ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ
ΠΜΣ «Μεθοδολογία Βιοϊατρικής Έρευνας, Βιοστατιστική και Κλινική Βιοπληροφορική»

Develop software in Python for performing Independent t-test and One-way ANOVA with post hoc test considering the Bonferroni's adjustment.

by

Boltsi Aggeliki

(August, 2015)

Supervisor: Axel Kowald

Evaluators: Elias Zintzaras, George Rachiotis

Abstract

The independent t-test, also called the two sample t-test, is a statistical test that determines whether there is a statistically significant difference between the means in two unrelated groups. The independent-samples t test is commonly referred to as a between-groups design, and can also be used to analyze a control and experimental group. With an independent-samples t test, each case must have values on two variables, the grouping (independent) variable and the test (dependent) variable. The grouping variable divides cases into two mutually exclusive groups or categories, such as boys or girls for the grouping variable gender, while the test variable describes each case on some quantitative dimension such as test performance. The t test evaluates whether the mean value of the test variable (e.g., test performance) for one group (e.g., boys) differs significantly from the mean value of the test variable for the second group (e.g., girls).

The one-way analysis of variance (ANOVA) is used to determine whether there are any significant differences between the means of two or more independent (unrelated) groups. For example, you could use a one-way ANOVA to understand whether exam performance differed based on test anxiety levels amongst students, dividing students into three independent groups (e.g., low, medium and high-stressed students). Also, it is important to realize that the one-way ANOVA is an *omnibus* test statistic and cannot tell you which specific groups were significantly different from each other; it only tells you that at least two groups were different. Since you may have three, four, five or more groups in your study design, determining which of these groups differ from each other is important. You can do this using a post-hoc test.

Post hoc tests are designed for situations in which the researcher has already obtained a significant omnibus F-test with a factor that consists of three or more means and additional exploration of the differences among means is needed to provide specific information on which means are significantly different from each other. We do individual comparisons between groups, e.g.: To compare a group with the group b, using the t-test and adjust the sig values with a Bonferroni adjustment. For Bonferroni adjustment the p value (which must be achieved for significance) is divided by the number of paired comparisons.

Table of Contents

Independent t Test

One-way ANOVA

Chapter 1 – Introduction

Hypothesis for the independent t-test:

The null hypothesis for the independent t-test is that the population means from the two unrelated groups are equal:

- H_0: $u_1 = u_2$

In most cases, we are looking to see if we can show that we can reject the null hypothesis and accept the alternative hypothesis, which is that the population means are not equal:

- H_A: $u_1 \neq u_2$

To do this, we need to set a significance level alpha that allows us to either reject or accept the alternative hypothesis. Most commonly, this value is set at 0.05.

Assume that we have two completely different (independent) groups of subjects that we want to compare and to determine if they are significantly different from one another: a between-groups design. A one sample t-test allows us to test whether a sample mean (of a normally distributed interval variable) significantly differs from a hypothesized value or population mean. An independent samples t-test is used when you want to compare the means of a normally distributed interval dependent variable for two independent groups. The classic example of this is when you have a sample and you randomly assign half of your subjects to the control condition and the other half to the experimental treatment condition. In this situation, we wish to compare the means of the two conditions/groups. We can no longer assume that we know a population mean / a hypothesized value and we must develop a new sampling distribution.

The hypotheses of interest in an ANOVA are as follows:

- H_0:$\mu_1 = \mu_2 = \mu_3 \ldots = \mu_k$ (H_0 is the null hypothesis and k is the number of conditions).
- H_1: Means are not all equal.

where k = the number of independent comparison groups.

The null hypothesis tested by ANOVA is that the population means for all conditions are the same.

Analysis of variance is a method for testing differences among means by analyzing variance. The test is based on two estimates of the population variance (σ^2). One estimate is called the mean square error (MSE) and is based on differences among scores within the groups. MSE estimates σ^2 regardless of whether the null hypothesis is true (the population means are equal). The second estimate is called the mean square between (MSB) and is based on differences among the sample means. MSB only estimates σ^2 if the population means are equal. If the population means are not equal, then MSB estimates a quantity larger than σ^2. Therefore, if the MSB is much larger than the MSE, then the population means are unlikely to be equal. On the other hand, if the MSB is about the same as MSE, then the data are consistent with the null hypothesis that the population means are equal.

