Hospital Expenses - Solution

(a)

We see that hospital A comes out favourable with a mean expense of 395, considerably less than the other two, which are not very different. A one-factor analysis of variance with hospital as the factor and year as repeats provides estimates with confidence limits and formal tests based on standard assumptions (of normality and equal variances):

One-way ANOVA: Expense versus Hospital

```
Source    DF     SS    MS      F      P
Hospital   2  18374  9187  23.03  0.002
Error      6   2394   399
Total      8  20768

S = 19.97   R-Sq = 88.47%   R-Sq(adj) = 84.63%


                         Individual 95% CIs For Mean Based on
                         Pooled StDev
Level  N    Mean  StDev  --------+---------+---------+---------+-
A      3  395.33  14.19  (------*------)
B      3  481.33  25.48                      (------*------)
C      3  498.67  18.61                           (------*------)
                         --------+---------+---------+---------+-
                               400       440       480       520

Pooled StDev = 19.97
```

The hypothesis of equal expected expense for the hospitals is rejected at 1% level (since P=0.002 < 0.01). We see from the individual confidence intervals that the one of hospital A is far from overlapping B and C, which in turn overlap to a large extent, supporting the conclusion apparent from the given table.

<u>Remarks.</u>

1. The assumption of normality and equal variance are always hard to check with so few data. Hospital B shows a wider variation, measured by the standard deviation (StDev), but it is not likely that this will upset the conclusion. As an alternative we may perform the non-parametric Kruskal-Wallis test.
2. The 95% individual confidence intervals do not account for multiple comparisons. Confidence intervals of expected expense differences between of the hospitals may be obtained with joint guarantee of 95% of not stating any difference when there are none. This is typically available as a subcommand choice, here as three pairwise comparisons. With our data the conclusion will not differ and the output is omitted here.

(b)

A two-factor analysis of variance with hospital as the first factor and year as the second factor provides estimates with confidence limits and formal tests based on standard assumptions (of normality and equal variances). With no repeats we have to assume additive model, i.e. no interaction between the factors hospital and year.

Two-way ANOVA: Expense versus Hospital; Year

```
Source    DF       SS       MS      F      P
Hospital   2  18374.2  9187.11  28.17  0.004
Year       2   1089.6   544.78   1.67  0.297
Error      4   1304.4   326.11
Total      8  20768.2

S = 18.06   R-Sq = 93.72%   R-Sq(adj) = 87.44%


                   Individual 95% CIs For Mean Based on
                   Pooled StDev
Hospital     Mean  -------+---------+---------+---------+--
A         395.333  (-----*-----)
B         481.333                   (-----*-----)
C         498.667                       (-----*-----)
                   -------+---------+---------+---------+--
                        400       450       500       550


               Individual 95% CIs For Mean Based on
               Pooled StDev
Year     Mean   -+---------+---------+---------+--------
1     450.333   (----------*-----------)
2     474.000            (-----------*----------)
3     451.000   (----------*-----------)
                -+---------+---------+---------+--------
               425       450       475       500
```

We see that the hypothesis of equal hospital expenses is rejected at the 1% level, and that the conclusion of hospital A having lower expected expenses remains the same as in (a). The hypothesis of equal level of expenses over the three years (regardless of hospital) is not rejected (P=0.297). The second year has somewhat higher mean than the other two, but may just as well be due to chance. The assumption of expense level not dependent on year taken in (a) seems so far to be justified.

(c)

The mean expense over the three years for each combination of hospital and unit were:

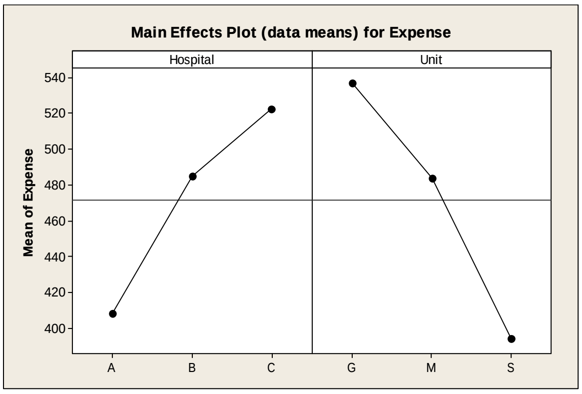
Tabulated statistics: Hospital; Unit

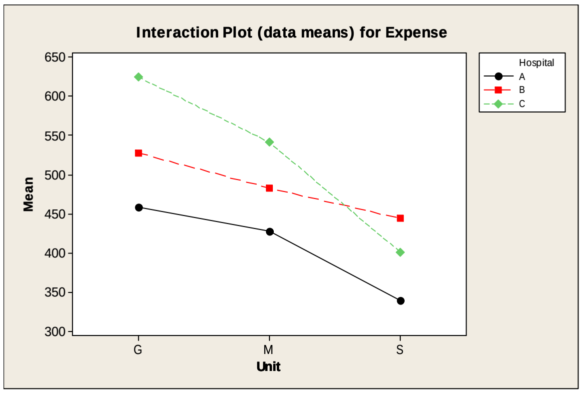
Rows: Hospital   Columns: Unit

	G	M	S	All
A	458.0	427.3	339.7	408.3
B	527.7	483.0	443.7	484.8
C	625.0	541.3	400.3	522.2
All	536.9	483.9	394.6	471.8

Cell Contents: Expense : Mean

This may be illustrated by graphs as follows:





We see from the main effect plot that the average expense over the three years at units G is on average higher than that of units M, which in turn is higher than on units S. From the second table in the case description we see that hospital A has lower activity (20%) at G and higher activity (45%) at S than the other two hospitals. This may be (partly) the reason why hospital A came out so favourable in our analysis in (a) and (b). From the interaction plot we see that hospital B and C are “out of line” with respect to unit S, taken as B having a higher expense and C lower expense at unit S than expected.

