

## 7. Analysis of Intervention Studies – I

### One-way Analysis of Variance (ANOVA)

This chapter and two more that follow build on Chapter 8 of Mother and Child Health: Research methods wherein the principles and the basic designs for intervention studies have been described.

Analysis of variance is a technique for assessing the effect of an explanatory categorical variable on a normally distributed continuous response variable. The categorical variable could be groups or different treatments, and the response variable a measure of their effects on a disease state. For example, in a clinical trial to assess the effectiveness of a new antihypertensive drug the categorical variable would include the new drug, the current treatment, and a placebo (i.e. three treatment groups). The subjects would be the patients enrolled in the study, and the level of blood pressure achieved would serve as the outcome measure. Another example is that of an observational study of sickle cell disease in which the investigator wishes to relate the level of anaemia to the type of haemoglobin – HbAS, HbSS and HbSC.

The general principles for designing such experiments or clinical trials are as follows:

1. Each subject should receive one treatment, chosen at random from the list of treatments to be compared. Care should be taken to ensure that subjects receiving one treatment do not differ from those receiving other treatments, and randomisation is a way of achieving this aim.
2. Each treatment should be used with several subjects (or experimental units) in order to get a proper measure of the variability of response between different subjects receiving the same treatment.

A new treatment that works well when special attention is being given, as indeed is the case in research institutions and university hospitals, may be less effective in ordinary routine conditions. Hence simplicity in design is essential. Simple designs are easy and straightforward to analyse.

A clinical trial in which the choice of treatment for each subject is completely random is called a **completely randomised** experiment. Such an approach is preferred when several treatments are to be compared, and where it is the level of response, which is important. When the number of subjects allocated to each treatment is the same the experimental design is **balanced**. Analysis of such studies is the subject of this chapter.

When several treatments are to be compared by the mean responses of the subjects, performing multiple *t* tests of significance for differences in outcome could be misleading. The reason is as follows:

Suppose two treatments were compared in a group of patients using the *t* test where the significant *P* value was 0.05. This means that there is a 5 per cent risk of the result arising by chance. Now if we were to next compare treatment one with treatment three in the same group of patients again using the *t* test with a significant *P* value of 0.05, the same possibility of 5 per cent of the result arising by chance occurs. Taking the results of the two tests together we run the risk of concluding that the treatments differ by 5 per cent + 5 per cent = 10 per cent of the time. In effect this means that our *P* value is now 0.10. The correct

approach is to use analysis of variance (ANOVA), which is the preferred method for comparisons involving several population means.

The ANOVA approach looks for group differences. Its basis is to compare the variance between groups receiving different treatments with the random variance that exists in nature within all groups. The first variance is on account of the treatment. The second is the natural variance in response because of variations in each individual's response. ANOVA is well suited to the analysis of response in classical experimental designs where the researcher wishes to compare the outcome in one or more experimental groups with that in the control group. The two groups are well matched on all possible counts. In clinical work assembling a large group of patients and randomly assigning them to different treatment groups, including one control group, commonly achieves this. The only difference between the groups is the treatment given, which becomes the independent variable and response to treatment is the dependent variable. The researcher is interested to see the response to different treatments in order to ascertain whether one treatment is better than the other. Why call it Analysis of Variance when the purpose of the study is to compare mean responses? ANOVA is justified because the test involves comparison of variances between groups as detailed below. Total variability is partitioned into components that correspond to the different influences on the data. In our example of treatment for hypertension the influences are due to the treatments. In our second example, these are due to the different types of haemoglobins.

The test is described as One Way Analysis of Variance, because there is one factor involved viz. the treatment. If we wish to see the difference in response between say males and females in the same group we use the Two Way Analysis of Variance because two factors are involved viz. treatments and sex.

