Statistics 106

Solutions for Homework 6

Due : Nov. 14, 2016, In Class

20.5

Brainstorming. A researcher investigated whether brainstorming is more effective for larger groups than for smaller ones by setting up four groups of agribusiness executives, the group sizes being two, three, four, and five, respectively. He also set up four groups of agribusiness scientists, the group sizes being the same as for the agribusiness executives. The researcher gave each group the same problem: ”How can Canada increase the value of its agricultural exports?” Each group was allowed 30 minutes to generate ideas. The variable of interest was the number or different ideas proposed by the group. The results, classified by type of group (factor A) and size of group (factor B), were:

		Factor B (size of group)			
	Factor A (type of group)	$j = 1$ Two	$j = 2$ Three	$j = 3$ Four	$j = 4$ Five
$i = 1$	Agribusiness executives	18	22	31	32
$i = 2$	Agribusiness scientists	15	23	29	33

Assume that no-interaction ANOVA model is appropriate.

a. Plot the data in the format of Figure 20.1 in the textbook (plot of the observations Y_{ij}). Does it appear that interaction effects are present? Does it appear that factor A and factor B main effects are present? Discuss.

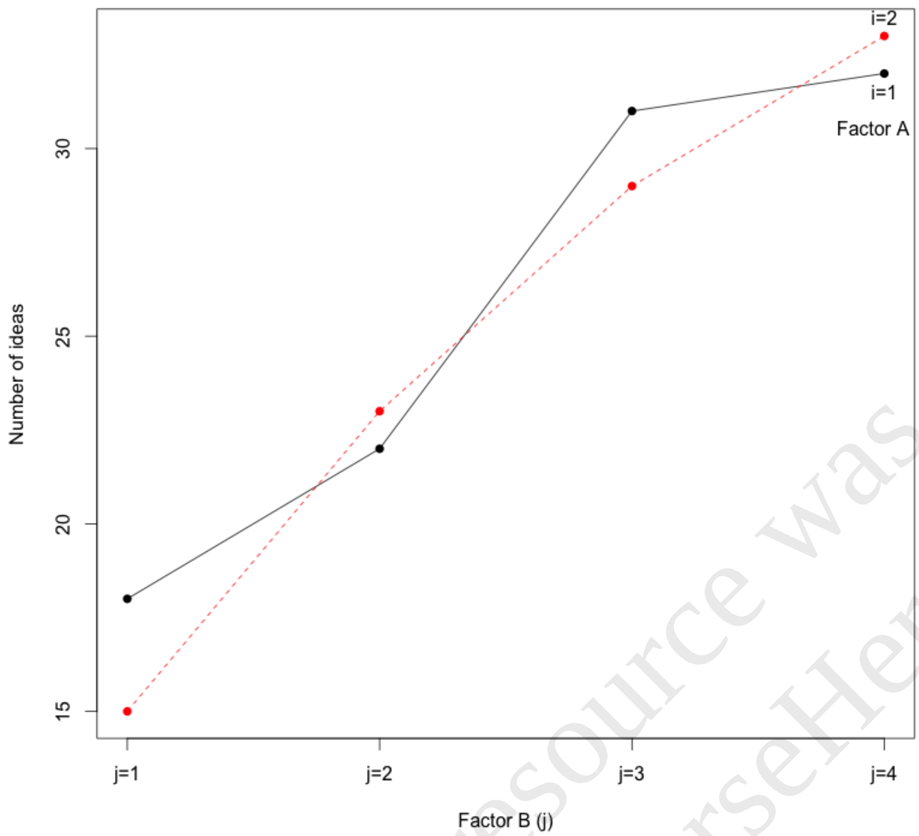
b. Conduct separate tests for type of group and size of group main effects. In each test, use level of significance $\alpha = .01$ and state the alternatives, decision rule, and conclusion. What is the P-value for each test?

c. Obtain confidence intervals for $D_1 = \mu_{.2} - \mu_{.1}$, $D_2 = \mu_{.3} - \mu_{.2}$, and $D_3 = \mu_{.4} - \mu_{.3}$; use the Bonferroni procedure with a 95 percent family confidence coefficient. State your findings.

d. Is the Bonferroni procedure used in pan (c) the most efficient one here? Explain.

