9. Analysis of Intervention Studies - III
Factorial Designs

In Chapters 7 and 8 the explanatory categorical variables were referred to as treatments or factors. These terms are often used interchangeably. So in a way we have already discussed simpler forms of factorial designs. More commonly the term ‘factorial design’ is used to describe situations where there are two or more factors. A *factor* is simply a categorical variable with two or more values, often referred to as *levels*. In situations where there are two or more explanatory variables (factors) the study is said to be factorial. If the factors have several different values they are referred to as levels. For example, in a clinical trial subjects may differ in terms of the treatments they receive (one factor), whether the medications are administered orally or by injection (second factor), and gender (third factor). If there are three different drugs (A, B and C), two different methods of administration (I and II), and two sexes (M and F), there will be in all $3 \times 2 \times 2 = 12$ groups of subjects for comparison of the outcome. By being able to combine a number of different research hypotheses into one intervention there is much saving of time and resource.

As we saw in Chapter 8, analysis involves considering the mean differences in outcome in different groups of subjects. First, the average difference in outcome corresponding to each factor can be compared; secondly, the interaction effect, which tells us how the outcomes would differ when one factor is changed to another (for example, oral administration to parenteral). The first type of difference is called the main effect; and the second type interaction effect. Thus, in addition to efficiency, factorial designs enable us to identify significant interactions, if they exist, amongst the interventions.

A further refinement in the design may be added by having each factor (or treatment) at two levels – low and high. For example, clinical trials in which treatments are administered at low and high doses. Such designs are not unusual in industry. Chemical reactions are often studied with catalysts and temperature at different levels to determine what combination of the reacting chemicals, the catalysts and temperature is best. A trial in which there are 3 factors and 2 levels (as in our example) is referred to as 2^3 factorial design. In order not to make interpretation too complicated it is usually recommended not to exceed 3 factors and 2 levels.

It has been implied in the discussion so far, and also in Chapter 7 and 8, that in designing intervention studies one has to balance the available resources (including time) against obtaining a sufficiently accurate answer to the research question. One way of doing this is by means of an appropriate blocking scheme. As we have seen, blocking is a way of obtaining the results of treatments in subjects who are homogeneous in their response. All comparisons are then made within blocks. When blocking has been done with careful planning, the researcher cuts out an extraneous source of variability in response, thereby improving precision.

The next consideration is how many subjects to include in each block. The more the number of subjects in each treatment the more accurate is the measurement of outcome. But the cost in time and money is also that much more. Hence various types of trade-offs are considered and sample size calculated accordingly. Random allocation of treatments within blocks is necessary to avoid bias.

Often there are unforeseen interactions between treatments, and this also has to be considered relative to the dose used. It is necessary to understand clearly what is meant by interaction between factors. If the outcome with any two dose levels of medication A are the same for all levels of factor B, and *vice versa*, then there is no interaction. Similarly, if the effects of the two drugs are uniformly additive for all levels (doses) of the drugs concerned there is no interaction. But if in changing from low to high dose of A the outcomes change for all levels of B, then we say that there is interaction. These definitions are important for correct interpretation of results.

To sum up, factorial designs as compared to one factor at a time approach have the following uses:

1. A greater precision is obtained in estimating the overall main factor effects.
2. Interactions between different factors can be explored.
3. Including additional factors can extend validity of conclusions derived.

However, there are several reasons why one should not plunge straight away into factorial designs, but instead keep them for the end. Firstly, there is the obvious consideration of simplicity. Studies with many factors at different levels and with a large number of subjects are difficult to organise and run smoothly. Secondly, simple preliminary studies are useful for identifying the more important factors and for determining the more promising ranges of each factor. Thirdly, in many clinical situations it is more profitable to first obtain a thorough insight into the pathophysiology. A chain of simple investigations, such that each is a result of preceding ones and is planned to throw further light on what was found before, is a good strategy to begin with. Once a sound background of information has been built with simpler studies, a multifactorial design with different levels of the factors involved can help to describe the situation by observing it under a wide range of conditions.

The Two-way ANOVA method of analysis does not extend to situations where there are more than two factors. Instead, use is made of a procedure called General Linear Model (GLM). It is a generic procedure to which One-way ANOVA, Two-way ANOVA and multiple linear regression belong.

