Statistics 512: Solution to Homework#10

1. (KNNL 19.4) In a two-factor study, the treamtent means $\mu_{i,j}$ are as follows:

	Factor B		
Factor A	B_1	B_2	B_3
A_1	34	23	36
A_2	40	29	42

   (a) Obtain the factor A level means.

   Solution: The factor A level means are

$$\begin{aligned}\mu_{1\cdot} &= \frac{34+23+36}{3} = 31\\ \mu_{2\cdot} &= \frac{40+29+42}{3} = 37.\end{aligned}$$

   (b) Obtain the main effects of factor A.

   Solution: The grand mean is

$$\hat{\mu} = \frac{31+37}{2} = 34;$$

   as a result, the main effects of factor A are

$$\begin{aligned}\alpha_1 &= 31 - 34 = -3\\ \alpha_2 &= 37 - 34 = 4.\end{aligned}$$

   (c) Does the fact that $\mu_{12} - \mu_{11} = -11$ while $\mu_{13} - \mu_{12} = 13$ imply the factors A and B interact? Explain.

   Solution: No. Factors A and B interact if $\mu_{12} - \mu_{11} \neq \mu_{22} - \mu_{mu21}$ or $\mu_{13} - \mu_{12} \neq \mu_{23} - \mu_{mu22}$.

   (d) Prepare a treatment means plot and determine whether the two factors interact. What do you find?

   Solution: Figure 1 shows that, as a result of the obvious parallelness between the trends, the interaction is absent.

2. Recall the machine filling data set used in Problem Set 9. In that data set, there were three columns (Y, machine, and carton). Is carton (with levels 1-20) a second treatment factor that one could have included in the model? Explain. (HINT: How many cartons are really used in this experiment, 20 or 120? Would that column be a meaningful variable to include in your analysis?).

   Solution: No. There are actually 120 cartons used in this experiment. The "carton" label is not actually a factor. It is simply a label to identify which of the 20 randomly selected cartons for a particular machine is referred to. However, carton 6 (say) from machine 1 has no relationship to carton 6 from machine 2.

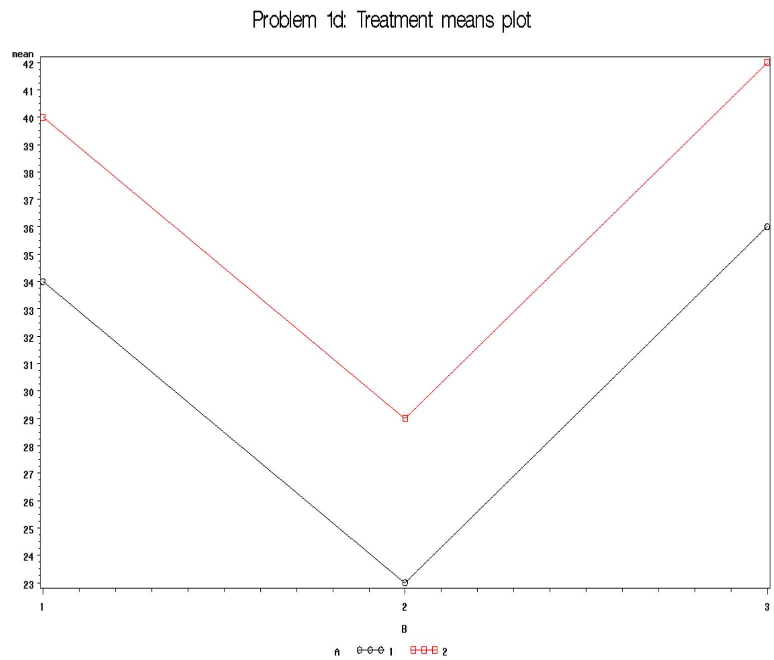

Figure 1: Interaction plot for Problem 1d

For the remaining questions use the disk drive data from problem 19.14 described on page 868 of the text.