## The statistics of ANOVA

Analysis of Variance may be looked upon as the multigroup generalisation of the  $t$  test. (In EPI - INFO vs.6 ANOVA is routinely performed for comparing several means). The basic assumption is that the samples to be compared are randomly drawn from a normally distributed population with the same standard deviations. The question the researcher is asking is "Are the differences in the means of the different treatment groups simply a reflection of random variation associated with the sampling process, or are they too large and therefore due to treatment?" The variability is partitioned into two groups viz.

- i). variability in outcome due to the intervention (i.e. variance between groups), and
  - ii). variability due to individual fluctuation within each group (i.e. variance within groups).
- The  $F$  statistic is computed by  $F = \text{Between group variance} \div \text{Within group variance}$ .

When  $F$  is large we conclude that at least one of the treatments had an effect different from the others, (or that one of the groups has a different level of haemoglobin).

## The Assumptions of ANOVA

Four assumptions are made in using ANOVA

1. Random samples have been taken to form the comparison groups of the subjects.
2. A value for the response has been recorded for each subject.
3. The response variable is normally distributed in each group.

4. The variance of the response variable is the same in each group.
5. A crucial assumption is that the effect of a treatment on a subject is to add a quantity, positive or negative or zero, to the overall response which would have been otherwise obtained.

In the clinical situation not all the assumptions are clearly satisfied, and yet one has to compare several outcome means. Generally speaking, one-way ANOVA is a robust method, and can be applied as long as none of the assumptions is badly violated.

One-way ANOVA test is now illustrated by means of two examples.

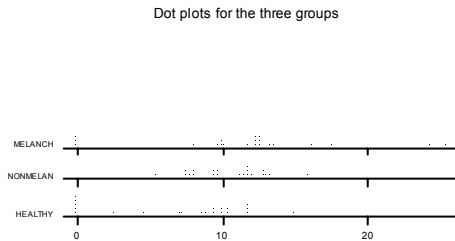
**Cortisol levels in depression**  
(*Arch. Gen. Psychiatry 1986; 43:849-858*)

Researchers wished to find out how depression in the individual affects adrenal function. They measured cortisol levels in the blood of healthy adults and in those diagnosed to be suffering from depression. The depressed subjects were of two types, those with more severe type called melancholic depression and the less severe non-melancholic. The data obtained are tabulated below. The investigators wish to know whether there are significant differences in the mean cortisol values between the three groups.

Healthy	Non-mel.	Melanch.
2.5	5.4	8.1
7.2	7.8	9.5
8.9	8.0	9.8
9.3	9.3	12.2
9.9	9.7	12.3
10.3	11.1	12.5
11.6	11.6	13.3
14.8	12.0	17.5
4.5	12.8	24.3
7.0	13.1	10.1
8.5	15.8	11.8
9.3	7.5	9.8
9.8	7.9	12.1
10.3	7.6	12.5
11.6	9.4	12.5
11.7	9.6	13.4
	11.3	16.1
	11.6	25.2
	11.8	
	12.6	
	13.2	
	16.3	

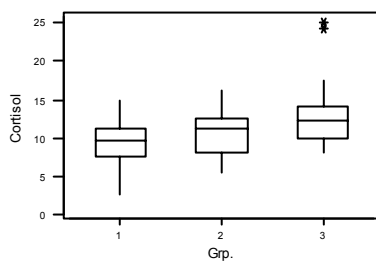
Variable	N	Mean	StDev	SE Mean
Healthy	16	9.200	2.931	0.733
Non-mel.	22	10.700	2.758	0.588
Melanch.	18	13.50	4.67	1.10

Univariate analysis has provided the above values for the means and standard deviations of cortisol levels in the three groups.



In the above plot cortisol levels tend to shift to the right (higher values) amongst the depressed groups, more so in the melancholic depressed. Are the mean differences significant?

Box plots showing the distribution of the data in the three groups



To answer the question of significance of difference between the means of the three groups we perform the following analysis.

