STATISTICS FOR ECONOMISTS: A BEGINNING

John E. Floyd
University of Toronto

July 2, 2010

PREFACE

The pages that follow contain the material presented in my introductory quantitative methods in economics class at the University of Toronto. They are designed to be used along with any reasonable statistics textbook. The most recent textbook for the course was James T. McClave, P. George Benson and Terry Sincich, *Statistics for Business and Economics*, Eighth Edition, Prentice Hall, 2001. The material draws upon earlier editions of that book as well as upon John Neter, William Wasserman and G. A. Whitmore, *Applied Statistics*, Fourth Edition, Allyn and Bacon, 1993, which was used previously and is now out of print. It is also consistent with Gerald Keller and Brian Warrack, *Statistics for Management and Economics*, Fifth Edition, Duxbury, 2000, which is the textbook used recently on the St. George Campus of the University of Toronto. The problems at the ends of the chapters are questions from mid-term and final exams at both the St. George and Mississauga campuses of the University of Toronto. They were set by Gordon Anderson, Lee Bailey, Greg Jump, Victor Yu and others including myself.

This manuscript should be useful for economics and business students enrolled in basic courses in statistics and, as well, for people who have studied statistics some time ago and need a review of what they are supposed to have learned. Indeed, one could learn statistics from scratch using this material alone, although those trying to do so may find the presentation somewhat compact, requiring slow and careful reading and thought as one goes along.

I would like to thank the above mentioned colleagues and, in addition, Adonis Yatchew, for helpful discussions over the years, and John Maheu for helping me clarify a number of points. I would especially like to thank Gordon Anderson, who I have bothered so frequently with questions that he deserves the status of mentor.

After the original version of this manuscript was completed, I received some detailed comments on Chapter 8 from Peter Westfall of Texas Tech University, enabling me to correct a number of errors. Such comments are much appreciated.

J. E. Floyd
July 2, 2010

Chapter 10

Analysis of Variance

Analysis of variance (ANOVA) models study the relationship between a dependent variable and one or more independent variables within the same framework as do linear regression models but from a different perspective. We begin by viewing from an ANOVA perspective the results of a regression explaining the response of Canadian real money holdings to Canadian real GNP and the interest rate on Canadian 90-day commercial paper.

10.1 Regression Results in an ANOVA Framework

The regression results were as follows:

Dependent Variable: Canadian Real Money Holdings

Constant	10.47	(3.21)
90-Day Paper Rate	-2.62	(0.38)
Real GNP	0.17	(0.01)
R-Squared	.91	
Standard Error ($\hat{\sigma}$)	6.70	
Number of Observations	40	
Degrees of Freedom	37	

The regression model can be seen as attempting to explain the total sum of squares of the dependent variable, real money holdings, using two independent variables, real GNP and the nominal interest rate. The residual sum of squares *SSE* represents the portion of the total sum of squares *SSTO* that cannot be explained by the independent variables. And the sum of squares due to the regression *SSR* represented the portion of the total sum of squares explained by the regressors. It will be recalled that the R^2 is the ratio of *SSR* to *SSTO*. The regression results above give the standard error of the regression $\hat{\sigma}$ which is a point estimate of σ—it is the square root of the mean square error *MSE*. The mean square error in the regression above is thus

$$MSE = \hat{\sigma}^2 = 6.70^2 = 44.89$$

so the sum of squared errors is

$$SSE = (n - K - 1)\, MSE = (37)(44.89) = 1660.93.$$

Since the coefficient of determination, R^2, equals

$$R^2 = \frac{SSR}{SSTO} = \frac{SSTO - SSE}{SSTO} = 1 - \frac{SSE}{SSTO}$$

it follows that

$$R^2 SSTO = SSTO - SSE \quad \Longrightarrow \quad (1 - R^2)\, SSTO = SSE,$$

so that, given R^2 and *SSE*, we can calculate *SSTO* from the relationship

$$SSTO = \frac{SSE}{1 - R^2} = \frac{1660.93}{1 - .91} = \frac{1660.93}{.09} = 18454.78.$$

The sum of squares due to regression then becomes

$$SSR = SSTO - SSE = 18454.78 - 1660.93 = 16793.85.$$

Now the variance of the dependent variable, real money holdings, is the total sum of squares divided by $(n - 1)$, the degrees of freedom relevant for calculating it—one observation out of the n available is used up calculating the mean of the dependent variable. And we have seen that the error variance is estimated by dividing the sum of squared errors by $(n - K - 1)$, the number of degrees of freedom relevant for its calculation—here we have used up K pieces of information calculating the regression coefficients of the independent variables and one piece of information to calculate the constant term, leaving only $(n - K - 1)$ independent squared residuals.

