http://wiki.stat.ucla.edu/socr/index.php/SOCR_Courses_2008_Thomson_ECON261

ANALYSIS OF VARIANCE - ANOVA Test

Grace S Thomson

Instructor

ANALYSIS OF VARIANCE- ANOVA Test

The analysis of Variance is a specific form of Hypothesis Test used when we are interested in testing the difference between more than 2 population parameters. When we studied Difference of means last week, we worked with 2 populations and 2 samples as estimators of such population, now we are able to work with as many populations and samples as we need.

The most important element in performing ANALYSIS OF VARIANCE is to define its type. There are two types of ANOVA tests: One Way ANOVA and PAIRED ANOVA TEST.

A One Way ANOVA test is performed when there is only ONE FACTOR that influences the behavior of the variables under study. For example, if you are interested in the average amount of sales in all your 10 locations, there is only 1 factor been analyzed, however you have 10 populations (N1….N10) to test. Since you won't be able to compute the information for the entire population, you will have to select samples in each one of these 10 locations (n1 ….to n10) with means for each location (mean 1, mean 2, ……, mean 10) and standard deviations for each one (s1……s10)

A Paired ANOVA Test is performed when there is more than one factor that influences the behavior of the variables under study. For example, if you are interested in the average amount of sales and the average amount of marketing to boost those sales in your 10 locations. There are 2 factors to be analyzed, so it is necessary to treat them as blocks and perform an analysis of Variance for paired differences as we learned before.

Let's start by ONE WAY ANOVA

ONE WAY ANOVA assumptions

1. All populations are normally distributed
2. Population variances are equal
3. Sampled observation are independent

CASE 1 - APPLICATION OF ONE-WAY ANOVA TEST

Fortune Relocation operates in three regions of the country, providing job search services, specialized training and resume development. The three regions of operation are west, southwest and northwest. The general manager has questioned whether the company's mean billing amount differed by region. He is interested in formulating an ANOVA test to determine this.

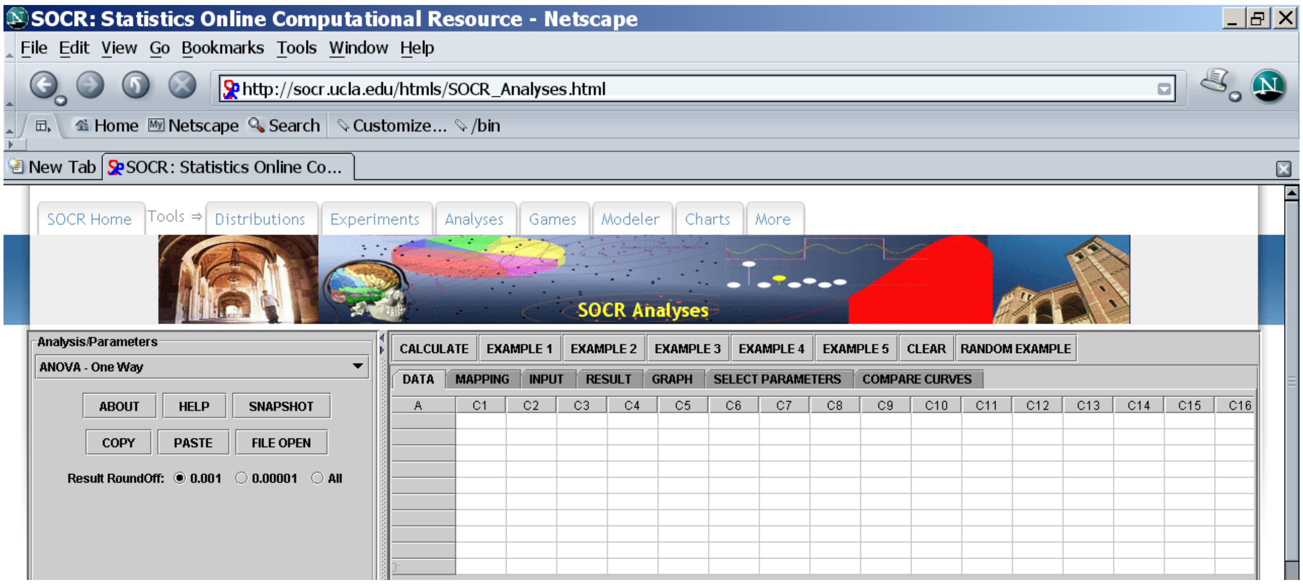
Simple random samples of employees served in these regions have been selected: 10 in the west, 8 in the southwest, and 12 in the northwest. The following sample data were collected:

