

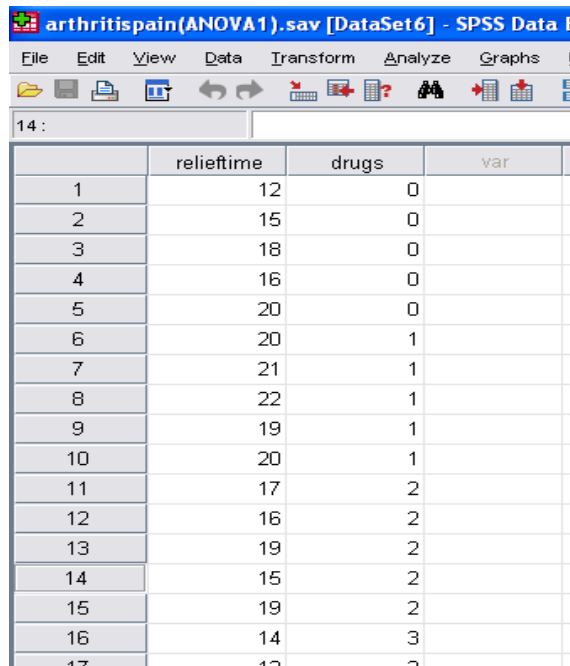
## One-way ANOVA tutorial

For one-way ANOVA we have 1 dependent variable and 1 independent variable [factor] which as at least 2 levels.

### Problem description

A pharmaceutical company is interested in the effectiveness of a new preparation designed to relieve arthritis pain. Three variations of the compound have been prepared for investigation, which differ according to the proportion of the active ingredients: T15 contains 15% active ingredients, T40 contains 40% active ingredients, and T50 contains 50% active ingredients. A sample of 20 patients is selected to participate in a study comparing the three variations of the compound. A control compound, which is currently available over the counter, is also included in the investigation. Patients are randomly assigned to one of the four treatments (control, T15, T40, T50) and the time (in minutes) until pain relief is recorded on each subject. The data is available at U:\\_MT Student File Area\hjkim\STAT375\SPSS tutorial\arthritispain.sav.

	Control	T15	T40	T50
1	12	20	17	14
2	15	21	16	13
3	18	22	19	12
4	16	19	15	14
5	20	20	19	11

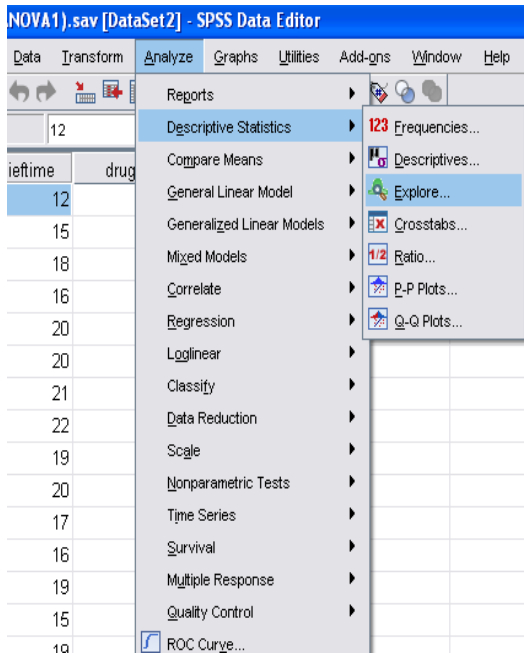


	relieftime	drugs	var
1	12	0	
2	15	0	
3	18	0	
4	16	0	
5	20	0	
6	20	1	
7	21	1	
8	22	1	
9	19	1	
10	20	1	
11	17	2	
12	16	2	
13	19	2	
14	15	2	
15	19	2	
16	14	3	
17	13	3	

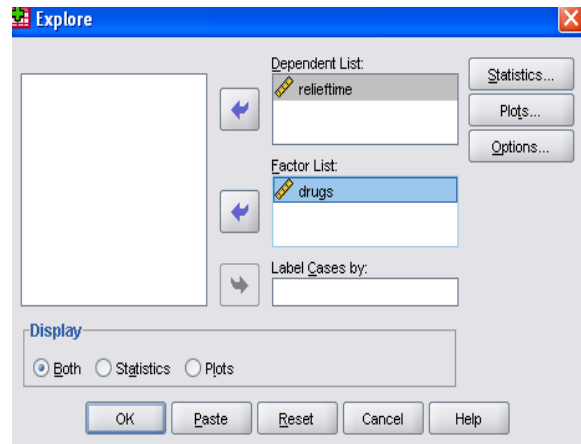
In SPSS, the data should be entered the following manner.