[In MINITAB Stat → ANOVA | One-way]

#### Analysis of Variance

Source	DF	SS	MS	F	P
Factor	2	164.7	82.3	6.61	0.003
Error	53	660.0	12.5		
Total	55	824.7			

Level	N	Mean	StDev
Healthy	16	9.200	2.931
Non-mel.	22	10.700	2.758
Melanch.	18	13.500	4.674

Pooled StDev = 3.529

Individual 95% CIs For Mean  
Based on Pooled StDev

## Interpreting the output of One-way Analysis of Variance

The research question is “Do the three sample means differ any more than we would expect just from random variation?” In other words “Are the differences between the sample means greater than the amount of variation we would expect to see given the spread of values in each group?”

The output begins with the analysis of variance table. In the table the total sum of squares is broken down into two sources - the variation due to the factor (i.e. belonging to a particular group) and random variation due to the uniqueness of each individual. Each sum of squares has a certain number of degrees of freedom associated with it. They are used to calculate the mean sum of squares, then the  $F$  ratio and finally the  $P$  value of  $F$ . Here it is 0.003, and significant. This means that at least one of the means is significantly different from that of the control group.

The next table summarises the data separately for each group, and provides the mean, the standard deviation, and the confidence interval. A plot showing the three means and the spread of the 95% confidence interval is next displayed.

There is considerable overlap between the confidence intervals of mean cortisol levels in the healthy and non-melancholic depressed, but no overlap between those for the healthy and the melancholic depressed. We conclude that the difference between the means for the healthy and the melancholic depressed are significant at  $P < 0.05$ .

Different computer software carry different procedures for making multiple comparisons. The commonly employed procedures are FISHER'S, TUKEY, MCB and DUNNETT's tests. All provide confidence intervals for differences between group means. All of them calculate a set or family of confidence intervals. The family error rate is the maximum probability of obtaining one or more confidence intervals that do not contain the true difference between group means. The individual error rate is the probability that a given confidence interval will not contain the true difference in means between groups. If an error rate is not specified the default error rate is 0.05. TUKEY and FISHER provide confidence intervals for all pairwise differences between the group means. DUNNETT provides a confidence interval for the difference between each group (or treatment) mean and a control mean. Dunnett's procedure is particularly useful when one wishes to compare several outcomes with a control group (Group 1 in our example). Tukey, Dunnett and MCB methods have protection against false positive built in to them.

In an intervention study the researcher usually has a control group in mind against which the effectiveness of other interventions will be measured. In such a situation DUNNETT's test is the preferred choice. The control group must be declared for the purpose of the analysis. In observational studies, on the other hand, the researcher collects the data and then looks through the data to find significant differences, if they exist. In this latter situation an entire set of comparisons are carried out. In multiple comparisons one runs the risk of making false positive conclusions (Type I error), and the tests are designed to protect against such errors by being more strict, e.g. TUKEY's test. On rare occasions the researcher may wish to compare all the means against a specified “best”, the “best” being either the smallest or the largest. Here the MCB method is more suitable.



**Fisher's pairwise comparisons**

Family error rate = 0.121  
 Individual error rate = 0.0500

Critical value = 2.006

Intervals for (column level mean) - (row level mean)

	1	2
2	-3.826 0.826	
3	-6.732 -1.868	-5.050 -0.550

The output has printed 3 confidence intervals for **Tukey**. Results are presented as a set of confidence intervals for the difference between pairs of means. Each has an individual error rate of 0.0194. Those confidence intervals, which include zero, have to be ignored because they do not help us to reject the hypothesis that the column level mean is equal to the row level mean. The family error rate is 0.05. This means that using Tukey we have a 0.05 probability that at least one interval will not contain the true difference in means.

**Fisher** gives a printout similar to Tukey.

It should be clear by now the importance of taking the family error rate into account when making multiple comparisons because the probability of making a type I error for a series of comparisons is greater than the error rate for any one of individual comparisons alone. For example, in the above printout Fisher's method is used to make three comparisons choosing individual error rate of 0.05. The chance of making at least one type I error in 3 pairwise comparisons is 0.121.

