitab Output

General Linear Model

```
Factor     Type Levels Values
Concentr  fixed      7  2  4  6  8 10 12 14
Days     random      7 1 2 3 4 5 6 7

Analysis of Variance for Strength, using Adjusted SS for Tests

Source     DF     Seq SS     Adj SS     Adj MS       F      P
Concentr    6    2037.62    1317.43     219.57   10.42  0.002
Days        6     394.10     394.10      65.68    3.12  0.070
Error       8     168.57     168.57      21.07
Total      20    2600.29
```

4.43. Analyze the data in Example 4.5 using the general regression significance test.

$$
\begin{array}{lllllllllll}
\mu: & 12\hat{\mu} & +3\hat{\tau}_1 & +3\hat{\tau}_2 & +3\hat{\tau}_3 & +3\hat{\tau}_4 & +3\hat{\beta}_1 & +3\hat{\beta}_2 & +3\hat{\beta}_3 & +3\hat{\beta}_4 & =870 \\
\tau_1: & 3\hat{\mu} & +3\hat{\tau}_1 & & & & +\hat{\beta}_1 & +\hat{\beta}_2 & & +\hat{\beta}_4 & =218 \\
\tau_2: & 3\hat{\mu} & & +3\hat{\tau}_2 & & & & +\hat{\beta}_2 & +\hat{\beta}_3 & +\hat{\beta}_4 & =214 \\
\tau_3: & 3\hat{\mu} & & & +3\hat{\tau}_3 & & +\hat{\beta}_1 & +\hat{\beta}_2 & +\hat{\beta}_3 & & =216 \\
\tau_4: & 3\hat{\mu} & & & & +3\hat{\tau}_4 & +\hat{\beta}_1 & & +\hat{\beta}_3 & +\hat{\beta}_4 & =222 \\
\beta_1: & 3\hat{\mu} & +\hat{\tau}_1 & & +\hat{\tau}_3 & +\hat{\tau}_4 & +3\hat{\beta}_1 & & & & =221 \\
\beta_2: & 3\hat{\mu} & +\hat{\tau}_1 & +\hat{\tau}_2 & +\hat{\tau}_3 & & & +3\hat{\beta}_2 & & & =224 \\
\beta_3: & 3\hat{\mu} & & +\hat{\tau}_2 & +\hat{\tau}_3 & +\hat{\tau}_4 & & & +3\hat{\beta}_3 & & =207 \\
\beta_4: & 3\hat{\mu} & +\hat{\tau}_1 & +\hat{\tau}_2 & & +\hat{\tau}_4 & & & & +3\hat{\beta}_4 & =218
\end{array}
$$

Applying the constraints $\sum \hat{\tau}_i = \sum \hat{\beta}_j = 0$, we obtain:

$$\hat{\mu} = 870/12,\ \hat{\tau}_1 = -9/8,\ \hat{\tau}_2 = -7/8,\ \hat{\tau}_3 = -4/8,\ \hat{\tau}_4 = 20/8,$$

$$\hat{\beta}_1 = 7/8,\ \hat{\beta}_2 = 24/8,\ \hat{\beta}_3 = -31/8,\ \hat{\beta}_4 = 0/8$$

$$R(\mu,\tau,\beta) = \hat{\mu}y_{..} + \sum_{i=1}^{4}\hat{\tau}_i y_{i.} + \sum_{j=1}^{4}\hat{\beta}_j y_{.j} = 63{,}152.75$$

with 7 degrees of freedom.