Finally, we can identify the degrees of freedom used in calculating the sum of squares due to regression (*SSR*). *SSR* is the sum of squared deviations of the fitted values of the dependent variable from the mean of the dependent variable—in terms of our regression notation,

$$SSR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2.$$

There are n fitted values $\hat{Y}$ that by the nature of the calculations are constrained to lie along the fitted line. The potential degrees of freedom in calculating this line are its K + 1 parameters—the slopes with respect to the K independent variables, and the intercept. One of these degrees of freedom is lost because only $n-1$ of the $(\hat{Y}_i - \bar{Y})$ are independent—the deviations must satisfy the constraint

$$\sum_{1=1}^{n}(\hat{Y}_i - \bar{Y}) = 0$$

so if we know any $n-1$ of these deviations we also know the remaining deviation. The sum of squares due to regression is thus calculated with K degrees of freedom (two in the above example). So we can calculate the variance due to the regression (i.e., the regression mean square) as

$$MSR = \frac{SSR}{K} = \frac{16793.85}{2} = 8396.925.$$

These analysis of variance results can be set out in the following ANOVA table:

Analysis of Variance: Canadian Real Money Holdings

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square
Regression	16793.85	2	8396.925
Error	1660.93	37	44.89
Total	18454.78	39	

Notice how the total degrees of freedom is the sum of the degrees of freedom for calculating the regression sum of squares and the degrees of freedom for calculating the sum of squared errors. And, as shown in the previous two chapters as well as above, the total sum of squares is equal to the sum of squares due to regression plus the error sum of squares. It is especially important to notice, however, that the mean square due to regression and the mean square error do not add up to equal the variance of the dependent variable, which in the case above is $18454.78/39 = 473.2$. The F-Statistic for testing the null hypothesis of no relationship between the regressors and the dependent variable is

$$\begin{aligned} F &= \frac{\sum(Y_i - \hat{Y})^2 - \sum e_i^2}{K} \div \frac{\sum e_i^2}{n - K - 1} \\ &= \frac{SST0 - SSE}{K} \div \frac{SSE}{n - K - 1} \\ &= \frac{MSR}{MSE} = \frac{8396.925}{44.89} = 185.72 \end{aligned}$$

which far exceeds the value of $F(2, 37)$ in the statistical tables for at any reasonable level of α.

10.2 Single-Factor Analysis of Variance

Let us now take a fresh problem and approach it strictly from an ANOVA perspective. Suppose we randomly select 5 male students and 5 female students from a large class and give each student an achievement test. Our objective is to investigate whether male students do better than their female counterparts on such a test. The resulting data are

	Student i				
Gender j	1	2	3	4	5
Male	86	82	94	77	86
Female	89	75	97	80	82

This is a *designed sampling experiment* because we control (and randomize) the selection of male and female participants. It would be an *observational sampling experiment* if we were to simply take a class of 10 students, half of whom turn out to be female, and give them an achievement test.

Analysis of variance has its own terminology. The achievement test score is the response or dependent variable as it would be in a linear regression. The independent variables, whose effects on the response variable we are interested in determining, are called *factors*. In the case at hand, there is a single factor, gender, and it is qualitative—i.e., not measured naturally on a numerical scale. We could add additional factors such as, say, the race of the student. The values of the factors utilized in the experiment are called *factor levels*. In this single factor experiment, we have two factor levels, male and female. In the single factor case the factor levels are also called *treatments*. In an experiment with more than one factor, the treatments are the factor-level combinations utilized. For example, if we take the race of the students as a second factor, the treatments might be male-white, female-white, male-non-white and female-non-white. The objects on which the response variables are observed—i.e., the individual students in the case considered here—are referred to as *experimental units.* These are called elements in regression analysis.

The objective of a completely randomized design is usually to compare the treatment means—these are the mean achievement scores of male and female students respectively. The means of the two treatments (male and female) are, respectively,

$$\frac{86 + 82 + 94 + 77 + 86}{5} = 85$$

and

$$\frac{89 + 75 + 97 + 80 + 82)}{5} = 84.6$$

and the overall mean is 84.8. Some thought suggests that if the response variable (achievement test score) is not much affected by treatment (i.e., by whether the student is male or female) the means for the two treatments will not differ very much as compared to the variability of the achievement test scores around their treatment means. On the other hand, if test score responds to gender, there should be a large degree of variability of the treatment means around their common mean as compared to the variability of the within-group test scores around their treatment means.

