# CHAPTER FIVE

# ANALYSIS OF VARIANCE

**Objectives:**

The aim of this chapter is to further expand the idea of hypothesis tests. We describe test for variance and then a test that simultaneously compares several means to determine if they come from equal population.

After completing this chapter, you should be able to:

* State the assumptions of ANOVA
* Describe the steps of ANOVA test
* Perform ANOVA test and interpret the result

**5.1 Introduction**

In the preceding chapter procedures for determining whether or not two populations have equal means were presented. However, management problems involve more than two populations, and decision makers want to know whether the means of these populations are, or not, equal.

**Case involving analysis of variance (ANOVA)**

Suppose that several methods are available for trading production workers. Then mgmt would want to know whether mean productivity (output per employee per day) is or is not the same for employees trained by different methods. So in this chapter you will learn how to perform tests of the equality of several population means.

First we will apply the test called the analysis of variance (ANOVA) test. A variance is the square of standard deviation. The test is called analysis of variance because, in performing it, we decide whether to accept or reject the hypothesis of equal population means by analyzing the variation (variance) in the sample means.

* 1. **ANOVA Assumptions**
1. The populations are normally distributed
2. The populations have equal standard deviations
3. The samples are selected independently.

**5.3 The ANOVA Concept**

Suppose that a type writer manufacturer prepared three different study manuals, for use by typists learning to operate in electronic word processing typewriter. Then each manual was studied by a simple random sample of n typists. The time to achieve proficiency (a typing rate of 80 words per minute) was recorded for each typist, and a sample mean learning times for manuals 1, 2 and 3 were  hours. The manufacturer wants to know whether the variation in sample means is large enough to show that the population mean learning times for the manuals are different.

Note that sample mean learning times would vary event if the typists in the samples studied from the same manual. That is because of the inherent variability of learning times. However if manuals do make a difference, mean learning times for different manuals will vary more than mean learning times for the same manual.

The variance is the ANOVA measure of variation. In our example we can form the ratio

 

* If the value of the ratio is larger than one mean that learning times for different manuals vary more than learning times for the same manual.
* If the value if the ratio is significantly large, we will conclude that population mean learning times for the manuals are different. i.e. manuals do make difference in learning times.

***The numerator*** of the above ratio, variance of sample means for different manuals, is simply the variance of the three means 

***The denominator*** poses a problem, because the three samples of typists studied different manuals. However within the manual 1 sample all n typists studied the same manual. You know that it is possible to compute an estimate of the standard deviation of sample means  when the standard deviation of only one sample of size n is available (see chapter 3)

So let *S* be a sample standard deviation. The relation between  and S is

 

Squaring both sides, we obtain the variance relation (*Note that a variance is the square of a standard deviation*)

 

Consequently, to obtain in estimate of the variance of sample means for typists studying the same manual, say manual 1, we compute the sample 1 variance  then we divide it by n to obtain

 

We can also compute  and  to estimate the variance of sample means for typists studying manual 2 and typists studying manual 3.

We now have three estimates of the variance in sample means for typists studying the same manual. In ANOVA it is assumed that the populations from which the samples are drawn have the same variance, (see the ANOVA assumptions) as a result  are estimates of the same quantity ; so we may average the three estimates to obtain the denominator of the ratio.

 

So our ratio will be

 

The above expression can be changed to

 

In general formula for the variance ratio is



The variance ratio is the test statistic in the ANOVA test. We will discuss the test after illustrating the computation of the variance ratio

* 1. Calculating Variances and The Sample F Statistic

In our learning-time example, suppose that three random samples of five typists were selected

Let k be the number of samples and

 n the size of each sample

So we have k = 3 samples of n = 5 typists per sample. Samples 1, 2 and 3 studied respectively from manuals 1, 2, and 3. The sample learning times are shown in table 6.1

### Table 5.1

### Learning times for k = 3 samples of n = 5

Typists per sample

|  |  |  |
| --- | --- | --- |
| manual 1 | manual 2 | manual 3 |
| 21 | 17 | 31 |
| 27 | 25 | 28 |
| 29 | 20 | 22 |
| 23 | 15 | 30 |
| 25 | 23 | 24 |
| Sum 125 | 100 | 135 |
|  |  |  |

The sample mean learning times,  are different but the question, as usual, is whether the population mean times are different. To answer this question, we must first compute the mean of the sample variance and the variance of the sample means.

**i.** **The mean of the sample variance**

 The standard deviation "s" of a sample of size n is computed by the formula:

 

 the variance would be

 

 The divisor in a sample variance calculation is called the number of degrees of freedom. Thus the variance for sample of size n has n-1 degrees of freedom.