$$\sum\sum y_{ij}^2 = 63{,}156.00$$

$$SS_E = \sum\sum y_{ij}^2 - R(\mu,\tau,\beta) = 63156.00 - 63152.75 = 3.25.$$

To test H_o: $\tau_i = 0$ the reduced model is $y_{ij} = \mu + \beta_j + \varepsilon_{ij}$. The normal equations used are:

$$\begin{array}{llllllll}
\mu: & 12\hat{\mu} & +3\hat{\beta}_1 & +3\hat{\beta}_2 & +3\hat{\beta}_3 & +3\hat{\beta}_4 & =870 \\
\beta_1: & 3\hat{\mu} & +3\hat{\beta}_1 & & & & =221 \\
\beta_2: & 3\hat{\mu} & & +3\hat{\beta}_2 & & & =224 \\
\beta_3: & 3\hat{\mu} & & & +3\hat{\beta}_3 & & =207 \\
\beta_4: & 3\hat{\mu} & & & & +3\hat{\beta}_4 & =218
\end{array}$$

Applying the constraint $\sum \hat{\beta}_j = 0$, we obtain:

$$\hat{\mu} = \frac{870}{12},\ \hat{\beta}_1 = \frac{7}{6},\ \hat{\beta}_2 = \frac{13}{6},\ \hat{\beta}_3 = \frac{-21}{6},\ \hat{\beta}_4 = \frac{1}{6}$$

$$R(\mu,\beta) = \hat{\mu}y_{..} + \sum_{j=1}^{4}\hat{\beta}_j y_{.j} = 63{,}130.00$$

with 4 degrees of freedom.

$$R(\tau|\mu,\beta) = R(\mu,\tau,\beta) - R(\mu,\beta) = 63152.75 - 63130.00 = 22.75 = SS_{Treatments}$$

with 7 – 4 = 3 degrees of freedom. $R(\tau|\mu,\beta)$ is used to test H_o: $\tau_i = 0$.

The sum of squares for blocks is found from the reduced model $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$. The normal equations used are:

Model Restricted to $\beta_j = 0$:

$$\begin{array}{llllllll}
\mu: & 12\hat{\mu} & +3\hat{\tau}_1 & +3\hat{\tau}_2 & +3\hat{\tau}_3 & +3\hat{\tau}_4 & =870 \\
\tau_1: & 3\hat{\mu} & +3\hat{\tau}_1 & & & & =218 \\
\tau_2: & 3\hat{\mu} & & +3\hat{\tau}_2 & & & =214 \\
\tau_3: & 3\hat{\mu} & & & +3\hat{\tau}_3 & & =216 \\
\tau_4: & 3\hat{\mu} & & & & +3\hat{\tau}_4 & =222
\end{array}$$

The sum of squares for blocks is found as in Example 4.5. We may use the method shown above to find an adjusted sum of squares for blocks from the reduced model, $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$.

4.44. Prove that $\dfrac{k\sum_{i=1}^{a} Q_i^2}{(\lambda a)}$ is the adjusted sum of squares for treatments in a BIBD.

We may use the general regression significance test to derive the computational formula for the adjusted treatment sum of squares. We will need the following:

$$\hat{\tau}_i = \frac{kQ_i}{(\lambda a)},\ kQ_i = ky_{i.} - \sum_{i=1}^{b} n_{ij} y_{.j}$$

$$R(\mu,\tau,\beta) = \hat{\mu} y_{..} + \sum_{i=1}^{a} \hat{\tau}_i y_{i.} + \sum_{j=1}^{b} \hat{\beta}_j y_{.j}$$

and the sum of squares we need is:

$$R(\tau|\mu,\beta) = \hat{\mu} y_{..} + \sum_{i=1}^{a} \hat{\tau}_i y_{i.} + \sum_{j=1}^{b} \hat{\beta}_j y_{.j} - \sum_{j=1}^{b} \frac{y_{.j}^2}{k}$$