Chapter 2 – Methods (Theory)

Independent t Test

Sampling Distribution of the Difference between the Means

To test for the potential statistical significance of a true difference between sample means, we need a sampling distribution of the difference between sample means (**Difference (D) = Mean 1 – Mean 2**). This would be a sampling distribution that will provide us with the probability that the difference between our two sample means differs from the null hypothesis population of sample mean differences: a population in which there is no difference between samples or, restated, the independent variable has no effect. The sampling distribution of the difference between the means can be created by taking all possible sample sizes of n1 and n2, calculating the sample means, and then taking the difference of those means. If you do this repeatedly for all of the possible combinations of your sample sizes, then you end up with a family of distributions of differences between the two means when they are randomly drawn from the same null hypothesis population.

- But, how confident are we that the difference between the means (D) deviates (is far way) from zero, i.e. is it significant? As we can see, in the following math formula, the t value, except from the difference between means, depends on the variability and the size of the trial n=n1+n2 patients (i.e. the error)

  Where n1 is the sample size of participants in the first group and n2 in the second group. - Error has two factors: variability and size of the trials.

  We must consider the difference between the two means (D) in conjunction with the error of this Difference (SE), i.e. the overall variability (the SD of the two statements) and the size of the trial (n).

  Then, we could test statistically whether the difference between the two means deviates from zero (i.e. is significant) using the t-test:

$$\mathbf{t}=\frac{(\mathbf{mean}\,1)-(\mathbf{mean}\,2)}{\mathbf{SE}}=\frac{\overline{\mathbf{x}}1-\overline{\mathbf{x}}2}{\mathbf{SE}}$$

$$\text{where } \mathbf{SE}=\sqrt{\mathbf{s}^2\left(\frac{1}{\mathbf{n}_1}+\frac{1}{\mathbf{n}_2}\right)} \text{ and } \mathbf{s}^2=\frac{\sum_{j=1}^{n1}(\mathbf{x}_j-\mathrm{X}1)^2+\sum_{i=1}^{n2}(\mathbf{x}_i-\mathrm{X}2)^2}{(\mathbf{n}_1-1)+(\mathbf{n}_2-1)}$$

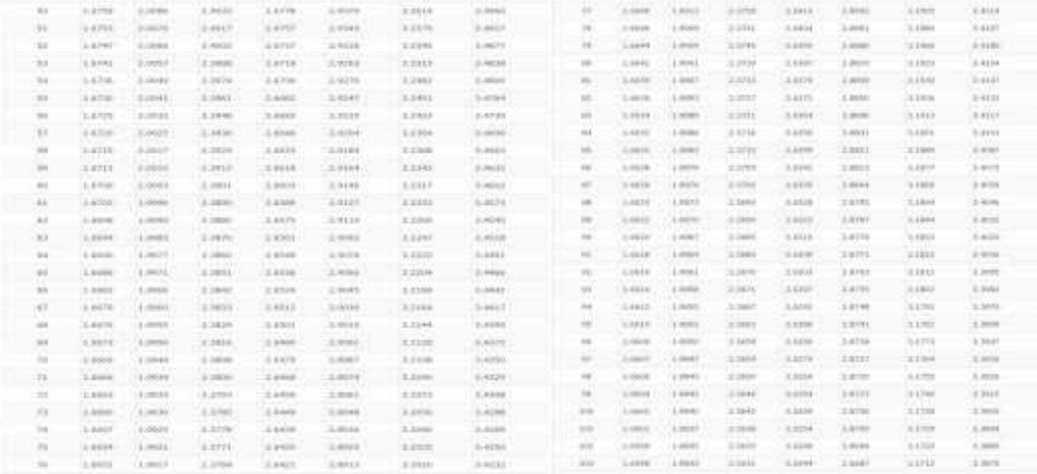
In fact, what we do is estimate the true population variability (or variance, s^2) by taking the average variance of our samples but weighted by their respective sample sizes. Sample size or degrees of freedom affects the accuracy of our variance estimates, so an estimate from a sample with a large sample size would be more accurate than an estimated variance from a smaller sample. So we need to weight our average variance by the respective sample sizes of each sample. In using this approach, we are going to make a new assumption—that the sample variances are estimating the same underlying population variance, the variance of the null hypothesis population.

