10. Analysis of Longitudinal Studies
Repeat-measures analysis

This chapter builds on the concepts and methods described in Chapters 7 and 8 of Mother and Child Health: Research methods.

In repeat-measures designs each subject is observed before and one or more times after an intervention. A common example is a pre- and post-test of the knowledge level of participants attending a course of lectures. The following examples illustrate a range of clinical situations where such studies may be considered:

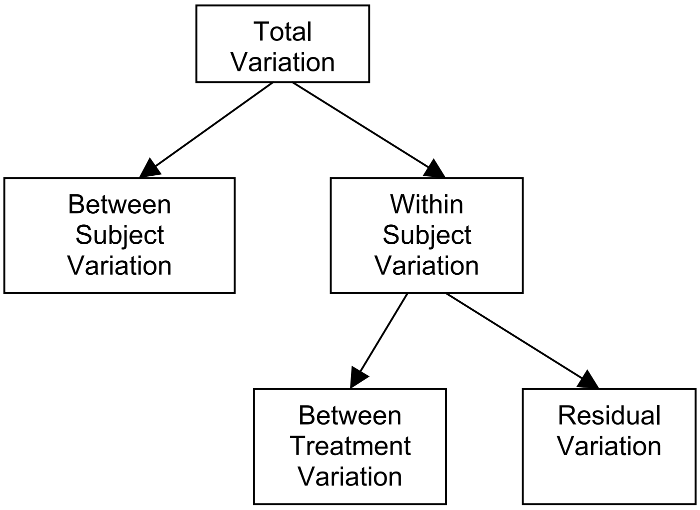
- In a clinical trial of pharmacokinetics of a drug blood samples may be taken hourly for 12 hours after its administration. So the data for an individual subject consist of a series of blood concentration levels.
- During the course of a study on treatments for the relief of asthma, a subject's forced expiratory volume (FEV) might be measured at weekly intervals to assess the efficacy of treatment.
- Preterm babies are weighed twice a week to monitor their rate of growth. A paediatrician may decide to put them on two different feeding regimes and compare growth rates in different groups.

Repeat-measure analysis is the generalisation of paired '*t*' test in the same way that One-way ANOVA is a generalisation of the two-sample '*t*' test. The main difference is that repeated observations or measurements are made on the same individuals. These are likely to be correlated necessitating an analysis that takes into account such correlation. For example, high systolic blood pressure in a subject on one occasion is likely to be followed by other high values.

Assessment of the effectiveness of treatments is more sensitive in repeat-measures study designs because they make it possible to measure how the treatment affects each individual. When control and treatment groups consist of different individuals the changes brought about by the treatment may be masked by the variability between subjects. By contrast, in repeat-measures designs each subject serves as own control. So in repeat-measures designs the variability between subjects can be isolated, and analysis can focus more precisely on treatment effects. A repeat-measures ANOVA puts each individual on an equal footing, and simply looks at how scores change with alternative treatments, or over time. The modification required to the standard method of performing analysis of variance is through partitioning of the total variation as shown in the illustration.

Reduction of random variability provides greater power to detect the effect of an intervention. Also, repeated measures designs are more economical, because each subject becomes own control and so fewer subjects are needed. But there are also disadvantages. In the simple ANOVA the accuracy of the *F* test depends upon the assumption that values of the outcome variable under different interventions are independent. In repeated measures this assumption is violated. For example if we are comparing the effectiveness of an anti-hypertensive drug against a placebo, those subjects who start with very high blood pressure will show the outcome levels of blood pressure higher than those who start with moderately high blood pressure. In other words, blood pressure levels obtained at the end of intervention are not independent of those at the beginning. And so another assumption has to be made. We assume that the relationship between pairs of values before/after intervention is roughly equal.

Statisticians call this **sphericity**. Sphericity refers to the equality of variances in the differences between interventions.



Repeat-measures studies are often referred to as *longitudinal studies*, and the data gathered are called *repeated-measures data*. The observations made on a subject during the duration of the study form a *cluster of responses*.

