Chapter 12: One-Way Analysis of Variance

First Read: Chapter 12 in the text, then read the notes and try the WEBCT assignment questions. If you need more practice, try the practice questions with answers available on the web.

Introduction

In earlier chapters we compared groups with box plots, stem and leaf displays and numerous numerical measures. Now that we have some background on the process of inference, we can, in this chapter, compare several groups to see if the differences in the group means are significant. We compared two groups in Chapter 7, using the two-group independent samples t test. We use the one-way analysis of variance technique in this chapter, which is an extension of the t test to compare several groups or means.

General Background

The One-way Analysis of Variance (ANOVA) is used to determine if mean differences exist for 2 or more treatments or groups.

Previously, in chapter 7, for 2 means we saw that the statistic.

$$t = \frac{\textbf{mean differences between samples}}{\textbf{differences expected from sampling error}}$$

$$= \frac{\textbf{mean differences}}{\textbf{standard error}}$$

Now, for the ANOVA we have the test statistic F which we define as:

$$F = \frac{\textbf{variance (differences) in means}}{\textbf{variance (differences) expected from sampling error}}$$

Note that the F ratio based on **variances** as we saw in Chapter 7 even though we are testing if there are differences in means.

Example

A researcher looks at student performance under 3 conditions of counseling (50, 70, 90 minutes). The hypothesis we want to examine, using the ANOVA technique, is written as follows.

$$H_0: \mu_1 = \mu_2 = \mu_3$$

(condition does not effect performance, means equal)

$$H_A: \mu_1 \neq \mu_2 \neq \mu_3$$

(at least one mean is different from the rest)

Five students are randomly assigned to each treatment. The data are given below. We want to see if the population means for each group differ.

Data:

Treat 1	Treat 2	Treat 3
50	70	90
0	4	1
1	3	2
3	6	2
1	3	0
0	4	0
$\bar{x} = 1$	$\bar{x} = 4$	$\bar{x} = 1$

Just as in Chapter 10, the data is broken into two pieces:

DATA = FIT + RESIDUAL

= MODEL + ERROR

The components of each piece are given below in the chart. In Chapter 10 the ANOVA analysis was used to determine if a line model was a reasonable one for the data. Here we use a model with a different mean for each group assuming there are group differences.

Note how the data is partitioned into two pieces. Each piece has a number of components that contribute to the test statistic F.

(DATA)
Total Variability

(FIT) **Between Treatments Variability**	**(RESIDUAL)** **Within Treatments Variability**
1) Treatment effect	
2) Individual differences	**2) Individual differences**
3) Experimental error	**3) Experimental Error**

The Between Treatment variability is composed of a treatment effect (if there is one), individual differences and experimental error. The within treatment variability is composed of individual differences and error. The ratio of these two variances is an F statistic.

$$F = \frac{\textbf{treatment effect (1) + individual differences (2) + experimental error (3)}}{\textbf{individual differences (2) + experimental error (3)}}$$

If H_o is true (all means equal, no treatment differences) then the F ratio is:

$$F = \frac{0 + 2) + 3)}{2) + 3)} = 1$$

If H_o is false (there are differences in the means) then the F ratio is:

$$\frac{1) + 2) + 3)}{2) + 3)} > 1$$

The F statistic is larger than 1 if we have differences in the group means. The technical details for the model are given below.

Statistical Model for One-Way ANOVA

Consider a simple random sample (SRS) from $N(\mu,\sigma)$ distribution $x_1, x_2,...,x_n$.
You could think of the observations varying about their mean:

$$y_j = \mu + e_j$$

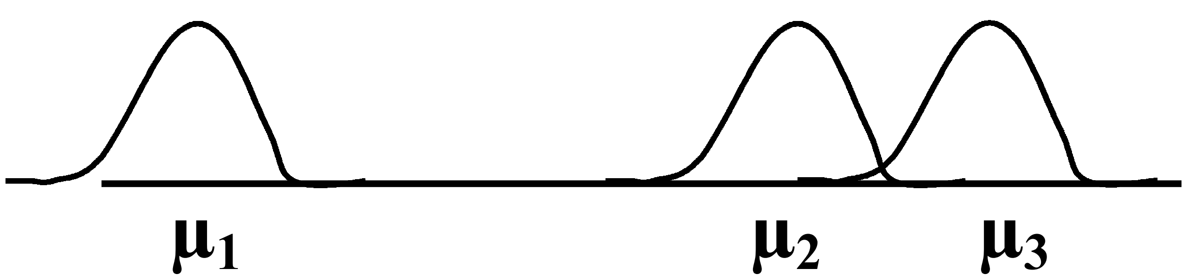
The value of j tracks the observations in the treatment. The data (x) can be viewed as composed of:

Data= Fit + Residual

Now suppose we have "I" different populations. Where "i" tracks each population.

$$y_{ij} = \mu_i + e_{ij}$$

and the populations (μ_i) are different. Three populations and their group means are drawn below:



The e_{ij} are a SRS from $N(0,\sigma)$.