We thus calculate the *Sum of Squares for Treatments* by squaring the distance between each treatment mean and the overall mean of all sample measurements, multiplying each squared difference by the number of sample measurements for the treatment, and adding the results over all treatments. This yields

$$SST = \sum_{j=1}^{p} (n_j)(\bar{x}_j - \bar{x})^2 = (5)(85 - 84.8)^2 + (5)(84.6 - 84.8)^2$$

$$= (5)(.04) + (5)(.04) = .2 + .2 = .4.$$

In the above expression $p = 2$ is the number of treatments, n_j is the number of sample elements receiving the j-th treatment, $\bar{x}_j$ is the mean response for the jth treatment and $\bar{x}$ is the mean response for the entire sample.

Next we calculate the *Sum of Squares for Error*, which measures the sampling variability within the treatments—that is, the variability around the treatment means, which is attributed to sampling error. This is computed by summing the squared distance between each response measurement and the corresponding treatment mean and then adding these sums of squared differences for all (both) treatments. This yields

$$SSE = \sum_{i=1}^{n_1}(x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^{n_2}(x_{i2} - \bar{x}_2)^2$$

$$= [(86 - 85)^2 + (82 - 85)^2 + (94 - 85)^2 + (77 - 85)^2 + (86 - 85)^2]$$

$$+[(89 - 84.6)^2 + (75 - 84.6)^2 + (97 - 84.6)^2 + (80 - 84.6)^2 + (82 - 84.6)^2]$$

$$= [1 + 9 + 81 + 64 + 1] + [19.36 + 92.16 + 153.76 + 21.16 + 6.76]$$

$$= 156 + 293.2 = 449.2.$$

Again, n_j is the number of sample measurements for the jth treatment and x_{ij} is the ith measurement for the jth treatment.

Finally, the *Total Sum of Squares* is the sum of squares for treatments plus the sum of squares for error. That is

$$SSTO = SST + SSE = .4 + 449.2 = 449.6.$$

Now we calculate the *Mean Square for Treatments* which equals the sum of squares for treatments divided by the appropriate degrees of freedom. We are summing p squared deviations (of each of the p treatment means from the overall mean) but only $p-1$ of these squared deviations are independent because we lose one piece of information in calculating the overall mean. So for the above example we have

$$MST = \frac{SST}{p-1} = \frac{0.4}{1} = 0.4.$$

Next we calculate the *Mean Square Error* which equals the sum of the squared deviations of the sample measurements from their respective treatment means for all measurements, again divided by the appropriate degrees of freedom. Here we have n cases (or sample measurements), where

$$n = n_1 + n_2 + n_3 + \ldots + n_p$$

but we had to calculate the p treatment means from the data, so the degrees of freedom will be $n - p$. We thus obtain

$$MSE = \frac{SSE}{n-p} = \frac{SSTO - SST}{n-p} = \frac{449.2}{10-2} = 56.15.$$

The above numbers can be used to construct the following ANOVA table:

Analysis of Variance: Achievement Test Scores

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square
Treatments	0.4	1	0.4
Error	449.2	8	56.15
Total	449.6	9	

The purpose of this whole exercise, of course, is to determine whether gender (given by treatments) has any effect on achievement test scores (the response variable). If there is no effect we would expect the error sum of squares to be nearly as big as the total sum of squares and the treatment sum of squares to be very small. This appears to be the case in the ANOVA table above. The sum of squares for treatments (which measures the variability of the treatment means around the overall mean) is extremely low relative to the error sum of squares. But is it low enough for us to conclude that there is no significant relationship of achievement scores to gender? Is the observed treatment sum of squares as high as it is purely because of sampling error?

The statistical test for significance is straight forward. From the discussions in the previous chapter it is evident that under the null hypothesis of no relationship

$$\frac{SST}{\sigma^2} = \frac{SSTO - SSE}{\sigma^2} = \chi^2(p-1)$$

where $(p-1)[= (n-1) - (n-p)]$ is the degrees of freedom for treatment and σ^2 is the common variance of the individual achievement scores around the overall mean achievement score and of the individual scores around their treatment means. The two types of variation have a common variance under the null hypotheses that the achievement test scores are independent of treatment. Also,

$$\frac{SSE}{\sigma^2} = \chi^2(n-p).$$

We can now apply the principle that the ratio of two independent χ^2 variables, each divided by its degrees of freedom, will be distributed according to the F-distribution with parameters equal to the number of degrees of freedom in the numerator and number of degrees of freedom in the denominator.

Thus we have

$$\frac{SSTO - SSE}{(n-1)-(n-p)} \div \frac{SSE}{n-p} = \frac{SST}{p-1} \div \frac{SSE}{n-p} = \frac{MST}{MSE} = F(p-1, n-p)$$

where the σ^2 terms cancel out. In the example under consideration, this yields

$$\frac{MST}{MSE} = \frac{.4}{56.15} = .007123778 = F(1, 8).$$

The critical value of F with one degree of freedom in the numerator and 8 degrees of freedom in the denominator for $\alpha = .1$ is 3.46. So we cannot reject the null hypothesis of no effect of gender on achievement test scores.

