8. Analysis of Intervention Studies – II
Two-way Analysis of Variance

In the data file about depression the subjects were divided by one factor viz. their mental state (healthy; non-melancholic depressed; melancholic depressed). Many clinical problems are investigated using such designs. As we saw in Chapter 7 designs of such types are referred to as completely randomised designs. But there are times when we may wish to divide the subjects according two factors. Subjects vary in their response by age, gender, social class, habits, and so on. Females may respond differently from males; younger subjects may have a different response to a medication compared to the elderly; and those who are in sedentary work respond differently to exercise challenge compared to those who are active. The precision of a study would be better if subjects were grouped according to such criteria. We can then expect homogeneity of response by groups, and less variation within groups. This is so because, generally speaking, subjects who are similar would show responses which are closer than subjects who are dissimilar. The technical term for such groups is **block**, and the study design is called *randomised block design.*

In order to avoid bias, blocks are formed first and then random allocation of intervention is carried out separately for each block. Thus, each subject within a block receives one treatment only, which is chosen at random from a list of treatments to be compared. Treatment comparisons should be restricted to subjects from the same block. In many clinical situations there is some background information available about the subjects which enables the investigator to predict which subjects are likely to respond in a similar manner. Such information is used for grouping of subjects.

In a randomised block design we have a two factors design, one factor is related to the basis on which blocks are formed and the other related to interventions. Each can have several levels. There can be more than two blocks, and likewise, more than two treatments. It all depends on the research question being asked.

The advantage of Two-way ANOVA is that the investigator can test three hypotheses simultaneously using the same data. For example, in the case of the "Depression" data set if the subjects would have been grouped by gender, the three hypotheses could have been:

1. Cortisol levels do not depend on the mental state of the individual (i.e. the same hypothesis as before)
2. Cortisol levels do not depend on the gender of the individual
3. The effect of the mental state on cortisol levels does not depend on the gender of the subject. (This third hypothesis is called the interaction effect, since one is looking for any possible interaction between the two explanatory variables).

The randomised block design is a more robust design than simply comparing the different group means in a random sample. By randomly selecting subjects the investigator makes sure that individual differences between subjects are randomly distributed across the groups. But within each group there can be variations, partly due to biological differences between individuals but also because of factors which have not been considered in the study design. Two-way ANOVA allows us to take into account simultaneously the effects of two factors on a variable of interest. As in all statistical analysis, because we are making use of more information we are producing more sensitive results.

Blocking is not to be thought of an extra complication that one has to incorporate into the design. Instead, it is advantageous to introduce a blocking structure, if possible, by arranging subjects in the most homogeneous way. We are thereby eliminating some of what would otherwise be extraneous variability. Even if it turns out that there is little variability between blocks rarely will anything be lost by blocking.

So, to recapitulate a randomised block design has the following characteristics:

1. Subjects are randomly selected.
2. Those subjects that are likely to have similar responses (homogeneity) are put together to form a block. Blocking is carried out before the intervention. It is part of the design of a trial.
3. Intervention is assigned at random to members of each block, such that each subject receives one treatment.
4. Comparisons of treatment outcomes are made within each block.

Since groups are formed along more than one dimension (e.g. gender and intervention) commonly referred to as factors, the differences among means are attributable to more than one source. Thus Two-way ANOVA works in the same manner as One-way ANOVA except that there is now an additional second explanatory variable. This is illustrated by the following example:

Example of a Two-way Analysis of Variance problem

In a medical school a new method of teaching in which professional actors played the roles of patients was introduced.
The test scores of male and female students who were taught by either the conventional method or by a new form of training using role-play are shown in the table below. The two methods are analysed using Two-way analysis of variance. An interaction effect between conventional training and the role-play method is sought.

