STATISTICS 402 – Assignment 7 – Sample Solution

The activities in this homework involve four treatments; A, B, C, and D and five blocks; white, blue, green, yellow and pink. The treatments could be food additives, the blocks litters of mice, and the response amount of weight gained. Alternatively, the treatments could be a combination of starting technique (without starting blocks, with starting blocks) and shoes (no spikes, spikes) with the response the time to run 400-meters, the blocks would be different individuals.

Activity 1: The first activity consists of two different randomizations, complete randomization and restricted randomization (blocking) and their subsequent analyses.

a) Complete Randomization

- Shuffle all 20 squares so that they are well mixed.
- Deal out 5 squares and record the values for Treatment A in the table below.
- Deal out 5 squares from the remaining 15 and record the values for Treatment B.
- Deal out 5 squares from the remaining 10 and record the values for Treatment C.
- There should be 5 squares remaining, record the values for Treatment D.

	Treatment			
	A	B	C	D
	45.6	**46.9**	**52.3**	**47.2**
	52.1	**46.3**	**42.7**	**56.3**
	54.8	**47.1**	**54.6**	**48.1**
	56.4	**50.5**	**50.3**	**55.0**
	48.9	**51.5**	**51.6**	**52.5**
Mean	**51.56**	**48.46**	**50.30**	**51.82**

b) Analysis of CRD

- Compute treatment means and construct an Analysis of Variance table.

Source	**df**	**SS**	**MS**	**F**	**Pr > F**
Treatment	**3**	**35.314**	**11.77**	**0.763**	**0.531**
Error	**16**	**246.932**	**15.43**		
C. Total	**19**	**282.246**			

- Report the value of the F-test and P-value for comparing treatments. What do these tell you about the differences between the various treatments?

  F = 0.763, P-value = 0.531. Because the P-value is not small we cannot reject the null hypothesis. There are no differences among the 4 treatment population means.

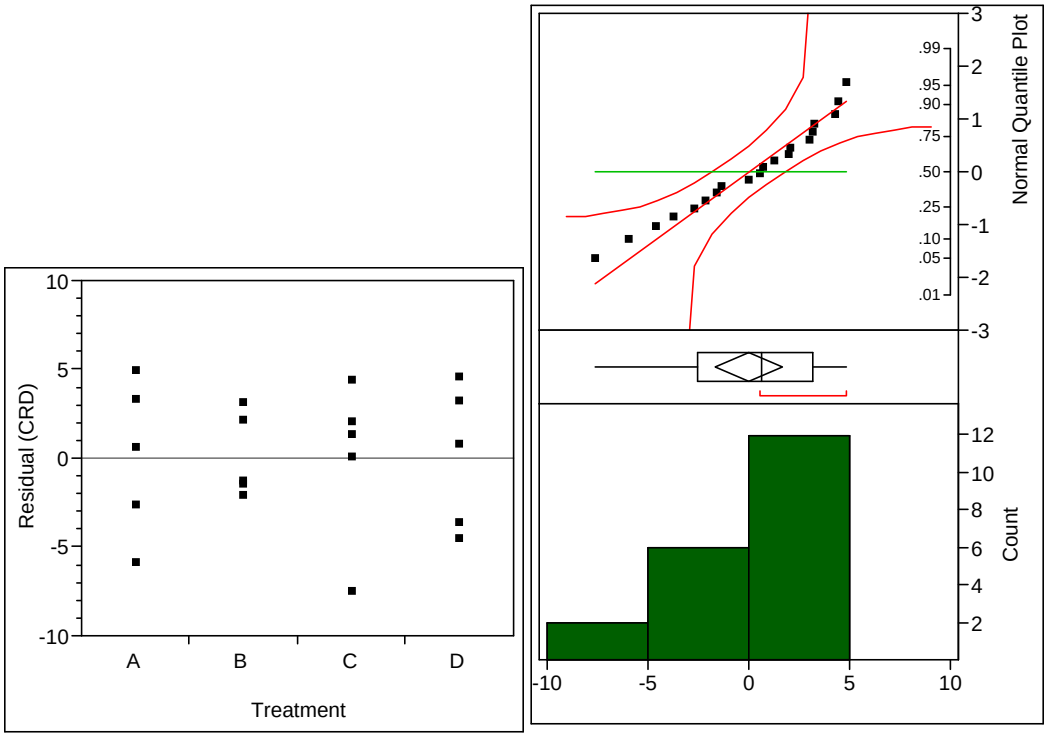
- Note particularly the value for the mean square error.

  Mean Square Error = 15.43.