(d)

A two-factor analysis of variance with hospital as the first factor and units as the second factor provides estimates with confidence limits and formal tests based on standard assumptions (of normality and equal variances). Taking years as repeats, partly justified in (b), we may allow the possibility of interaction between the factors hospital and unit.

Two-way ANOVA: Expense versus Hospital; Unit

```
Source       DF      SS       MS      F      P
Hospital      2   60650  30324.8  35.52  0.000
Unit          2   93145  46572.3  54.56  0.000
Interaction   4   17438   4359.6   5.11  0.006
Error        18   15366    853.7
Total        26  186599

S = 29.22   R-Sq = 91.77%   R-Sq(adj) = 88.11%


                   Individual 95% CIs For Mean Based on
                   Pooled StDev
Hospital     Mean  ---+---------+---------+---------+------
A         408.333  (----*----)
B         484.778                    (----*----)
C         522.222                             (-----*----)
                   ---+---------+---------+---------+------
                    400       440       480       520


               Individual 95% CIs For Mean Based on
               Pooled StDev
Unit     Mean  -----+---------+---------+---------+----
G     536.889                               (---*---)
M     483.889                     (---*---)
S     394.556  (---*---)
               -----+---------+---------+---------+----
                  400       450       500       550
```

From the P-values of the ANOVA table we see that:

i. The hypothesis of equal expected expenses at the hospitals (adjusted for units) is rejected.

ii. The hypothesis of equal expected expenses at the units (adjusted for hospitals) is rejected.
iii. The hypothesis of no interaction effect between hospital and unit is rejected.

The latter means that there are differences between the units depending on which hospital we are at, or equivalently differences between the hospitals depending on which unit we are at. The character of this interaction is already made clear from the interaction plot above.

Finally, it may be worthwhile to perform a three-way analysis of variance, taking hospital, unit and year as factors. With only one observation for the 27 factor combinations we have to stay with an additive model with no interactions.

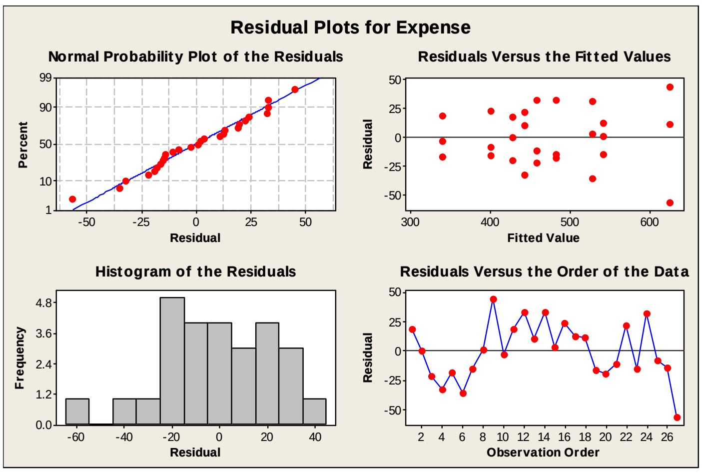
ANOVA: Expense versus Hospital; Unit; Year

```
Factor     Type   Levels  Values
Hospital  fixed        3  A; B; C
Unit      fixed        3  G; M; S
Year      fixed        3  1; 2; 3

Analysis of Variance for Expense

Source    DF      SS     MS      F      P
Hospital   2   60650  30325  20.70  0.000
Unit       2   93145  46572  31.79  0.000
Year       2    3506   1753   1.20  0.323
Error     20   29298   1465
Total     26  186599


S = 38.2743   R-Sq = 84.30%   R-Sq(adj) = 79.59%
```

We see that the year did not come out as a significant level effect after taking hospital and unit into account, thus supporting the decision to leave out year as factor in the explanatory model.

(f)

Going back to the two-way analysis with interaction it makes sense with 27 observations to look into the analysis of variance assumptions of normality and constant variance of the error terms. Here is a “four in one” graph based on the computes residuals.



We see that the histogram of the residuals does not look normal, but the deviation from normality is not serious, which is supported by a normal probability plot with points lining up fairly linearly. The plot of residuals versus fitted values indicates some increase in variance with the response level, but not to a degree that invalidates any of our analysis. The plot of the residuals versus order of the data shows more frequent positive residuals in the middle region, thus indicating a higher variance in year 2. This is not likely to affect our analysis to any degree.

Closing remarks. The expenses given here are not likely to be a useful measure of cost-effectiveness. It may just as well reflect the character of the activity or the way it is reported within each medical speciality (this may be an issue to look into). The apparent interaction may also be an interesting observation to look closer into for the supervisory body. It may also be of interest for the authorities to see whether the increased variances the second year have some explanation that may be used for routine improvements.