### Graphical Analysis [stem and leaf plots/boxplots/normal plots]

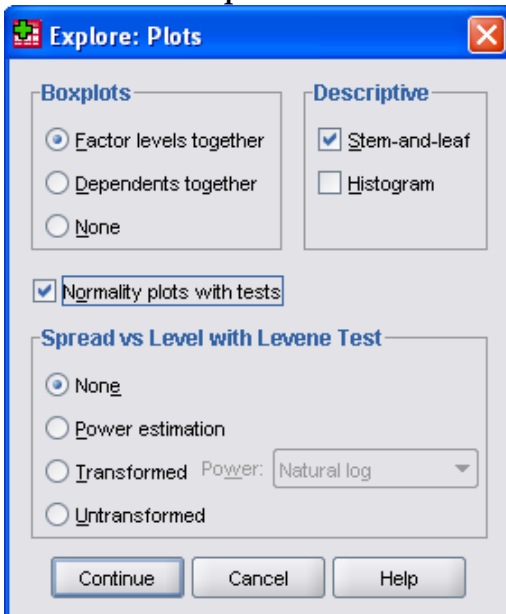
We will begin the ANOVA by assessing the necessary assumption of normality and equal variance. To do this we will need to create boxplots, stem and leaf plots, and normal plots. These options can be found by going to the **Analyze-Descriptive Statistics-Explore** pull down menus.



This will take you to the **Explore** window. We need to place “relieftime” in the dependent list as it is our response. “drug” should be placed in the factor field as it is our independent variable/factor in this experiment.



Having defined our dependent/response variable and our independent/factor variable we can now set up our plots. By pushing on the **Plots** button, found on the image above, we move into the **Explore Plots** window.



We need only to make sure that the three boxes are checked and then we can hit the **continue** button. This returns us to the Explore window where we click on the **OK** button and review the output from the analysis.

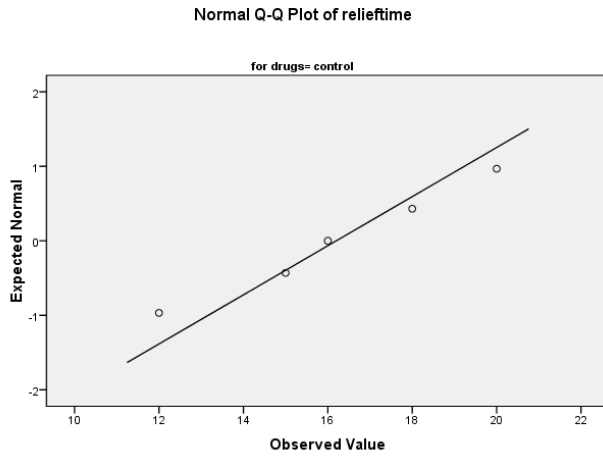
You should find a lot of output in the SPSS output window. Among the information are the stem & leaf plots, normal plots and boxplots [partial examples shown below].

relieftime Stem-and-Leaf Plot for  
drugs= control

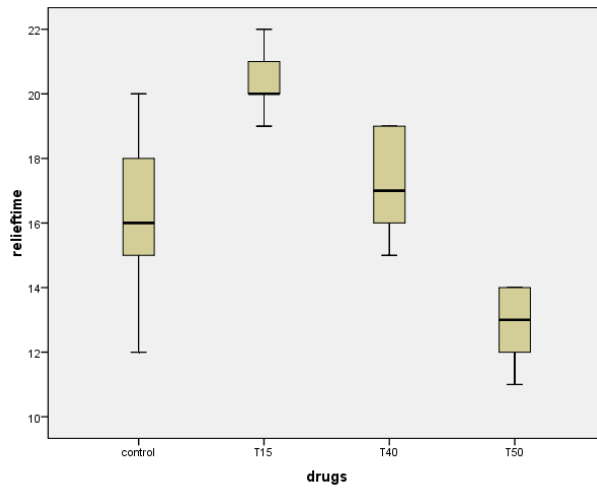
Frequency Stem & Leaf

1.00	1 . 2
3.00	1 . 568
1.00	2 . 0

Stem width: 10  
Each leaf: 1 case(s)