You might recognize the similarity of this analysis of variance test to the tests we did in Chapter 6 for differences in the means of two populations. Indeed, the tests are identical. In Chapter 6 we expressed the difference between the two population means as

$$E\{\bar{Y} - \bar{X}\} = E\{\bar{Y}\} - E\{\bar{X}\} = \mu_2 - \mu_1$$

and the variance of the difference between the two means as

$$\sigma^2\{\bar{Y} - \bar{X}\} = \sigma^2\{\bar{Y}\} + \sigma^2\{\bar{X}\},$$

using

$$s^2\{\bar{Y} - \bar{X}\} = s^2\{\bar{Y}\} + s^2\{\bar{X}\}$$

as an unbiased point estimator of $\sigma^2\{\bar{Y} - \bar{X}\}$. We then used in this formula the expressions for the variances of the means,

$$s^2\{\bar{Y}\} = s^2\{Y/n\}$$

and

$$s^2\{\bar{X}\} = s^2\{X/n\}.$$

The difference in means in the case above is $85 - 84.6 = 0.4$. The sample population variances can be obtained by noting that the sums of the squared deviations of the achievement scores of the male and female students around their respective means are, respectively, 156 and 293.2. Dividing each of these by the degrees of freedom relevant for their calculation ($n_i - 1 = 5 - 1 = 4$), we obtain sample population variances for male and female students of 39 and 73.3 respectively. Imposing the condition that the true variances of the two groups are the same, we then obtain a pooled estimator of this common variance by calculating a weighted average of the two estimated variances with the weights being the ratios of their respective degrees of freedom to the total. That is

$$s_P^2 = \frac{(4)(39) + (4)(73.3)}{8} = \frac{156 + 293.2}{8} = 56.15$$

which, you will note, equals *MSE*. The variance of the difference between the two means (which we denote using the subscripts m for male and f for female) equals

$$\sigma_{m-f}^2 = \frac{\sigma_m^2}{n_m} + \frac{\sigma_f^2}{n_f} = s_P^2 \left[\frac{1}{n_m} + \frac{1}{n_f}\right] = (56.15)(.2 + .2) = 22.46.$$

The standard deviation of the difference between the two means then equals $\sqrt{22.46} = 4.739198$. Given the point estimate of the difference in the means of 0.4, the t-statistic for testing the null-hypothesis of no difference between the means is

$$t^* = \frac{.4}{47392} = .08440246.$$

This statistic will be within the acceptance region for any reasonable level of significance. The result is the same as we obtained from the analysis of variance.

As a matter of fact, this test and the analysis of variance test are identical. Squaring t^*, we obtain .007123778 which equals the F-statistic obtained in the analysis of variance procedure. This is consistent with the principle, already noted, that when there is one degree of freedom in the numerator, $F = t^2$.

A third way of approaching this same problem is from the point of view of regression analysis. We have $n = 10$ observations on gender and want to determine the response of achievement test score to gender. Gender is a qualitative variable which we can introduce as a dummy variable taking a value of 0 for elements that are male and 1 for elements that are female. Our regression model becomes

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

where Y_i , $i = 1 \ldots 10$, is the test score for the i-th student, and X_i is the dummy variable taking a value of zero for male students and unity for female students. The regression results obtained are:

Dependent Variable: Achievement Test Score

Constant	85	(3.35112)
Female Dummy	-.40	(4.73920)

R-Squared	.000889
Standard Error ($\hat{\sigma}$)	7.4933
Number of Observations	10
Degrees of Freedom	8

The dummy variable for female indicates that the 'constant term for females' is $85 - 0.4 = 84.6$, which is the treatment mean for females obtained by the analysis of variance procedure. The $t-$ratio for the hypothesis that the female dummy is zero (i.e., the female treatment mean equals the male treatment mean) is -.4/473920, which is the same as was obtained for the above test for difference between the means. And the square of $\hat{\sigma}$ is 56.15, the mean squared error obtained in the analysis of variance procedure.

Now let us take a more complicated problem. Suppose we randomly divide fifteen male students enrolled in a mathematics course into three groups of five students each. We then randomly assign each group to one of three instructional modes: (1) programmed text, (2) video-taped lecture-style presentation, and (3) interactive computer programs. These modes are all designed to augment a standard textbook which is the same for all three groups. At the end of the course, we give all students the same achievement test, with the following results:

	Student i				
Mode j	1	2	3	4	5
1	86	82	94	77	86
2	90	79	88	87	96
3	78	70	65	74	63

Again we have a *designed sampling experiment* because we were able to control the details of the instructional modes for the three groups and make sure that students were randomly assigned to groups and groups were randomly assigned to instructional modes. The experiment is completely *randomized* because the allocation of the students to the three groups is random and the allocation of the groups to the instructional modes is random. In contrast, an *observational sampling experiment* would be one where we, for example, observe the test scores of three groups of students, perhaps of different sizes, who for reasons beyond our control happen to have been instructed in accordance with three alternative instructional modes of the above types. In this single factor study there are three factor levels or treatments representing the three modes of instruction.