Repeat-measures designs have the obvious advantage of eliminating the difference in responses between individuals. But there are also some problems which one should be aware of. These are carry-over effect, latent effect, and learning effect.

Carry-over effect occurs when a treatment is administered before the effects of the previous treatment have worn off. The carry-over effect is taken care of by allowing sufficient time between treatments.

Latent effect occurs when one treatment can activate the dormant effects of a previous treatment. When latency effect is suspected it is best to avoid repeat-measures design.

Learning effect occurs in situations where response improves each time subjects take a test e.g. I.Q. testing, and in all pre and post-test situations.

Examples to clarify terminology

Young infants often bring up feeds. When this happens frequently enough to cause parental anxiety, and does not settle with simple measures like burping after feeds, treatment may be offered. One popular medication is alginate with antacid – "gaviscon". Let us assume that a clinical trial with gaviscon is conducted in which parents maintain a diary of the frequency of bringing up of feeds before and during treatment with gaviscon. Each infant is measured on the two levels of the treatment factor (no gaviscon / on gaviscon). Since each subject gets measured on every level of the treatment factor it is called a *crossover* factor.

Now suppose the trial consisted of two groups of infants such that one group received gaviscon and another a placebo, and a similar diary of frequency of bringing up of feeds is kept. Each infant gets measured on just one level of the treatment factor viz. on gaviscon / on placebo. In this situation the treatment factor is referred to as *nest* factor.

Let us consider another example. A lecturer teaching a course on research methods decides to evaluate the teaching by means of a pre- and post-course assessment of knowledge. Since each participant is measured on both the levels of the variable 'knowledge'. The variable 'knowledge' is a *crossover* factor. The lecturer had previously classified the participants by whether they were research fellows or clinical house officers. The variable 'participant' is one for which each subject is observed at just one level. So 'participant' is a *nest* factor.

The distinction between crossover and nest factors is made because the partitioning of variance in repeated-measures ANOVA depends on which factors are crossover and which are nest. The variance attributable to nest factors is derived from the between subject partition, whereas that attributable to crossover factors is derived from within subject partition.

In longitudinal studies where repeated observations (usually measurements) are made over time, a crossover factor is also referred to as *time dependent* variable, and a nest factor is called *time-independent* variable. A time dependent variable will take on values that may change for different observations on the same subject. In the pre- and post- example the scores on each subject are expected to be different for the two assessments. By contrast, a time independent variable has the same value for all observations on the same subject.

Besides the crossed/nested designation of variables, there is another designation which some computer packages ask for. A factor can be either *fixed* or *random*. If the research interest is in each individual subject with regard to the test results, then the subject factor is fixed. If the subjects can be considered to have been drawn at random from a larger population and the research interest is in the population rather than the individual, then the factor is random. Some software packages assume the factors to be fixed unless specified otherwise.

Balanced and unbalanced repeat-measures designs.

Since the repeated-measures design is an extension of the paired '*t*' test, by definition it is a balanced design. For most laboratory experiments and clinical trials on in-patients this should work well. But reality is different. Subjects get excluded because of side effects of drugs, non-compliance or self-discharge. In such cases the study design becomes unbalanced.

In this section we first discuss analysis of balanced designs, and then go on to consider strategies for situations where the study has become unbalanced.

Assumptions on which analysis of repeated-measures designs is based.

- **Normality**. Each population of scores should have a normal distribution.
- **Random selection**. Samples should have been independently and randomly selected from the population of interest.
- **Homogeneity of variance**. Different scores should have homogeneous variances. This is assessed prior to analysis by obtaining the variances of each group, and dividing the largest variance by the smallest to obtain an F max value. If this value turns out to be greater than 3 then the assumption has been violated, and the resulting F ratio must be evaluated at a more conservative level of significance.

- **Sphericity**. The variances of the population difference scores for any two conditions should be the same as the variance of population difference for any other two conditions.