3. Give a table of sample sizes, means, and standard deviations for the nine different treatment combinations.

   Solution: In this case, we can refer directly to SAS output.

Level of A	Level of B	N	relief Mean	relief Std Dev
1	1	5	59.8000000	7.85493475
1	2	5	47.8000000	7.46324326
1	3	5	58.4000000	8.53229160
2	1	5	48.4000000	6.76756973
2	2	5	61.2000000	7.32802838
2	3	5	56.2000000	8.04363102
3	1	5	60.2000000	7.32802838
3	2	5	60.8000000	6.30079360
3	3	5	49.6000000	4.50555213

4. Write the factor effects model for this analysis, and estimate the parameters of this model under the zero-sum constraint system (i.e. the one described on page 832 of the text). Also demonstrate that your estimates do in fact satisfy these constraints.

   Solution: The factor effects model is

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j} + \epsilon_{i,j,k}$$

The grand mean $\hat{\mu}$ is 55.8222 . The main effects are

$$\begin{aligned}
\hat{\alpha}_1 = &\quad 55.3333 - 55.8222 = & -0.4889 \\
\hat{\alpha}_2 = &\quad 55.2667 - 55.8222 = & -0.5556 \\
\hat{\alpha}_3 = &\quad 55.8667 - 55.8222 = & +1.0444
\end{aligned}$$

and

$$\begin{aligned}
\hat{\beta}_1 = &\quad 56.1333 - 55.8222 = & +0.3111 \\
\hat{\beta}_2 = &\quad 56.6000 - 55.8222 = & +0.7778 \\
\hat{\beta}_3 = &\quad 54.7333 - 55.8222 = & -1.0889.
\end{aligned}$$

The interaction terms are

$$\begin{aligned}
\left(\hat{\alpha\beta}\right)_{1,1} &= 59.8 - (55.8222 - 0.4889 + 0.3111) = & 4.1556 \\
\left(\hat{\alpha\beta}\right)_{1,2} &= 47.8 - (55.8222 - 0.4889 + 0.7778) = & -8.3111 \\
\left(\hat{\alpha\beta}\right)_{1,3} &= 58.4 - (55.8222 - 0.4889 - 1.0889) = & 4.1556 \\
\left(\hat{\alpha\beta}\right)_{2,1} &= 48.4 - (55.8222 - 0.5556 + 0.3111) = & -7.1778 \\
\left(\hat{\alpha\beta}\right)_{2,2} &= 61.2 - (55.8222 - 0.5556 + 0.7778) = & 5.1556 \\
\left(\hat{\alpha\beta}\right)_{2,3} &= 56.2 - (55.8222 - 0.5556 - 1.0889) = & 2.0222 \\
\left(\hat{\alpha\beta}\right)_{3,1} &= 60.2 - (55.8222 + 1.0444 + 0.3111) = & 3.0222 \\
\left(\hat{\alpha\beta}\right)_{3,2} &= 60.8 - (55.8222 + 1.0444 + 0.7778) = & 3.1556 \\
\left(\hat{\alpha\beta}\right)_{3,3} &= 49.6 - (55.8222 + 1.0444 - 1.0889) = & -6.1778.
\end{aligned}$$

These same values can also be obtained from SAS:

Obs	A	B	AB	muhat	alpha	beta	alphabeta
1	1	1	11	55.8222	-0.48889	0.31111	4.15556
6	1	2	12	55.8222	-0.48889	0.77778	-8.31111
11	1	3	13	55.8222	-0.48889	-1.08889	4.15556
16	2	1	21	55.8222	-0.55556	0.31111	-7.17778
21	2	2	22	55.8222	-0.55556	0.77778	5.15556
26	2	3	23	55.8222	-0.55556	-1.08889	2.02222
31	3	1	31	55.8222	1.04444	0.31111	3.02222
36	3	2	32	55.8222	1.04444	0.77778	3.15556
41	3	3	33	55.8222	1.04444	-1.08889	-6.17778