- If the F-test indicates that there are some statistically significant differences follow up with an appropriate multiple comparison procedure. That is, compute a value for the adjusted Least Significant Difference (Bonferroni), or Kramer-Tukey HSD, and use it to determine which of the treatment means differ significantly.

  Because the F statistic is not significant, no further analysis is appropriate.

- Look at plots of residuals versus treatments and the distribution of residuals. What do these plots tell you about the conditions necessary for the analysis above?



There appear to be differences in variation for the four treatments. Treatment B has the smallest variation and treatments A and C have the largest. The equal variance condition may not be met. The distribution of residuals is skewed left (histogram is a stair step to the left and the box plot shows longer intervals to the left as well as the mean being less than the median). The Normal quantile plot shows a definite curve (above, below, above the diagonal line representing a normal model). The normally distributed errors condition may not be met.

c) Restricted Randomization, Blocking

- Take the 4 white squares and shuffle them so they are well mixed.
- Deal out one white square and record the value for treatment A.
- Deal out another white square and record the value for treatment B.
- Deal out another white square and record the value for treatment C.
- For the remaining white square record the value for treatment D.
- Repeat the above steps for the blue, green, yellow and pink squares.

	Treatment			
	A	B	C	D
White	**52.1**	**50.5**	**49.5**	**52.5**
Blue	**56.4**	**56.6**	**50.3**	**55.0**
Green	**54.3**	**53.6**	**53.8**	**56.3**
Yellow	**46.9**	**47.1**	**43.6**	**50.9**
Pink	**46.3**	**45.2**	**42.7**	**48.1**
Mean	**51.20**	**50.60**	**47.98**	**52.56**

d) Analysis of RCBD

- Compute treatment means and construct an Analysis of Variance table. Be sure to include a block effect in your model.

Source	**df**	**SS**	**MS**	**F**	**Pr > F**
Treatment	**3**	**55.326**	**18.44**	**9.60**	**0.0016**
Block	**4**	**274.553**	**68.64**		
Error	**12**	**23.047**	**1.921**		
C. Total	**19**	**352.926**			

- Report the value of the F-test and P-value for comparing treatments. What do these tell you about the difference between the various treatments?

  F = 9.60, P-value = 0.0016. Because the P-value is so small we should reject the null hypothesis. There are differences among some of the 4 treatment population means.

- Note particularly the value for the mean square error.

  Mean Square Error = 1.921.

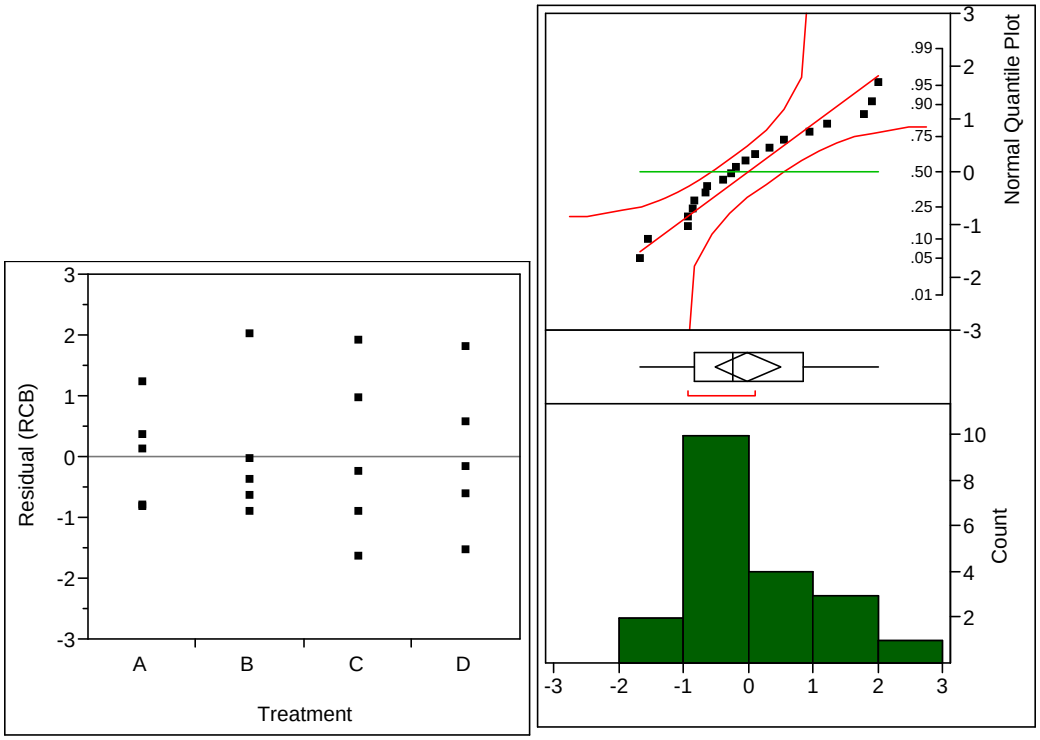
- If the F-test indicates that there are some statistically significant differences among the treatments follow up with a multiple comparison procedure. That is, compute a value for the adjusted Least Significant Difference (Bonferroni), or Kramer-Tukey HSD, and use it to determine which of the treatment means differ significantly.