Example of repeated-measures design with one crossover factor

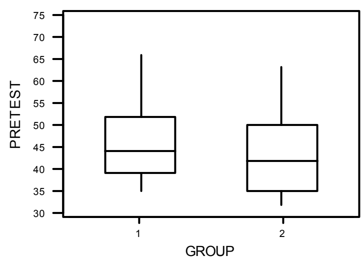
A lecturer in a course on research methods decides to assess his teaching by measuring the knowledge scores of the participants by means of a pre and post-test. There were 2 groups of participants – house officers and research fellows.

The data set for the knowledge scores of participants of a course is given below

GROUP	SUBJ	PRE	POST
1	1	51	56
1	2	35	49
1	3	66	70
1	4	40	56
1	5	39	52
1	6	46	59
1	7	52	62
1	8	42	64
2	9	34	48
2	10	40	59
2	11	34	57
2	12	36	60
2	13	38	49
2	14	32	57
2	15	44	69
2	16	50	67
2	17	60	75
2	18	63	76
2	19	50	68
2	20	42	59
2	21	43	67

Group 1 = House officers
Group 2 = Research fellows

We begin with exploring the data by looking at each test separately first.

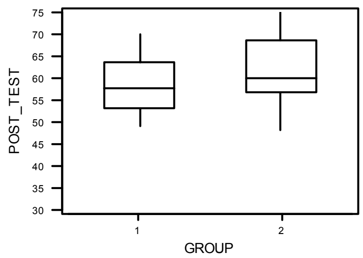


The mean score for Group 1 is a little higher than that of Group 2. Is the difference significant? We carry out the following analysis:

Source	DF	Seq SS	Adj SS	Adj MS	F	P
GROUP	1	39.85	39.85	39.85	0.41	0.528
Error	19	1833.11	1833.11	96.48		
Total	20	1872.95				

The difference in cores between the two groups is not significant (P = 0.528)

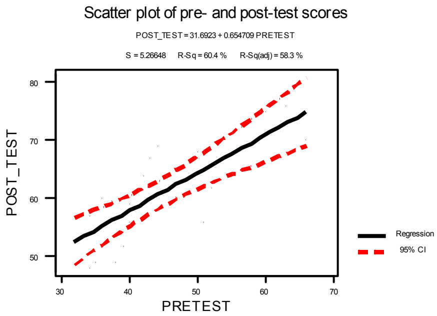
We next look at the post-test results:



Both the groups have improved their scores after the course but group 2 has done better. Is the difference significant? We check this by means of analysis of variance.

Source	DF	Seq SS	Adj SS	Adj MS	F	P
GROUP	1	74.73	74.73	74.73	1.13	0.301
Error	19	1255.08	1255.08	66.06		
Total	20	1329.81				

The difference is not significant. Is this due to a masking effect due to the results of pre-test? How closely the pre-test and post-test scores are correlated? This can be found out by means of a scatter plot as follows:



There is a significant correlation. Those scoring low on pre-test also score low on the post-test.

Pre-test scores act as covariate which could be masking the true effect of the course on post-test scores. This can be checked by means of ANCOVA.