**Table 5.2**

**Sample variance calculation**

|  |  |  |
| --- | --- | --- |
| manual 1 learning times x |  |  |
| 21 | 21 - 25 - -4 | 16 |
| 27 | 27 - 25 =2 | 4 |
| 29 | 29 - 25 = 4 | 16 |
| 23 | 23 - 25 = -2 | 4 |
| 25 | 25 - 25 = 0 | 0 |
|  Sum = 125 |  |  |
|    |

 by the same calculation ,and 

The denominator of the variance ratio is then

  for mean of k sample

 variances 

 mean of sample variance = 

Here there are n - 1 = 4 degrees of freedom in each of the K = 3 sample variance computations. Consequently, the mean of the three sample variances has k(n-1) = 3(4) = 12 degrees of freedom. In general, the mean of the sample variances has k(n-1) degrees of freedom if all samples are of size n.

**ii.** **The variance of the sample mean**

 In table 6.1, the sample means are

 

The mean of the numbers in all samples, denoted by  is called the grand mean. For k samples of the same size, n, the grand mean it’s the mean of the samples means. That is

 

 So in our example

 

The variance of k sample means is

 

 

 

  = 13 …with 2 degree of freedom

**The sample F statistic**

The variance ratio we discuss is called the sample F statistic



The numerator of the sample F has k - 1 degrees of freedom, and the denominator has k(n-1) degrees of freedom. These separate degrees of freedom are denoted by

 

 

For our example,

 Sample 

 

 

and the sample F has and  degrees of freedom.

Let's summarize all the calculations illustrated so far in Table 6.3 below.

# Table 5.3

Calculations of variances and the sample F

|  |  |  |
| --- | --- | --- |
| manual 1 | manual 2 | manual 3 |
| X |  | x |  | x |  |
| 21 | 16 | 17 | 9 | 31 | 16 |
| 27 | 4 | 25 | 25 | 28 | 1 |
| 29 | 16 | 20 | 0 | 22 | 25 |
| 23 | 4 | 15 | 25 | 30 | 9 |
| 25 | 0 | 23 | 9 | 24 | 9 |
| Sum= 125 | 40 | 100 | 68 | 135 | 60 |

1. sample means 
2. sample variances 

   

1. mean of sample variances 

 

1. Grand mean: 

 

e) Variance of sample means: 

 

 f) sample 

 

Now we must refer to a table of right-tail F values to determine whether the sample F = 4.64 is or is not, large enough to lead us to reject, at significance level  , the hypothesis that the population means are equal.

**5.5 The Distribution of the Sample F Statistic**

The probability distribution used in this chapter is the F distribution. It was named to honor Sir Ronald Fisher one of the founders of modern day statistics. It is applied when we want to compare several population means simultaneously. This simultaneous comparison of several population means is called analysis of variance (ANOVA)

The ANOVA test is based on the assumption that simple random samples are drawn independently from normal distribution that has the same variance. From these assumptions, mathematicians derived the probability distribution of the sample F statistics. The distribution has two parameters. They are the two degrees of freedom for the sample F,  and 

 = degrees of freedom for the numerator

  = degrees of freedom for the denominator

When the k samples are of the sample size n

 

 

where k is number of samples and

 n sample size

There is a separate F distribution for each pair of degrees of freedom. The distribution have one mode and are skewed to the right as is illustrated in the figure below

Less than 1% when v1 and v2 are at least 10.

 Fig 5.1

In general, the larger v1 and v2 are the smaller is the F value where the right rail almost touches the horizontal axis. in particular, when v1 and v2 are both at least 10, less than 1 percent of the area (probability) will be in the tail to the right of F = 5

**5.6 Some Characteristics of F Distribution**

1. F cannot be negative and it is a continuous distribution this is so because variances are never negative and n cannot be negative.
2. the F distribution is positively skewed
3. Large value of sample F indicates relatively large variation in the sample means
4. Its value range from 0 to  . As the value increase, the curve approaches the

 x-axis but never touches it.

The decision rule at significance level , will be to reject the null hypothesis of equal population means if the sample F exceeds.

i.e. Sample F > 

The symbol means the value of F for a right tail area of  when there are  and  degrees of freedom (Table at the back of the module) at the end of this module contain values for ) for our example v1 = 2 and v2 = 12 and for 

 

We use this value to formulate the ANOVA test decision rule.

**Figure 5.2**

* 1. **Steps In Performing ANOVA Test**

**1. State the hypotheses**

 

 the k population means are not equal

**2. Formulate the decision rule**

 the decision rule at significance level  is

 Reject if sample 

**3. Compute the value of sample statistic (sample F)**

 Sample 

**4. Make decision (accept or reject)**

 if we accept , we conclude that population means are equal.

In the context of our learning-time example we have

1.  : mean learning times for the three manuals are equal

  mean learning times for the three manuals are not all equal

2. k = 3 samples of n = 5 typists so

 

 

Taking  as significance level we find form table at the back that

 

Hence the decision rule is

 Reject  if sample F > 3.89

3. Sample F = 4.64 (see table 6.3)

4. The sample F = 4.64 exceeds the value 3.89 in the decision rule so

 We reject 

We conclude that mean learning times are not the same for the three manuals



**Figure 5.3**