Score	Sex	Tchng.Mthd
64.0000	0	0
75.6000	0	0
60.6000	0	0
69.3000	0	0
63.7000	0	0
53.3000	0	0
55.7000	0	0
70.4000	0	0
37.7000	0	1
53.5000	0	1
33.9000	0	1
78.6000	0	1
46.0000	0	1
38.7000	0	1
65.8000	0	1
68.4000	0	1
41.9000	1	0
55.0000	1	0
32.1000	1	0
50.1000	1	0
52.1000	1	0
56.6000	1	0
51.8000	1	0
51.7000	1	0
25.6000	1	1
23.1000	1	1
32.8000	1	1
43.5000	1	1
12.2000	1	1
35.4000	1	1
28.0000	1	1
41.9000	1	1

```
Sex 0 = Male             Tchng. Mthd. 0 = Conventional
    1 = Female                        1 = Role Play
```

Variable	sex	N	Mean	StDev
score	0	16	58.45	13.70
	1	16	39.61	13.13

Variable	Tchmthd	N	Mean	StDev
score	0	16	56.49	10.87
	1	16	41.57	17.67

The results of Two-way Analysis of variance are displayed below. In doing a Two-way ANOVA there is a choice between fitting an additive model, or a model which would show the interaction between the two factors. The latter is the default option. We examine the outputs obtained by fitting both the model types in sequence.

[In MINITAB → Stat → ANOVA → TWO-way │ In Response box "Score" In Row Factor box "Sex" In Column factor box "Tchmthd"]

(i). Two-way ANOVA using the additive model

Analysis of Variance for score

Source	DF	SS	MS	F	P
sex	1	2839	2839	22.75	0.000
Tchmthd	1	1782	1782	14.28	0.001
Error	29	3619	125		
Total	31	8240			

```
                       Individual 95% CI
sex           Mean   --------+---------+---------+---------+---
0             58.5                           (------*------)
1             39.6   (-------*------)
                     --------+---------+---------+---------+---
                          40.0      48.0      56.0      64.0
```

```
                         Individual 95% CI
Tchmthd          Mean   ---------+---------+---------+---------+--
0                56.5                         (-------*-------)
1                41.6   (-------*--------)
                        ---------+---------+---------+---------+--
                              42.0      49.0      56.0      63.0
```

(ii). Two-way ANOVA with interaction

Analysis of Variance for score

Source	DF	SS	MS	F	P
sex	1	2839	2839	22.64	0.000
Tchmthd	1	1782	1782	14.21	0.001
Interaction	1	108	108	0.86	0.361
Error	28	3511	125		
Total	31	8240			

```
                         Individual 95% CI
sex              Mean   --------+---------+---------+---------+---
0                58.5                           (------*------)
1                39.6   (-------*------)
                        --------+---------+---------+---------+---
                             40.0      48.0      56.0      64.0

                         Individual 95% CI
Tchmthd          Mean   ---------+---------+---------+---------+--
0                56.5                         (-------*-------)
1                41.6   (-------*--------)
                        ---------+---------+---------+---------+--
                              42.0      49.0      56.0      63.0
```

Interpretation of the results

In the same way as we did with One-way ANOVA we can express the total variation in the data as the sum of the variance from several sources as shown below:

Total variance = Variance due to gender + variance due to the teaching method
+ random variance within groups.

Blocking in ANOVA is the extension of the paired '*t*' test. By blocking and thereby making the groups homogeneous we have reduced the unexplained variability so that the treatment effects are made more striking.

What Two-way ANOVA has done is to separate out the blocking effect on the sum of squares. If the variance due to gender is much greater than the random variance within groups we should get statistically significant evidence of a difference between male and female subjects. Similarly, if the variance due to teaching method is much greater than that due to random error, we will have statistically significant evidence of a difference due to the teaching method. The model with interaction additionally looks for variance due to interaction between the factors. In this case, it is gender and teaching method. One sex may be more sensitive to role-play.

We note that the confidence intervals for sex do not overlap; in other words there is a significant difference between the two mean scores for gender. Similarly, there is also a significant difference between the mean scores for the two types of teaching methods. So this analysis, so far, helps us to test two hypotheses viz.

1. There is no difference in scores between men and women. We reject the hypothesis. There is a difference.
2. There is no difference in scores between students taught by the conventional or the role playing method. We reject this hypothesis also. There is a difference.