The Linear Model for our problem with 3 populations is:

$$\mathbf{y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e}$$

x_i is a ***dummy variable*** that has the value 1 when y is in group i, and 0 otherwise. β_i is the mean of y when x = i.

We want to make inferences about the β_i or means of the populations. The sample mean estimates the population mean and the deviation of the observation from its mean estimates the error in the system.

$$\overline{y}_i \textbf{ estimates } \mu_i$$

$$y_{ij} - \overline{y}_i \textbf{ estimates } e_{ij}$$

In order to test the hypothesis that all the means are equal:

H_0: $\mu 1 = \mu 2 = \mu 3$ VERSUS H_a: $\mu_1 \neq \mu_2 \neq \mu_3$
The alternative hypothesis that at least one differs from the rest.

researchers construct an ANOVA table.

One-Way ANOVA Table for Testing I Means

The **Total Row** of the table refers to the variance of the data, the **Between Groups** row refers to the amount of variance accounted by the model and the **Within Groups** row refers to the error variance of the model. The ANOVA table is usually calculated by computer.

Source	Sums of Squares	Degrees of Freedom	Mean Square	F
Between Groups	$\sum n_i(y_i-\bar{y})^2$	**I – 1**	**SSG/DFG**	**MSG/ MSE**
Within Groups	$\sum(n_i-1)s_i^2$	**N – 1**	**SSE/DFE**	
Total	$\sum(y_{ij}-\bar{y})^2$	**N – 1**		

The F ratio is defined as the ratio of mean square of groups over the mean square within groups or error.

$$F = \frac{MSG}{MSE} = \frac{\text{variance between treatments}}{\text{variance within treatments}}$$

Just as in Chapter 10, we can calculate R squared for the model:

$$R^2 = \frac{SSG}{SST} = \text{coefficient of determination}$$

Example: The ANOVA calculations below are based on our data with 3 groups (50, 70, 90). The calculations are given to aid in your understanding of the technique, however in practice it is best to use a computer to obtain your results.

1	2	3
50	**70**	**90**
$T_1 = 5$	$T_2 = 20$	$T_3 = 5$ **(totals)**
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$ **(number of obs.)**
$\bar{y}_1 = 1$	$\bar{y}_2 = 4$	$\bar{y}_3 = 1$ **(means)**

Overall, T = 30 (total for all observations ($T_1+T_2+T_3$))

$N = 15$ (total number of observations ($n_1+n_2+n_3$))

$\bar{y} = 2$ (overall mean)

$I = 3$ (number of groups)

Sums of squares are calculated for each row of the table:

SSG $= 5(1-2)^2 + 5(4-2)^2 + 5(1-2)^2$
$= 5 + 20 + 5$
$= 30$

SSE $= 6 + 6 + 4$
$= 16$

SST $= 46$

The degrees of freedom are calculated for each row

Total degrees of freedom **15 - 1 = 14**

Error degrees of freedom **15 - 3 = 12**

Group degrees of freedom **3 - 1 = 2**

The corresponding mean squares are calculated below:

Mean Square = SS/DF

MSG = 30/2 = 15

MSE = 16/12 = 1.33 or

S^2_p = pooled estimate of variance = 1.33

In the two sample t (pooled), two estimates of the variance are combined or pooled, whereas in the ANOVA for several groups, several variance estimates are pooled.