Which of the methods depicted above should one use? The decision is guided by two principles:

- The difference between Tukey's method and Fisher's is which error rate one has selected. In Tukey's method select the family error rate and the individual error rates are adjusted accordingly. In the case of Fisher's method select the individual error rate.
- For a given family error rate Dunnett's method and Hsu method (not depicted here) are more likely to detect an effect (a difference) than Tukey's method. But these methods provide fewer comparisons, so we must decide beforehand what comparisons are required and choose the method accordingly.

In this first example the differences in mean cortisol levels between 3 groups of subjects have been counted as 'treatment effects', with the 'treatment factor' being the disease state.

Our next example is to demonstrate the use of One-way ANOVA for the analysis of data collected during a laboratory experiment.

#### **The effect of diet and drug on atherosclerosis.**

In an experiment to see whether changing diet can slow the progression of atherosclerosis or giving verapamil (a calcium channel blocking drug), groups of rabbits were administered different treatments as follows:

All rabbits received a high cholesterol diet for 12 weeks. They were then divided into treatment groups as follows:

1. **Immediate study.** They were to serve as controls. ( $n = 8$ ).
2. 12 further weeks of **high cholesterol diet.** ( $n = 5$ ).
3. 12 further weeks of **normal diet.** ( $n = 6$ ).
4. 12 further weeks of **normal diet + verapamil** ( $n = 6$ ).
5. 12 further weeks of **high cholesterol diet + verapamil.** ( $n=5$ ).

At the end of the experiment the animals were sacrificed and the percentages area of the aorta covered with atheromatous plaques were measured.

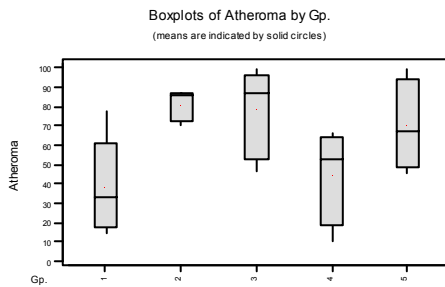
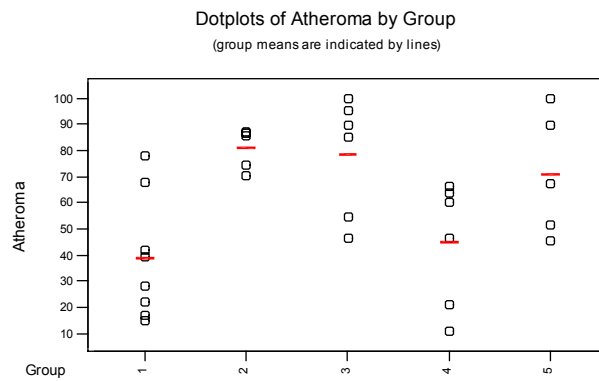
Thus the investigators were measuring the development of atheroma on the aorta of the experimental animals under controlled high cholesterol diet conditions and various combinations of high cholesterol or normal diet and the use of verapamil. The objective of the experiment was to determine whether normalizing the diet can halt the progression of atherosclerosis, or does giving verapamil after the atherosclerosis has started achieve the same purpose? (*Cardiovascular Drugs Therapy. 1987; 1: 65-69*).

The data are presented below, where Atheroma = Percent of aorta covered by atheromatous plaque.

Atheroma	Group
	1
22.00	1
14.9	1
16.8	1
68	1
78	1
42	1
39.4	1
28	1
85.7	2
87	2
74.6	2
86.7	2
70.6	2
46.4	3
100	3
89.6	3
95.4	3
85	3
54.8	3
66.4	4
63.9	4
46.5	4
10.8	4
21	4
60	4
45.7	5
67.1	5
89.6	5
100	5
51.7	5

The research question is: “Is there any difference in the extent of atheroma observed in the treatment groups compared to that seen after 12 weeks of high cholesterol diet (Group 1)?