Solution:

a. The plot of the observations are shown below:



It appears that interaction effects are very little, since the two curves are nearly parallel. Factor A main effects are also very little, since the two curves are nearly overlapped. Factor B main effects are present, since the two curves have large departure from horizontal.

b. We first construct the ANOVA table. Note that the error sum of squares is given by $SSAB$, which can be obtained by $SSAB = SSTO - SSA - SSB$ in R.

Source	SS	df	MS
Type of group	1.125	1	1.125
Size of group	318.375	3	106.125
Error	6.375	3	2.125
Total	325.875	7	

F test for type of group main effects:
$H_0: \alpha_1 = \alpha_2 = 0$, H_a: not both α_1 and α_2 are 0. $F^* = MSA/MSAB = 1.125/2.125 = .53$, $F(1-\alpha; a-1, (a-1)(b-1)) = F(.99; 1, 3) = 34.1$. If $F^* < 34.1$, conclude H_0, otherwise H_a. Here conclude H_0. P-value=.52.
$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$, H_a: not all β_i's are 0. $F^* = MSB/MSAB = 106.125/2.125 = 49.94$, $F(1-\alpha; b-1, (a-1)(b-1)) = F(.99; 3, 3) = 29.5$. If $F^* < 29.5$, conclude H_0, otherwise H_a. Here conclude H_a. P-value=.005.

c. $\hat{D}_1 = \bar{Y}_{.2} - \bar{Y}_{.1} = 6$, $\hat{D}_2 = \bar{Y}_{3.} - \bar{Y}_{2.} = 7.5$, $\hat{D}_3 = \bar{Y}_{.4} - \bar{Y}_{.3} = 2.5$. $s(\hat{D}_1) = s(\hat{D}_2) = s(\hat{D}_3) = \sqrt{2MSAB/a} = 1.4577$. $B = t(1-\alpha/(2g);(a-1)(b-1)) = t(.99167;3) = 4.857$.
Thus, the 95% family-wise CI for D_1, D_2 and D_3 using the Bonferroni procedure are $6 \pm 4.857(1.4577) = [-1.08, 13.08]$, $7.5 \pm 4.857(1.4577) = [.42, 14.58]$ and $2.5 \pm 4.857(1.4577) = [-4.58, 9.58]$.

d. Scheffe's procedure and Tukey's procedure are also applicable in part (c).
$S = \sqrt{(b-1)F(1-\alpha;b-1,(a-1)(b-1))} = \sqrt{3F(.95;3,3)} = 5.275$, $T = \frac{1}{\sqrt{2}}q(1-\alpha;b,(a-1)(b-1)) = \frac{1}{\sqrt{2}}q(.95;4,3) = 4.822$.
Since $T < B < S$, the Tukey's procedure is the most efficient in part (c).

20.6
Refer to **Brainstorming** Problem 20.5. It is desired to estimate μ_{14}. Obtain a point estimate of μ_{14}.
Solution:
$\hat{\mu}_{14} = \bar{Y}_{1.} + \bar{Y}_{.4} - \bar{Y}_{..} = 25.75 + 32.5 - 25.375 = 32.875$.

20.7
Refer to **Brainstorming** Problem 20.5. Conduct the Tukey test for additivity; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. If the additive model is not appropriate, what might you do?
Solution:
By R, $SSAB^* = 1.4688$, $SSRem^* = SSTO - SSA - SSB - SSAB^* = 4.9062$.
$H_0: D = 0$, $H_1: D \neq 0$. $F^* = \frac{SSAB^*/1}{SSRem^*/(ab-a-b)} = \frac{1.4688/1}{4.9062/2} = .60$, $F(1-\alpha;1,ab-a-b) = F(.99;1,2) = 98.5$. If $F^* < 98.5$, conclude H_0, otherwise H_a. Here conclude H_0.
If the additive model is not appropriate, efforts should be made to remove the interactions so that the analysis can be utilized. One possibility is to try simple transformations of the response variable, such as a square root or a logarithmic transformation. Another possibility is to search in the family of power transformations on Y in connection with the Box-Cox transformations.