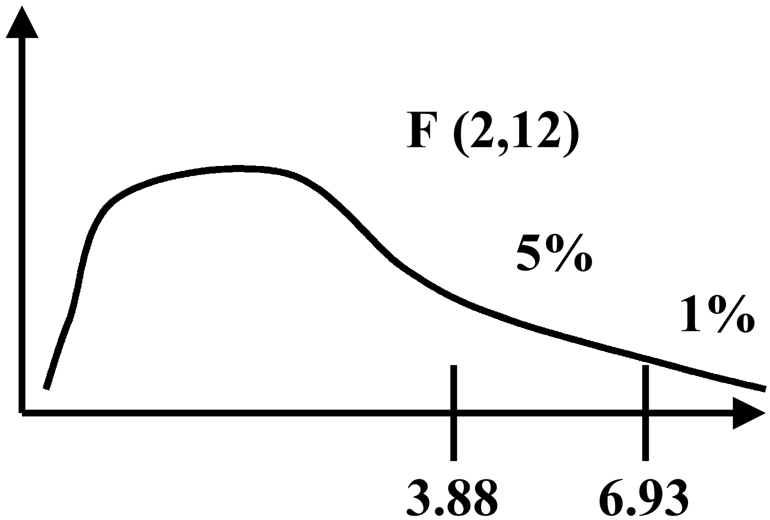
The F = 15/1.33 = 11.28. Remember if there is no difference in means, we would expect a value of approximately 1.0.

Note R^2 = SSG/SST = 30/46 = .65 or 65%. In other words, 65% of the variance in performance is accounted for by counseling.

We display the above calculations in ANOVA table below. Please see the text for more examples.

Source	Sums of Squares	Degrees of Freedom	Mean Square	F
Between Groups	30	2	15	11.28
Within Groups	16	12	1.33	
Total	46	14		

<u>Note</u>: The F = 11.28 statistic is much larger than 1 as previously discussed. We look up the F distribution with (2,12) degrees of freedom.



and find the F statistic significant the at the 5% level (F=3.88) and the 1% level (F=6.93). We have a p-value = .0001.

We reject H_o that the means are equal. We have very strong evidence against $H_{o.}$

We conclude that mean performance differs in the three groups.

However, we do not know the location of the differences from the ANOVA test.

All we can say is that at least one mean differs from the rest.

We examine the sample means using contrasts to find out where the differences in means occur.

Example

A researcher examines 3 methods of reading instruction for novice readers . The methods are:

1) BASAL 2) DRTA 3) STRAT

TABLE 12.1 Pretest reading scores

Group	Subject	Score	Group	Subject	Score	Group	Subject	Score
Basal	1	4	DRTA	23	7	Strat	45	11
Basal	2	6	DRTA	24	7	Strat	46	7
Basal	3	9	DRTA	25	12	Strat	47	4
Basal	4	12	DRTA	26	10	Strat	48	7
Basal	5	16	DRTA	27	16	Strat	49	7
Basal	6	15	DRTA	28	15	Strat	50	6
Basal	7	14	DRTA	29	9	Strat	51	11
Basal	8	12	DRTA	30	8	Strat	52	14
Basal	9	12	DRTA	31	13	Strat	53	13
Basal	10	8	DRTA	32	12	Strat	54	9
Basal	11	13	DRTA	33	7	Strat	55	12
Basal	12	9	DRTA	34	6	Strat	56	13
Basal	13	12	DRTA	35	8	Strat	57	4
Basal	14	12	DRTA	36	9	Strat	58	13
Basal	15	12	DRTA	37	9	Strat	59	6
Basal	16	10	DRTA	38	8	Strat	60	12
Basal	17	8	DRTA	39	9	Strat	61	6
Basal	18	12	DRTA	40	13	Strat	62	11
Basal	19	11	DRTA	41	10	Strat	63	14
Basal	20	8	DRTA	42	8	Strat	64	8
Basal	21	7	DRTA	43	8	Strat	65	5
Basal	22	9	DRTA	44	10	Strat	66	8

Click SPSS Example One to see how to read the data into SPSS and,

- obtain boxplots of data
- perform one way ANOVA and examine the residuals for normality
- perform contrasts of means
- examine LSD and Bonferonni Multiple Comparison Procedure to see what means differ.

Now click the SPSS example below and see how to complete the above by computer.

SPSS EXAMPLE ONE

Contrasts

Contrasts are used after the ANOVA technique in order to see what group means differ. The researcher conducts an ANOVA–rejects H_0–knows now that differences in means exist, but where? What means differ?

She can examine *pre-planned* comparisons to see where differences lie. The researcher, before examining the data, determines which means to compare, for example in our previous example 3 reading groups are compared.

The researcher wishes to compare the average of DRTA and STRAT groups with the BASAL Group.

H_{01}: $(\mu_D + \mu_S)/2 = \mu_B$ (now multiply both sides by 2)

: $1\mu_D + 1\mu_S = 2\mu_B$ (rewrite the equation to equal zero)

Or: $1\mu_D + 1\mu_S - 2\mu_B = 0$ (you can write the equation without the 1)

Or: $\mu_D + \mu_S - 2\mu_B = 0$

H_{A1}: $1\mu_D + 1\mu_S - 2\mu_B \neq 0$ (the alternative is not equal to zero)

Or: $\mu_D + \mu_S - 2\mu_B \neq 0$

Note: If the sum of the coefficients in each combination of means is zero we call it ***a contrast.***

In the above example, the coefficients do add to zero:
$+ 1 + 1 - 2 = 0$, therefore, **we have a contrast**.