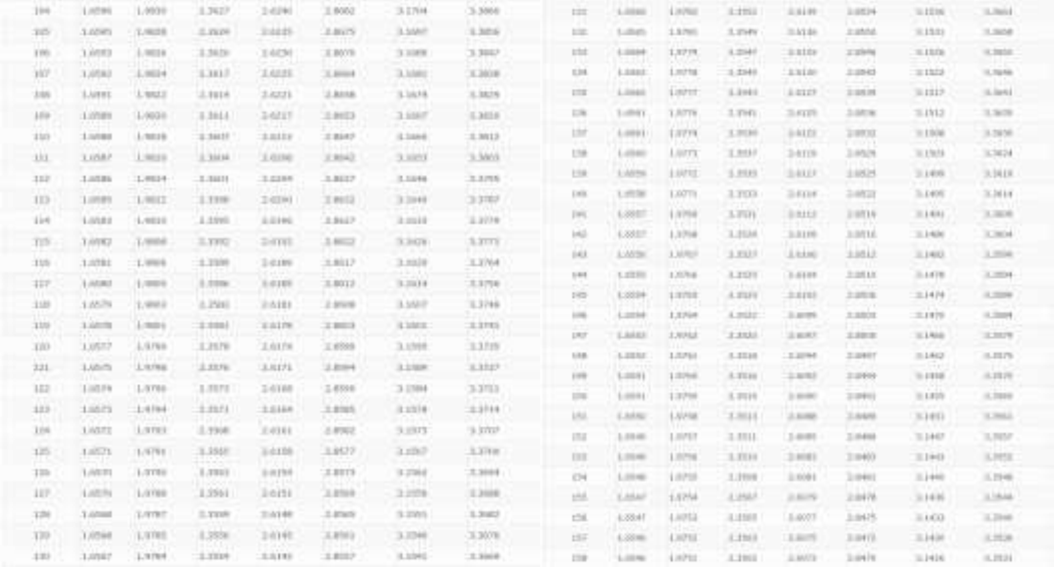
<u>Significance of the difference, P-value</u>

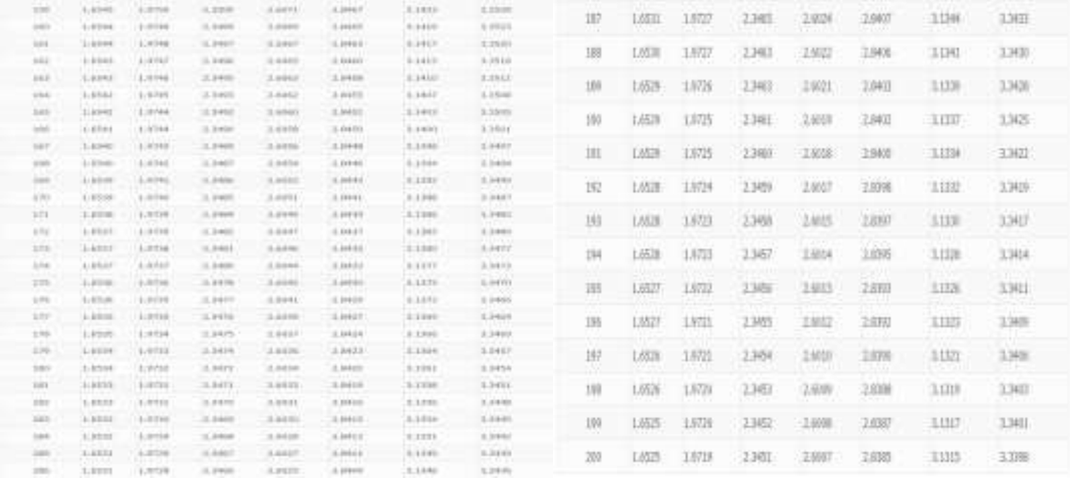
- We have to answer the following question: How confident are we that the value t is different from zero, i.e. significant; alternatively, what is the error probability (i.e. the P-value, or the probability of false-positive result) for claiming that the t is significant?

  If t is different from zero then, we claim the difference between means (taking into account the variability of the data and the size of the trial) is significant (i.e. different from zero).
  We could answer the question be comparing the value t (the sign is ignored) with the value 5% point of the t-distribution with n1+n2-2 df (see Table of t-distribution)

T Distribution Critical Values Table







If we simulate the study 10000 times by randomly permuting all the numbers and after every permutation we calculate the t-test, we expect 5% of the random t-tests to be greater of + 5%point or less - 5%point.

If the value t is larger than 5%point, we can claim that the t is not a random value, but a significant one (i.e. different from zero), with a probability error (P-value) $P<0.05$ (i.e. a small error probability).

Thus, we may argue that the difference between the two means is significant with a probability error $P<0.05$.

Confidence interval (CI) of the difference between two means.