But is there an effect between teaching method and gender? In other words, is one type of method more effective with one gender than the other? To answer this question let us look at the results of the Two-way ANOVA with interaction. In the first table headed analysis variance for score there occurs an additional term 'interaction'. However, the F statistic of 0.86 is not significant (*P* = 0.361). This means that there is no significant evidence of interaction between gender and teaching method.

Regression Approach to Two-way ANOVA

Recall that in the data set on "Score" for gender the coding is 0 = Male; 1 = Female.
And for teaching method it is 0 = conventional; 1 = Role-play.
We create two dummy (indicator) variables X_1 and X_2 such that
$X_1 = 0$ when gender is male and $X_1 = 1$ when gender is female. Similarly,
$X_2 = 0$ when the teaching method is conventional; and $X_2 = 1$ when the teaching method is role-play.

The regression model may now be written as
Score = $\beta_0 + \beta_1$Gender + β_2Tchng mthd.
For male students who were taught by the conventional method only (no role-play) the mean score is $\beta_0 + 0 + 0 = \beta_0$
For women students who were taught by conventional method only (no role-play) the mean score is given by $\beta_0 + \beta_1 + 0 = \beta_0 + \beta_1$
For men students who had role-play the mean score is given by $\beta_0 + 0 + \beta_2 = \beta_0 + \beta_2$
For women students who had role-play the mean score is given by $\beta_0+\beta_1+\beta_2$

The results of regression analysis are displayed below

Regression Analysis

```
The regression equation is
score = 65.9 - 18.8 X1 - 14.9 X2

Predictor        Coef       StDev          T        P
Constant       65.913       3.420      19.27    0.000
X1            -18.838       3.950      -4.77    0.000
X2            -14.925       3.950      -3.78    0.001

S = 11.17       R-Sq = 56.1%     R-Sq(adj) = 53.1%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2      4620.9      2310.4     18.51    0.000
Residual Error    29      3619.0       124.8
Total             31      8239.8

Source       DF      Seq SS
X1            1      2838.8
X2            1      1782.0

Unusual Observations
Obs         X1      score        Fit   StDev Fit    Residual    St Resid
 12       0.00      78.60      50.99        3.42       27.61       2.60R

R denotes an observation with a large standardized residual
```

For men students who had only conventional teaching the mean score is 65.9 = β_0.
For women students who had only conventional teaching the mean score is
65.9 – 18.8 = 47.1 = $\beta_0 + \beta_1$.
For men students who were taught by the role-play method the score is
65.9 – 14.9 = 51 = $\beta_0 + \beta_2$.
For women students who were taught by the role-play method is
65.9 – 18.8 –14.9 = 32.2 = $\beta_0 + \beta_1 + \beta_2$.

We next perform the same regression analysis but with an interaction term included.

```
The regression equation is
score = 64.1 - 15.2 X1 - 11.2 X2 - 7.35 X1 * X2

Predictor        Coef       StDev          T        P
Constant       64.075       3.959      16.18    0.000
X1            -15.163       5.599      -2.71    0.011
X2            -11.250       5.599      -2.01    0.054
X1 * X2        -7.350       7.918      -0.93    0.361

S = 11.20       R-Sq = 57.4%     R-Sq(adj) = 52.8%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         3      4728.9      1576.3     12.57    0.000
Residual Error    28      3510.9       125.4
Total             31      8239.8

Source       DF      Seq SS
X1            1      2838.8
X2            1      1782.0
X1 * X2       1       108.0

Unusual Observations
Obs         X1      score         Fit   StDev Fit    Residual    St Resid
 12       0.00      78.60       52.83        3.96       25.77       2.46R
```

In both the regression analyses we notice a difference in the values of various means as compared to what was obtained with Two-way ANOVA. This is because in Two-way ANOVA the means displayed are for each factor. In other words they answer the questions “Are there differences in mean scores by gender?” and “Are there differences in mean scores by teaching method?” In the case of regression analysis the means are for the two factors taken together as we saw. In addition, in the second regression analysis all means are reduced by 7.35, the interaction effect. Again, there is no evidence of significance for interaction between the gender and teaching method since the t-ratio is not significant ($P = 0.361$).