The normal equation for β is, from equation (4.35),

$$\beta : k\hat{\mu} + \sum_{i=1}^{a} n_{ij}\hat{\tau}_i + k\hat{\beta}_j = y_{.j}$$

and from this we have:

$$ky_{.j}\hat{\beta}_j = y_{.j}^2 - ky_{.j}\hat{\mu} - y_{.j}\sum_{i=1}^{a} n_{ij}\hat{\tau}_i$$

therefore,

$$R(\tau|\mu,\beta) = \hat{\mu} y_{..} + \sum_{i=1}^{a} \hat{\tau}_i y_{i.} + \sum_{j=1}^{b}\left[\frac{y_{.j}^2}{k} - \frac{k\hat{\mu} y_{.j}}{k} - \frac{y_{.j}\sum_{i=1}^{a} n_{ij}\hat{\tau}_i}{k} - \frac{y_{.j}^2}{k}\right]$$

$$R(\tau|\mu,\beta) = \sum_{i=1}^{a} \hat{\tau}_i\left(y_{i.} - \frac{1}{k}\sum_{i=1}^{a} n_{ij} y_{.j}\right) = \sum_{i=1}^{a} Q_i\left(\frac{kQ_i}{\lambda a}\right) = k\sum_{i=1}^{a}\left(\frac{Q_i^2}{\lambda a}\right) \equiv SS_{Treatments(adjusted)}$$

4.45. An experimenter wishes to compare four treatments in blocks of two runs. Find a BIBD for this experiment with six blocks.

Treatment	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6
1	X	X	X			
2	X			X	X	
3		X		X		X
4			X		X	X

Note that the design is formed by taking all combinations of the 4 treatments 2 at a time. The parameters of the design are $\lambda = 1$, $a = 4$, $b = 6$, $k = 3$, and $r = 2$

4.46. An experimenter wishes to compare eight treatments in blocks of four runs. Find a BIBD with 14 blocks and $\lambda = 3$.

The design has parameters $a = 8$, $b = 14$, $\lambda = 3$, $r = 2$ and $k = 4$. It may be generated from a 2^3 factorial design confounded in two blocks of four observations each, with each main effect and interaction successively confounded (7 replications) forming the 14 blocks. The design is discussed by John (1971, pg. 222) and Cochran and Cox (1957, pg. 473). The design follows:

Blocks	1=(*I*)	2=*a*	3=*b*	4=*ab*	5=*c*	6=*ac*	7=*bc*	8=*abc*
1	X		X		X		X	
2		X		X		X		X
3	X		X			X		X
4		X		X	X		X	
5	X	X			X	X		
6			X	X			X	X
7	X	X					X	X
8			X	X	X	X		
9	X	X	X	X				
10					X	X	X	X
11	X			X		X	X	
12		X	X		X			X
13	X			X	X			X
14		X	X			X	X	

4.47. Perform the interblock analysis for the design in Problem 4.40.

The interblock analysis for Problem 4.33 uses $\hat{\sigma}^2 = 0.91$ and $\hat{\sigma}_\beta^2 = 2.63$. A summary of the interblock, intrablock and combined estimates is:

Parameter	Intrablock	Interblock
τ_1	2.20	-1.80
τ_2	0.73	0.20
τ_3	-0.20	-5.80
τ_4	-0.93	9.20
τ_5	-1.80	-1.80

4.48. Perform the interblock analysis for the design in Problem 4.42.

The interblock analysis for Pproblem 4.42 uses $\hat{\sigma}^2 = 21.07$ and

$$\sigma_\beta^2 = \frac{\left[MS_{Blocks(adj)} - MS_E\right](b-1)}{a(r-1)} = \frac{[65.68 - 21.07](6)}{7(2)} = 19.12 .$$

A summary of the interblock, intrablock, and combined estimates is give below

Parameter	Intrablock	Interblock	Combined
τ_1	-12.43	-11.79	-12.38
τ_2	-8.57	-4.29	-7.92
τ_3	2.57	-8.79	1.76
τ_4	10.71	9.21	10.61
τ_5	13.71	21.21	14.67
τ_6	-5.14	-22.29	-6.36
τ_7	-0.86	10.71	-0.03

4.49. Verify that a BIBD with the parameters a = 8, r = 8, k = 4, and b = 16 does not exist.

These conditions imply that $\lambda = \frac{r(k-1)}{a-1} = \frac{8(3)}{7} = \frac{24}{7}$, which is not an integer, so a balanced design with these parameters cannot exist.