21.7
Fat in diets. A researcher studied the effects of three experimental diets with varying fat contents on the total lipid (fat) level in plasma. Total lipid level is a widely used predictor of coronary heart disease. Fifteen male subjects who were within 20 percent of their ideal body weight were grouped into five blocks according to age. Within each block, the three experimental diets were randomly assigned to the three subjects. Data on reduction in lipid level (in grams per liter) after the subjects were on the diet for a fixed period of time follow.

		Fat Content of Diet		
Block		$j = 1$	$j = 2$	$j = 3$
i		Extremely Low	Fairly Low	Moderately Low
1	Ages 15-24	.73	.67	.15
2	Ages 25-34	.86	.75	.21
3	Ages 35-44	.94	.81	.26
4	Ages 45-54	1.40	1.32	.75
5	Ages 55-64	1.62	1.41	.78

a. Why do you think that age of subject was used as a blocking variable?

b. Obtain the residuals for randomized block model (21.1) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What are your findings?

c. Plot the responses Y_{ij} by blocks in the format or Figure 21.2 in the textbook. What does this plot suggest about the appropriateness of the no-interaction assumption here?

d. Conduct the Tukey test for additivity of block and treatment effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

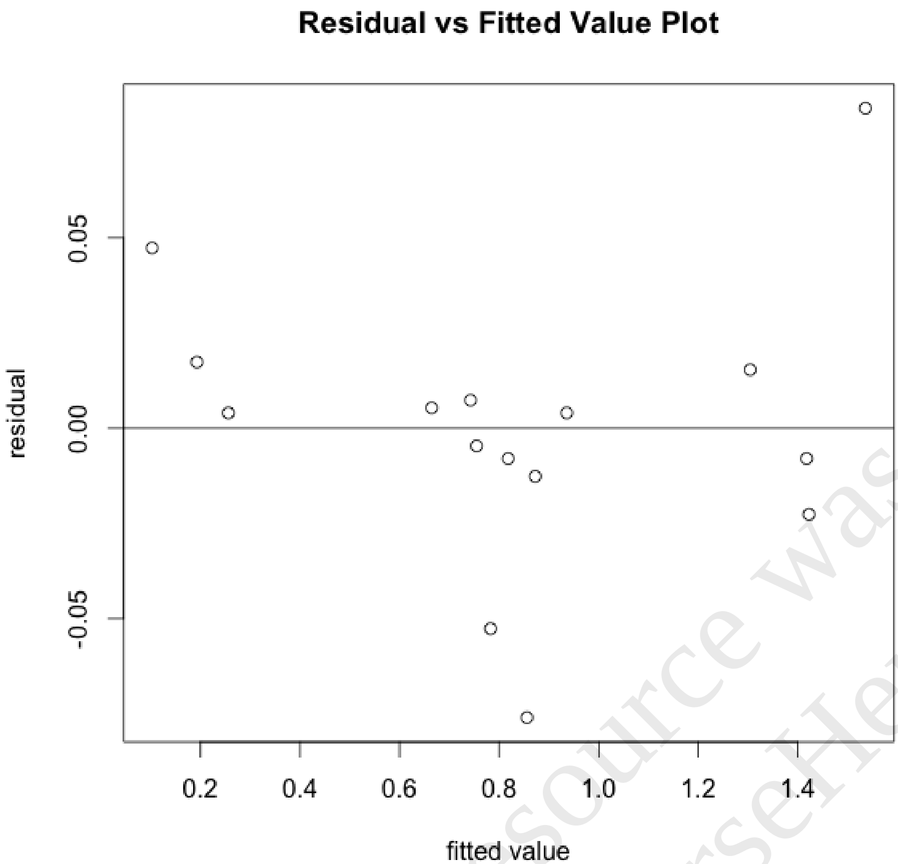
Solution:

a. The total lipid level may vary according to age, so age is a confounding factor.

