Assignment #9 Solutions – M. Lacey, 04/16/02

Chapter 12, Problem #3:

For a one-way analysis of variance with I=2 treatment groups, show that the F statistic is t^2, where t is the usual t statistic for a two-sample case.

$$\text{For } I = 2, F = \frac{SS_B/(I-1)}{SS_w/[I(J-1)]} = \frac{SS_B}{SS_w/[2(J-1)]} = \frac{J\left[(\bar{Y}_{1.} - \bar{Y}_{..})^2 + (\bar{Y}_{2.} - \bar{Y}_{..})^2\right]}{\sum_{j=1}^{J}\left[(Y_{1j} - \bar{Y}_{1.})^2 + (Y_{2j} - \bar{Y}_{2.})^2\right]/[2(J-1)]}.$$

$$\text{Since } \bar{Y}_{..} = \frac{\sum_{j=1}^{J}(Y_{1j} + Y_{2j})}{2J} = \frac{\bar{Y}_{1.} + \bar{Y}_{2.}}{2}, (\bar{Y}_{1.} - \bar{Y}_{..})^2 = \left(\frac{\bar{Y}_{1.} - \bar{Y}_{2.}}{2}\right)^2 = (\bar{Y}_{2.} - \bar{Y}_{..})^2, \text{ and thus } SS_B = J/2 * (\bar{Y}_{1.} - \bar{Y}_{2.})^2.$$

$$\text{We can also recognize that } SS_w = \sum_{j=1}^{J}\left[(Y_{1j} - \bar{Y}_{1.})^2 + (Y_{2j} - \bar{Y}_{2.})^2\right] = (J-1)(J)(s^2_{\bar{Y}_{1.} - \bar{Y}_{2.}}).$$

$$\text{So } \frac{SS_B}{SS_w/[2(J-1)]} = \frac{J/2 * (\bar{Y}_{1.} - \bar{Y}_{2.})^2}{(J-1)(J)(s^2_{\bar{Y}_{1.} - \bar{Y}_{2.}})/[2(J-1)]} = \frac{(\bar{Y}_{1.} - \bar{Y}_{2.})^2}{s^2_{\bar{Y}_{1.} - \bar{Y}_{2.}}}.$$

To see that this is equivalent to t^2, let $\bar{X} = \bar{Y}_{1.}, \bar{Y} = \bar{Y}_{2.}$.

Chapter 12, Problem #14:

Prove the sum of squares identity for the two-way layout: $SS_{TOT} = SS_A + SS_B + SS_{AB} + SS_E$

The outline of this proof is provided on p. 459, which begins by adding and subtracting terms to write the expression $Y_{ijk} - \bar{Y}_{...} = (Y_{ijk} - \bar{Y}_{ij.}) + (\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{.j.} - \bar{Y}_{...}) + (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})$. Squaring both sides, we get

$$\begin{aligned}
(Y_{ijk} - \bar{Y}_{...})^2 &= \left[(Y_{ijk} - \bar{Y}_{ij.}) + (\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{.j.} - \bar{Y}_{...}) + (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})\right]^2 \\
&= (Y_{ijk} - \bar{Y}_{ij.})^2 + (\bar{Y}_{i..} - \bar{Y}_{...})^2 + (\bar{Y}_{.j.} - \bar{Y}_{...})^2 + (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 + \\
&\quad 2(Y_{ijk} - \bar{Y}_{ij.})\left[(\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{.j.} - \bar{Y}_{...}) + (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})\right] + \\
&\quad 2(\bar{Y}_{i..} - \bar{Y}_{...})\left[(\bar{Y}_{.j.} - \bar{Y}_{...}) + (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})\right] + 2(\bar{Y}_{.j.} - \bar{Y}_{...})(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}) \\
&= (Y_{ijk} - \bar{Y}_{ij.})^2 + (\bar{Y}_{i..} - \bar{Y}_{...})^2 + (\bar{Y}_{.j.} - \bar{Y}_{...})^2 + (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 + 2(Y_{ijk} - \bar{Y}_{ij.})(\bar{Y}_{ij.} - \bar{Y}_{...}) + \\
&\quad 2(\bar{Y}_{i..} - \bar{Y}_{...})(\bar{Y}_{ij.} - \bar{Y}_{i..}) + 2(\bar{Y}_{.j.} - \bar{Y}_{...})(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}).
\end{aligned}$$