We first plot the data to visualise the spread of the data points as is shown next:



There are clear differences between the means. Groups 2 and 3 show obviously high mean values, and Groups 4 and 5 slightly larger mean values compared to the control Group 1. How significant are these differences? We check by carrying out the ANOVA test as follows:

[In MINITAB Stat → ANOVA | One-Way ]

Analysis of Variance for Atheroma

Source	DF	SS	MS	F	P
Group	4	9876	2469	5.32	0.003
Error	25	11601	464		
Total	29	21477			

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
1	8	38.64	23.46	(-----*-----)
2	5	80.92	7.74	(-----*-----)
3	6	78.53	22.38	(-----*-----)
4	6	44.77	23.61	(-----*-----)
5	5	70.82	23.53	(-----*-----)

Pooled StDev = 21.54

25                      50                      75                      100

The confidence intervals of the means for Groups 2 and 3 do not overlap with the confidence interval of the mean for Group 1, indicating that the differences are significant. This is not the case with the confidence intervals of the means of Groups 4 and 5. This is borne out by the Dunnett's test below.

Dunnett's intervals for treatment mean minus control mean

Family error rate = 0.0500  
Individual error rate = 0.0144

Critical value = 2.63

Control = level (1) of Group

Level	Lower	Center	Upper	Dunnett's Interval
2	9.98	42.28	74.59	(-----*-----)
3	9.29	39.90	70.50	(-----*-----)
4	-24.48	6.13	36.73	(-----*-----)
5	-0.12	32.18	64.49	(-----*-----)

0                      30                      60

Dunnett's test shows that the intervals for treatment means minus control mean contain zero for Groups 4 and 5 (though borderline for group 5), which is not the case for Groups 2 and 3.

**Tukey's pairwise comparisons**

Family error rate = 0.0500  
 Individual error rate = 0.00706

Critical value = 4.15

Intervals for (column level mean) - (row level mean)

	1	2	3	4
2	-78.32 -6.24			
3	-74.04 -5.76	-35.89 40.66		
4	-40.27 28.01	-2.12 74.43	-2.73 70.26	
5	-68.22 3.86	-29.88 50.08	-30.56 45.99	-64.33 12.22

**Fisher's pairwise comparisons**

Family error rate = 0.268  
 Individual error rate = 0.0500

Critical value = 2.060

Intervals for (column level mean) - (row level mean)

	1	2	3	4
2	-67.58 -16.98			
3	-63.86 -15.93	-24.48 29.26		
4	-30.09 17.84	9.28 63.02	8.15 59.39	
5	-57.48 -6.88	-17.97 38.17	-19.16 34.58	-52.92 0.82

From this analysis it can be concluded that continuing with the high cholesterol diet for further 12 weeks makes the atheroma worse. A change to normal diet does not help much (Group3). Perhaps 12 weeks is a short time to see any appreciable change. Verapamil is effective, much more so if the diet is also changed to normal (Group 4), than high in cholesterol.

**Comment**

The F-test in ANOVA is testing the existence or otherwise of true differences between population means. If we do find a difference between the populations, then we know that a score on the dependent variable (cortisol level in the first example, and atheroma in the second) will be at least partly determined by group membership. But how much is “partly”? This can be determined by the extent to which the independent variable accounts for the variance of the dependent variable. It is calculated by the ratio of between group sum of squares and the total sum of squares. In the first example the ratio works out to be  $164.7/824.7 = 0.1997$  (or about 20%), and  $11601/21477 = 0.54$  (Or 54%) in the second example.

