CHAPTER 10

Analysis of Variance: One-Way Classification

Summary

Analysis of variance (ANOVA) is another NHST technique that allows you to draw conclusions about populations from sample data. The null hypothesis of a one-way ANOVA is that the means of the populations that the two or more samples come from are equal. If the data allow you to reject this null hypothesis, you can tell a story about the effects of the various levels of the independent variable on the dependent variable.

The *rationale of ANOVA* is important. To really grasp the rationale, you must have clear ideas of how the *two estimates of the population variance* are found. Here is the rationale of ANOVA presented in two different ways, first in list form and then with paragraphs and pictures.

Rationale of ANOVA---I

A. Assume that the variances of the populations from which the samples were taken are equal. Estimate this variance by calculating a variance for each sample and averaging them. Put this estimate of the population variance aside for now.

B. When the null hypothesis is true:

1. When the null hypothesis is true, the sample means will not vary much from each other. The variability among the means can be measured with a variance of the means. Calculate this variance and multiply it by a factor that makes it equal to the population variance that was estimated previously (in A above).
2. Construct a ratio. Put the variance that measures the variability among sample means in the numerator and the variance that measures the population variance in the denominator. Expect that this ratio (called *F*) will be about 1.00, although some variability in the ratio is expected because of sampling variation. The variation of this *F* ratio is a sampling distribution of *F*. *F* values with probabilities of .05 and .01 are in the table in your text.

C. When the null hypothesis is false:

1. When the null hypothesis is *false*, the sample means will be different from one another. This variability can be measured with a variance, producing a value that will be larger than the one you get when the samples come from the same population.
2. Construct a ratio. Put the variance produced by sample means drawn from different populations in the numerator and the variance that measures population variance in the denominator. Expect that the ratio will be larger than 1.00.

D. Using data from an experiment, calculate the *F* value. The probability of this *F* value when the null hypothesis is true is given in Table F in your text. If this probability is low (.05 or less), reject the null hypothesis. If the probability is not low (.051 or greater), retain the null hypothesis.

Rationale of ANOVA---II

When the Null Hypothesis is True: Look at Population A in Figure 10.1. Suppose you drew three samples from Population A. For each sample, a variance, ^{2} can be calculated. Each ^{2} is an estimate of σ^2, the population variance. Pooling the three sample variances provides an even better estimate of σ^2.

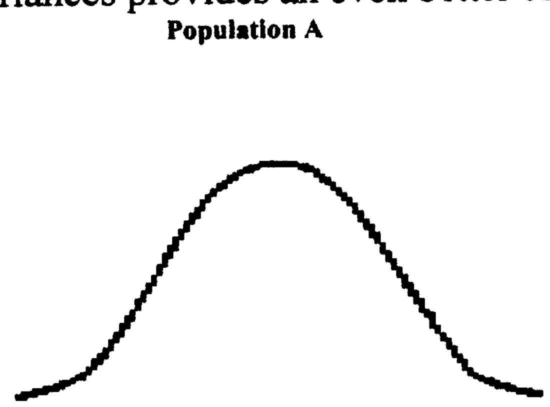

Figure 10.1 Population A with variance, $\mathbf{a^2}$. Used to illustrate ANOVA when the null hypothesis is true

The three samples also yield three $\overline{X}$'s. The three means, which will be fairly close together, provide another way to estimate σ^2. The sample means, *N*, and algebra produce another accurate estimate of σ^2.

Finally, if the second estimate of σ^2 is divided by the first estimate of σ^2, a ratio (the F ratio) will be about 1.00. This is the result when the null hypothesis is true, which is the case when all the samples are drawn from Population A.

When the Null Hypothesis is False: Look at Figure 10.2, which shows three populations, B, C, and D, whose variances are equal. Suppose you drew a sample from each population. For each sample, a variance, ^{2}, can be calculated. Each ^{2} is an estimate of the variance, σ^2, the variance that is the same for all three populations.

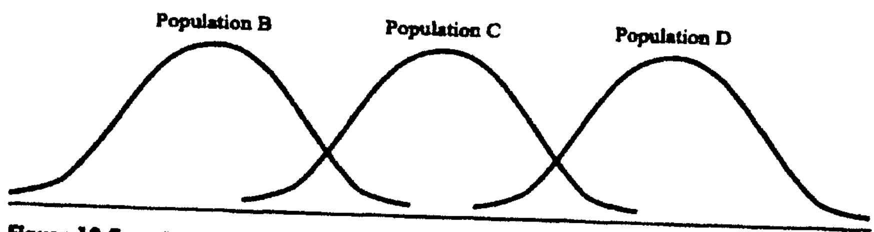

Figure 10.2 Populations B, C, and D, which have variances that are equal but means that are different. Used to illustrate ANOVA when the null hypothesis is false.