To verify constraints, notice that for α,

$$\sum \hat{\alpha}_i = -0.4889 - 0.5556 + 1.0444 = 0;$$

for β,

$$\sum \hat{\beta}_j = 0.3111 + 0.7778 - 1.0889 = 0;$$

and for $\alpha\beta$,

$$\begin{aligned}
\sum_j \hat{\alpha\beta}_{1,j} &= \quad 4.1556 - 8.3111 + 4.1556 = \quad 0 \\
\sum_j \hat{\alpha\beta}_{2,j} &= \quad -7.1778 + 5.1556 + 2.0222 = \quad 0 \\
\sum_j \hat{\alpha\beta}_{3,j} &= \quad 3.0222 + 3.1556 - 6.1778 = \quad 0 \\
\sum_i \hat{\alpha\beta}_{i,1} &= \quad 4.1556 - 7.1778 + 3.0222 = \quad 0 \\
\sum_i \hat{\alpha\beta}_{i,2} &= \quad -8.311 + 5.1556 + 3.1556 = \quad 0 \\
\sum_i \hat{\alpha\beta}_{i,3} &= \quad 4.1556 + 2.0222 - 6.1778 = \quad 0.
\end{aligned}$$

5. Perform the two-way analysis of variance for this data set. State the null and alternative hypotheses for main and interaction effects in terms of the factor effects model parameters. For each test, give the test statistic with degrees of freedom and p-value, and your conclusion for each null hypothesis.

Solution: The null and alternative hypotheses for the main effect A are

$$\begin{aligned}
\mathrm{H}_{0A}: &\quad \alpha_1 = \alpha_2 = \alpha_3 = 0 \\
\mathrm{H}_{1A}: &\quad \text{at least one } \alpha_i \neq 0
\end{aligned}$$

The F-statistic in this case is 0.24 with $(2, 36)$ degrees of freedom. Since the p-value is large (0.7908), we fail to reject H_{0A}.

The null and alternative hypotheses for the main effect B are

$$\begin{aligned}
\mathrm{H}_{0B}: &\quad \beta_1 = \beta_2 = \beta_3 = 0 \\
\mathrm{H}_{1B}: &\quad \text{at least one } \beta_i \neq 0
\end{aligned}$$

The F-statistic in this case is 0.27 with $(2, 36)$ degrees of freedom. Since the p-value is large (0.7633), we fail to reject H_{0B}.

The null and alternative hypotheses for the interatction effect AB are

$$\begin{aligned}
\mathrm{H}_{0AB}: &\quad \alpha\beta_{1,1} = \alpha\beta_{1,2} = \alpha\beta_{1,3} = \alpha\beta_{2,1} = \alpha\beta_{2,2} = \alpha\beta_{2,3} = \alpha\beta_{3,1} = \alpha\beta_{3,2} = \alpha\beta_{3,3} = 0 \\
\mathrm{H}_{1AB}: &\quad \text{at least one } \alpha\beta_{i,j} \neq 0
\end{aligned}$$

The F-statistic in this case is 5.84 with $(4, 36)$ degrees of freedom. Since the p-value is 0.0010, we reject H_{0AB}.

```
                                  The GLM Procedure
Dependent Variable: service
                                          Sum of
Source                       DF          Squares     Mean Square    F Value    Pr > F
Model                         8      1268.177778      158.522222       3.05    0.0101
Error                        36      1872.400000       52.011111
Corrected Total              44      3140.577778
```

R-Square	Coeff Var	Root MSE	service Mean
0.403804	12.91936	7.211873	55.82222

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	2	24.577778	12.288889	0.24	0.7908
B	2	28.311111	14.155556	0.27	0.7633
A*B	4	1215.288889	303.822222	5.84	0.0010

6. Make an interaction plot of the cell means with the level of Factor 1 (technician) on the x-axis. Use a different line for each level of Factor 2 (make). Describe the plot in terms of main and interection effects.