The various points raised so far would now be illustrated by taking an example of a two factor experiment in which a further twist has been added by having each factor at more than one level.

In order to study the role of dietary protein in promoting growth, researchers fed equal numbers of laboratory mice on three dietary regimens in which the protein was derived either from cereal, or beef, or pork. There were 20 mice in each of the three groups. In each group half the number were fed the diet in which the study protein was in small amounts, and the other half on a diet in which the quantity of the study protein was much greater. Growth was measured by weighing the mice at the end.

(*Source: Senedecore GW,Cochrane WG. Statistical methods 6th edition. 1967. Iowa State University Press*).

The data set is given below:

Weight(g)	Protein type	Amount
107	1	1
95	1	1
97	1	1
80	1	1
98	1	1
74	1	1
74	1	1
67	1	1
89	1	1
58	1	1
98	1	2
74	1	2
56	1	2
111	1	2
95	1	2
88	1	2
82	1	2
77	1	2
86	1	2
92	1	2
90	2	1
76	2	1
90	2	1
64	2	1
86	2	1
51	2	1
72	2	1
90	2	1
95	2	1
78	2	1
73	2	2
102	2	2
118	2	2
104	2	2
81	2	2
107	2	2
100	2	2
87	2	2
117	2	2
111	2	2
49	3	1
82	3	1
73	3	1
86	3	1
81	3	1
97	3	1
106	3	1
70	3	1
61	3	1
82	3	1
94	3	2
79	3	2
96	3	2
98	3	2
102	3	2
102	3	2
108	3	2
91	3	2
120	3	2
105	3	2

Protein type 1 = Cereal
2 = Beef
3 = Pork

Amount
1 = Low
2 = High

[In MINITAB Stat ANOVA General Linear Model In response box "weight(g)". In box marked Model "Protein" "Low/High".]

```
General Linear Model

Factor     Type Levels Values
Protein   fixed      3 1 2 3
Low/High  fixed      2 1 2

Analysis of Variance for Weight(g) using Adjusted SS for Tests

Source     DF     Seq SS     Adj SS     Adj MS       F      P
Protein     2      266.5      266.5      133.3    0.58  0.561
Low/High    1     3168.3     3168.3     3168.3   13.90  0.000
Error      56    12764.1    12764.1      227.9
Total      59    16198.9


Unusual Observations for Weight(g)

Obs  Weight(g       Fit   StDev Fit  Residual   St Resid
  1   107.000    77.633       3.898    29.367      2.01R
 13    56.000    92.167       3.898   -36.167     -2.48R
 26    51.000    82.333       3.898   -31.333     -2.15R
 41    49.000    81.833       3.898   -32.833     -2.25R

R denotes an observation with a large standardized residual.

Dunnett 95.0% Simultaneous Confidence Intervals
Response Variable Weight(g
Comparisons with Control Level
Protein = 1 subtracted from:

Protein     Lower    Center     Upper  --+---------+---------+---------+----
2          -6.134     4.700     15.53    (-----------------*-----------------)
3          -6.634     4.200     15.03   (-----------------*-----------------)
                                       --+---------+---------+---------+----
                                       -6.0       0.0       6.0      12.0


Dunnett Simultaneous Tests
Response Variable Weight(g
Comparisons with Control Level
Protein = 1 subtracted from:

Level       Difference       SE of              Adjusted
Protein       of Means  Difference    T-Value    P-Value
2                4.700       4.774     0.9845     0.5164
3                4.200       4.774     0.8797     0.5859

Dunnett 95.0% Simultaneous Confidence Intervals
Response Variable Weight(g
Comparisons with Control Level
Low/High = 1 subtracted from:

Low/High     Lower    Center     Upper  -------+---------+---------+---------
2            6.724     14.53     22.34  (---------------*---------------)
                                        -------+---------+---------+---------
                                             10.0      15.0      20.0


Dunnett Simultaneous Tests
Response Variable Weight(g
Comparisons with Control Level
Low/High = 1 subtracted from:

Level        Difference       SE of              Adjusted
Low/High       of Means  Difference    T-Value    P-Value
2                 14.53       3.898      3.728     0.0005
```