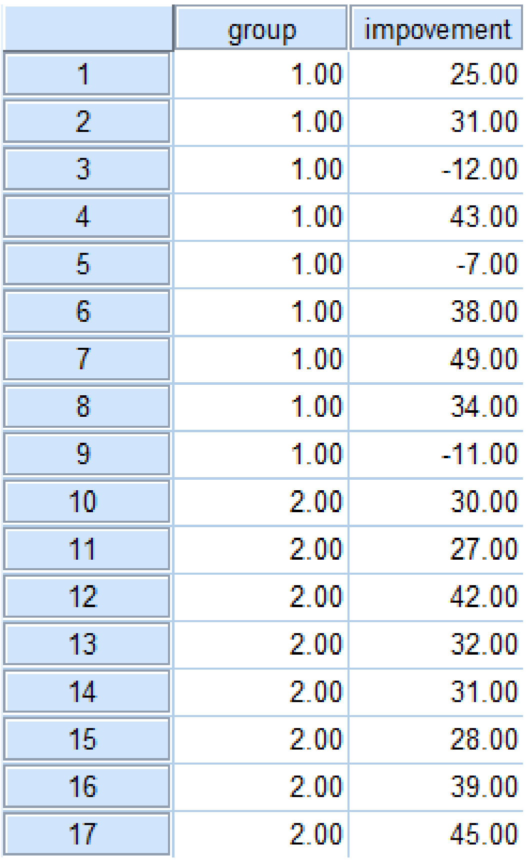
The significance of the difference between the two means D can also be assessed using the 95% CI. The 95% CI is defined as:

(D-t*SE, D+t*SE) t is the 5% point of the t-distribution for n1+n2-2 df

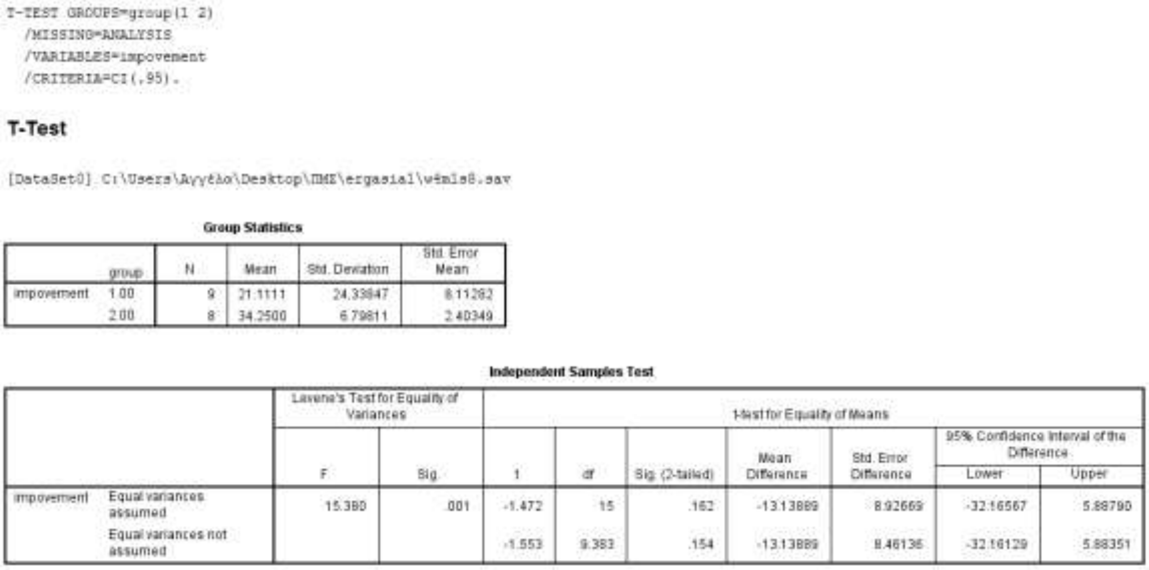
If zero is not included in the 95% CI, there is significance difference between the two treatments.