   Solution: The lines (Figure 2) interact in a way that makes it difficult to see a pattern. Each pair of factor levels intersects. It is obvious that for each level of Factor A, the means do not vary much relative to the interaction variability.

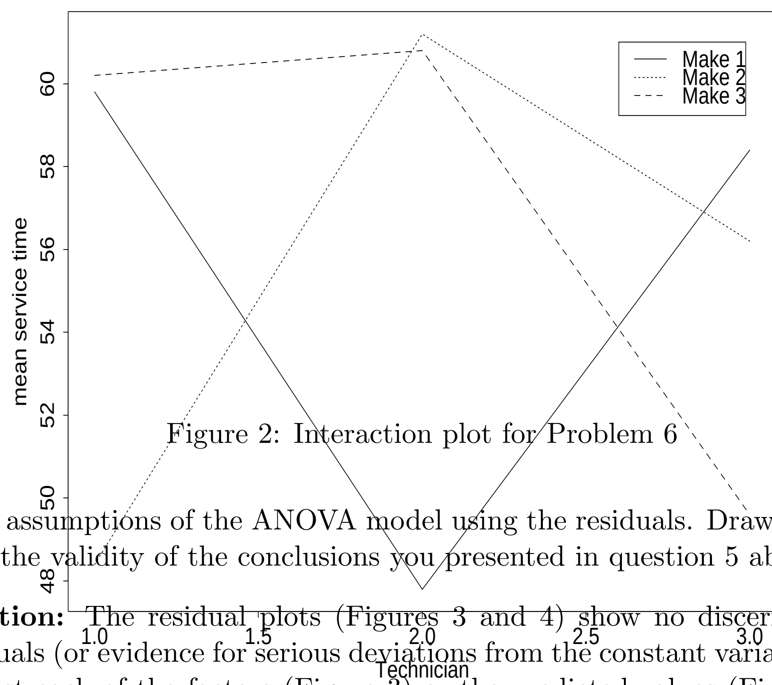

Figure 2: Interaction plot for Problem 6

7. Check the assumptions of the ANOVA model using the residuals. Draw an overall conclusion regarding the validity of the conclusions you presented in question 5 above.

   Solution: The residual plots (Figures 3 and 4) show no discernable pattern of residuals (or evidence for serious deviations from the constant variance assumption) against each of the factors (Figure 3) or the predicted values (Figure 4). The qq-plot and history of residuals (Figure 5) show that the residuals are approximately normally distributed. Our assumptions appear to be met, and so we are satisfied with the conclusions in the above problem.