### Regression Approach to Analysis of Variance problems

ANOVA estimates the  $F$  statistic for testing the significance of relationships between a continuous  $Y$  variable and categorical  $X$  variable. The same  $F$  statistic is also obtained with regression analysis. In fact, ANOVA is a subset of regression methods and provides a simpler approach for certain types of analytical problems. Some authors believe that regression gives a clearer view of the estimates of the different parameters of a data set. In the case of ANOVA all the explanatory variables must be treated as categorical. In the case of regression analysis any mixture of variable types is allowed.

In applying the method of regression for an ANOVA problem the categorical explanatory variable with say  $k$  categories needs to be transformed into  $k-1$  dummy variables. All computer programmes carry a set of commands to perform the transformation of a categorical variable into dummy (or indicator) variables. For example, if we were to compare several treatments then 'Treatments' comprise the factors. If we were comparing  $k$  such treatments then each is included in a regression model by defining  $k-1$  dummy (indicator) variables. The treatment left out is the reference (or control) treatment with which all other treatments are to be compared.

Let us now see how this technique works with the "Depression" data set (first example). There are 3 groups, and so we need to create 2 dummy variables  $X_1$  and  $X_2$ , such that  $X_1 = 1$  if the group is 'non-melancholic depressed' and  $X_1 = 0$  otherwise;  $X_2 = 1$  if the group is 'melancholic depressed' and  $X_2 = 0$  otherwise. The group that is left out viz. 'healthy' becomes the reference group.

The arrangement of the data file is shown below.

Corti.	Grp.	$X_1$			$X_2$
		Healthy	Non-mela.	Melanch.	Melanch.
2.5	1	1	0	0	
7.2	1	1	0	0	
8.9	1	1	0	0	
9.3	1	1	0	0	
9.9	1	1	0	0	
10.3	1	1	0	0	
11.6	1	1	0	0	
14.8	1	1	0	0	
4.5	1	1	0	0	
7.0	1	1	0	0	
8.5	1	1	0	0	
9.3	1	1	0	0	
9.8	1	1	0	0	
10.3	1	1	0	0	
11.6	1	1	0	0	
11.7	1	1	0	0	
5.4	2	0	1	0	
7.8	2	0	1	0	
8.0	2	0	1	0	
9.3	2	0	1	0	
9.7	2	0	1	0	
11.1	2	0	1	0	
11.6	2	0	1	0	
12.0	2	0	1	0	
12.8	2	0	1	0	
13.1	2	0	1	0	
15.8	2	0	1	0	
7.5	2	0	1	0	
7.9	2	0	1	0	
7.6	2	0	1	0	
9.4	2	0	1	0	
9.6	2	0	1	0	
11.3	2	0	1	0	

		0	1	0
Corti	Grp	Healthy	Non-mel.	Melan.
11.6	2	0	1	0
11.8	2	0	1	0
12.6	2	0	1	0
13.2	2	0	1	0
16.3	2	0	1	0
8.1	3	0	0	1
9.5	3	0	0	1
9.8	3	0	0	1
12.2	3	0	0	1
12.3	3	0	0	1
12.5	3	0	0	1
13.3	3	0	0	1
17.5	3	0	0	1
24.3	3	0	0	1
10.1	3	0	0	1
11.8	3	0	0	1
9.8	3	0	0	1
12.1	3	0	0	1
12.5	3	0	0	1
12.5	3	0	0	1
13.4	3	0	0	1
16.1	3	0	0	1
25.2	3	0	0	1

The results of the regression analysis next follow.

The regression equation is  
 Cortisol = 9.20 + 1.50 Non-mela. + 4.30 Melanch.