Complete Example in SPSS

To test the effectiveness of treatment of RRMS in two groups of patients and whether it differs, we are checking the annual frequency of relapses with t-test for two independent samples.



	group	impovement
1	1.00	25.00
2	1.00	31.00
3	1.00	-12.00
4	1.00	43.00
5	1.00	-7.00
6	1.00	38.00
7	1.00	49.00
8	1.00	34.00
9	1.00	-11.00
10	2.00	30.00
11	2.00	27.00
12	2.00	42.00
13	2.00	32.00
14	2.00	31.00
15	2.00	28.00
16	2.00	39.00
17	2.00	45.00



```
T-TEST GROUPS=group(1 2)
  /MISSING=ANALYSIS
  /VARIABLES=impovement
  /CRITERIA=CI(.95).
```

T-Test

```
[DataSet0] C:\Users\Αγγέλο\Desktop\ΠΜΣ\ergasial\v4mls8.sav
```

Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
impovement	1.00	9	21.1111	24.33847	8.11282
	2.00	8	34.2500	6.79811	2.40349

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
impovement	Equal variances assumed	15.380	.001	-1.472	15	.162	-13.13889	8.92669	-32.16567	5.88790
	Equal variances not assumed			-1.553	9.383	.154	-13.13889	8.46136	-32.16129	5.88351

Remark:

We see from the table above the 95% CI includes zero and the value of P = 0.162> 0.05 so our initial hypothesis that the two treatments did not differ in their effectiveness is correct.

One-way ANOVA

Analysis of Variance is a test that looks at the variance, or ways in which data is different, for more than two groups. One-way ANOVA is used for groups that only have one independent variable. It checks if the scores within each group vary about the mean, which is called within group variance. It also checks if the means between the groups vary, which is called between group variance. When you do an one-way ANOVA and get significant results it means that there is some difference between the groups, you must do further tests to determine where the difference is.

The basic logic behind the ANOVA:

If we have 4 groups, group 1 could differ from groups 2-4, groups 2 and 4 could differ from groups 1 and 3, group 1 and 2 could differ from 3, but not 4, etc.

Since our hypothesis should be as precise as possible (presuming you're researching something that isn't completely new), you will want to determine the precise nature of these differences.

There are two sources of variation here, the *between group* and the *within group variation*. This gives us the basic layout for the ANOVA table.

Source	SS	df	MS	F
Between				
Within				
Total				

SS stands for Sum of Squares. It is the sum of the squares of the deviations from the means. In other words, each number in the SS column is a variation.

df stands for degrees of freedom.

MS stands for Mean Square. It is a kind of "average variation" and is found by dividing the variation by the degrees of freedom. So, each number in the MS column is found by dividing the number in the SS column by the number in the df column and the result is a *variance*.

F stands for an F variable. F was the ratio of two independent chi-squared variables divided by their respective degrees of freedom. So the F column will be found by dividing the two numbers in the MS column.

Filling in the ANOVA's table

Sum of Square = Variations

You can add up the two sources of variation, the *between group* and the *within group*.

df = Degrees of Freedom

Total degrees of freedom, N-1 ,where N is the number of treatments or groups.

If k groups were there in the problem (*we are comparing* between *the group)* there are k-1 degrees of freedom. In general, that is one less than the number of groups, since *k* represents the number of groups, that would be *k*-1.

This raises the question of how many degrees of freedom there are *within* the groups. Well, if there are N-1 degrees of freedom altogether, and k-1 of them were between the groups, then (N-1) - (*k*-1) = N-1-*k*+1=N-*k* of them are within the groups.

Mean Squares = Variances

The variances are found by dividing the variations by the degrees of freedom, so divide the SS(between) by the df(between) to get the MS (between) and divide the SS(within) by the df(within) to get the MS(within) .

There is no total variance. Well, there is, but no one cares what it is, and it isn't put into the table.

F

Once you have the variances, you divide them to find the F test statistic.

So, divide MS(between) by MS(within) to get F .

Ideally we want to ***maximize*** $MS_{between}$ or $MS_{treatment}$, because we're predicting that our treatment will differentially effect our groups.

MS_{within} or MS_{error} = average variance among subjects in the same group

Ideally we want to ***minimize*** MS_{error}, because -ideally- our treatment influences everyone equally – everyone improves, and does so at the same rate (i.e. variability is low) .