Factor	Type	Levels	Values
GROUP	fixed	2	1 2

Analysis of Variance for POST_TES, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
PRETEST	1	802.83	894.91	894.91	44.73	0.000
GROUP	1	166.82	166.82	166.82	8.34	0.010
Error	18	360.16	360.16	20.01		
Total	20	1329.81				

```
Term          Coef   SE Coef        T      P
Constant    29.031     4.803     6.04  0.000
PRETEST     0.6987    0.1045     6.69  0.000


Unusual Observations for POST_TES

Obs  POST_TES       Fit      SE Fit  Residual   St Resid
  8   64.0000   55.4431      1.6462    8.5569      2.06R
 13   49.0000   58.5148      1.3689   -9.5148     -2.23R

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

In the analysis of variance table the total unexplained variability (error term) is reduced from 1255.08 to 360.16 so that the variability explained by Group is now 166.82, and we get a significant p value of 0.01.

Uses of ANCOVA

ANCOVA is used for 2 main purposes:

1. To increase the sensitivity of the test for main effects and interactions by reducing the error term. This was the case in the above example where the unexplained variability (error) is reduced. The reduction and adjustment for it is according to the correlation between the outcome variable (post-test) and the covariate (pre-test). In practice this is the most common use.
2. A second use is to adjust the means of the outcome variable to what they would be if all subjects scored equally on the covariate.

What to do with unbalanced designs?

By definition repeat-measures designs are balanced since each individual subject becomes own control. But reality in the clinical world can be different. If a group of subjects is being followed for a considerable period of time, the chances of gaps in the record of an individual increase with the number of follow-up visits planned. Most computer packages simply leave out of the analysis subjects with gaps in their data. This can lead to loss of considerable amount of data. New approaches are being evolved, and are being included into the more recent versions of some software.

Analysis of Covariance

A special problem sometimes arises in the analysis of repeated measures studies. Say for example, a group of subjects were given an aptitude test before they joined a study programme. Half the group attended conventional type classroom teaching and the other half had distance learning. Immediately after the study all the participants took a test, and were tested again a month later to assess retention of knowledge. The investigators are interested in knowing which teaching method was superior. At first glance this may appear to be a straightforward case of comparing different means. But what about the aptitude test? Does aptitude have a bearing on how a subject will respond to a learning experience? If so, is it the training method or the aptitude that resulted in the differences noted? Particularly if the scores of the aptitude test show a linear relationship with the test scores of the candidates it stands to reason that the aptitude test scores should be taken into account in making a decision about the test scores of the participants. The statistical procedure employed in such a situation is Analysis of Covariance. It may be thought of as a statistical matching procedure. The effects of aptitude score are taken into account by introducing a measure of aptitude test as a **covariate** into the ANOVA model. Any significant difference noticed in the test scores between the two methods can then be correctly attributed to the teaching method (Treatment effect). Continuous variables that are not part of the intervention but have an influence on the outcome variable are known as covariates.

The need for covariate analysis is even more strongly felt in observational studies because unlike intervention studies rigorous control of extraneous factors is not possible. In performing analysis by ANOVA we assess significance by comparing the total variability in the data (total sum of squares) with the variability that can be explained. With the help of ANCOVA we try to explain part of the unexplained variance in terms of the covariate(s).

Any number of covariates can be included in the analysis, but it should be recognised that the larger the number of covariates the more difficult becomes the interpretation.

In principle, ANCOVA represents a combination of ANOVA and linear regression. In the analysis the programme first makes an estimate of the regression of the depend variable on the covariate. The scores of the dependent variable are then adjusted to take into account the influence of the covariate. A standard ANOVA is then performed on these adjusted values. In this way the linear effect of the covariate gets accounted for. Since linear regression is involved all covariate terms must be continuous measures.

Analysis of Covariance (ANCOVA) is demonstrated in the following example:

Parents attending an Under-Fives' Clinic were first assessed on a scale of 0 to 10 to find out whether they held traditional beliefs regarding the causation of illness or were receptive to modern ideas. They then attended several sessions of health education. One month after the completion of training their knowledge was assessed by a scoring method on a scale of 0 to 100.

The investigators wish to know whether parents' belief system influences their acceptance of new ideas besides age, sex and years of schooling.

The data are given below.

Belief System	Sex	Yrs.at School	Age	Knowledge Score
1	1	16	21.0	80
2	1	9	27.0	70
6	0	0	29.0	65
6	1	0	36.5	67
3	1	5	31.5	60
2	1	10	28.0	80
7	0	0	26.0	57
3	0	8	17.0	70
2	1	12	26.0	83
3	1	5	33.5	65
2	1	11	26.5	73
8	0	0	22.5	40
3	0	6	26.0	75
7	0	1	34.5	61
3	1	14	33.5	71
8	1	3	28.0	65
9	0	0	40.0	59
3	0	5	28.5	76
8	0	0	31.5	54
1	1	12	20.5	89

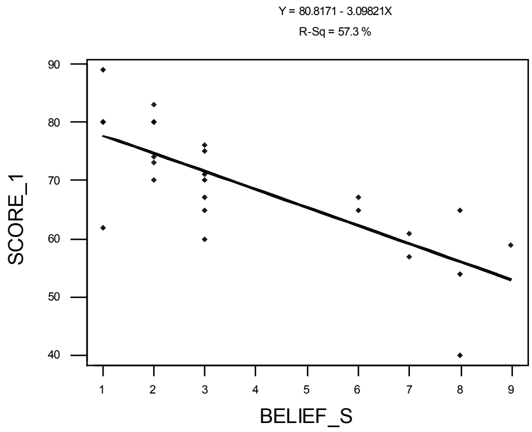
2	1	9	30.5	80
3	0	5	30.5	67
2	1	11	33.5	74
1	1	7	25.0	62
1	1	13	21.5	80

Belief System 1 = Receptive of modern ides
10 = Extreme Traditional

Sex 1 = Male
0 = Female

We first check whether there is a linear relationship between knowledge scores and the parents' belief system by plotting knowledge scores against the scores obtained regarding the cultural beliefs about disease causation.

Knowledge Scores plotted against Scores on belief system



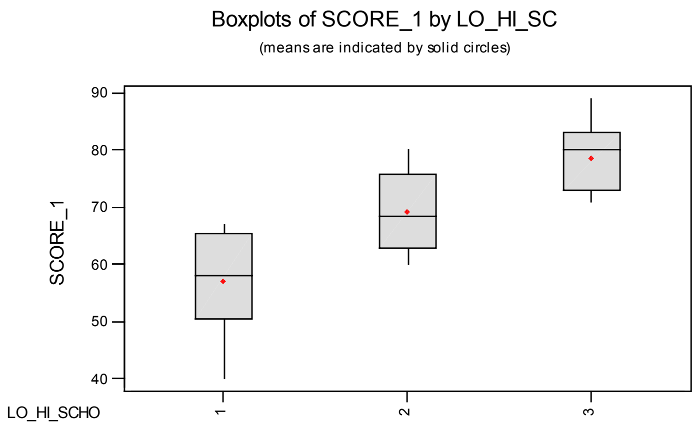
The plot shows a clear relationship with *R sq*. = 57%. It would appear that in considering the knowledge scores achieved after completion of the health education sessions we should make allowance for the individual's belief system.

In order to demonstrate the differences we first perform analysis of variance and then ANCOVA. For the purpose we group the variable YEARS AT SCHOOL into 3 groups 1 = No schooling; 2 = 10 years of school attendance; 3 = >10 years of school attendance. We also group the variable AGE as 1 = 0 – <25 years; 2 = 25 – 40 years.

Mean Knowledge Scores by Schooling