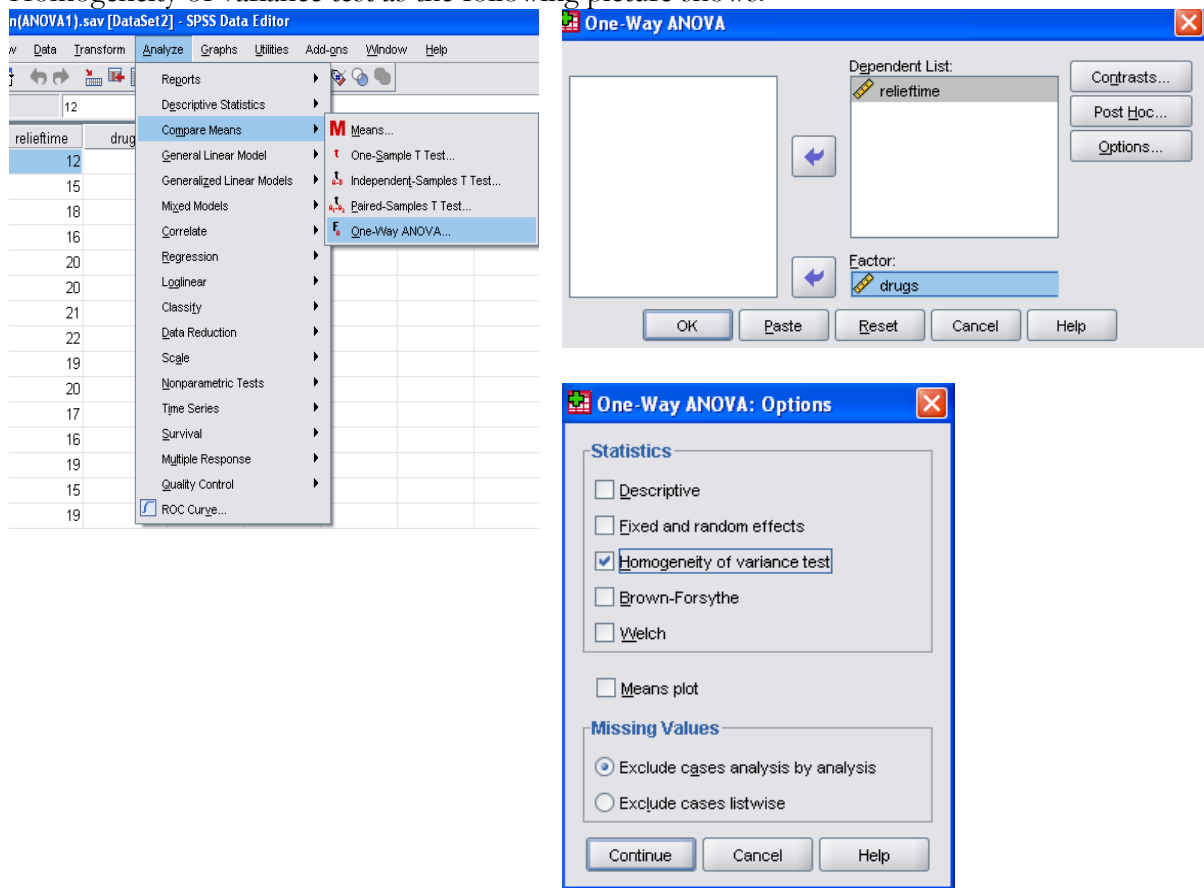


Boxplot



You can click on any of the outputs to bring up an editor window if you would like to add text or edit the images. Additionally, if you wish to cut and paste individual parts you need only to click on the image and use the copy command from the main tool bar.

For homogeneous variance assumption click the options under ANOVA window and check Homogeneity of variance test as the following picture shows.



This generates the output and p-value is .197 that is larger than 0.05. We fail to reject the null hypothesis of equal variances.