Interpretation of the results

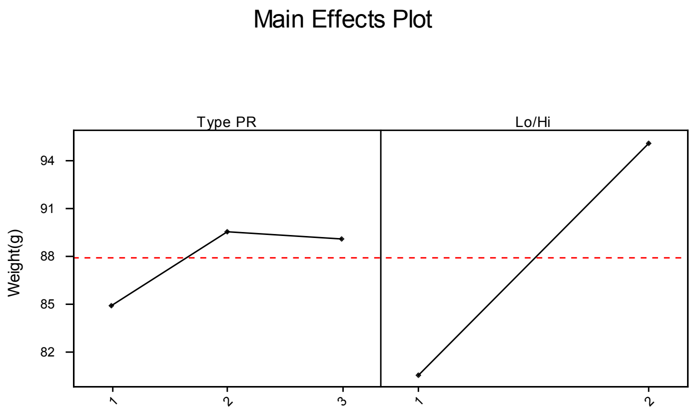
The first table displays the factors with their number of levels and the level values. Recall that for protein there are 3 values (1= cereal; 2= beef; 3=pork). For amount (Low/High) there are 2 levels; 1=low and 2=high.

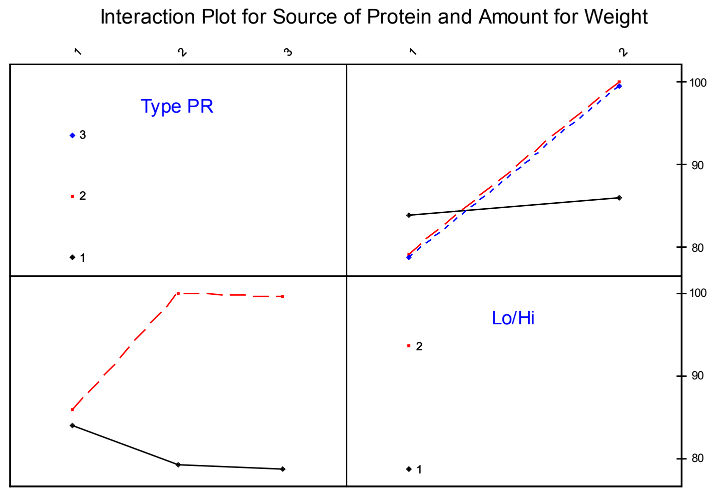
The second table is the analysis of variance table. The *P* value for Low/High is significant but not for protein.

Next is a table of unusual observations, followed by a series of Dunnett's tests. The first of the series is comparing the means of weights by the levels of protein. The test is not significant because zero is contained in the interval. Finally, there is a test for amount (Low/High), and it is significant.

The conclusion that we can draw from these results is that there is no significant difference in weight gain by the type of protein, but there is by the amount.

We next obtain a plot of the means for the main effects, and another one for interaction between the two factors.





The graph in the top right hand quadrant shows that the lines are not parallel. This means that there is an interaction. Recall that interaction is said to occur when the values of the outcome variable for each level of one explanatory variable do not show the same pattern for all values of another explanatory variable. In our example, the pattern of weight gain achieved for beef and pork derived proteins is similar for low and high amounts fed since the lines are parallel, but this is not the case between cereal and animal protein. The two ends of the lines show the means of weight gained on feeding the low and high amounts respectively. At the low amount of feeding the mean weight gained on cereal protein is better than that achieved with beef or pork derived protein. At the high amount of feeding the mean weight gain on animal protein is far greater compared to cereal protein. This explains why the lines are not parallel.

An example of a study with 3 factors each at 2 levels (i.e. 2^3 factorial design).

We next consider a 3-factors, each at 2 levels, in the next example.

In a chemical plant a pilot experiment was carried out to investigate the yield of a product using 3 factors. These were the concentration of the substrates, the temperature at which reaction took place, and 2 different types of catalysts. Two levels of concentration of the substrate were used viz. 20% and 40%. Two levels of temperatures were used viz. 160^0 C and 180^0 C. Each combination was repeated once, so there are two runs per combination. The data are listed below.