```
Analysis of Variance for SCORE_1
Source      DF        SS        MS        F        P
LO_HI_SC     2    1505.9     752.9    13.08    0.000
Error       22    1266.0      57.5
Total       24    2771.8
                                    Individual 95% CIs For Mean
                                    Based on Pooled StDev
Level        N      Mean     StDev  ----------+---------+---------+------
1            6    57.000     9.654   (-----*-----)
2           12    69.250     7.124              (---*----)
3            7    78.571     6.347                      (-----*-----)
                                    ----------+---------+---------+------
Pooled StDev =     7.586                     60        70        80
```

There is a significant difference between levels 1 and 2 and 1 and 3



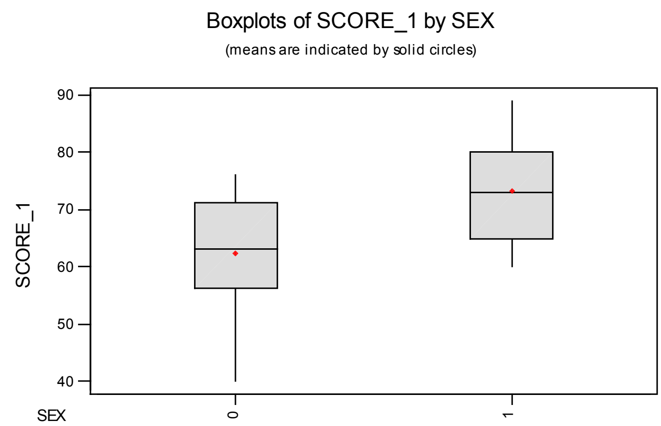
Mean Knowledge Scores by Age

No significant difference was noted. The mean score for the 0 – 25 years old was 70.17 ± 17.46; the mean score for the 25 – 40 years old was 68.53 ± 8.28.

Mean Knowledge Scores by Gender