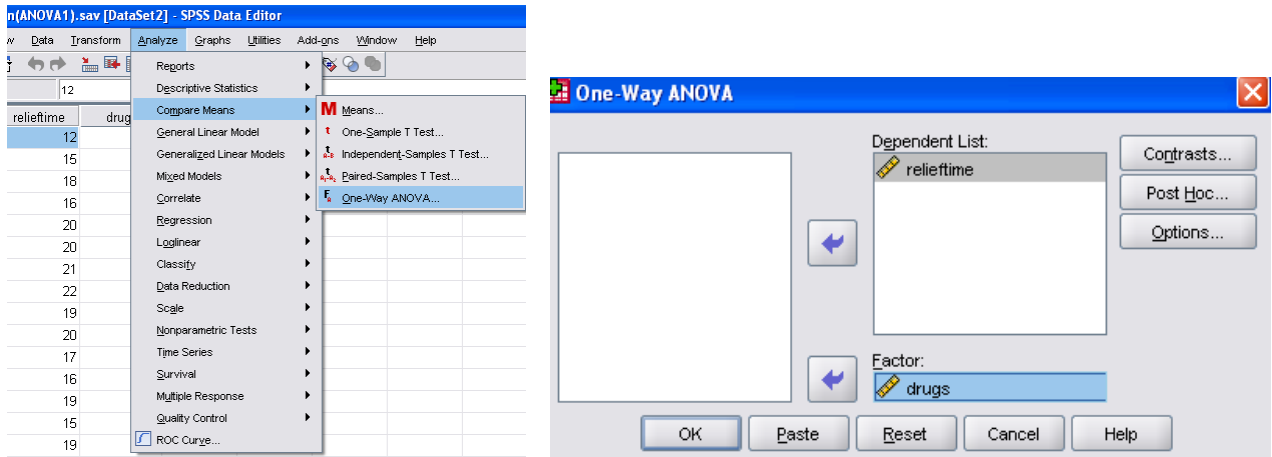
#### Test of Homogeneity of Variances

Relieftime

Levene Statistic	df1	df2	Sig.
1.751	3	16	.197

#### Analysis of Variance - ANOVA

Assuming not problems with our assumptions we can continue by running the one-way ANOVA. We click on **Analyze-Compare Means-One way ANOVA** in the pull down menus to start the process. Once again we need to define which of our variables is the dependent variable and which is the factor variable. After moving each of them to the proper field we hit OK and the analysis window will appear with the ANOVA output.



### Output

Recall that the null hypothesis in ANOVA is that the means of all the groups are the same and the alternative is that at least one is different . So for our example with 4 treatment groups

$$H_o : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_A : \text{at least one mean is different}$$

To check the hypothesis the computer compares the value for the observed F [12.723] to the expected value for F-observed given the number of groups and the sample sizes. If this is a rare event [it will be unusually large is Ho is not true] we will reject Ho. To determine if F-observed is unusual, we need to look at the (Sig)nificance of the F value [often called the P-value]. If this value is less than .05, it means that a score this large would occur less than 5% of the time [or 1 in every 20 trials] and we will consider it sufficiently rare. The smaller this value gets the more rare the score and the more certain we can be that the null hypothesis is incorrect. In the case of above example, we can be confident that if we reject the null hypothesis, there is less than a 1 in a 1000 chance that we would be incorrect. We limit ourselves to 1 in a 1000 as [Sig=.000] does not mean that the probability of getting a score this large is zero, just that its equal to zero at three significant figures [it could be Sig=.0004].

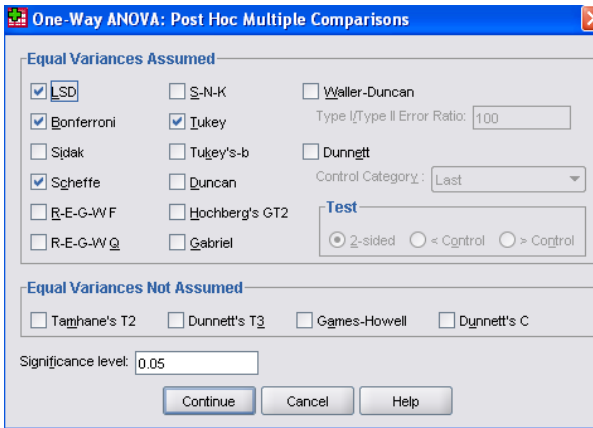
ANOVA

relieftime	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	146.950	3	48.983	12.723	.000
Within Groups	61.600	16	3.850		
Total	208.550	19			

### Multiple Comparisons

#### Post hoc Tests

Using ANOVA table we reject the null hypothesis and conclude that at least one mean is different from the others. The next question is how they are different. To answer this question, SPSS provide post hoc tests. Click Post hoc in ANOVA window, then the following window will appear.