Our objective in this completely randomized design is to compare the treatment means—the mean achievement scores of the students in the three groups taught using the different instructional modes. The means of the three modes are

$$\frac{86+82+94+77+86}{5} = 85$$
$$\frac{90+79+88+87+96}{5} = 88$$
$$\frac{78+70+65+74+63}{5} = 70$$

And the overall mean is 81. Again we note that if the response variable (achievement test score) is not much affected by treatment (instructional mode) the means for the three treatments will not differ very much as compared to the variability of the achievement test scores around their treatment means. On the other hand, if test score responds to instructional mode, there should be a large degree of variability of the treatment means around their common mean as compared to the variability of the within-group test scores around their treatment means.

We again calculate the *Sum of Squares for Treatments* by squaring the distance between each treatment mean and the overall mean of all sample measurements, multiplying each squared distance by the number of sample measurements for the treatment, and adding the results over all treatments.

$$SST = \sum_{j=1}^{p} n_j\,(\bar{x}_j - \bar{x})^2 = (5)(85-81)^2 + (5)(88-81)^2 + (5)(70-81)^2$$

$$= (5)(16) + (5)(49) + (5)(121) = 80 + 245 + 605 = 930.$$

In the above expression $p = 3$ is the number of treatments, $\bar{x}_j$ is the mean response for the jth treatment and $\bar{x}$ is the mean response for the entire sample.

Next we calculate the *Sum of Squares for Error*, which measures the sampling variability within the treatments—the variability around the treatment means that we attribute to sampling error. This is computed by summing the squared distance between each response measurement and the corresponding treatment mean and then adding the squared differences over all measurements in the entire sample.

$$SSE = \sum_{i=1}^{n_1}(x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^{n_2}(x_{i2} - \bar{x}_2)^2 + \sum_{i=1}^{n_3}(x_{i3} - \bar{x}_3)^2$$

$$= (86-85)^2 + (82-85)^2 + (94-85)^2 + (77-85)^2 + (86-85)^2$$
$$+(90-88)^2 + (79-88)^2 + (88-88)^2 + (87-88)^2 + (96-88)^2$$
$$+(78-70)^2 + (70-70)^2 + (65-70)^2 + (74-70)^2 + (63-70)^2$$
$$= [1+9+81+64+1] + [4+81+0+1+64] + [64+0+25+16+49]$$
$$= 156 + 150 + 154 = 460.$$

Again, n_j is the number of sample measurements for the jth treatment, which turns out to be 5 for all treatments, and x_{ij} is the ith measurement for the jth treatment.

Finally, the *Total Sum of Squares*, which equals the sum of squares for treatments plus the sum of squares for error, is

$$SSTO = SST + SSE = 930 + 460 = 1390.$$

Now we calculate the *Mean Square for Treatments* which equals the sum of squares for treatments divided by the appropriate degrees of freedom. We are summing 3 squared deviations from the overall mean but only 2 of these squared deviations are independent because we lose one piece of information in calculating the overall mean. So we have

$$MST = \frac{SST}{p-1} = \frac{930}{2} = 465.$$

Finally, we calculate the *Mean Square Error* which equals the sum of the squared deviations of the sample measurements from their respective treatment means for all measurements, again divided by the appropriate degrees of freedom. Here we have 15 cases (or sample measurements), but we had to calculate the 3 treatment means from the data, so the degrees of freedom will be 12. We thus obtain

$$MSE = \frac{SSE}{n-p} - \frac{460}{12} = 38.333.$$

The above numbers can be used to construct the following ANOVA table:

Analysis of Variance: Achievement Test Scores

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square
Treatments	930	2	465
Error	430	12	38.33
Total	1390	14	

Our goal is to determine whether mode of instruction (given by treatments) has any effect on achievement test score (the response variable). If there is no effect we would expect the error sum of squares to be nearly as big as the total sum of squares and the treatment sum of squares to be very small. It turns out that the sum of squares for treatments (which measures the variability of the treatment means around the overall mean) is quite high relative to the error sum of squares. But is it high enough for us to conclude that there is a significant response of achievement scores to instructional mode? We answer this question by doing an F-test. The F-statistic obtained is

$$\frac{MST}{MSE} = \frac{465}{38.33} = 12.13 = F(2,12),$$

which is well above the critical value of 6.93 for $\alpha = .01$. We reject the null hypothesis of no effect of instruction mode on achievement test score.