```
Analysis of Variance for SCORE_1
Source      DF        SS        MS        F        P
SEX          1     708.5     708.5     7.90    0.010
Error       23    2063.3      89.7
Total       24    2771.8
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level        N      Mean     StDev  -------+---------+---------+---------
0           10    62.400    10.772   (---------*---------)
1           15    73.267     8.531                     (--------*--------)
                                    -------+---------+---------+---------
Pooled StDev =     9.472                 60.0      66.0      72.0
```

There is a difference significant at $P = 0.010$



In summary, significant differences in knowledge scores are demonstrable by level of schooling and gender.

We come back to the original question. What influence does a person's cultural attitude towards disease causation have on knowledge gained through the health education sessions?

We attempt to answer this by doing ANCOVA wherein the variable Belief System is included in the analysis, as shown below:

[IN SPSS → Simple Factorial ANOVA │ In box marked "Dependent" → "Score". In box marked factors → "Young Old" "Lo Hi Schooling" "Sex". In box marked "Covariates" → Belief System.]

```
                              * * *  C E L L   M E A N S  * * *
```

Part I

```
               SCORE_1
            by YOUNG_OL
               LO_HI_SC
               SEX


Total Population

    68.92
 (    25)


YOUNG_OL
        1          2

    70.17      68.53
 (      6)  (     19)


LO_HI_SC
        1          2          3

    57.00      69.25      78.57
 (      6)  (     12)  (      7)


SEX
        0          1

    62.40      73.27
 (     10)  (     15)



          LO_HI_SC
                  1          2          3
YOUNG_OL
        1     40.00      66.00      83.00
           (      1)  (      2)  (      3)

        2     60.40      69.90      75.25
           (      5)  (     10)  (      4)



          SEX
                  0          1
YOUNG_OL
        1     55.00      77.75
           (      2)  (      4)


        2     64.25      71.64
           (      8)  (     11)
```

```
          SEX
                 0         1
LO_HI_SC
        1     55.00     67.00
           (      5)  (      1)

        2     69.80     68.86
           (      5)  (      7)

        3       .00     78.57
           (      0)  (      7)
```

Part II

```
            * * *  A N A L Y S I S   O F   V A R I A N C E  * * *

                 SCORE_1
            by   YOUNG_OL
                 LO_HI_SC
                 SEX
            with BELIEF_S

                 EXPERIMENTAL sums of squares
                 Covariates entered FIRST
```

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Covariates	1587.275	1	1587.275	32.197	.000
BELIEF_S	1587.275	1	1587.275	32.197	.000
Main Effects	247.894	4	61.974	1.257	.321
YOUNG_OL	91.497	1	91.497	1.856	.189
LO_HI_SC	167.264	2	83.632	1.696	.210
SEX	2.730	1	2.730	.055	.816
Explained	1835.169	5	367.034	7.445	.001
Residual	936.671	19	49.298		
Total	2771.840	24	115.493		

```
Covariate   Raw Regression Coefficient

BELIEF_S       -3.098


25 cases were processed.
0 cases (.0 pct) were missing.


Due to empty cells or a singular matrix,
higher order interactions have been suppressed.
```

Part III

```
* * *  M U L T I P L E   C L A S S I F I C A T I O N   A N A L Y S I S  * * *

          SCORE_1
     by   YOUNG_OL
          LO_HI_SC
          SEX
     with BELIEF_S

Grand Mean =    68.92                           Adjusted for
                                                Independents
                                    Unadjusted  + Covariates
Variable + Category           N     Dev'n  Eta  Dev'n   Beta

YOUNG_OL
    1                         6      1.25       -3.72
    2                        19      -.39        1.18
                                            .07          .20

LO_HI_SC
    1                         6    -11.92       -4.12
    2                        12       .33       -1.28
    3                         7      9.65        5.73
                                            .74          .36