HSD = 2.96883(0.87649) = 2.60

Treatment	Mean	
D	52.56	A
A	51.20	A
B	50.60	A
C	47.98	B

Treatment means not connected by the same letter at statistically different.

- Look at plots of residuals versus treatments and the distribution of residuals. What do these plots tell you about the conditions necessary for the analysis above?



Treatment A has the least amount of variation but there is much less variation overall. The condition of equal variances is probably met. The distribution of residuals is skewed right (histogram is a stair step to the right, mean is bigger than the median and the Normal quantile plot curves below, above, below the line representing the normal model). The normally distributed errors condition may be in doubt.

e) Comparison of Designs

- Compare the results of the two analyses. With each design/analysis indicate what treatment means are significantly different.

Completely Randomized	**Randomized Complete Block**
Found no statistically significant differences among the treatment means.	**Found statistically significant differences between treatment C and all the other treatments.**

- Compare the estimates of error variability (e.g. Mean Square Error) for the two designs, complete randomization and restricted randomization. What do you notice?

Completely Randomized	**Randomized Complete Block**
Mean Square Error = 15.43	**Mean Square Error = 1.92**

The randomized complete block design resulted in a much smaller Mean Square Error.

- Write a one-paragraph explanation of the importance of blocking in this experiment?

With a completely randomized design, the experiment was unable to detect any differences among treatment means. The mean square error, which estimates the random error variance, was very large owing to differences in the experimental material (e.g. the colors of the tiles). With a randomized complete block design, the estimate of random error variance was much smaller, allowing the experiment to detect some differences in treatment mean. The randomized complete block design and subsequent analysis, where block is included in the model, was able to account for variation in experimental material separate from random error.

Activity 2: The second activity looks at the effect of block size. So that there will be balance in the simulated experiments use only the Blue, Green, Pink and Yellow squares for this activity.

a) 4 Blocks

To save you some time, rather than redo the randomization using four blocks, use the data from Activity 1, part c) but exclude the white block, i.e. use only the data for the Blue, Green, Pink and Yellow blocks.

	Treatment			
	A	B	C	D
Blue	**56.4**	**56.6**	**50.3**	**55.0**
Green	**54.3**	**53.6**	**53.8**	**56.3**
Yellow	**46.9**	**47.1**	**43.6**	**50.9**
Pink	**46.3**	**45.2**	**42.7**	**48.1**
Mean	**50.975**	**50.625**	**47.6**	**52.575**

- Compute treatment means and construct an Analysis of Variance table. Be sure to include a block effect in your model.

Source	**df**	**SS**	**MS**	**F**	**Pr > F**
Treatment	**3**	**51.776875**	**17.2589**	**7.495**	**0.0081**
Block	**3**	**272.956875**	**90.9856**		
Error	**9**	**20.725625**	**2.3028**		
C. Total	**15**	**345.459375**			

- Report the value of the F-test and P-value for comparing treatments. What do these tell you about the difference between the various treatments?

  F = 7.495, P-value = 0.0081. Because the P-value is so small we should reject the null hypothesis. There are differences among some of the 4 treatment population means.

- Note particularly the value for the mean square error.

  Mean Square Error = 2.303.

- If the F-test indicates that there are some statistically significant differences among the treatments follow up with a multiple comparison procedure. That is, compute a value for the adjusted Least Significant Difference (Bonferroni), or Kramer-Tukey HSD, and use it to determine which of the treatment means differ significantly.

HSD = 3.12182(1.07304) = 3.35

Treatment	**Mean**		
D	**52.575**	**A**	
A	**50.975**	**A**	
B	**50.625**	**A**	**B**
C	**47.600**		**B**

Treatment means not connected by the same letter at statistically different.

b) 2 Blocks – Blue/Green and Pink/Yellow

Shuffle the Blue and Green squares together until they are well mixed. Deal out two and record the values for A. Deal out two from the remaining 6 and record the values for B. Deal out two from the remaining 4 and record the values for C. For the remaining two squares record the values for D. Repeat this process with the Pink and Yellow squares.