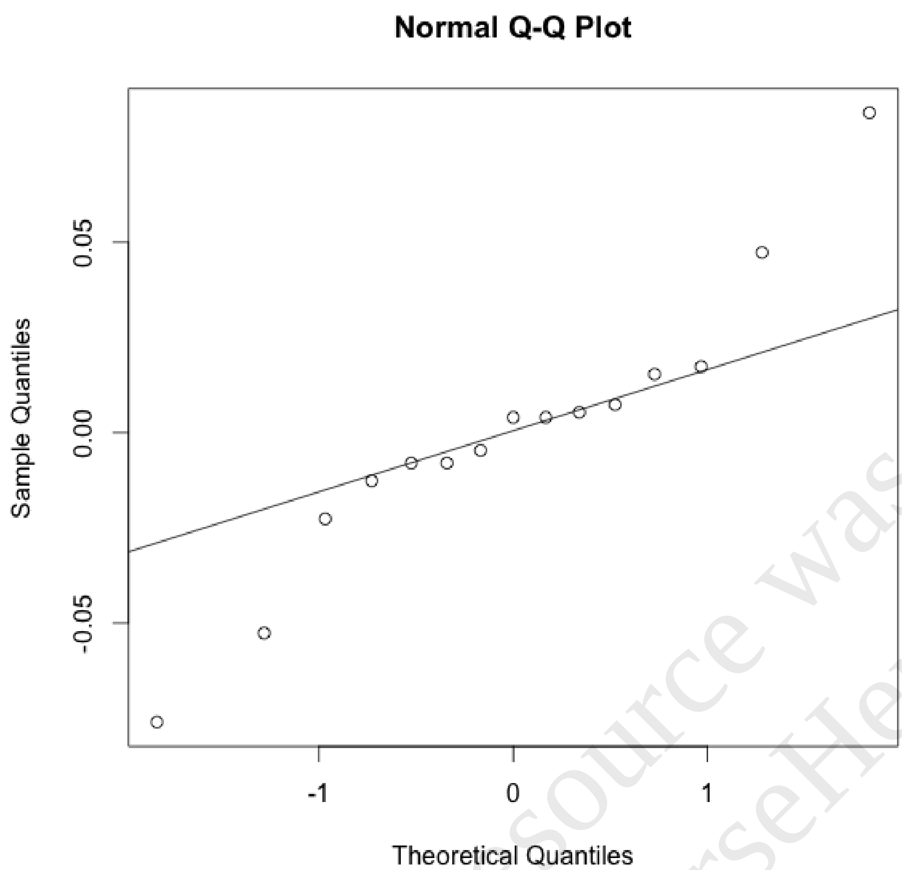
b. The residuals are

$$\mathbf{e_1} = (-0.0527, -0.0127, 0.004, -0.0227, 0.084),$$
$$\mathbf{e_2} = (0.0053, -0.0047, -0.008, 0.0153, -0.008),$$
$$\mathbf{e_3} = (0.0473, 0.0173, 0.004, 0.0073, -0.076),$$

