(4) One-way Analysis of Variance test (ANOVA)

Use One-Way ANOVA (analysis of variance) to do the following when you have one categorical factor and a continuous response:

- Determine whether the means of two or more groups differ.
- Obtain a range of values for the difference between the means for each pair of groups.

For example, a carpet manufacturer wants to determine whether there are differences in durability among several types of carpet.

Where to find this analysis

Stat> ANOVA > One-Way ANOVA

The test steps

1) **The null hypothesis**

$$H_0 : \mu_1 = \mu_2 = ... = \mu_k$$

The Alternative hypothesis

$$H_1 : \mu_i \neq \mu_j \quad \forall \ i \neq j$$

2) **Test statistics: Calculate F0 from the ANOVA Table**

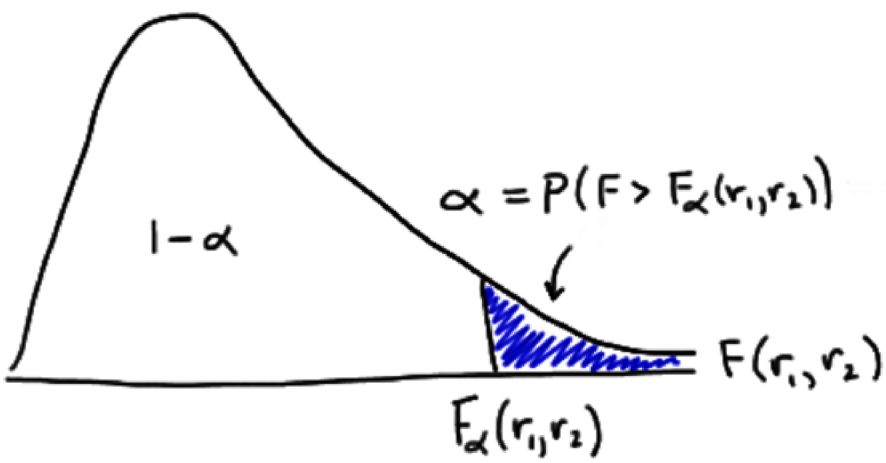
Source Variation	DF	SS	MS	F0
Treatments	k-1	SSTr	MSTr	MSTr/MSE
Error	n-k	SSE	MSE	
Total	n-1	SSTOT		

where

$$n = \sum_{j=1}^{k} n_j,$$

n_k **: is the number of the observation the j-th treatment.**

3) **Critical region**



where

$$r_1 = k - 1,\ r_2 = n - k$$

4) **Decision: Reject H0 when the value of the test statistic F0 belongs to the shaded area.**

Or one can use p-value approach.

Example

You design an experiment to assess the durability of four experimental carpet products. You place a sample of each of the carpet products in four homes and you measure durability after 60 days. Because you wish to test the equality of means and to assess the differences in means, you use the one-way ANOVA procedure (data in stacked form) with multiple comparisons. Generally, you would choose one multiple comparison method as appropriate for your data. However, two methods are selected here to demonstrate Minitab's capabilities.

1 Open the worksheet EXH_AOV.MTW.

2 Choose **Stat > ANOVA > One-Way**.

3 In **Response**, enter *Durability*. In **Factor**, enter *Carpet*.

4 Click **Comparisons**. Check **Tukey's, family error rate**. Check **Hsu's MCB, family error rate** and enter *10*.

5 Click **OK** in each dialog box.

Session window output

One-way ANOVA: Durability versus Carpet

Source	DF	SS	MS	F	P
Carpet	3	146.4	48.8	3.58	0.047
Error	12	163.5	13.6		
Total	15	309.9			

From the results, we see that

p-value=0.047<0.05 , So we cannot reject H0, that there is no significance differences between the durability means of the carpet products.

(5) Chi-square test

Test of independence

Use a test of independence to determine whether the observed value of one variable depends on the observed value of a different variable. For example, for an election, you might want to determine whether the candidate that a person votes for is independent of the gender of the voter.

Example

Is gender independent of education level? A random sample of 395 people were surveyed and each person was asked to report the highest education level they obtained. The data that resulted from the survey is summarized in the following table:

	High School	Bachelors	Masters	Ph.d.	Total
Female	60	54	46	41	201
Male	40	44	53	57	194
Total	100	98	99	98	395

Question: Are gender and education level dependent at 5% level of significance? In other words, given the data collected above, is there a relationship between the gender of an individual and the level of education that they have obtained?
Here's the table of expected counts:

	High School	Bachelors	Masters	Ph.d.	Total
Female	50.886	49.868	50.377	49.868	201
Male	49.114	48.132	48.623	48.132	194
Total	100	98	99	98	395

So, working this out, $\chi^2=(60-50.886)^2/50.886+\cdots+(57-48.132)^2/48.132=8.006$χ2=(60−50.886)2/50.886+⋯+(57−48.132)2/48.132=8.006

The critical value of $\chi 2$ with 3 degree of freedom is 7.815. Since 8.006 > 7.815, therefore we reject the null hypothesis and conclude that the education level depends on gender at a 5% level of significance.

Using Minitab

We can enter the data into Minitab and request that the 'Chi-square test' be conducted for the above hypotheses. The Minitab output for this example is shown below:

Chi-Square Test: High School, Bachelors, Masters, Ph.d.

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected
counts

          High
        School  Bachelors  Masters  Ph.d.  Total
    1       60         54       46     41    201
         50.89      49.87    50.38  49.87
         1.632      0.342    0.380  1.577

    2       40         44       53     57    194
         49.11      48.13    48.62  48.13
         1.691      0.355    0.394  1.634

Total      100         98       99     98    395

Chi-Sq = 8.006, DF = 3, P-Value = 0.046
```

(6) Correlation

have verbal and math SAT scores and first-year college grade-point averages for 200 students and we wish to investigate the relatedness of these variables. We use correlation with the default choice for displaying p-values.

1 Open the worksheet GRADES.MTW.

2 Choose **Stat > Basic Statistics > Correlation**.

3 In **Variables**, enter *Verbal Math GPA*. Click **OK**.

Session window output

Correlations: Verbal, Math, GPA

	Verbal	Math
Math	0.275	
	0.000	
GPA	0.322	0.194
	0.000	0.006

(7) Regression

The simple linear regression model is given by

$$Y_i = \alpha + \beta x_i + \epsilon_i.$$

The parameters are estimated using the least square method to be

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \quad \text{and} \quad \hat{\beta} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}.$$

Minitab can be used to calculate the estimation of the model as:

Stat > Regression > Regression

You can use Regression to perform simple regression using least squares .

Example

The following data represents the income and expenditure of 10 families. Estimate the simple linear regression model of the data and interpret the results.

Expenditure	Income
2400	4120
2650	5010
2350	5200
4950	6600
3100	4450
2500	3770
5106	7350
3100	3750
2900	5670
1750	3560

The Minitab results are:

Regression Analysis: Expenditure versus Income

```
The regression equation is
Expenditure = - 495 + 0.723 Income


Predictor    Coef  SE Coef      T      P
Constant   -495.0    839.4  -0.59  0.572
Income     0.7226   0.1648   4.39  0.002
```

From the results, we see that when the income increase by one SR, the expenditure increases by 0.723 SR.

When there is no income still one needs to borrow 495 SR for expenditure