If $F = MS_{treatment}/ MS_{error}$, then making $MS_{treatment}$ large and MS_{error} small will result in a large value of F

Like t, a large value corresponds to small p-values, which makes it more likely to reject H_o

However, before we calculate MS, we need to calculate what are called *sums of squares*, or SS

Example for how we fill the ANOVA's table

Treatment	Measures
X	1 2 2
Y	5 6 5
Z	2 1

Step1:

Meanx = (1+2+2)/3=1.667
Meany= 5.333
Meanz= 1.5

Overall mean = μ = (1+2+2+5+6+5+2+1)/8=3

Estimated effects = Estimated treatment mean - Estimated overall mean

A1=Meanx-μ=-1.333
A2=2.333
A3=-1.5

Step 2: The ANOVA table

Cause of the variation	df SS MS	F
Treatment	….. …… …….	….
Residuals	….. …… …….	
Total	…… ……	

dftreat=3-1=2
dftot=8-1=7
dfres=7-2=5

SStreat =”sum of squares between treatment groups”

$$= \sum A_i^2 * n_i$$

= (−1.33)^2 · 3 + (2.33)^2 · 3 + (1.5)^2 · 2 = 26.17

SSres =”sum of squares within treatment groups”

$$= \sum_i \sum_j (y_{ij} - mean_i)^2 = \sum_i ss_{rowi}$$

$$= (1 - 1.667)^2 + (2 - 1.667)^2 + (2 - 1.667)^2 + [0.667] + [0.5] = 1.83$$

SStot =”Total sum of squares”

$$= \sum_{ij} (y_{ij} - \mu)^2$$

$$= (1 - 3)^2 + (2 - 3)^2 + \ldots + (1 - 3)^2 = 28$$

Remark:

The total ”SS” is always equal to the sum of the other ”SS”. SStot = SStreat + SSres

MS = SS/df, then:

MStreat = SStreat/ dftreat = 26.17/ 2 = 13.08

MSres = SSres /dfres = 1.83 /5 = 0.37

The F-value is just given by: F = MStreat /MSres = 13.08 /0.37 = 35.68

Interpretation:

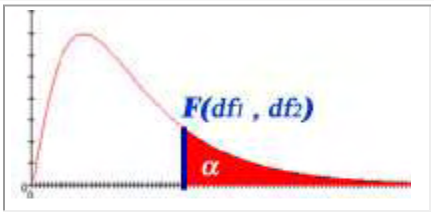
The F −value says us how far away we are from the hypothesis ”we cannot distinguish between error and treatment”, i.e. ”Treatment is not relevant according to our data”! A big F−value implies that the effect of the treatment is relevant!

The significance of the value F is determined in a similar manner to t-test (ie. Simulate random 10000 times the study, assuming no different on the treatments and calculate the 10000 F-tests, which form the F-distribution, and we find the rate of F-tests that are larger than the F)

If F > 5% point of F-distribution we can claim that the treatments differ significantly (with a small probability of error P<0.05)

F-distribution

F Table for α = 0.05



/	df₁=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
df₂=1	161.4476	199.5000	215.7073	224.5832	230.1619	233.9860	236.7684	238.8827	240.5433	241.8817	243.9060	245.9499	248.0131	249.0518	250.0951	251.1432	252.1957	253.2529	254.3144
2	18.5128	19.0000	19.1643	19.2468	19.2964	19.3295	19.3532	19.3710	19.3848	19.3959	19.4125	19.4291	19.4458	19.4541	19.4624	19.4707	19.4791	19.4874	19.4957
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.5720	8.5494	8.5264
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.7170	5.6877	5.6581	5.6281
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3985	4.3650
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6689
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7067
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.9130	2.8450	2.7740	2.7372	2.6996	2.6609	2.6211	2.5801	2.5379
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536	2.7876	2.7186	2.6464	2.6090	2.5705	2.5309	2.4901	2.4480	2.4045
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.3410	2.2962
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.5342	2.4630	2.3879	2.3487	2.3082	2.2664	2.2229	2.1778	2.1307
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.4753	2.4034	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0658
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.1040	2.0584	2.0107	1.9604
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.3080	2.2341	2.1555	2.1141	2.0712	2.0264	1.9795	1.9302	1.8780

20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660	2.3210	2.2504	2.1757	2.0960	2.0540	2.0102	1.9645	1.9165	1.8657	1.8117
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.2258	2.1508	2.0707	2.0283	1.9842	1.9380	1.8894	1.8380	1.7831
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201	2.2747	2.2036	2.1282	2.0476	2.0050	1.9605	1.9139	1.8648	1.8128	1.7570
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.1834	2.1077	2.0267	1.9838	1.9390	1.8920	1.8424	1.7896	1.7330
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7684	1.7110
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	2.1479	2.0716	1.9898	1.9464	1.9010	1.8533	1.8027	1.7488	1.6906
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7306	1.6717
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360	2.1900	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6541
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6376
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6223
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5089
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401	1.9926	1.9174	1.8364	1.7480	1.7001	1.6491	1.5943	1.5343	1.4673	1.3893
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.4290	1.3519	1.2539
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	1.8307	1.7522	1.6664	1.5705	1.5173	1.4591	1.3940	1.3180	1.2214	1.0000