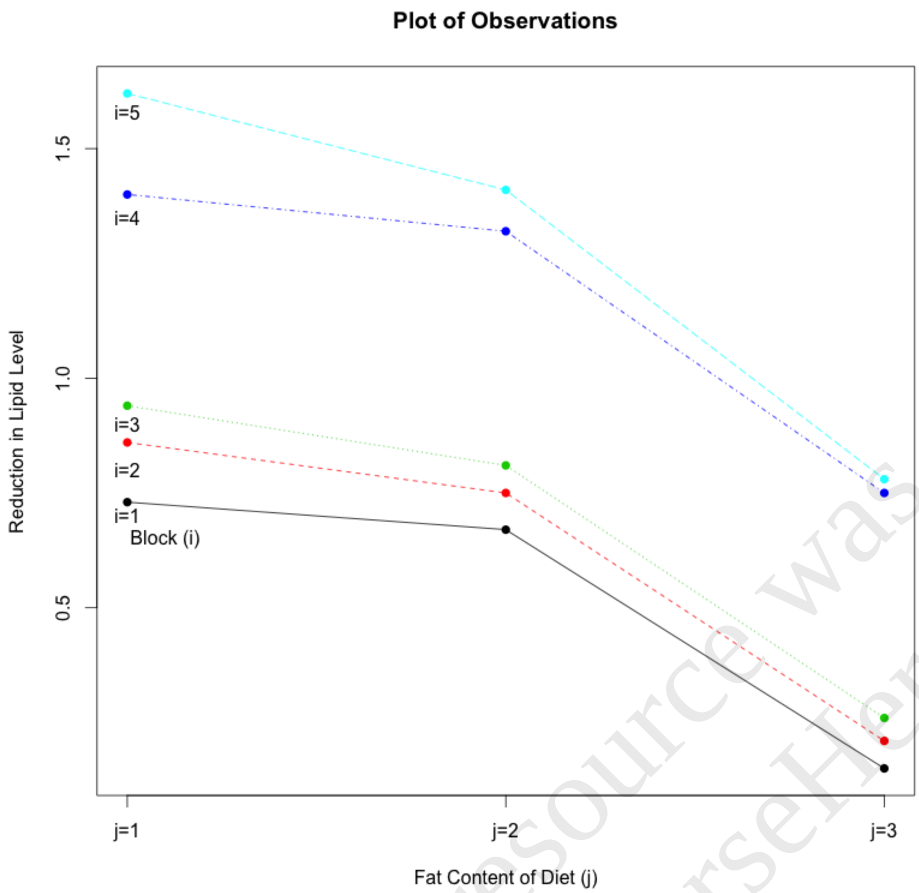
where $\mathbf{e_i} = (e_{1i}, e_{2i}, e_{3i}, e_{4i}, e_{5i})$. The residuals vs fitted values plot is shown below:



The normal Q-Q plot is shown below:



c. The plot of the observations is shown below:



Since the five curves corresponding to five blocks are nearly parallel, there is little interaction between block and treatment. Thus the no-interaction assumption is appropriate here.

d. H_0: $D = 0$, H_a: $D \neq 0$. $SSBL.TR^* = .0093$, $SSRem^* = .038$.
$F^* = (.0093/1)/(.038/(3-3-5)) = 6.50$, $F(.99; 1, 3-3-5) = 12.2$.
If $F^* < 12.2$ conclude H_0, otherwise H_a. Conclude H_0. P-value=.038.

21.8

Refer to **Fat in diets** Problem 21.7. Assume that randomized block model (21.1) is appropriate.

a. Obtain the analysis of variance table.

b. Prepare a bar-interval graph of the estimated treatment means, using 95 percent confidence intervals. Does it appear that the treatment means differ substantially here?

c. Test whether or not the mean reductions in lipid level differ for the three diets; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Estimate $L_1 = \mu_{.1} - \mu_{.2}$ and $L_2 = \mu_{.2} - \mu_{.3}$ using the Bonferroni procedure with a 95 percent family confidence coefficient. State your findings.

e. Test whether or not blocking effects are present; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

f. A standard diet was not used in this experiment as a control. What justification do you think the experimenters might give for not having a control treatment here for comparative purposes?

Solution:

a. The ANOVA table is given below:

Source	SS	df	MS
Blocks	1.41896	4	.35474
Fat content	1.32028	2	.66014
Error	.01932	8	.002415
Total	2.75856	14	

b. $\bar{Y}_{.1} = 1.11$, $\bar{Y}_{.2} = .992$, $\bar{Y}_{.3} = .43$, $s(\bar{Y}_{.i}) = \sqrt{MSBL.TR/5} = .022$, $i = 1, 2, 3$, $t(.975; 8) = 2.306$.
95 percent confidence intervals for treatment means are:

$$\bar{Y}_{.1} \pm t(.975; 8)s(\bar{Y}_{.1}) = 1.11 \pm 2.306(.022) = [1.059, 1.161],$$
$$\bar{Y}_{.2} \pm t(.975; 8)s(\bar{Y}_{.2}) = .992 \pm 2.306(.022) = [.941, 1.043],$$
$$\bar{Y}_{.3} \pm t(.975; 8)s(\bar{Y}_{.3}) = .43 \pm 2.306(.022) = [.379, .481].$$