Based on the results of the regression analysis, we can state that the average scores for men who are taught by the conventional method are the highest, followed by those of women taught in the conventional way. Role-play method reduces the average test scores for both sexes. There is no evidence of interaction between the two factors – gender and teaching method.

There is a discrepancy between the main effects as seen with Two-way ANOVA and regression analysis. The reason is the type of coding used for the coding the dummy variables.

The kind of coding we have used so far is called **reference coding**. An alternative form of coding is **effect coding**. What is the difference between the two methods of coding dummy variables?

If there are k explanatory categorical variables (factors), k – 1 dummy variables are created in both methods. Then in effect coding a value of –1 is assigned to the control group (instead of 0), to the category of interest (e.g. new treatment) is assigned a value of +1, and 0 for all others. Dichotomous variables are effect coded as –1 and +1, instead of 0 and +1.

In our example, taking the gender male as control group, the code assigned is –1. (The gender female is coded as +1 being the category of interest).

As regards the method of teaching, we take conventional form of teaching as control, and assign the code –1. Role-play is coded as +1. There are no other categories.

The results of regression analysis with effect coding and including the interaction term are shown below

```
The regression equation is
Score = 49.0 - 9.42 Gender - 7.46 Mthd - 1.84 Interact

Predictor      Coef          StDev         T             p
Constant       49.031        1.980         24.77         0.000
Gender         -9.419        1.980         -4.76         0.000
Mthd           -7.462        1.980         -3.77         0.001
Interact       -1.837        1.980         -0.93         0.361
S = 11.20      R-sq = 57.4%  R-sq(adj) = 52.8%

Analysis of variance

Source         DF     SS            Ms            F             p
Regression     3      4728.9        1576.3        12.57         0.000
Residual Error 28     3510.9        125.4
Total          31      8239.8

Source       DF      Seq SS
Gender        1      2838.8
Mthd          1      1782.0
Interact      1       108.0

Unusual Observations
Obs     Gender      score         Fit   StDev Fit    Residual    St Resid
 12      -1.00      78.60       52.83        3.96       25.77        2.46R

R denotes an observation with a large standardized residual
```

Interpretation of the different means is slightly different, but not difficult at all. The intercept or β_0 is 49.0 indicating that the overall mean score is 49. For gender the β coefficient is − 9.42. Men students were coded as –1. For men students we add −9.42 x (− 1) = + 9.42 to the overall mean score of 49, to obtain a score of 52.42. Likewise, for women students (coded as +1) we deduct 9.42 from the overall mean score of 49, and obtain 39.58. (Men students scored on average 9.42 above the mean score of 49.0, and women scored on average 9.4 points below the overall mean score). In the same way for conventional teaching method (coded as –1) we add 7.46 to the overall mean score of 49, and obtain 56.46; and deduct 7.46 for role-play method (coded as +1) to get 41.54. (Role-play as a teaching method resulted in an average score of 7.46 below the mean score, and the conventional form of teaching raised the score on average by 7.46)

The data set with effect coding is shown below for comparing with that on page 82.

score	Sex	TchngMthd
64.0000	-1	-1
75.6000	-1	-1
60.6000	-1	-1
69.3000	-1	-1
63.7000	-1	-1
53.3000	-1	-1
55.7000	-1	-1
70.4000	-1	-1
37.7000	-1	1
53.5000	-1	1
33.9000	-1	1
78.6000	-1	1
46.0000	-1	1
38.7000	-1	1
65.8000	-1	1
68.4000	-1	1
41.9000	1	-1
55.0000	1	-1
32.1000	1	-1
50.1000	1	-1
52.1000	1	-1
56.6000	1	-1
51.8000	1	-1
51.7000	1	-1
25.6000	1	1
23.1000	1	1
32.8000	1	1
43.5000	1	1
12.2000	1	1
35.4000	1	1
28.0000	1	1
41.9000	1	1

Comment

When we analyse data that can be grouped according to two or more factors (in our example the outcome variable Score was grouped by Sex and Teaching method) we use Two-way ANOVA. This approach allows us to test the hypothesis about each main effect (e.g. gender and teaching method), as well as to test for interaction between the two main effects.