The natural question to ask at this point is: Which of the instructional modes are responsible for the significant overall relationship? All our analysis of variance results tell us is that there is a significant effect of at least one of the three modes of instruction, compared to the other two, on achievement test score. We have not established the relative importance of these modes in determining students' achievement test scores. To investigate this, we can approach the problem from the point of view of regression analysis.

The dependent variable for our regression is achievement test score in a sample of 15 students. Taking the programmed text instructional mode as a reference, we create two dummy variables—one that takes a value of 1 when the instructional mode is video-taped lecture and zero otherwise, and a second that takes a value of 1 when the mode is interactive computer programs and zero otherwise. The effect of programmed text, the reference treatment, is thus measured by the constant terms and the differences in the effects of the other two treatments from the reference treatment are measured by the coefficients of their respective dummy variables. Our regression model is therefore

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

where Y_i , $i = 1 \ldots 15$, is the test score for the i-th student, and X_{1i} is the dummy variable for video-taped lecture and X_{2i} is the dummy variable for computer programs. The regression results obtained are:

Dependent Variable: Achievement Test Score

Constant	85	(2.77)
Dummy-video	3	(3.92)
Dummy-computer	-15	(3.92)
R-Squared	.67	
Standard Error ($\hat{\sigma}$)	6.19139	
Number of Observations	15	
Degrees of Freedom	12	

The mean score for students using programmed text is equal to the constant term, 85. And the mean score for students receiving video-taped lectures is 3 points higher than that for students using programmed text—i.e., 85 + 3 = 88. Finally, the mean score for students using computer programs is 15 points less than those using programmed text—i.e., 85 - 15 = 70. These correspond to the means calculated earlier. The t-statistic for testing the null hypothesis of no difference between the means for programmed text and video-taped lectures—that is $\beta_1 = 0$—is

$$t^* = \frac{3}{3.92} = .765,$$

which is well with any reasonable acceptance region. So we cannot reject the null hypothesis of no difference between the means for programmed text and video-taped lecture. The t-statistic for testing the null hypothesis of no difference between computer program and programmed text is

$$t^* = \frac{-15}{3.92} = -3.83,$$

leading us to conclude that mean test score under computer programmed learning is significantly below that of programmed text—the critical value of $t(12)$ for $\alpha = .005$ is 3.055.

The question arises as to whether there is a significant difference between the test scores under video-taped lecture vs. computer programmed learning. This would seem to be the case. To check this out we rerun the regression letting video-taped lecture be the reference—that is, including dummies for programmed text and computer program but no dummy variable for video-taped lecture. This yields

Dependent Variable: Achievement Test Score

Constant	88	(2.77)
Dummy-text	-3	(3.92)
Dummy-computer	-18	(3.92)
R-Squared	.67	
Standard Error ($\hat{\sigma}$)	6.19139	
Number of Observations	15	
Degrees of Freedom	12	

The computer program dummy is clearly statistically significant, having a $t-$statistic of -4.59. We have to reject the null hypothesis of no difference between the mean test scores under video-taped lecture and computer programmed learning.

Notice how the difference between the coefficient of Dummy-video and Dummy-computer in the regression that uses programmed text as the reference treatment is exactly the same as the coefficient of Dummy-computer in the regression that uses video-taped lectures as the reference treatment, and that the standard errors of the dummy coefficients are the same in both regressions. It would appear that instead of running the second regression we could have simply subtracted the coefficient of Dummy-computer from the coefficient of Dummy-video (to obtain the number 18) and then simply divided that difference by the variance of all dummy coefficients to obtain the correct t-statistic for testing the null hypothesis of no difference between the coefficients of the two dummy variables.

This suggests that we might have approached the problem of testing for a significant difference between the two coefficients in the same way as we approached the problem of comparing two population means in Chapter 6. In the problem at hand, however, the required computations are different than we used in Chapter 6 for two reasons. First, the regression coefficients we are comparing represent the mean responses of the dependent variable to the respective independent variables, so their variances are the variances of means rather than population variances. We therefore do not need to divide these variances by n. Second, the coefficients of the independent variables in linear regressions are not necessarily statistically independent, so we cannot obtain the variance of the difference between two coefficients simply by adding their variances—we must subtract from this sum an amount equal to twice their covariance. The variance-covariance matrix of the coefficients in the regression that used programmed text as the reference treatment is[1]

	b_0	b_1	b_2
b_0	7.6666	-7.6666	-7.6666
b_1	-7.6666	15.3333	7.6666
b_2	-7.6666	7.6666	15.3333

The variance of the difference between the coefficient estimates b_1 and b_2 is

$$Var\{b_1 - b_2\} = Var\{b_1\} + Var\{b_2\} - 2\,Cov\{b_1, b_2\}$$

$$= 15.3333 + 15.3333 - (2)(7.6666) = 15.3333 + 15.3333 - 15.3333 = 15.3333.$$

The standard deviation of the difference between the two coefficients is therefore equal to the square root of 15.3333, which equals 3.91578, the standard error of the coefficients of both dummy variables. So we can legitimately test whether the coefficients of Dummy-video and Dummy computer differ significantly by taking the difference between the coefficients and dividing it by their common standard error to form an appropriate t-statistic.