West	Southwest	Northwest
$3,700	$3,300	$2,900
2.900	2,100	4,300
4,100	2,600	5,200
4,900	2,100	3,300
4,900	3,600	3,600
5,300	2,700	3,300
2,200	4,500	3,700
3,700	2,400	2,400
4,800		4,400
3,000		3,300
		4,400
		3,200

Apply the steps for ANOVA one way test at a 0.05 level of significance, to determine whether there is a statistically significant difference between the 3 regions.

Now let's try this using the tools from the Statistics Online Computational Resource (www.SOCR.ucla.edu)

1. Go to SOCR Analyses: **http://socr.ucla.edu/htmls/SOCR_Analyses.html**



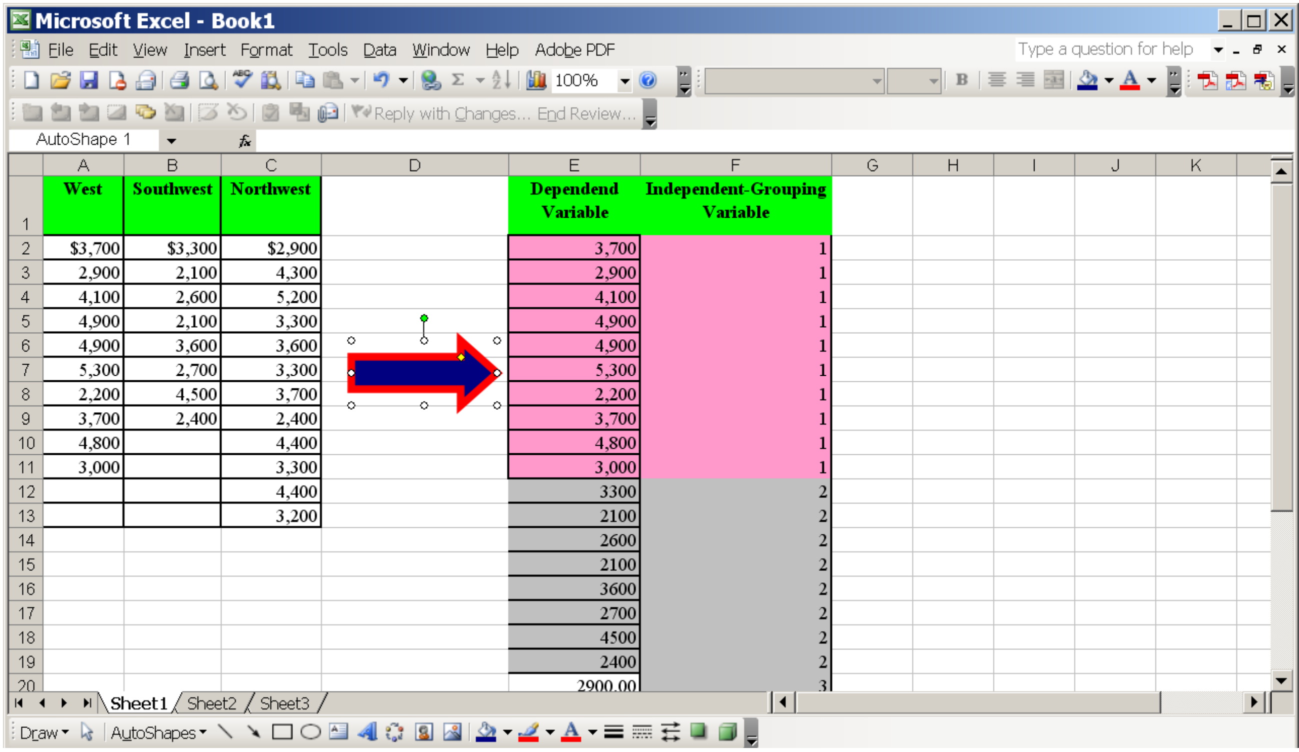
You can find help on how to use SOCR Analyses and see many examples here:

http://wiki.stat.ucla.edu/socr/index.php/SOCR_EduMaterials_AnalysesActivities

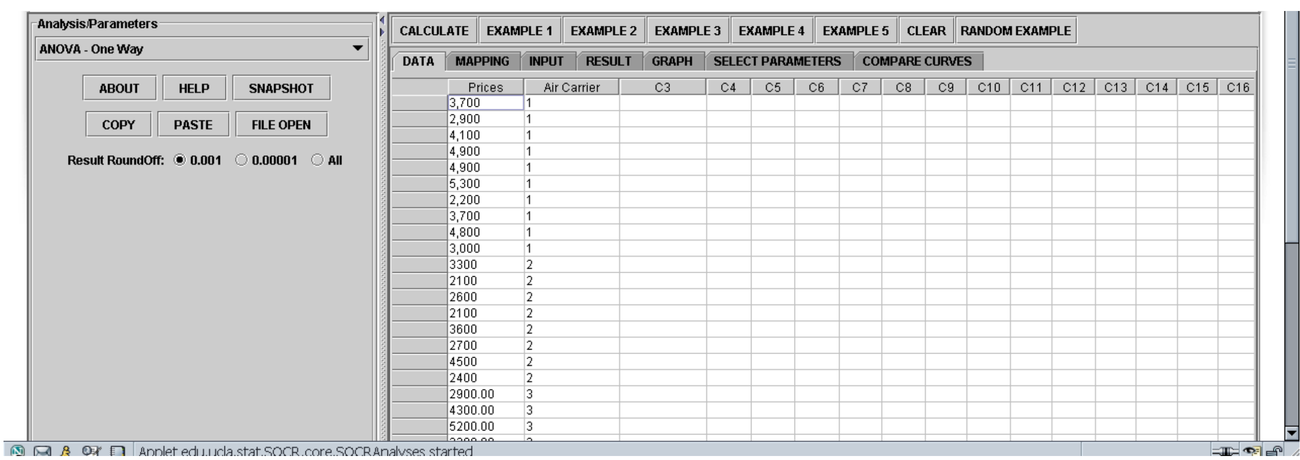
http://wiki.stat.ucla.edu/socr/index.php/Help_pages_for_SOCR_Analyses

http://wiki.stat.ucla.edu/socr/index.php/SOCR_EduMaterials_AnalysisActivities_ANOVA_1

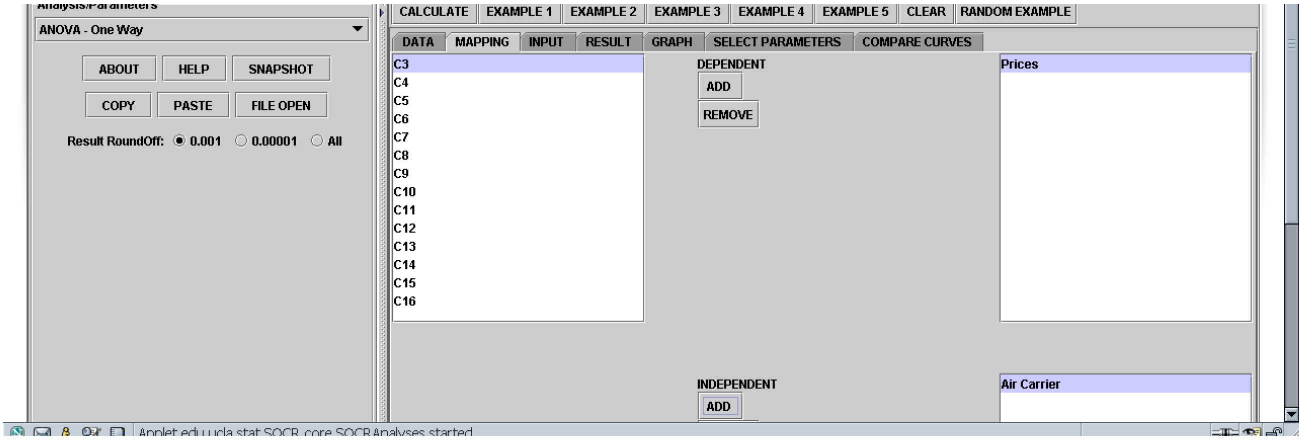
2. Copy the data from the table above and paste it in an Excel spreadsheet, using CTR-C + CTR-V and your mouse – you need to reconfigure the data to be in 2 columns (one for the observations and one for the grouping variable