Bonferroni adjustment

For Bonferroni adjustment you divide the p value to be achieved for significance by the number of paired comparisons to be made. So if you have three groups, and the overall test of significance between them comes out significant (using .05 as the threshold value), you might want to compare all the pairs to see which are significantly different. How many possible pairs are there to compare? The formula k(k-1)/2 tells you – where k is the number of groups or conditions. So here k(k-1)/2 = 3 * 2 / 2 = 3. So any pair has to achieve a sig value on a paired test smaller than .05/3 = .017 to be sig at the .05 level. With four groups, and again wanting to compare all possible pairs (k(k-1)/2 = 6), then p for any pair has to be smaller than .05/6 = .0083 to be sig. Put simply, the adjusted p value (or alpha level as it is sometimes called) for n paired comparisons is:

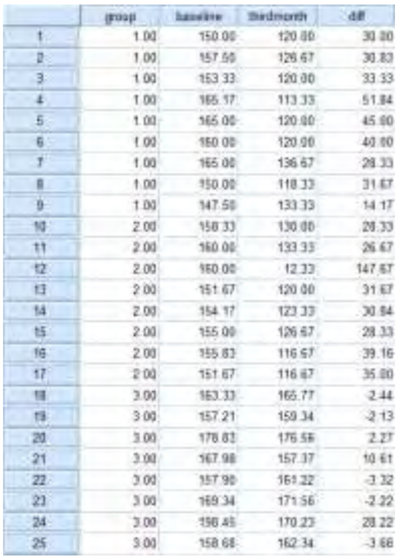
$$\frac{\text{Target p value}}{n}$$

Bonferroni adjustment is often seen as a bit too conservative, however, I.e. it protects against the danger of overclaiming the number of significant differences between pairs of values/ conditions/ groups when doing multiple followup comparisons, but it does so at the cost of possibly underclaiming.

Post Hoc Tests

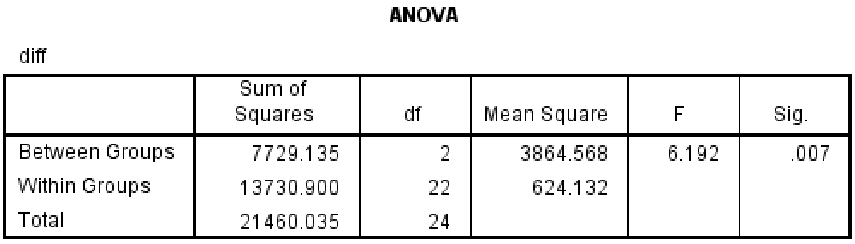
If the ANOVA show that there are significant differences between the groups then we can make individual comparisons between groups, e.g. To compare a group with the group b, using the t-test. However, this t-test differs from the previous in SE [here calculated using the random variation, the error].

A complete example in SPSS



	group	baseline	thirdmonth	diff
1	1.00	150.00	120.00	30.00
2	1.00	157.50	126.67	30.83
3	1.00	153.33	120.00	33.33
4	1.00	165.17	113.33	51.84
5	1.00	165.00	120.00	45.00
6	1.00	160.00	120.00	40.00
7	1.00	165.00	136.67	28.33
8	1.00	150.00	118.33	31.67
9	1.00	147.50	133.33	14.17
10	2.00	158.33	130.00	28.33
11	2.00	160.00	133.33	26.67
12	2.00	160.00	12.33	147.67
13	2.00	151.67	120.00	31.67
14	2.00	154.17	123.33	30.84
15	2.00	155.00	126.67	28.33
16	2.00	155.83	116.67	39.16
17	2.00	151.67	116.67	35.00
18	3.00	163.33	165.77	-2.44
19	3.00	157.21	159.34	-2.13
20	3.00	178.83	176.56	2.27
21	3.00	167.98	157.37	10.61
22	3.00	157.90	161.22	-3.32
23	3.00	169.34	171.56	-2.22
24	3.00	198.45	170.23	28.22
25	3.00	158.68	162.34	-3.66

Oneway



ANOVA

diff

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7729.135	2	3864.568	6.192	.007
Within Groups	13730.900	22	624.132		
Total	21460.035	24			

We observe that the Pvalue = 0.007 < 0.05 so our assumption is wrong since there is a difference in the effectiveness between drugs and placebo. In the post-hoc tests we will find detailed comparisons between drugs and placebo.