	Treatment			
	A	B	C	D
Blue/Green	**56.5** **54.3**	**56.6** **54.3**	**51.6** **50.3**	**55.0** **56.8**
Pink/Yellow	**46.1** **48.9**	**44.7** **45.2**	**44.9** **44.3**	**47.6** **49.1**
Mean	**51.45**	**50.20**	**47.775**	**52.125**

- Compute treatment means and construct an Analysis of Variance table. Be sure to include a block effect in your model.

Source	**df**	**SS**	**MS**	**F**	**Pr > F**
Treatment	**3**	**44.0325**	**14.6775**	**7.325**	**0.0057**
Block	**1**	**260.8225**	**260.8225**		
Error	**11**	**22.0425**	**2.0039**		
C. Total	**15**	**326.8975**			

- Report the value of the F-test and P-value for comparing treatments. What do these tell you about the difference between the various treatments?

  F = 7.325, P-value = 0.0057. Because the P-value is so small we should reject the null hypothesis. There are differences among some of the 4 treatment population means.

- Note particularly the value for the mean square error.

  Mean Square Error = 2.004.

- If the F-test indicates that there are some statistically significant differences among the treatments follow up with a multiple comparison procedure. That is, compute a value for the adjusted Least Significant Difference (Bonferroni), or Kramer-Tukey HSD, and use it to determine which of the treatment means differ significantly.

  HSD = 3.00951(1.00097) = 3.01

Treatment	**Mean**		
D	**52.125**	**A**	
A	**51.450**	**A**	
B	**50.200**	**A**	**B**
C	**47.775**		**B**

  Treatment means not connected by the same letter at statistically different.

c) 2 Blocks – Blue/Pink and Green/Yellow

Shuffle the Blue and Pink squares together until they are well mixed. Deal out two and record the values for A. Deal out two from the remaining 6 and record the values for B. Deal out two from the remaining 4 and record the values for C. For the remaining two squares record the values for D. Repeat this process with the Green and Yellow squares.

	Treatment			
	A	B	C	D
Blue/Pink	**56.4** **46.3**	**56.6** **46.1**	**50.3** **42.7**	**55.0** **47.2**
Green/Yellow	**45.6** **47.1**	**53.6** **48.9**	**52.3** **44.9**	**56.3** **56.8**
Mean	**48.85**	**51.30**	**47.55**	**53.825**

- Compute treatment means and construct an Analysis of Variance table. Be sure to include a block effect in your model.

Source	**df**	**SS**	**MS**	**F**	**Pr > F**
Treatment	**3**	**92.256875**	**30.7523**	**1.288**	**0.327**
Block	**1**	**1.500625**	**1.5006**		
Error	**11**	**262.726875**	**23.8843**		
C. Total	**15**	**356.484375**			

- Report the value of the F-test and P-value for comparing treatments. What do these tell you about the difference between the various treatments?

  F = 1.288, P-value = 0.327. Because the P-value is not small we cannot reject the null hypothesis. There are no differences among the 4 treatment population means.

- Note particularly the value for the mean square error.

  Mean Square Error = 23.88.

- If the F-test indicates that there are some statistically significant differences among the treatments follow up with a multiple comparison procedure. That is, compute a value for the adjusted Least Significant Difference (Bonferroni), or Kramer-Tukey HSD, and use it to determine which of the treatment means differ significantly.

  Because the F statistic is not significant, no further analysis is appropriate.

d) Comparison

- Compare the results of the three analyses. With each design/analysis indicate what treatment means are significantly different.

RCB(4 blocks)	RCB(2 blocks – BG/PY)	RCB(2 blocks – BP/GY)
Treatments A and D are different from C.	**Treatments A and D are different from C.**	**No differences amongst treatments were found.**

- Compare the estimates of error variability (e.g. Mean Square Error) for the three designs, blocking with 4 blocks and the two experiments blocking with 2 blocks. What do you notice?

RCB(4 blocks)	RCB(2 blocks – BG/PY)	RCB(2 blocks – BP/GY)
MS_{Error} = 2.303	**MS_{Error} = 2.004**	**MS_{Error} = 23.884**

  The design with BG/PY blocks gave about the same Mean Square Error as the design with 4 blocks. The design with BP/GY blocks gave a much larger (factor of 10) Mean Square error than the other two designs.

- Which design was most effective in reducing the Mean Square Error?

  The design with 2 blocks – BG/PY gave the smallest Mean Square Error but this was very close to the Mean Square Error for the 4 block design.