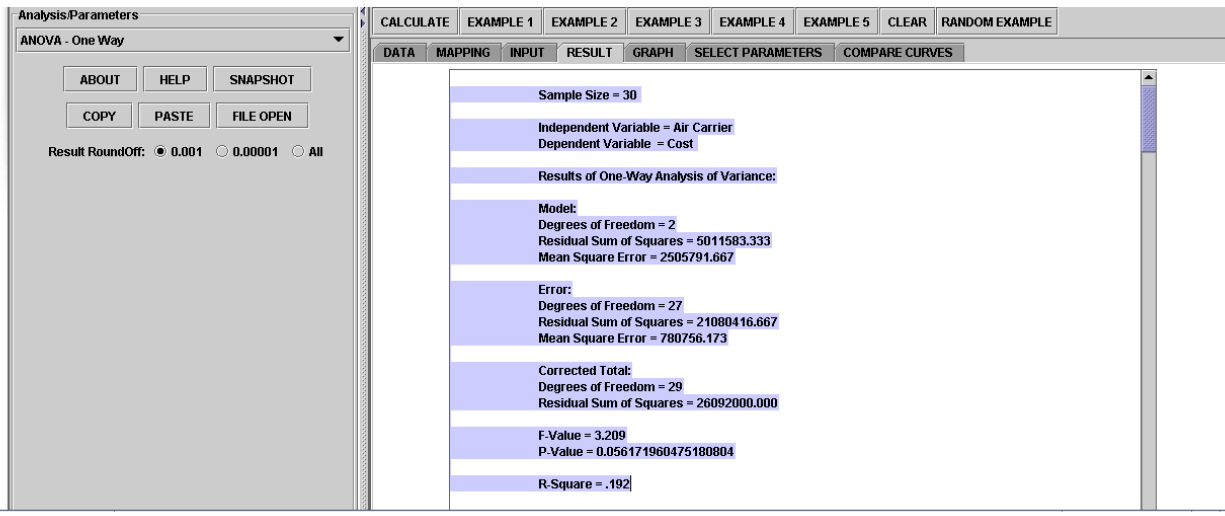
3. Copy the 2-column data from the Excel table and paste it in the SOCR spreadsheet, using CTR-C + CTR-V and your mouse



4. Map the columns (this provides the software with instructions which columns to use for the ANOVA analysis):



5. Click the **CALCULATE** button and go to the **RESULTS** tab to see the output of the ANOVA:

6. 

Notice the F-statistics (3.209) and the P-Value (0.056171960475180804)!

1. Specify parameter of interest

Mean dollars billed in each region

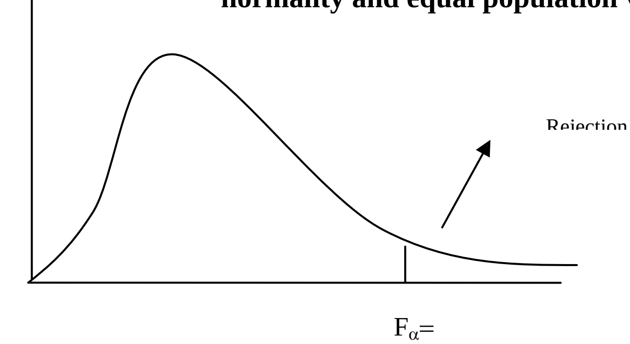
2. Formulate the null and alternative hypothesis