The three samples also yield three $\overline{X}$'s, which again can be used to estimate σ^2. This time, however, the means will not be close together. Now, the estimate of σ^2 that is based on sample means will not be accurate; it will *overestimate* the actual size of σ^2.

Finally, if the second estimate of σ^2 (based on sample means) is divided by the first estimate of σ^2 (based on averaging $\hat{s}^2$), the ratio will be greater than 1.00. This is the result when the null hypothesis is false (the three samples are from different populations).

Thus, if the ratio of the two variances is much greater than 1.00, there is evidence that the null hypothesis is false. How big is "much greater?" The F distribution shows F values that occur 5% and 1% of the time when the null hypothesis is true. If your experiment produces a calculated F value that is larger than the tabled F value, reject the null hypothesis and conclude that the samples did not all come from the same population.

The degrees of freedom for the numerator of the F ratio is K-1, where K is the number of groups. The df for the denominator is $N_{tot} - K$.

Chapter 10

Tests Subsequent to ANOVA

If the data allow you to reject the null hypothesis, further tests are informative. These tests are either *a priori* or *post hoc* tests. *A priori* tests require that a limited number of comparisons be chosen on logical grounds prior to data analysis. *Post hoc* tests allow you to make statistical comparisons after examining the data. One *post hoc* test, the Tukey Honestly Significant Difference (Tukey HSD) allows you to make all *pairwise comparisons*.

An effect size index, f, gives you additional information that the F test doesn't provide. The text's formula for f is appropriate when the sample N's are equal. Values of f of 0.10, 0.25 and 0.40 indicate the size of the effect of the independent variable is small, medium or large, respectively.

After conducting an ANOVA, applying subsequent tests, and finding an effect size estimate, your final task is to explain the results of the experiment using the terms of the experiment. This is perhaps the most important part of data analysis!

The ANOVA technique described in your text is appropriate for analyzing quantitative data from independent samples if certain assumptions about the populations hold true. These assumptions are that the population variances are equal and that the populations are normally distributed. In addition, the technique requires that the participants be randomly assigned to levels of the independent variable if cause and effect conclusions between the independent and dependent variables are drawn.

Multiple-Choice Questions

1. The person who developed ANOVA was
   (1) W. S. Gosset, a businessman;
   (2) John W. Tukey, a statistician;
   (3) "student," a pseudonym;
   (4) Ronald A. Fisher, a biologist.

2. An *F* distribution is a
   (1) normal distribution;
   (2) *t* distribution;
   (3) sampling distribution;
   (4) none of the above.

3. The null hypothesis tested by ANOVA is that
   (1) all samples have the same mean;
   (2) each sample is drawn from a different population;
   (3) the populations from which the samples are drawn have the same mean;
   (4) one or more of the populations from which the samples are drawn has a mean that is different from the others.

4. The ANOVA technique described in the text can be used on
   (1) paired-samples designs;
   (2) independent-samples designs;
   (3) both (1) and (2);
   (4) neither (1) nor (2).

5. If the null hypothesis is true, __________ will be a good estimate of the population variance.
   (1) the error mean square;
   (2) the treatment mean square;
   (3) both (1) and (2);
   (4) neither (1) nor (2).

6. If the null hypothesis is false, __________ will be a good estimate of the population variance.
   (1) the error mean square;
   (2) the treatment mean square;
   (3) both (1) and (2);
   (4) neither (1) nor (2).

7. The larger the population variance, the larger ______________ is (are).
   (1) F;
   (2) df_{treat};
   (3) MS_{error};
   (4) all of the above.

8. If a tabled value of F is 10.00 and the F obtained from the data is only 9. you should
   (1) retain the null hypothesis;
   (2) reject the null hypothesis;
   (3) calculate the F value again, such a number is not possible;
   (4) not enough information is given.

9. Suppose MS_{treat} is calculated for three samples that are drawn from a population with a mean μ. Under which condition below would MS_{treat} certainly become larger?
   (1) the addition of a sample from a population with the same mean μ;
   (2) the addition of a sample from a population with a mean $\mu + \mu$;
   (3) both (1) and (2);
   (4) the removal of one of the three samples from the calculations.