4.50. Show that the variance of the intra block estimators $\{\hat{\tau}_i\}$ is $\frac{k((a-1))\sigma^2}{(\lambda a^2)}$.

Note that $\hat{\tau}_i = \frac{kQ_i}{(\lambda a)}$, and $Q_i = y_{i.} - \frac{1}{k}\sum_{j=1}^{b} n_{ij} y_{.j}$, and $kQ_i = ky_{i.} - \sum_{j=1}^{b} n_{ij} y_{.j} = (k-1)y_{i.} - \left(\sum_{j=1}^{b} n_{ij} y_{.j} - y_{i.}\right)$

$y_{i.}$ contains r observations, and the quantity in the parenthesis is the sum of $r(k$-1) observations, not including treatment i. Therefore,

$$V(kQ_i) = k^2 V(Q_i) = r(k-1)^2\sigma^2 + r(k-1)\sigma^2$$

or

$$V(Q_i) = \frac{1}{k^2}\left[r(k-1)\sigma^2\{(k-1)+1\}\right] = \frac{r(k-1)\sigma^2}{k}$$

To find $V(\hat{\tau}_i)$, note that:

$$V(\hat{\tau}_i) = \left(\frac{k}{\lambda a}\right)^2 V(Q)_i = \left(\frac{k}{\lambda a}\right)^2 \frac{r(k-1)}{k}\sigma^2 = \frac{kr(k-1)}{(\lambda a)^2}\sigma^2$$

However, since $\lambda(a-1) = r(k-1)$, we have:

$$V\left(\ddot{\varphi}_i\right)=\frac{k(a-1)}{\lambda a^2}\sigma^2$$

Furthermore, the $\{\ddot{\varphi}_i\}$ are not independent, this is required to show that $V\left(\ddot{\varphi}_i-\ddot{\varphi}_j\right)=\frac{2k}{\lambda a}\sigma^2$

4.51. ***Extended incomplete block designs.*** Occasionally the block size obeys the relationship $a < k < 2a$. An extended incomplete block design consists of a single replicate or each treatment in each block along with an incomplete block design with $k^* = k\text{-}a$. In the balanced case, the incomplete block design will have parameters $k^* = k\text{-}a$, $r^* = r\text{-}b$, and λ^*. Write out the statistical analysis. (Hint: In the extended incomplete block design, we have $\lambda = 2r\text{-}b+\lambda^*$.)

As an example of an extended incomplete block design, suppose we have a=5 treatments, b=5 blocks and k=9. A design could be found by running all five treatments in each block, plus a block from the balanced incomplete block design with $k^* = k\text{-}a$=9-5=4 and λ^*=3. The design is:

Block	Complete Treatment	Incomplete Treatment
1	1,2,3,4,5	2,3,4,5
2	1,2,3,4,5	1,2,4,5
3	1,2,3,4,5	1,3,4,5
4	1,2,3,4,5	1,2,3,4
5	1,2,3,4,5	1,2,3,5

Note that r=9, since the augmenting incomplete block design has r*=4, and $r= r^* + b$ = 4+5=9, and $\lambda = 2r\text{-}b+\lambda^*$=18-5+3=16. Since some treatments are repeated in each block it is possible to compute an error sum of squares between repeat observations. The difference between this and the residual sum of squares is due to interaction. The analysis of variance table is shown below:

Source	*SS*	*DF*
Treatments (adjusted)	$k\sum\frac{Q_i^2}{a\lambda}$	a-1
Blocks	$\sum\frac{y_{.j}^2}{k}-\frac{y_{..}^2}{N}$	b-1
Interaction	Subtraction	(a-1)(b-1)
Error	[SS between repeat observations]	b(k-a)
Total	$\sum\sum y_{ij}^2-\frac{y_{..}^2}{N}$	N-1