Summing over i, j, and k gives

$$\begin{aligned}
SS_{TOT} &= SS_E + SS_A + SS_B + SS_{AB} + \\
&\quad 2\sum_i \sum_j \sum_k \left[(Y_{ijk} - \bar{Y}_{ij.})(\bar{Y}_{ij.} - \bar{Y}_{...}) + (\bar{Y}_{i..} - \bar{Y}_{...})(\bar{Y}_{ij.} - \bar{Y}_{i..}) + (\bar{Y}_{.j.} - \bar{Y}_{...})(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})\right].
\end{aligned}$$

To prove the identity, we must show that the remaining terms cancel out. For the first term,

$$\sum_i \sum_j \sum_k \left[\left(Y_{ijk} - \overline{Y}_{ij.}\right)\left(\overline{Y}_{ij.} - \overline{Y}_{...}\right) \right] = \sum_i \sum_j \left[\left(\overline{Y}_{ij.} - \overline{Y}_{...}\right) \sum_k \left(Y_{ijk} - \overline{Y}_{ij.}\right) \right]$$, and since $\sum_k \left(Y_{ijk} - \overline{Y}_{ij.}\right)$ is the sum of the k deviations Y_{ijk} from the mean $\overline{Y}_{ij.}$, this term is equal to 0.

Similar arguments apply demonstrate that the second and third terms are also equal to 0, and this proves the identity.

Chapter 12, Problem #19:

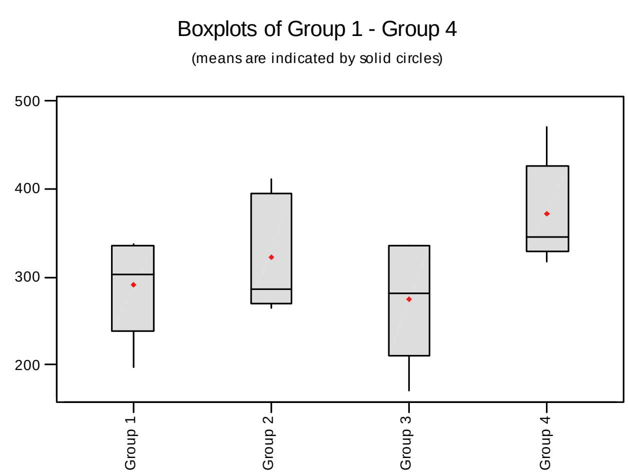
Analysis of worm counts in four control groups of rats.

First, we may review the descriptive statistics and look at the corresponding boxplots for the four groups:

Variable	N	Mean	Median	StDev
Group 1	5	290.4	303.0	57.0
Group 2	5	323.2	286.0	67.0
Group 3	5	274.8	282.0	68.0
Group 4	5	371.2	346.0	60.2

Variable	Minimum	Maximum	Q1	Q3
Group 1	198.0	338.0	238.5	336.0
Group 2	265.0	412.0	270.0	395.0
Group 3	172.0	335.0	211.0	335.0
Group 4	318.0	471.0	329.0	426.0

Side-by-side boxplots (produced in MINITAB)



ANOVA table and F-test:

One-way Analysis of Variance

Analysis of Variance

Source	DF	SS	MS	F	P
Factor	3	27234	9078	2.27	0.119
Error	16	63954	3997		
Total	19	91188			

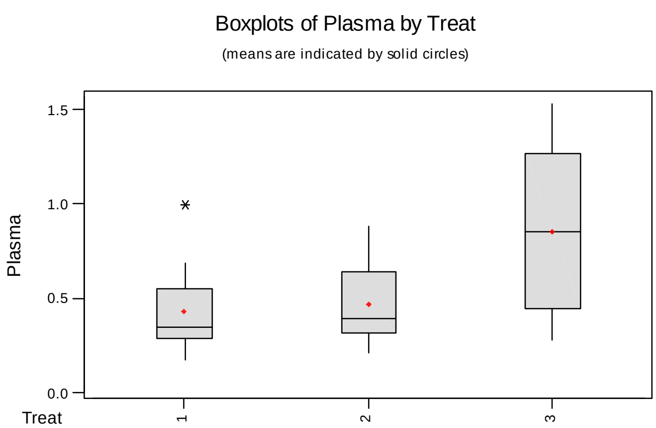
The appropriate distribution for the test statistic is $F_{3,16}$, and from Table 5 the critical value for $\alpha = 0.1$ is 2.46, so we would not reject the null hypothesis that there is no difference among the groups (the exact p-value is 0.12, as shown in the table above).