Predictor	Coef	StDev	T	P	VIF
Constant	9.2000	0.8822	10.43	0.000	
Non-mela	1.500	1.159	1.29	0.201	1.4
Melanch.	4.300	1.213	3.55	0.001	1.4

S = 3.529      R-Sq = 20.0%      R-Sq(adj) = 16.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	164.67	82.34	6.61	0.003
Residual Error	53	660.02	12.45		
Total	55	824.69			

Source	DF	Seq SS
Non-mela	1	8.05
Melanch.	1	156.62

Unusual Observations

Obs	Non-mela	Cortisol	Fit	StDev Fit	Residual	St Resid
47	0.00	24.300	13.500	0.832	10.800	3.15R
56	0.00	25.200	13.500	0.832	11.700	3.41R

R denotes an observation with a large standardized residual

Durbin-Watson statistic = 1.24

In the above data set the variable ‘Group’ has 3 levels. The lowest level (healthy) is left out in making the indicator variables. Level 2 (non-melancholic) has been assigned indicator variable  $X_1$ . Level 3 (melancholic) has been assigned indicator variable  $X_2$ . This is the pattern used by most computer programmes.

The regression equation may be written as  
 $cortisol = 9.20 + 1.50X_1 + 4.30X_2$

In the case of members of the ‘healthy’ group  $X_1 = 0$  and  $X_2 = 0$ , and so cortisol = 9.20, which is the mean value of cortisol for Group 1 (healthy).

For members of the non-melancholic depressed group  $X_1 = 1$  and  $X_2 = 0$ . Hence cortisol =  $9.20 + 1.50 = 10.70$ , which is the mean value for Group 2.

Similarly, the mean value of cortisol for Group 3 (melancholic depressed) works out as  $9.20 + 4.30 = 13.50$

These values are the same as obtained with ANOVA. The analysis of variance table is also the same. A significant  $F$  statistic indicates that at least one of the means is significantly different from the base line mean in the healthy group.

Regression analysis provides all of the ANOVA results, and also additional information in the form of coefficients, standard errors, and  $t$  tests. Each  $\beta$  coefficient adds a value relevant to the group to the base line mean value of group 1 (healthy). We are also informed that the type of depression accounts for 17 per cent of the variance in cortisol values. The  $t$ -ratio of the melancholic group at 3.55 has a significant  $P$  value of 0.001, but this is not the case for the melancholic group.

We next observe how these methods and techniques extend to 5 groups in the second example. The data set with the indicator variables is provided below. Group 1 is taken as the reference group, and 4 indicator variables are created such that

$X_1 = 1$  if group is 2 and 0 otherwise

$X_2 = 1$  if group is 3 and 0 if otherwise

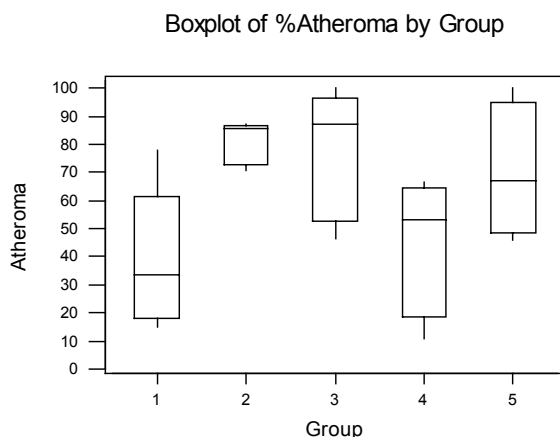
$X_3 = 1$  if group is 4 and 0 if otherwise

$X_4 = 1$  if group is 5 and 0 otherwise

(all the 4 indicator variables would be equal to zero if the group is 1).

	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	
%Ather.	Gp2	Gp3	Gp4	Gp5	Gp.
22.0	0	0	0	0	1
14.9	0	0	0	0	1
16.8	0	0	0	0	1
68.0	0	0	0	0	1
78.0	0	0	0	0	1
42.0	0	0	0	0	1
39.4	0	0	0	0	1
28.0	0	0	0	0	1
85.7	1	0	0	0	2
87.0	1	0	0	0	2
74.6	1	0	0	0	2
86.7	1	0	0	0	2
70.6	1	0	0	0	2
46.4	0	1	0	0	3
100.0	0	1	0	0	3
89.6	0	1	0	0	3
95.4	0	1	0	0	3
85.0	0	1	0	0	3
54.8	0	1	0	0	3
66.4	0	0	1	0	4
63.9	0	0	1	0	4
46.5	0	0	1	0	4
10.8	0	0	1	0	4
21.0	0	0	1	0	4
60.0	0	0	1	0	4
45.7	0	0	0	1	5
67.1	0	0	0	1	5
89.6	0	0	0	1	5
100.0	0	0	0	1	5
51.7	0	0	0	1	5

A boxplot of the data and results of regression analysis next follow



The boxplot indicates that the mean atheroma per cent are different for groups 2 to 5 from those for Group 1.