4.52. Suppose that a single-factor experiment with five levels of the factor has been conducted. There are three replicates and the experiment has been conducted as a complete randomized design. If the experiment had been conducted in blocks, the pure error degrees of freedom would be reduced by (choose the correct answer):

(c) 2

4.53. Physics graduate student Laura Van Ertia has conducted a complete randomized design with a single factor, hoping to solve the mystery of the unified theory and complete her dissertation. The results of this experiment are summarized in the following ANOVA display:

Source	*DF*	*SS*	*MS*	*F*
Factor	-	-	14.18	-
Error	-	37.75	-	
Total	23	108.63		

The completed ANOVA is as follows:

Source	*DF*	*SS*	*MS*	*F*	*P*
Factor	5	70.88	14.18	6.76	0.00104
Error	18	37.75	2.10		
Total	23	108.63			

Answer the following questions about this experiment.

(a) The sum of squares for the factor is 70.88.

(b) The number of degrees of freedom for the single factor in the experiment is 5.

(c) The number of degrees of freedom for the error is 18.

(d) The mean square for error is 2.10.

(e) The value of the test statistic is 6.67.

(f) If the significance level is 0.05, your conclusions are not to reject the null hypothesis. No.

(g) An upper bound on the *P*-value for the test statistic is 0.001.

(h) A lower bound on the *P*-value for the test statistic is 0.0001.

(i) Laura used 6 levels of the factor in this experiment.

(j) Laura replicated this experiment 4 times.

(k) Suppose that Laura had actually conducted this experiment as a random complete block design and the sum of squares for the blocks was 12. Reconstruct the ANOVA display above to reflect this new situation. How much has the blocking reduced the estimate of the experimental error?

Source	*DF*	*SS*	*MS*	*F*	*P*
Block	3	12.00	4.00		
Factor	5	70.88	14.18	9.91	0.00011
Error	18	25.75	1.43		
Total	23	108.63			

The blocking reduced the SS_{error} by 12 and the MS_{error} by 0.67 (32%).

4.54. Consider the direct mail marketing experiment in Problem 4.8. suppose that this experiment has been run as a complete randomized design, ignoring potential regional differences, but that exactly the same data was obtained. Reanalyze the experiment under this new assumption. What difference would ignoring the blocking have on the results and conclusions?

The solution for Problem 4.8 used a square root transformation, so the solution below also includes this same transformation. The results below are similar to Problem 4.8 in that the the difference in designs is statistically significant; however, the F value changed from 60.46 to only 7.02. The corresponding P value increased from 0.0001 to 0.0145.

Response: Number of responses Transform: Square root Constant: 0

ANOVA for Selected Factorial Model

Analysis of variance table [Terms added sequentially (first to last)]

Source	**Sum of Squares**	**DF**	**Mean Square**	**F Value**	**Prob > F**	
Model	60.73	2	30.37	7.02	0.0145	significant
A-Design	*60.73*	*2*	*30.37*	*7.02*	*0.0145*	
Pure Error	38.90	9	4.32			
Cor Total	99.64	11				

The Model F-value of 7.02 implies the model is significant. There is only
a 1.45% chance that a "Model F-Value" this large could occur due to noise.

Std. Dev.	2.08	R-Squared	0.6095
Mean	18.52	Adj R-Squared	0.5228
C.V. %	11.23	Pred R-Squared	0.3058
PRESS	69.16	Adeq Precision	4.803

Treatment Means (Adjusted, If Necessary)

	Estimated Mean	**Standard Error**
1-1	17.17	1.04
2-2	21.69	1.04
3-3	16.69	1.04

Treatment	**Mean Difference**	**DF**	**Standard Error**	**t for H_0 Coeff=0**	**Prob > |t|**
1 vs 2	-4.52	1	1.47	-3.07	0.0133
1 vs 3	0.48	1	1.47	0.33	0.7525
2 vs 3	4.99	1	1.47	3.40	0.0079