However, unlike One-way ANOVA Two-way ANOVA requires that the two groups be of equal size. This is referred to as *balanced* design. A balanced design may not always be the case in clinical work because of patient non-compliance or dropouts. Also Two-way ANOVA is affected by missing data values. In such cases multiple regression analysis can help. The model can be extended to handle additional categorical independent variables.

Why the two methods of coding the dummy variables? Some time the investigator is interested in finding out about deviation of each group from the overall mean, particularly if there is no control group. For One-way ANOVA it does not matter how the dummy variables are coded. The study design is relatively straightforward and interpretation of results is easy. In two factors (or higher order designs) unravelling the results can pose a problem. In such cases effect coding makes the interpretation of different parameters easy. For anything other than one factor design effect coding is often resorted to. Some authors say that the more complicated the analysis the more necessary it is to use effect coding (i.e. 1, 0, –1) rather than reference coding (0, 1).

With effect coding the regression coefficients are halved but the P values do not change.

In reference coding (i.e. 0, 1 coding) the regression coefficients provide the estimates of the differences in the means of various treatment groups and the control (or reference) group. In effect coding (i.e. -1 for control group, +1 for group of interest and 0 for others), the regression coefficients provide deviation of each treatment group from the overall mean response instead of deviation from the control group mean.

What does testing for interaction tell us? It gives the researcher an insight into whether the main effects act independently, or whether one influences the action of the other. In some diseases multiple drug regimens are necessary. Testing for interaction provides an insight into their interaction.

In using regression analysis one employs indicator variables to create binary ("dummy") variables to represent all but one of the categories for each of the categorical explanatory variables. In our example of teaching method and student scores we have two categorical variables viz. gender and teaching method. Each of these has two categories. For each of them one indicator variable was created. In the regression analysis, using reference coding the constant represented the mean scores of the two excluded categories. In our example it represented the expected mean scores for those individuals who were male and received conventional teaching. Each partial regression coefficient represented the additive effect of the corresponding explanatory variable controlling for the effect of the other explanatory variable.

Although the additive model provides fairly accurate prediction of the scores of individual students, it is possible that the effects of gender and teaching method are not strictly additive. It may be possible to explain more of the variance in scores with the inclusion of an interaction variable that is a multiplicative function of "Gender" and "Teaching method". In other words the model with the interaction term included considers the score of each individual as being the result of gender, the effect of the teaching method, and of belonging to a particular gender **and** being taught by a specified teaching method. The significance of the single interaction term is tested using the '*t*' test and turned out to be not significant.

In the regression analysis using effect coding (i.e. -1, +1, 0) because the indicator ("dummy") variables were coded differently the values of the regression coefficients, their standard errors and the sum of squares are different from those obtained with reference coding (i.e.+1, 0). The intercept obtained is 49.031 indicating that the mean of all scores is 49. The β coefficient for Gender is −9.431 indicating that the mean score for all male students (coded as −1) was 9.4 points above the overall mean. That for all female students (coded as +1) was 9.4 points below the overall mean.

Sometimes one is interested in deviations of each group from an overall mean particularly if there is no control (or reference) group. In the case of One-way ANOVA it makes no difference what method of coding is used, and reference coding is recommended. For anything more complex than One-way ANOVA it helps to obtain the overall mean from which the various other parameters can be calculated. If there is interaction, effect coding helps to protect against multicollinearity and reference coding does not causing uncertainties in parameter estimates.

To summarise, in analyzing data with a Two-way or higher order analysis of variance and using multiple regression, if one includes interaction effect, one should use effect coding to ensure that correct sums of squares are obtained. If one does not include an interaction effect the coding does not matter. But rarely do we have prior information to justify excluding interaction.