Ho: $\mu_w = \mu_{sw} = \mu_{nw}$

HA: Not all populations have the same mean

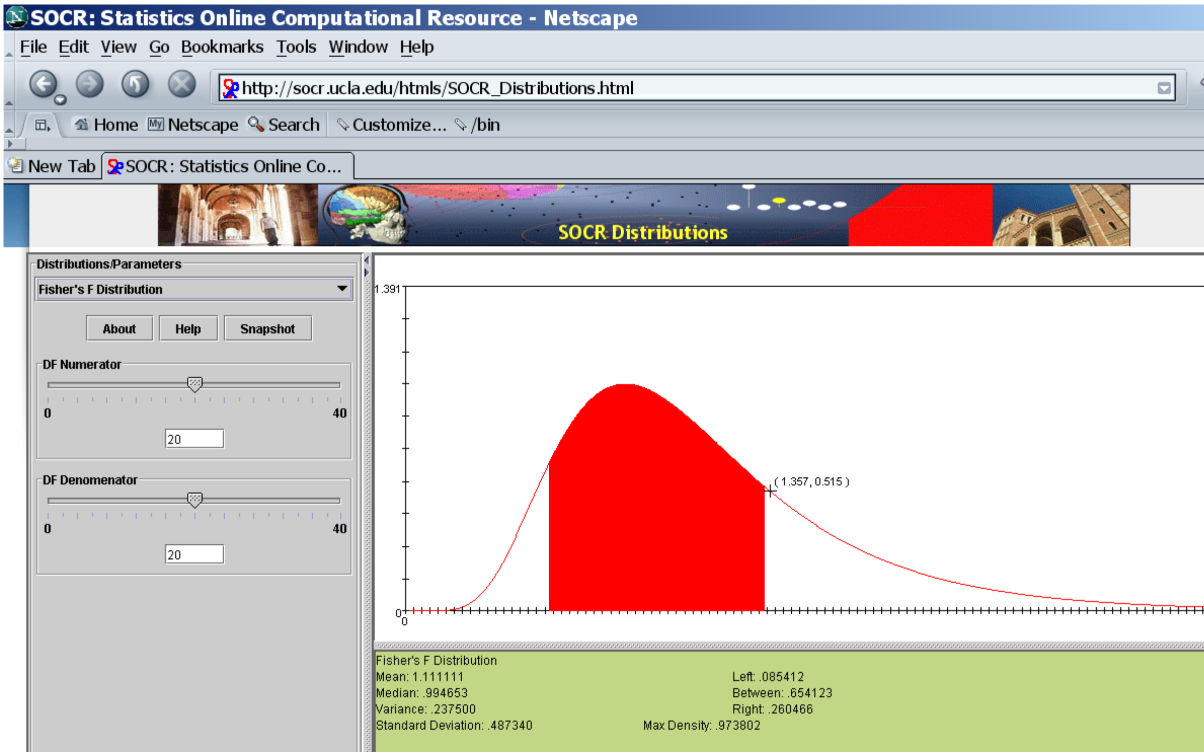
3. Assume normality and equal population variances.

Define rejection area.



If F test > F critical, Reject Ho

For Interactive demonstration of the F-distribution go to SOCR Distributions (http://socr.ucla.edu/htmls/SOCR_Distributions.html) and select Fisher's F-dsitribution (choose appropriate parameters). Then click on the graph and drag the limits to obtain the desired areas (some will correspond to the probability values in access of a critical score (see graph below).