10. According to the data analyzed and interpreted by your text, the effect of different schedules of reinforcement is to produce different
    (1) speeds of learning during training;
    (2) degrees of persistence during extinction;
    (3) degrees of forgetting over time;
    (4) rates of responding when a maze-learning task is.

11. For a one-way ANOVA an effect size index that qualifies as large is
    (1) 0.20;
    (2) 0.40;
    (3) both (1) and (2);
    (4) neither (1) nor (2).

12. A group of 72 subjects was equally divided into four groups. A Tukey HSD test produced a value that led to the conclusion that Mean 1 was significantly larger than Mean 2, $p < .05$. Which of the following situations would lead to such a conclusion?
    (1) $\bar{X}_1 = 7$, $\bar{X}_2 = 0$, $MS_{error} = 80$;
    (2) $\bar{X}_1 = 24$, $\bar{X}_2 = 13$, $MS_{error} = 180$;
    (3) both (1) and (2);
    (4) neither (1) nor (2).

13. A *priori* and *post hoc* are terms that refer to
    (1) whether the null hypothesis should be rejected;
    (2) whether the assumptions of ANOVA have been met;
    (3) kinds of tests used after an ANOVA;
    (4) all of the above.

14. The Tukey Honestly Significant Difference test is a(n) __________ test.
    (1) a *priori*;
    (2) *post hoc*;
    (3) both (1) and (2);
    (4) neither (1) nor (2).

15. A group of 36 subjects was equally divided into three groups. A Tukey HSD produced a value that led to the conclusion that Mean 1 was significantly larger than Mean 2, $p < .05$. Which of the following situations would lead to such a conclusion?
    (1) $\bar{X}_1 = 9$, $\bar{X}_2 = 2$, $MS_{error} = 50$;
    (2) $\bar{X}_1 = 24$, $\bar{X}_2 = 14$, $MS_{error} = 84$;
    (3) both (1) and (2);
    (4) neither (1) nor (2).

16. Suppose the following F values were calculated from different experiments. If
    $\alpha = .01$, which of them would lead to rejection of the null hypothesis?
    (1) $F = 18.50$, $df = 2, 2$;
    (2) $F = 2.55$, $df = 10, 17$;
    (3) $F = 2.41$, $df = 20, 42$;
    (4) none of the above.

17. Compare the *F* test and the *t* test.
    (1) The *F* test can be used with only two groups; the *t* test can be used wit more than two groups;
    (2) The *t* test can be used with only two groups; the *F* test can be used wit more than two groups;
    (3) There is no difference in number of groups.

18. How many degrees of freedom is there for the numerator and denominator respectively if there are four groups with eight participants in each?
    (1) 4; 8;
    (2) 3; 8;
    (3) 3; 21;
    (4) 4; 21.

19. How many degrees of freedom is there for numerator and denominator respectively if there are 3 groups with 10 participants in each?
    (1) 3, 10;
    (2) 2, 9;
    (3) 3, 30;
    (4) 2, 27.

20. The difference between the Tukey HSD and *a priori* comparisons is related to
    (1) when you decided to make the comparisons;
    (2) the type of statistic used;
    (3) the power of the statistic;
    (4) all of the above.

Interpretation ____________________

1. Florence Nightingale (1820-1910) was instrumental in reforming medical care. Her methods were based on her experience administering hospitals foı British soldiers who were casualties when England was fighting in Crimea (peninsula in the Black Sea). Part of Nightingale's success can be attributed her pioneering use of statistics and graphs. (Nightingale was especially appreciative of the earlier work of Quetelet. See Cohen [1984].)

Chapter 10

The "improvement scores" below will produce conclusions like those that Nightingale found when she compared patients in her military hospital in Crimea with civilian patients in English and French hospitals.

Begin by identifying the independent and dependent variable. Next, examine the summary data and calculate *F*, *f*, and HSD values. Finally, write your conclusions.

	Hospitals Located in		
	Crimea	England	France
Means	15	9	8
N	12	12	12

Source	*Df*	*MS*
Hospitals	2	124.52
Error	33	20.62

2. What affects thinking? Alice Isen's participants worked on seven difficult problems that required creative thinking. One group prepared by exercising; a second group watched a comedy video; a third group simply began working on the problems. The summary statistics that follow include the mean number of problems worked by each group, the ANOVA summary table, *f*, and the three HSD values. Identify the independent and dependent variable, and write an interpretation of the results.