It should be noted, however, that although we could have obtained an appropriate test of the difference between the coefficients of the two dummy variables in this case by simply dividing the difference between the coefficients by their common standard error and comparing the resulting t-statistic with the critical values in the table at the back of our textbook, this will not necessarily work under all circumstances. We have not investigated what would be the best procedure to follow when, for example, the numbers of sample elements receiving each of the three treatments differ. We always have to take account of the fact that the covariance between estimated regression coefficients will not in general be zero.

[1] This was obtained from XlispStat, the computer program used to calculate the regression. Using the matrix notation we very briefly developed in Chapter 9, the variance covariance matrix can be written (see page 5) as $s^2(X'X)^{-1}$.

10.3 Two-factor Analysis of Variance

In ending this chapter we examine briefly a *two-factor designed experiment.* We add fifteen randomly selected female students to the fifteen male students in the above single factor experiment. These fifteen female students are also randomly divided into three groups of 5 students each. One group is instructed by programmed text, one by video-taped lecture and one by computer programs. In this two factor experiment the number of treatments expands from three to six according to the six factor combinations—male-text, male-video, male-computer, female-text, female-video and female-computer. The best way to approach this problem for our purposes is to use a regression analysis of the sort immediately above. In setting up the regression, we obviously need a dummy variable to separate the genders—we let it take a value of 0 if the student is male and 1 if the student is female. Letting programmed text be the reference, we also need dummy variables for video-taped lecture (taking the value of 1 if the instructional mode is video-taped lecture and zero otherwise) and for computer programmed learning (taking a value of 1 if the instructional mode is computer programs and zero otherwise). This would give us the following regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i$$

where Y_i is test score, X_{1i} is the female dummy, X_{2i} the video dummy and X_{3i} the computer dummy. The mean test scores identified in the model are as follows:

Males-text	β_0
Females-text	$\beta_0 + \beta_1$
Males-video	$\beta_0 + \beta_2$
Males-computer	$\beta_0 + \beta_3$
Females-video	$\beta_0 + \beta_1 + \beta_2$
Females-computer	$\beta_0 + \beta_1 + \beta_3$

But this imposes the condition that the effects of the different modes of instruction on achievement test scores be the same for males as for females—using video-taped-lectures instead of programmed text will increase the test scores by an amount equal to β_2 for both males and females, and using computer programs instead of programmed text will increase their test scores uniformly by β_3.

This formulation is inadequate because we should be taking account of whether mode of instruction has a differential effect on the achievement test scores of females and males. We do this by adding *interaction dummy variables* constructed by multiplying the female dummy by the mode-of-instruction dummy. Our regression model then becomes

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{1i} X_{2i} + \beta_5 X_{1i} X_{3i} + \epsilon_i.$$

The first three independent variables are the same as before—female dummy, the video dummy, and the computer dummy. The fourth independent variable is the product of the female dummy and the video dummy—it will take the value 1 if the student is both female and using video-taped lecture instruction and 0 otherwise. And the fifth independent variable is the product of the female dummy and computer dummy, which will be equal to 1 if the student is both female and using computer programmed instruction and 0 otherwise. Notice that the five dummy variables together with the constant term represent the six treatments in the two-factor experiment. The mean test scores identified in the model for the six treatments are now

Males-text	β_0
Females-text	$\beta_0 + \beta_1$
Males-video	$\beta_0 + \beta_2$
Males-computer	$\beta_0 + \beta_3$
Females-video	$\beta_0 + \beta_1 + \beta_2 + \beta_4$
Females-computer	$\beta_0 + \beta_1 + \beta_3 + \beta_5$

The regression results obtained are as follows:

Dependent Variable: Achievement Test Score

Constant	85	(2.86)
Dummy-female	-0.4	(4.05)
Dummy-video	3	(4.05)
Dummy-computer	-15	(4.05)
Dummy-video-female	1	(5.73)
Dummy-computer-female	14	(5.73)
R-Squared	.54	
Standard Error ($\hat{\sigma}$)	6.40182	
Number of Observations	30	
Degrees of Freedom	24	