SEX
    0                        10     -6.52         .54
    1                        15      4.35        -.36
                                            .51          .04

Multiple R Squared                                      .662
Multiple R                                              .814
```

Interpretation of the Results

Different parts of the output are numbered from I to III in order to help with following the interpretation provided.

Part I.

The first part of the output describes the mean scores of the various groups together with a statement of the number of subjects in each cell.

It begins with the overall mean score of 68.92 for all the scores of n=25 subjects pooled together.

The mean score for the young (0 - <25 years) group was 70.17 (with n=6) and for the older group (25 – 40 years) it was 68.5 (with n=19), and so on for all the other groups. These mean scores are the same as those obtained on page 95. Finally part I ends with cross-tabulations of mean scores by different groups.

Part II.

Part II is the analysis of variance table. The effects of the covariate are entered first. A very small value of *P* for the significance of F in the case of the variable ‘Belief System’ tells us that this variable has a strong influence on the scores. The values of F for the main effects (Young/Old; School groups; gender) are not significant. They have been calculated using the means adjusted for the effects of ‘Belief System’.

At the bottom of the table is given the Raw Regression Coefficient of the covariate 'Belief System' and the dependent variable 'Test Score at 1 month'.

Part III.

Part III is the multiple classification analysis table. It commences with a reminder of the overall mean score of 68.92.

In the body of the table there are 2 columns labelled 'Unadjusted deviation' and 'Adjusted for Independent + covariates'. Unadjusted deviation is the difference between each group mean and the overall mean. Thus 70.17 – 68.92 = 1.25; 68.53 – 68.92 = –0.39, and so on.

The adjacent column provides similar information, but this time each group mean is adjusted for the effect of the covariate. For the young group (0 - <25 years) the adjusted mean is 65.3 (65.3 – 68.92 = –3.62). For the older group (25 – 40 years) the adjusted means is 70.06. (70.06 – 68.92 = 1.14), and so on. The adjustment performed on each group mean is in accordance with the Beta value or the 'slope'.

Conclusion.

There are significant differences in the mean test scores of young and old parents, as well as by their years of schooling. However, these are heavily influenced by each subject's belief in traditional explanations of disease causation.

Comment

Analysis of Covariance (ANCOVA) is an extension of analysis of variance. The main effects and interactions of the explanatory variables are estimated for the response variable, as is the case with ANOVA, but *after adjusting for* the influence of one or more continuous variables (i.e. covariates). This additional variable is one that could possibly be affecting the scores on the outcome variable. The main question being asked by the investigator is essentially the same for both ANOVA and ANCOVA viz. "Are the mean differences in the response likely to have occurred by chance?"

ANCOVA helps to increase the sensitivity of the main effects and interactions by adjusting the residual sum of squares for the relationship between the response variable and the covariate. Such an adjustment increases the power of the F statistic. The mean responses for the different groups also get adjusted to what they would be if all subjects scored identically on the covariate. Differences between subjects relating to the covariate(s) are removed so that the only differences that remain are the effects of the explanatory variables. What ANCOVA does is to first perform regression of the covariate(s) on the response variable. Then the individual responses and the means are adjusted to remove the linear effects of the covariate(s) before going on to performing analysis of variance on the adjusted values of the response variable. ANCOVA may be used as part of one-way ANOVA (one explanatory variable – one outcome variable); Two-way ANOVA (two explanatory variables – one outcome variable); or multivariate ANOVA analysis.

Caution must be exercised in the interpretation of such adjusted means. They do not represent the real world. All that they indicate is what could have been expected if all the subjects had the same scores on the covariate.

ANCOVA is based on the following assumptions:

1. Measurement of the covariate must be done before the intervention.
2. The covariate must be measured reliably choosing the most accurate tools available.
3. If there are more covariates than one they should not be strongly correlated with one another. (r <0.8).
4. There should be a linear relationship between the outcome variable and the covariate in each group of subjects. (Checked by means of scatter plots to rule out a curvilinear relationship).
5. This linear relationship between the outcome variable and the covariate should be same for each of the groups.