How significant are the differences between these means?

The regression equation is

$$\text{Atheroma} = 38.6 + 42.3 \text{ Group2} + 39.9 \text{ Group3} + 6.1 \text{ Group4} + 32.2 \text{ Group5}$$

Predictor	Coef	StDev	T	P	VIF
Constant	38.638	7.616	5.07	0.000	
Group2	42.28	12.28	3.44	0.002	1.4
Group3	39.90	11.63	3.43	0.002	1.4
Group4	6.13	11.63	0.53	0.603	1.4
Group5	32.18	12.28	2.62	0.015	1.4

S = 21.54      R-Sq = 46.0%      R-Sq(adj) = 37.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	9875.8	2468.9	5.32	0.003
Residual Error	25	11601.2	464.0		
Total	29	21476.9			

Source	DF	Seq SS
Group2	1	2562.7
Group3	1	3966.0
Group4	1	160.3
Group5	1	3186.8

Durbin-Watson statistic = 1.71

The  $F$  value for the regression model is similar to that obtained with One-way ANOVA. It is significant at  $P = 0.003$ . We thereby conclude that there is at least one mean, which is significantly different.

We next turn our attention to the table of coefficients. The value for the constant (Group 1 – 12 weeks on high cholesterol diet) is 38.638, which indicates that the mean plaque coverage with atheroma is about 39%. The  $\beta$  value for Group 2 is 42.28, which means that the mean plaque coverage for Group 2 (12 weeks + 12 weeks on high cholesterol diet) is  $38.638 + 42.28 = 80.9$  or about 81 per cent, and so on.

The  $t$ -ratios in the case of Groups 2, 3 (12 weeks on high cholesterol + 12 weeks on normal diet) and 5 (12 weeks + 12 weeks on high cholesterol + verapamil) have significant  $P$  values. The  $R$ -sq value of 46% indicates that the various experimental interventions account for the extent of atheroma formation that has been observed in the experiment.

We can conclude from these results that high cholesterol diet causes atheroma to develop on the aorta of the experimental subjects. Switching to a normal diet does not help it much, but the addition of verapamil halts the process. If the high cholesterol diet is continued the benefits of verapamil are not so obvious. A normal diet and verapamil are more effective in the control of atheroma.

### **Comment**

In this chapter we first considered the details of One-Way Analysis of Variance, which is an extension of the two sample  $t$  test for comparison of means to 2 + n samples. Like the  $t$  test ANOVA is also built on similar assumptions, as described on page 58.

ANOVA is a robust procedure and would be affected only when there are serious violations of the assumptions. Such problems arise when the sample sizes are too small, and one should aim for at least 20 subjects per group, and preferably more. The main difficulty with small sample sizes arises from subjects with unusual values, and checking for outliers before performing the analysis is always beneficial.

If there is evidence that the values are not normally distributed then one chooses to perform either the log or other transformation, or employ the non-parametric Kruskal-Wallis test instead.

With regard to equality of variances, if the number of subjects in each group is equal or very similar, the problem does not arise. The more uneven the group sizes the more is the likelihood of any inequality in group variance becoming a problem.

In the multiple linear regression approach we saw the application of the principle of additivity. The constant ( $\beta_0$ ) represents the overall mean response. Then belonging to a particular treatment group adds a contribution as represented by the  $\beta$  coefficient for that group.