It appears that the treatment means differ substantially here.

c. H_0: $\tau_1 = \tau_2 = \tau_3 = 0$, H_a: not all τ_j's are 0.
$F^* = MSTR/MSBL.TR = .66014/.002415 = 273.35$, $F(.95; 2, 8) = 4.46$.
If $F^* < 4.46$ conclude H_0, otherwise H_a. Conclude H_a. P-value≈ 0.

d. $\hat{L}_1 = .118$, $\hat{L}_2 = .562$, $s(\hat{L}_i) = \sqrt{2MSBL.TR/5} = .03108$, $i = 1, 2$, $B = t(.9875; 8) = 2.7515$.
95 percent familywise confidence intervals are:

$$\hat{L}_1 \pm B \cdot s(\hat{L}_1) = .118 \pm 2.7515(.03108) = [.032, .204],$$
$$\hat{L}_2 \pm B \cdot s(\hat{L}_2) = .562 \pm 2.7515(.03108) = [.476, .648].$$

e. H_0: all ρ_i equal 0 ($i = 1, \cdots, 5$), H_a: not all ρ_i equal 0.
$F^* = MSBL/MSBL.TR = .35474/.002415 = 146.89$, $F(.95; 4, 8) = 3.84$.
If $F^* < 3.84$ conclude H_0, otherwise H_a. Conclude H_a. P-value≈ 0.

Problem

Refer to the "Decision making" example in Lecture 19.

1. Derive the ANOVA table for the single factor ANOVA model using the ANOVA table for the RCBD.

2. Conduct the F test of no treatment effect under the single factor ANOVA model and compare the result with that under RCBD. What do you observe? Explain your observation briefly.

From lecture 19:

RCBD ANOVA table :

Source of Variation	SS	df	MS
Blocks	$SSBL = 171.3$	$n_b - 1 = 4$	$MSBL = 42.8$
Treatments	$SSTR = 202.8$	$r - 1 = 2$	$MSTR = 101.4$
Error	$SSBL.TR = 23.9$	$(n_b - 1)(r - 1) = 8$	$MSBL.TR = 2.99$
Total	$SSTO = 398.0$	$n_b r - 1 = 14$	

- Test $H_0 : \tau_1 = \tau_2 = \tau_3$.
- F ratio: $F^* = \frac{MSTR}{MSBL.TR} = \frac{101.4}{2.99} = 33.9$.
- P-value: $p = P(F_{2,8} > 33.9) = 0.0001$.
- The result is highly significant. Thus we reject H_0 and conclude that the average confidence ratings of these three methods differ.

Solution:

1. Under the single factor CRD model, use the decomposition of $SSE = SSBL + SSBL.TR \geq SSBL.TR$. Also $SSTR$ is the same quantity in both CRD and RCBD models in this case because $n_b = n_i = 5,$ for $i = 1, \ldots, 3$. Thus,

Source of Variation	SS	df	MS
Treatments	$SSTR = 202.8$	$r - 1 = 2$	$MSTR = 101.4$
Error	$SSE = SSBL + SSBL.TR = 195.2$	$n_T - I = 12$	$MSE = 16.2\bar{6}$
Total	$SSTO = 398.0$	$n_T - 1 = 14$	

- Test $H_0 : \mu_1 = \mu_2 = \mu_3$.
- F ratio: $F^* = \frac{MSTR}{MSE} = \frac{101.4}{16.2\bar{6}} = 6.2336$.
- P-value: $p = P(F_{2,12} > 6.2336) = 0.01391812$.
- The result under CRD model (larger p-value) is not as significant as the RCBD model (smaller p-value). The blocking variation represented in $SSBL$ is sufficiently large that it is worth trading smaller degrees of freedom in the error (i.e. $df(MSBL.TR) < df(MSE)$) such that we gain greater statistical efficiency.