4. Compute the test statistic F (from ANOVA TABLE)

Tools/ Data Analysis/ ANOVA: Single factor

Define data range

Specify α level of significance

From the table we can extract the F-test value= 3.209442

And the value of F-critical = 3.354131

Anova:

Single Factor

SUMMARY

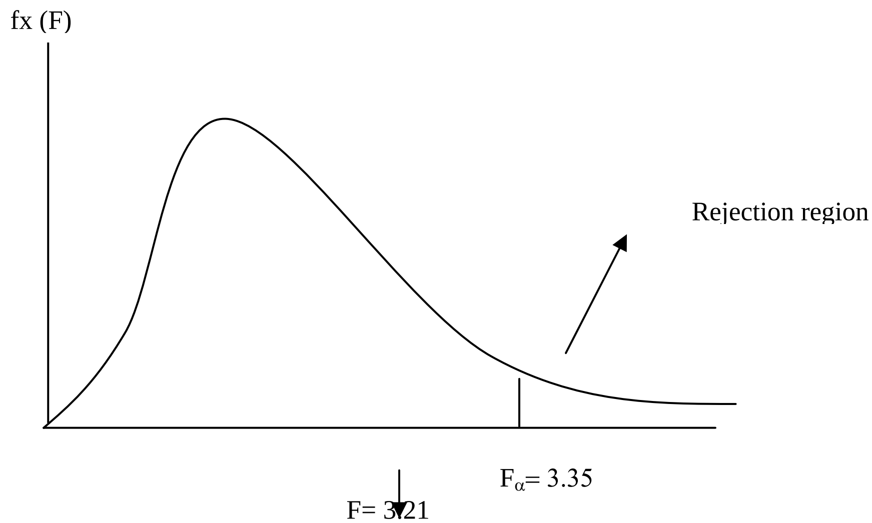
Groups	Count	Sum	Average	Variance
West	10	39500	3950	1062778
Southwest	8	23300	2912.5	695535.7
Northwest	12	44000	3666.667	604242.4

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	5011583	2	2505792	3.209442	0.056172	3.35413
Within Groups	21080417	27	780756.2			
Total	26092000	29				

5. Reach a decision: Reject or Not Reject

Because 3.209442 < 3.354 We Do **not Reject** null hypothesis of equality of means



6. Draw a conclusion: Is there difference or no difference?

"We are not able to detect a difference in the mean billing per customer per region"

SUMMARIZING

In order to perform an ANOVA test, you will need to do the following:

1. To confirm normal distribution of means → Draw a histogram
2. To confirm equality of variances → assume equality
3. To test equality of means → Use F statistic vs. F critical

If F test statistic > critical F_{α} → Reject Ho of equal means

Or use p-value

If p-value < α → Reject Ho of equal means

When you reject Ho, you are saying that the means are not equal.

- If you reject the Hypothesis of equality, your next step is to test what pair of means is not equal.

Case 2

Accubrakes makes disc brakes for automobiles and the Research & Development department tested four brake systems to determine if there is a difference in the average stopping distance among them. 40 identical mid-sized cars were driven on a test track. Then cars were fitted with Brake A, 10 with Brake B, and so forth. The number of feet required to bring the car to a full stop was recorded.

Here is the table with the recorded information:

Car	Brake A	Brake B	Brake C	Brake D
1	274	277	264	283
2	259	267	258	270
3	275	271	257	281
4	276	267	264	259
5	278	279	269	258
6	283	272	257	259
7	262	287	265	270
8	272	269	264	259
9	275	262	257	257
10	269	261	268	255

Column A: Sample (car)

Column B: Stopping distance for brake A

Column C: Stopping distance for brake B

Column D: Stopping distance for brake C

Column E: Stopping distance for brake D

Formulate a hypothesis test to determine whether the four brake systems have the same or different mean stopping distances. If a significant difference is found among the mean stopping distances, run a post-test to determine which populations have different means.

SOLUTION

Anova: Single Factor

SUMMARY

Groups	***Count***	***Sum***	***Average***	***Variance***
Brake A	10	2723.59	272.359	49.90007
Brake B	10	2713.299	271.3299	61.85569
Brake C	10	2623.14	262.314	21.73556
Brake D	10	2652.357	265.2357	106.4385

ANOVA

Source of Variation	***SS***	***df***	***MS***	***F***	***P-value***	***F crit***
Between Groups	699.1636	3	233.0545	3.885378	0.016685	2.866266
Within Groups	2159.369	36	**59.98246**			
Total	2858.532	39				

See you in class!