There are many options to choose from for the post hoc. In this class we will consider LSD, Tukey, and Scheffe. LSD is the most liberal post hoc test and Scheffe is the most conservative post hoc test. Click Continue and OK, and we will obtain the following output.

### Homogeneous Subsets

		Relieftime			
		N	Subset for alpha = 0.05		
drugs			1	2	3
Tukey HSD <sup>a</sup>	T50	5	12.80		
	control	5	16.20	16.20	
	T40	5		17.20	17.20
	T15	5			20.40
	Sig.		.063	.851	.085
Scheffe <sup>a</sup>	T50	5	12.80		
	control	5	16.20	16.20	
	T40	5		17.20	17.20
	T15	5			20.40
	Sig.		.096	.884	.126

Means for groups in homogeneous subsets are displayed.

This means by HSD and Scheffe, T50 group mean has the least response time but this is not significantly different from Control group, and T15 group mean has the most response time but this is also not significantly different from T40 group. Notice that LSD does not provide the homogeneous subsets. But multiple comparison table provides this information. \* means two groups are significantly different based on given post hoc test.

### Contrast

A priori contrast: comparisons identified before the experiment has been performed. For this example we will compare 3 different contrasts

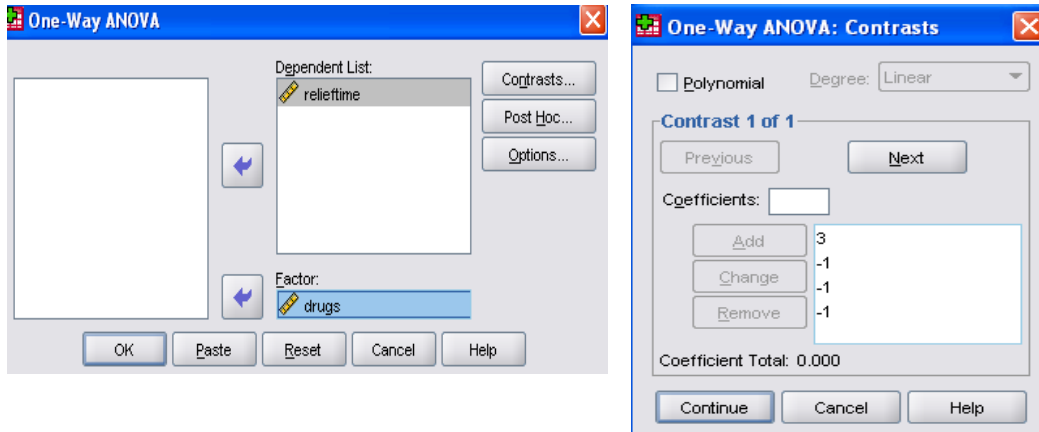
1. Control vs. Compound
2. Low compound (T15) vs. High compound (T40, T50)
3. T40 vs. T50

Creating contrasts: Only requirement is that the sum of the coefficients = 0

1. (1, -1/3, -1/3, -1/3) → (3, -1, -1, -1)
2. (0, 1, -1/2, -1/2) → (0, 2, -1, -1)
3. (0, 0, 1, -1) → (0, 0, 1, -1)

### Entering contrasts

To enter the contrasts into SPSS we bring up the One-Way ANOVA screen and click on the Contrasts button. SPSS menu will only allow you to enter 2 decimal places on your coefficients which results in some round-off problems. To circumvent the problem we will convert all of the coefficients into whole numbers. (This will not affect the statistical significance of the contrasts.) By clicking Next button, you can add other contrasts.



After entering all three contrasts, click Continue and OK. This will generate the following output.

**Contrast Coefficients**

Contrast	drugs			
	control	T15	T40	T50
1	3	-1	-1	-1
2	0	2	-1	-1
3	0	0	1	-1

**Contrast Tests**

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
relieftime	Assume equal variances	1	-1.80	3.040	-.592	16	.562
		2	10.80	2.149	5.025	16	.000
		3	4.40	1.241	3.546	16	.003
	Does not assume equal variances	1	-1.80	4.219	-.427	4.611	.689
		2	10.80	1.421	7.599	10.158	.000
		3	4.40	.990	4.445	7.315	.003

Using p-values we can conclude that there is significant difference between low compound and high compound and between T40 and T50. However, there is no significant difference between control and compound group.