Chapter 12, Problem #24:

Concentrations of plasma epinephrine for ten dogs under three different types of anesthesia. To determine whether there is a treatment effect, we begin by looking at boxplots and descriptive statistics:

Descriptive statistics for plasma by anesthesia type:

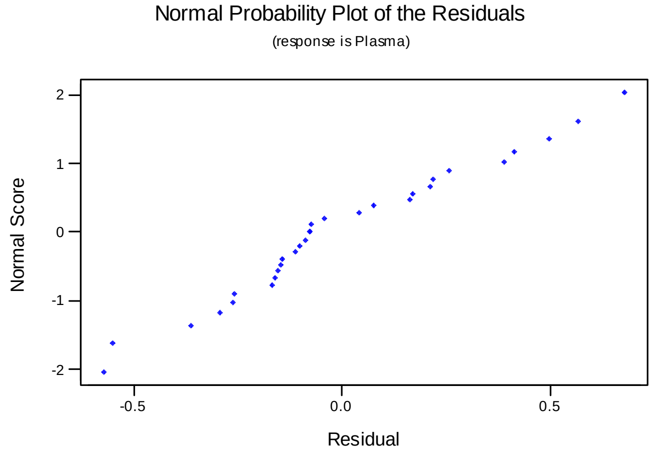
```
Variable    Treat    N      Mean     Median     StDev
Plasma      1       10    0.4340     0.3450    0.2443
            2       10    0.4690     0.3900    0.2053
            3       10    0.853      0.855     0.448

Variable    Treat          Minimum    Maximum        Q1         Q3
Plasma      1              0.1700     1.0000     0.2875     0.5550
            2              0.2100     0.8800     0.3150     0.6425
            3              0.280      1.530      0.443      1.268
```



It appears that there are differences among the three treatment groups. To determine whether the treatment effect is statistically significant, we produce a one-way ANOVA table:

```
Analysis of Variance for Plasma
Source     DF        SS        MS        F        P
Treat       2     1.081     0.540     5.35    0.011
Error      27     2.725     0.101
Total      29     3.806
```

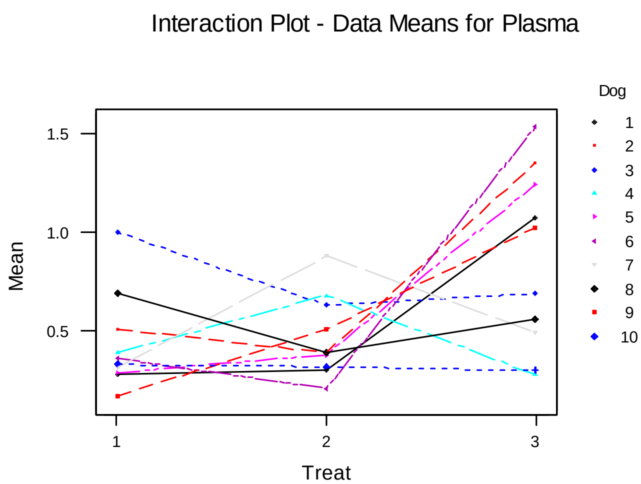
The distribution of the test statistic for anesthesia effect is $F_{2,27}$, and from Table 5 we can check that $F_{.975}(2,27) = 4.24 < 5.35$, so we can reject the null hypothesis at the 0.025 level (as shown in the table above, the exact *p*-value is 0.011). A normal probability plot of the residuals (below) indicates that the model provides a reasonably good fit to the data.



Note: If we wish to consider possible "dog" effects, a two-way ANOVA table gives the following results:

```
Analysis of Variance for Plasma

Source      DF        SS         MS       F       P
Dog          9    0.5169     0.0574    0.47   0.877
Treat        2    1.0808     0.5404    4.41   0.028
Error       18    2.2078     0.1227
Total       29    3.8055
```

While the "dog" effect is negligible, the treatment effect is no longer significant at the 0.025 level, which indicates some degree of interaction between the dogs and the treatment type in plasma levels. As the graph shown below illustrates, the treatment effect is not consistent across the group of 10 dogs.



Chapter 12, Problem #27:

Survival times for poisoned animals.

Part (a): Two-way ANOVA for survival times.