The coefficient of Dummy-video, which is a point estimate of β_2, measures the difference in male scores under video-taped lecture instruction as compared to programmed text. Its t-ratio of

$$t^* = \frac{3}{4.05} = .74071$$

indicates that β_2 is not significantly different from zero. The coefficient of Dummy-computer is a point estimate of β_3 and measures the difference in male scores under computer-programmed instruction as compared to programmed text. Its t-ratio is

$$t^* = \frac{-15}{4.05} = -3.7037,$$

indicating a significant negative effect of computer programs over programmed text as a method of instruction. The coefficient for Dummy-female is a point estimate of β_1, measuring the effect of being female rather than male on achievement test scores when programmed text is the method of instruction. It should be obvious that β_1 is not significantly different from zero. The point estimate of β_4, the coefficient of Dummy-video-female, measures the estimated effect of being female rather than male when taking video-lecture instruction. The relevant t-ratio, .17452, indicates no significant effect. Finally, the coefficient of Dummy-computer-female, which measures the effect of being female rather than male when taking computer-programmed instruction, is plus 14 with a t-statistic of

$$t^* = \frac{14}{5.73} = 2.4433.$$

This indicates a significantly positive effect of being female rather than male when taking computer programmed instruction. It is clear that females do significantly better than males when the instruction mode is computer programs. In fact, it can be seen from a comparison of the coefficient of Dummy-computer with that of Dummy-computer-female that the negative effect of computer programmed instruction on learning, which is statistically significant for male students, almost vanishes when the student is female.

10.4 Exercises

1. In a completely randomized design experiment with one factor the following data were obtained for two samples:

Sample 1: 5 5 7 11 13 13
Sample 2: 10 10 12 16 18 18

Test the null hypothesis that the two samples were drawn from populations with equal means and draw up the appropriate ANOVA table.

2. A clinical psychologist wished to compare three methods for reducing hostility levels in university students. A certain psychological test (HLT) was used to measure the degree of hostility. High scores on this test indicate great hostility. Eleven students obtaining high and nearly equal scores were used in the experiment. Five were selected at random from among the eleven problem cases and treated by method A. Three were taken at random from the remaining six students and treated by method B. The other three students were treated by method C. All treatments continued throughout a semester. Each student was given the HLT test again at the end of the semester, with the results shown below:

Method A	Method B	Method C
73	54	79
83	74	95
76	71	87
68		
80		

Do the data provide sufficient evidence to indicate that at least one of the methods of treatment produces a mean student response different from the other methods? What would you conclude at the $\alpha = .05$ level of significance?

3. Is eating oat bran an effective way to reduce cholesterol? Early studies indicated that eating oat bran daily reduces cholesterol levels by 5 to 10%. Reports of these studies resulted in the introduction of many new breakfast cereals with various percentages of oat bran as an ingredient. However, a January 1990 experiment performed by medical researchers in Boston, Massachusetts cast doubt on the effectiveness of oat bran. In that study,

20 volunteers ate oat bran for breakfast and another 20 volunteers ate another grain cereal for breakfast. At the end of six weeks the percentage of cholesterol reduction was computed for both groups:

Oat Bran	Other Cereal
14	3
18	3
4	8
9	11
4	9
0	7
12	12
2	13
8	18
12	2
10	7
11	5
12	1
6	5
15	3
17	13
12	11
4	2
14	19
7	9

What can we conclude at the 5% significance level?

4. Prior to general distribution of a successful hardcover novel in paperback form, an experiment was conducted in nine test markets with approximately equal sales potential. The experiment sought to assess the effects of three different price discount levels for the paperback (50, 75, 95 cents off the printed cover price) and the effects of three different cover designs (abstract, photograph, drawing) on sales of the paperback. Each of the nine combinations of price discount and cover design was assigned at random to one of the test markets. The dependent variable was sales, and the independent variables were the discount off cover price, a dummy variable taking a value of 1 if the design was photograph and 0 otherwise, and a dummy variable taking a value of 1 if the design was drawing and 0 otherwise.

The regression results were as follows:

Dependent Variable: Sales

Constant	6.03685	(0.753114)
Discount	0.18363	(0.009418)
Photo-dummy	-0.68333	(0.424682)
Drawing-Dummy	1.60000	(0.424682)
R-Squared	.98970	
Standard Error ($\hat{\sigma}$)	0.520126	
Number of Observations	9	
Degrees of Freedom	5	

The numbers in brackets are the standard errors of the respective coefficients and $\hat{\sigma}$ is the standard error of the regression, a point estimate of the standard deviation of the error term.

a) Is there good evidence that discounting the price increases sales?

b) Is there good evidence that using an abstract cover rather than putting on a photograph or drawing results in less sales?

c) Is the overall regression relationship statistically significant?

d) What would be the expected level of sales if the discount is 75 cents off the printed cover price and a drawing is put on the cover?