Yield	Temp.	Concent.	Catalyst
59	160	20	1
61	160	20	1
74	180	20	1
70	180	20	1
50	160	40	1
58	160	40	1
69	180	40	1
67	180	40	1
50	160	20	2
54	160	20	2
81	180	20	2
85	180	20	2
46	160	40	2
44	160	40	2
79	180	40	2
81	180	40	2

The results of the analysis are as follows:

Factor	Type	Levels	Values
Catalyst	fixed	2	1 2
Temperat	fixed	2	160 180
Cocentra	fixed	2	20 40

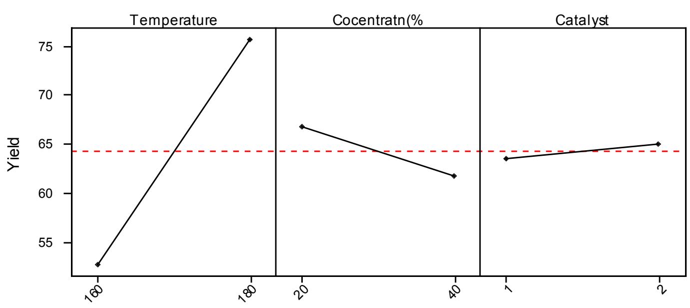
Analysis of Variance for Yield, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Catalyst	1	9.00	9.00	9.00	0.23	0.642
Temperatr.	1	2116.00	2116.00	2116.00	53.57	0.000
Cocentratn	1	100.00	100.00	100.00	2.53	0.138
Error	12	474.00	474.00	39.50		
Total	15	2699.00				

From the F test results we see that temperature has a significant effect.

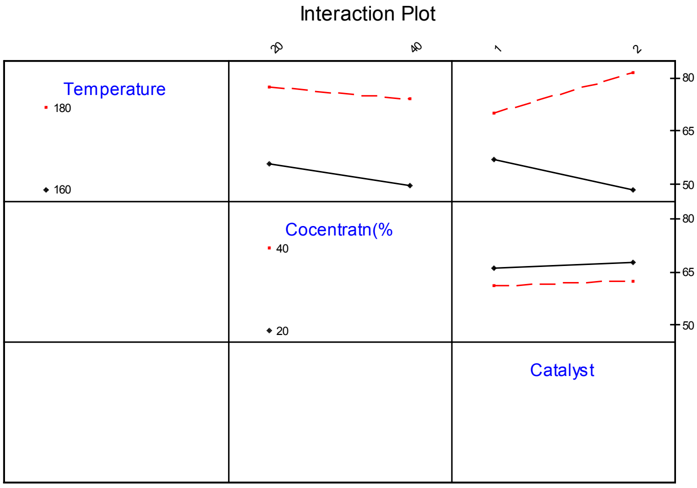
We follow this up with plots of main effects and for interaction

Main Effects Plot



In the main effects plot the difference in the mean yield at 160^{0}C (52 units) and at

180^0C (76 units) is clear to see. The mean yield at concentration of 20% is 66.75 units, and at 40% it is 61.75 units. Increasing the concentration from 20% to 40% reduces the mean yield by 5 units. Changing from catalyst 1 to 2 makes no appreciable difference.



In the interaction plot, in the top right hand corner we notice an interaction between temperature and catalysts (the two lines are not parallel as elsewhere). Increasing the temperature produces a greater yield with catalyst 2 than with catalyst 1.

Multiple Regression Approach to analysis of factorial designs

We next consider how the same two data sets may be analysed using the multiple linear regression approach.

(i) Data set of weight gain in laboratory rats on feeding three types of proteins

There are two factors viz. protein source and amount. Dummy variables need to be created for each factor. We decide to use weight gain on cereal protein fed in low amounts as the control group.

Effect coding of the dummy variables is used as follows:

For the three levels of protein source (viz. cereal, beef and pork) we create two dummy variables (X_1 and X_2) such that:

$X_1 = 1$ for beef protein
$X_1 = 0$ for pork protein
$X_1 = -1$ for cereal protein
and
$X_2 = 0$ for beef protein
$X_2 = 1$ for pork protein
$X_2 = -1$ for cereal protein

For the two levels of the factor 'amount' (viz. low and high) we create one dummy variable (X_3) such that

$X_3 = -1$ for low
$X_3 = +1$ for high.

We would look for interaction between the two factors by including the terms $X_1 * X_3$ and $X_2 * X_3$ in the regression equation.