Suppose we also want to compare the DRTA and STRAT groups.

H_{02}: $\mu_D = \mu_S$ (rewrite so the equation equals zero)
: $\mu_D - \mu_S = 0$

H_{A2}: $\mu_D \neq \mu_S$ (the alternative is not equal to zero)
: $\mu_D - \mu_S \neq 0$

The coefficients add to zero, 1 - 1 = 0, therefore, we have a contrast. We estimate the population means in the contrast using the sample means. The sample contrast is tested using a t statistic. For example, H_{01} is tested as follows: multiply the means by the coefficients of the contrast.

$$\mathbf{C_1 = -2(41.05) + 46.73 + 44.37 = 4.45}$$

Standard error is:

$$\mathbf{SE_{ci} = 6.314\sqrt{\frac{(-2)^2}{22} + \frac{(1)^2}{22} + \frac{(1)^2}{22}} = 1.65}$$

The test statistic T is the contrast divided by the standard error, which under the null hypothesis should be a reasonable value from the t distribution.

$$\mathbf{T = \frac{c_1}{SE_{c1}} = \frac{4.45}{1.65} = 2.70}$$

The degrees of freedom for error in our ANOVA is 63 so t(63) gives a p=.005. We have very strong evidence against H_{01}.

We conclude that:

The average of the DRTA and STRAT groups differs from the average of the BASAL group.

We want our contrasts to be independent or orthogonal. To see if two contrasts are orthogonal, check if the cross product of the coefficients adds to zero.

<u>Example</u>: Write out the coefficients for the means for our two contrasts below in a table.

	μ_D	μ_S	μ_B
Contrast Ψ_1	1	1	-2
Contrast Ψ_2	1	-1	0

Calculate the Cross Products $= (1*1) + (1*-1) + (-2*0)$
$= 1 + -1 + 0$
$= 0$

Contrasts are independent or orthogonal since the sum of the cross products of the coefficients is equal to zero.

Multiple Comparisons

Researchers cannot always pre-plan comparisons (contrasts) of means as previously discussed in the last section. We want to find out which pairs of means differ when we have not pre-planned our comparisons. Consider the problem of 5 groups. We have 20 pairwise comparisons of means. Even if H_0 is true (all the means are equal) some of the pairwise comparisons of means will be declared significant.

For example, consider performing the pairwise comparisons of means at α=.05, and the probability that the test is not significant = 1 - .05 = .95, assuming the tests are independent–the probability of at least one significant result is one minus all tests that are not significant, $1 - (.95)^{20}$, which equals .23 when 5 groups are compared pairwise (note 20 comparisons) and there are no differences in the means.

This probability 23% is high. How can we control this problem of declaring means different even though they are not? One can choose a critical value that is lower so the probability of at least one false positive is small (i.e., .01). The problem of multiple comparisons is controlling the probability of rejection so the researcher has confidence in her conclusions.

Multiple Comparisons are used only after rejecting the ANOVA H_0.
Two methods will be examined using the computer:

1. **LSD–least significant difference method**
2. **Bonferroni Method**

Click the SPSS example below for an illustration of the multiple comparison technique used by SPSS.

SPSS EXAMPLE ONE

View the Video – Case Study

Chapter 12 Summary

The **ANOVA** technique is used to compare the means of several populations. The **null hypothesis** states that all the means are equal. If the **alternative hypothesis** is true, this indicates a difference in the means.

The ANOVA technique partitions the total variation in the data into two pieces, **Between Groups or Model** and **Within Groups or Error**. The **F-statistic** and corresponding p-value are used to test if the model piece is larger than the error piece. If comparisons of means are constructed before the data is collected, they can be expressed as **contrasts**. If comparisons are constructed after examination of the data and the ANOVA null hypothesis is rejected, **multiple comparisons** can be used to compare means.

Remember: Tips for Success

1) Read the text.

2) Read the notes.

3) Try the assignment.

4) If needed, try the exercise questions.

5) Try the simulations and view the videos if you need more help with a concept.

6) Try the self tests for practice on each chapter of the text at **www.whfreeman.com/ips.**

7) Steady Work = **SUCCESS**