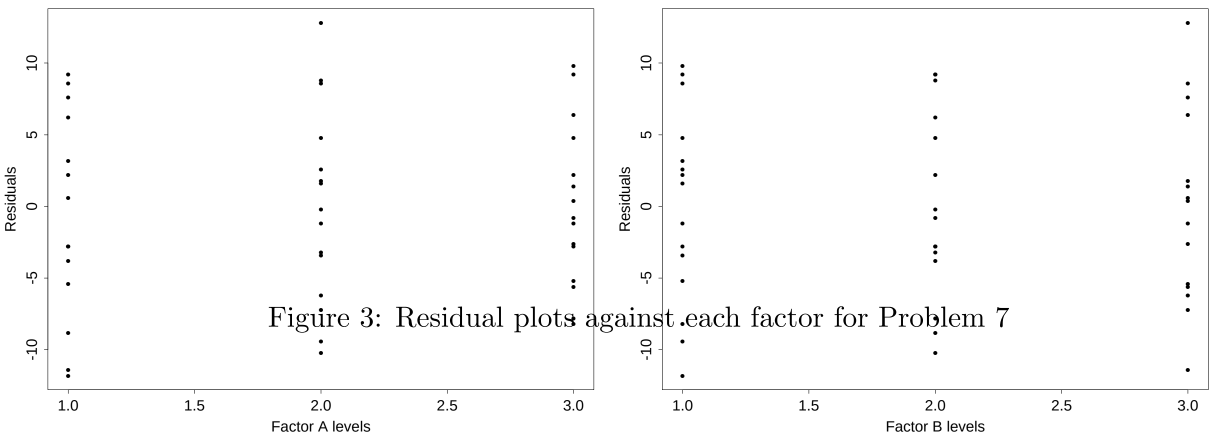

Figure 3: Residual plots against each factor for Problem 7

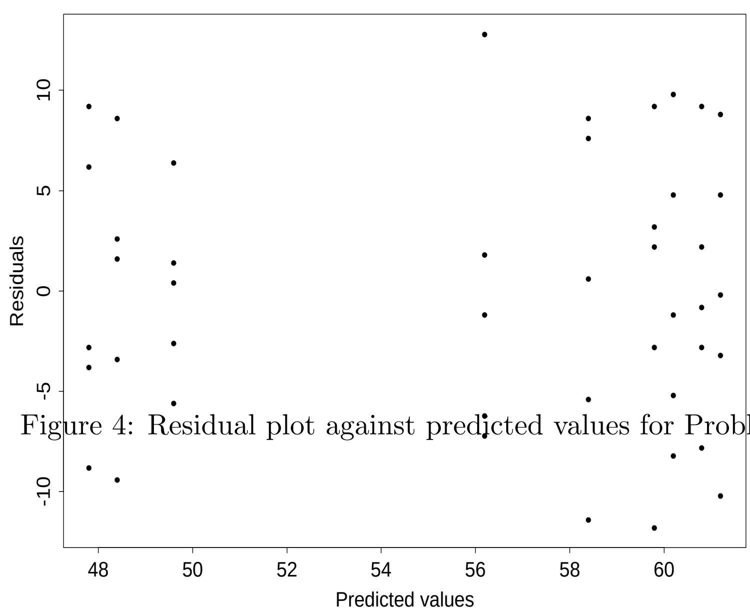

Figure 4: Residual plot against predicted values for Problem 7

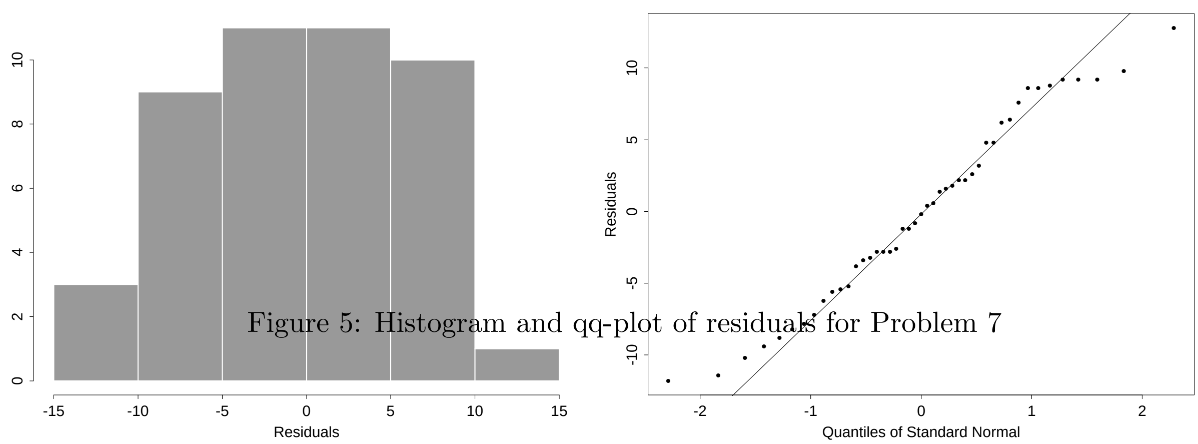

Figure 5: Histogram and qq-plot of residuals for Problem 7