The result of the regression analysis is shown below:

```
The regression equation is
Weight(g) = 87.9 + 1.73 X1 + 1.23 X2 + 7.27 X3 + 3.13 X1*x3 + 3.13 X2*X3

Predictor        Coef       StDev          T        P
Constant       87.867       1.891      46.47    0.000
X1              1.733       2.674       0.65    0.520
X2              1.233       2.674       0.46    0.647
X3              7.267       1.891       3.84    0.000
X1*x3           3.133       2.674       1.17    0.246
X2*X3           3.133       2.674       1.17    0.246

S = 14.65       R-Sq = 28.5%     R-Sq(adj) = 21.9%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         5      4612.9       922.6      4.30    0.002
Residual Error    54     11586.0       214.6
Total             59     16198.9

Source       DF      Seq SS
X1            1       220.9
X2            1        45.6
X3            1      3168.3
X1*x3         1       883.6
X2*X3         1       294.5

Unusual Observations
Obs        X1   Weight(g         Fit   StDev Fit    Residual    St Resid
 13     -1.00      56.00       85.90        4.63      -29.90      -2.15R
 26      1.00      51.00       79.20        4.63      -28.20      -2.03R
 41      0.00      49.00       78.70        4.63      -29.70      -2.14R

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

The amount of protein fed makes a significant difference (T-ratio 3.84; $P = 0.000$)

We next proceed to calculating the mean weight gain on different combinations of protein and amounts fed. Recall that for cereal protein X_1 and $X_2 = -1$; and for low amount $X_3 = -1$, since cereal protein fed in low quantity is taken as the control.

Mean weight gain on low amount of cereal protein $= 87.9 - 1.73 - 1.23 - 7.27 + 3.13 + 3.13$
$= 83.93$
Mean weight gain on high amount of cereal protein $= 87.9 - 1.73 - 1.23 + 7.27 - 3.13 - 3.13$
$= 85.95$
Mean weight gain on low amount of beef protein $= 87.9 + 1.73 + 0 - 7.27 - 3.13 + 0$
$= 79.23$
Mean weight gain on high amounts of beef protein $= 87.9 + 1.73 + 0 + 7.27 + 3.13 + 0$
$= 100.03$

Mean weight gain on low amount of pork protein = 87.9 + 0 + 1.23 – 7.27 + 0 – 3.13
= 78.73
Mean weight gain on high amount of pork protein = 87.9 + 0 + 1.23 + 7.27 + 0 + 3.13
= 99.53

We can conclude that in the case of cereal protein weight gain is better than that for beef or pork protein when proteins are fed in small amounts. When fed in larger amounts animal protein has an advantage. There is little difference between beef and pork protein at both levels of the amount fed.

(ii). We next go on to analyse the second data set .

The data is about the yield of a chemical using different combinations of 2 levels each of temperature, concentration of substrate and two catalysts.

There are 3 factors, each with two levels. This makes the task easy. We create 1 dummy variable respectively for each factor (X_1, X_2, and X_3) such that

X_1 represents the factor 'Temperature'; 160^0 c = –1 (control); and 180^0 c = +1
X_2 represents the factor 'Concentration'; 20% = –1(control); and 40% = +1
X_3 represents the factor 'Catalyst'; 1 = –1(control); 2 = +1

The results of the regression analysis are as follows:

```
The regression equation is
Yield = 64.3 + 11.5 X1Temp. - 2.50 X2 Concentr. + 0.75 X3catalyst

Predictor        Coef       StDev          T        P
Constant       64.250       1.571      40.89    0.000
X1Temp.        11.500       1.571       7.32    0.000
X2 Conce       -2.500       1.571      -1.59    0.138
X3cataly        0.750       1.571       0.48    0.642

S = 6.285       R-Sq = 82.4%     R-Sq(adj) = 78.0%


Analysis of Variance

Source            DF          SS          MS         F        P
Regression         3     2225.00      741.67     18.78    0.000
Residual Error    12      474.00       39.50
Total             15     2699.00

Source       DF      Seq SS
X1Temp.       1     2116.00
X2 Conce      1      100.00
X3cataly      1        9.00
```

We notice that the T-ratio for the factor 'Temperature' is significant (*P* 0.000). The factors 'Concentration' and 'Catalyst' each have a non-significant T-ratio.

It may be concluded that temperature at which the reaction occurs has a maximum influence on the yield.

The mean yield at different levels of temperature and concentration of the substrate using one or the other catalyst can be worked out as in the previous example. However, this was not the research question. In this pilot study all that the investigators wished to find out was what combination of temperature, concentration of substrate and catalyst would give the best yield.