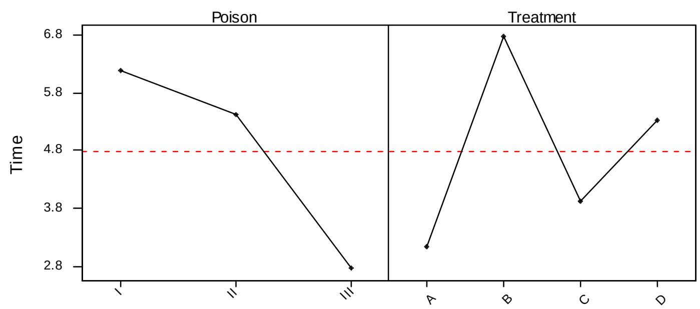
Table of mean survival times by poison type:

Poison	N	Time
I	16	6.1750
II	16	5.4313
III	16	2.7625

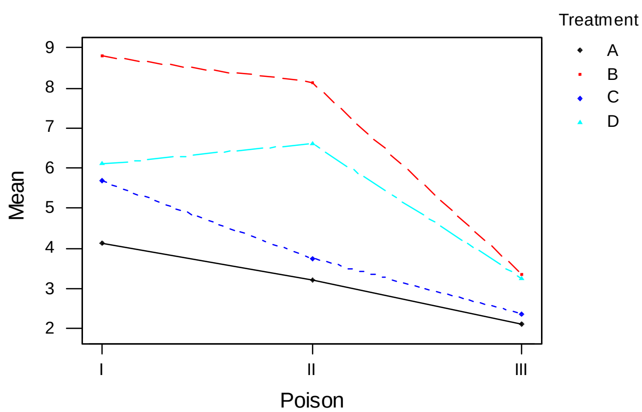
Table of mean survival times by treatment type:

Treatment	N	Time
A	12	3.1417
B	12	6.7667
C	12	3.9250
D	12	5.3250

Main Effects Plot - Data Means for Time



Interaction Plot - Data Means for Time



```
Analysis of Variance for Time

Source              DF        SS        MS       F      P
Poison               2   103.043    51.521   23.57  0.000
Treatment            3    91.904    30.635   14.01  0.000
Poison*Treatment     6    24.745     4.124    1.89  0.110
Error               36    78.692     2.186
Total               47   298.385
```

From the ANOVA table and the plots shown above, there appear to be significant effects due to both Poison type and Treatment. The interaction effect is marginally significant, although the null hypothesis of no interaction cannot be rejected at the 0.1 level.

Part (b): Two-way ANOVA for the reciprocals of the survival times

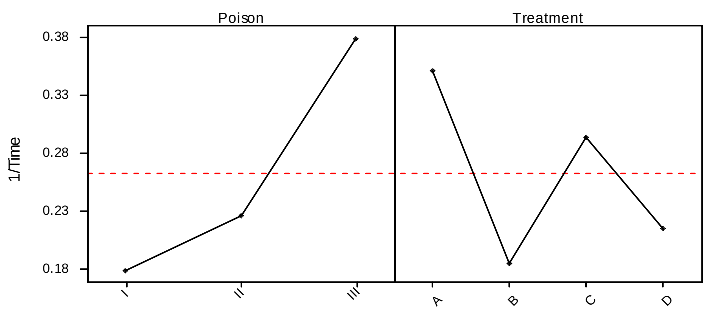
Table of mean reciprocal survival times by poison type:

Poison	N	1/Time
I	16	0.18007
II	16	0.22706
III	16	0.37971

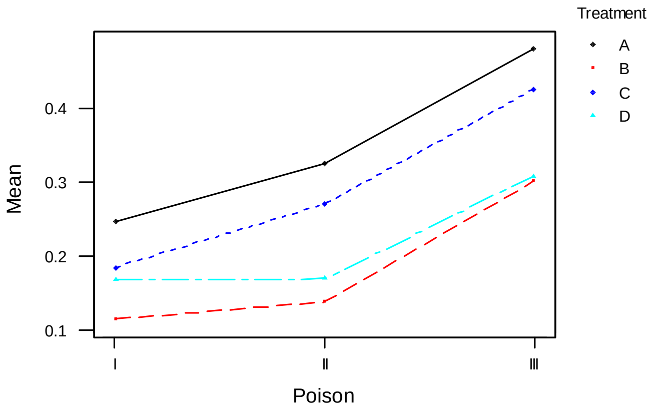
Table of mean reciprocal survival times by treatment type:

Treatment	N	1/Time
A	12	0.35193
B	12	0.18619
C	12	0.29472
D	12	0.21626

Main Effects Plot - Data Means for 1/Time



Interaction Plot - Data Means for 1/Time

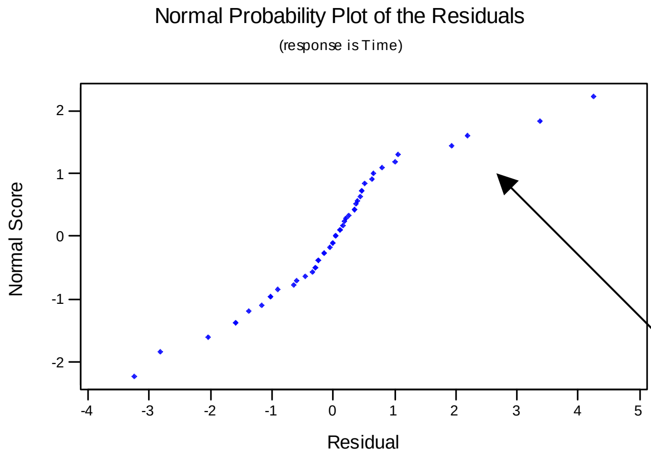


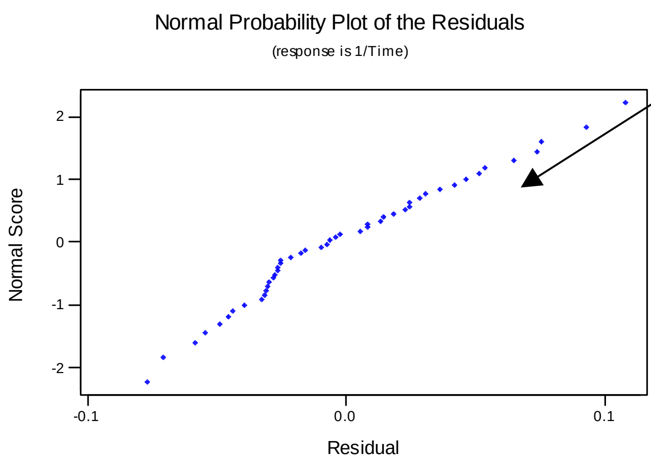
Analysis of Variance for 1/Time

Source	DF	SS	MS	F	P
Poison	2	0.348633	0.174316	72.84	0.000
Treatment	3	0.203962	0.067987	28.41	0.000
Poison*Treatment	6	0.015666	0.002611	1.09	0.386
Error	36	0.086151	0.002393		
Total	47	0.654412			

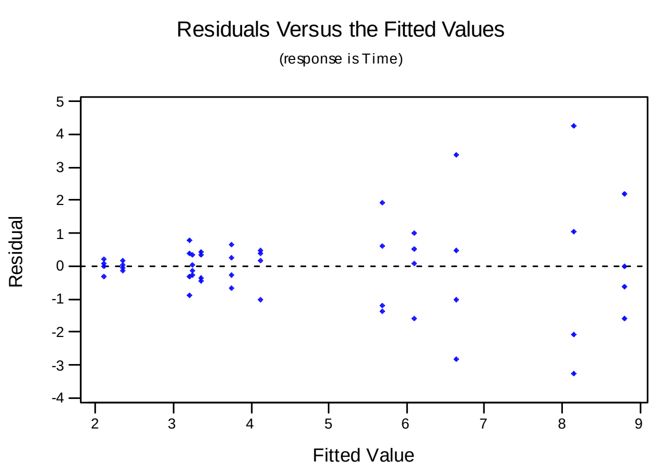
From the ANOVA table, both Poison and Treatment effects are highly significant for the reciprocal survival times, while the interaction effect is no longer significant.

To compare the results from (a) with the results from (b), we can look at plots of the residuals from each model:





A normal probability plot of the residuals from the first model (time) indicates a skewed distribution, while the residuals for the second model (1/time) seem to be more reasonably normal.



A plot of the residuals against the fitted values for part (a) shows that the model provides a poor fit to longer survival times. The reciprocal transformation corrects for this problem, and thus the two-way ANOVA model provides a better fit for model (b).