	Exercise	Comedy	Control
Group Means	2.4	5.1	1.9

Source	*SS*	*df*	*MS*	*F*	*p*
Treatments	43.14	2	21.57	8.94	$< .01$
Error	43.43	18	2.41		
Total	86.57	20			

$f = 0.50$

HSD (Exercise and Comedy) = 4.62
HSD (Exercise and Control) = 0.97
HSD (Comedy and Control) = 5.57

3. A vintner wanted to market a new red wine that would be a blend of several varieties grown in his vineyards. He developed four blends, and he wanted to decide which was the best. He employed the services of eight wine tasters and had each one rate each of the four wines on a 7-point scale, ranging from abominable (1) to exquisite (7). Explain why the ANOVA method described in Chapter 10 is inappropriate for the analysis of the data.

4. Assume that five groups of rats each receive different dosages of a drug. You believe that the dosages are going to result in different running rates among the rats. Further assume that you believe that the rats getting the highest dosage of the drug will run faster than rats receiving the lowest dosage of the drug. What type of analysis will allow you to determine if your belief is correct?

5. What is the difference between ANOVA and the *t* test? Is there a case where you could use either one? Why or why not?

Problems

1. As you know from the text, the time required to extinguish a response depends on the schedule of reinforcement during learning. Extinction time also depends on the predictability of the reinforcement during learning. The following data are patterned after a classic study by Hulse (1973) who reported on the effects of predictability. In this study all pigeons pecked an average of four times for each reinforcement, but the predictability of a reinforcement varied for the three groups. The Very Predictable group was on an FR4 schedule—every fourth response was reinforced. (This is one of the schedules used in the text problem.) The Fairly Predictable group got a reinforcement after two, then four, then six responses. The pattern then repeated. The Unpredictable group was on a schedule produced by a random number generator. It was programmed, however, so that on the average, every fourth response was reinforced. (The name of this schedule is variable

ratio—4, abbreviated as VR4.) After ten days of training, extinction began (responses were never again reinforced). The time to extinction in minutes was recorded. Identify the independent and dependent variable. Analyze the data as completely as you can and write a conclusion about the predictability of reinforcement and persistence.

Very Predictable	Fairly Predictable	Unpredictable
8	16	18
13	11	19
11	15	22
8		16
		15

2. Farmer Marc A., who also serves as his community's Shakespearean promoter, delivered a plea at the county fair, asking farmers from the county's three groups to lend him unshelled corn. (After examining the groups, you might be able to figure out how he would phrase his plea.) The number of bushels offered by the farmers is shown below. Analyze the data as thoroughly as you can, and write a conclusion about the three groups.

Friends	Romans	Countrymen
7	8	10
9	7	12
3	4	16
5	7	14

3. In most hospital delivery rooms, newborn infants are evaluated at one minute of age and again at five minutes of age. The evaluation is based on the *Apgar Scale*, which uses certain criteria to rate the infant's heart rate, respiratory effort, crying, muscle tone, and color. Scores from 0 to 10 are possible, with 10 indicating the highest general well-being. For the study that is described, the neonate's *Apgar* Score is the dependent variable.

The independent variable is the type of anesthetic given to the mother: twilight sleep induced by sedatives, spinal block, opiate, and the Lamaze method (no drugs). Below are summary data for the five-minute *Apgar* scores for the newborns.

Perform an ANOVA, calculate *F*, and make any appropriate HSD tests. Write an interpretation of the results.

	Sedatives	Spinal	Opiate	Lamaze
ΣX	49	74	78	83
ΣX^2	269	574	636	711
N	10	10	10	10

4. Walking is valuable exercise. Is there any relationship between the amount of walking people do and the size of the place where they live? The following summary data (in miles per day) are entirely hypothetical, but they represent what may be the case. Identify the independent and dependent variables, and analyze the scores with an ANOVA and HSD tests. Write a conclusion based on the data. Do you agree with the hypothesis about city/town size and walking? (The place names are ordered from large to small and are from specific sources—can you identify a source or two?)

	Gotham	Middletown	Grovers Corners
ΣX	16	24	32
ΣX^2	70	60	140
N	4	12	8

5. Let's say you are interested in determining if students learn better by hearing about material prior to lecture rather than the traditional way of reading prior to class. Three classes participate in the study. One class has no lecture. They simply read the material. Class two is lectured to after reading the material and class three is lectured to first. You obtain the following data.

Chapter 10

No lecture	Lecture after	Lecture before
.26	.40	.85
.29	.47	.88
.33	.44	.70
.15	.37	.79
.28	.57	.90

What analysis would be appropriate? Would you support the hypothesis that harder work results in better memory? Why or why not?