Chapter 3 – Results in Python

Software

Our application works as follows:

Window 1:

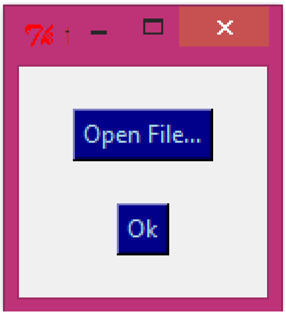
The user can press one of the 2 Buttons Independent T-test or One-Way Anova:

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

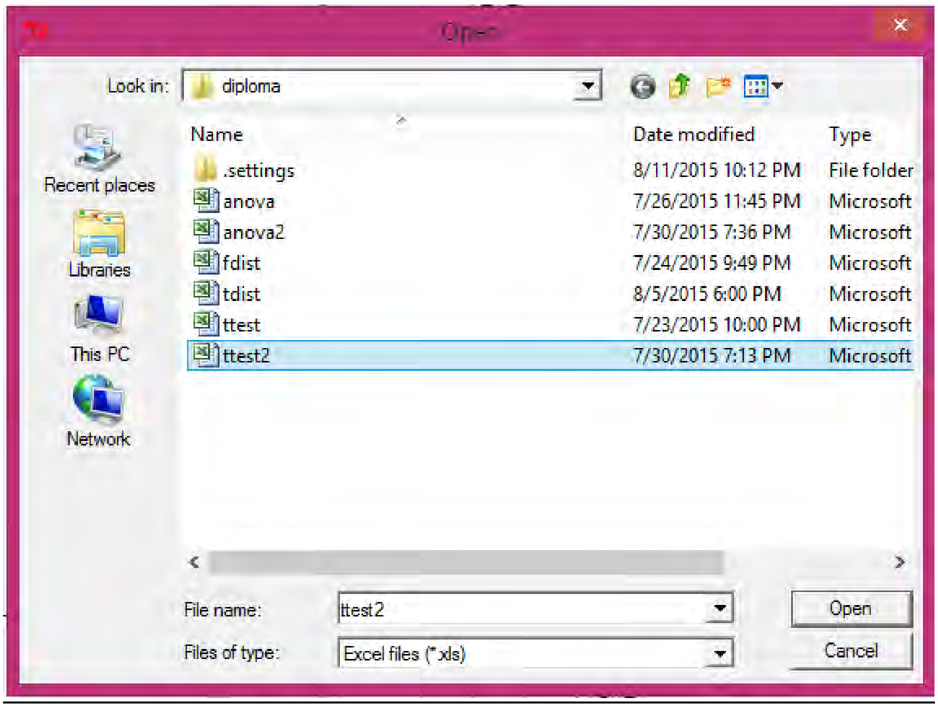
This software performs independent T-test & One Way Anova with post hoc test considering the Bonferonni's adjustment

We saved our measures in a excel file. The second window, which automatically opens when the user press one of the 2 buttons above, helps us to browse in our hard disk(window 3) and open it then click Ok. With this click, we close the window 1 and appears the following window 3, the application keeps the file's path and then we can select between the two buttons for a t-test or an one-way anova.

Window 2:



Window 3:



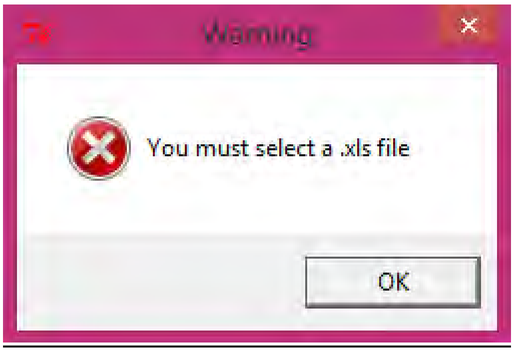
Look in: diploma

Name	Date modified	Type
.settings	8/11/2015 10:12 PM	File folder
anova	7/26/2015 11:45 PM	Microsoft
anova2	7/30/2015 7:36 PM	Microsoft
fdist	7/24/2015 9:49 PM	Microsoft
tdist	8/5/2015 6:00 PM	Microsoft
ttest	7/23/2015 10:00 PM	Microsoft
ttest2	7/30/2015 7:13 PM	Microsoft

File name: ttest2

Files of type: Excel files (*.xls)

If we press the OK button (window 2) without selecting a file at first, a warning message appears that reminds us to select one file if we want to continue.

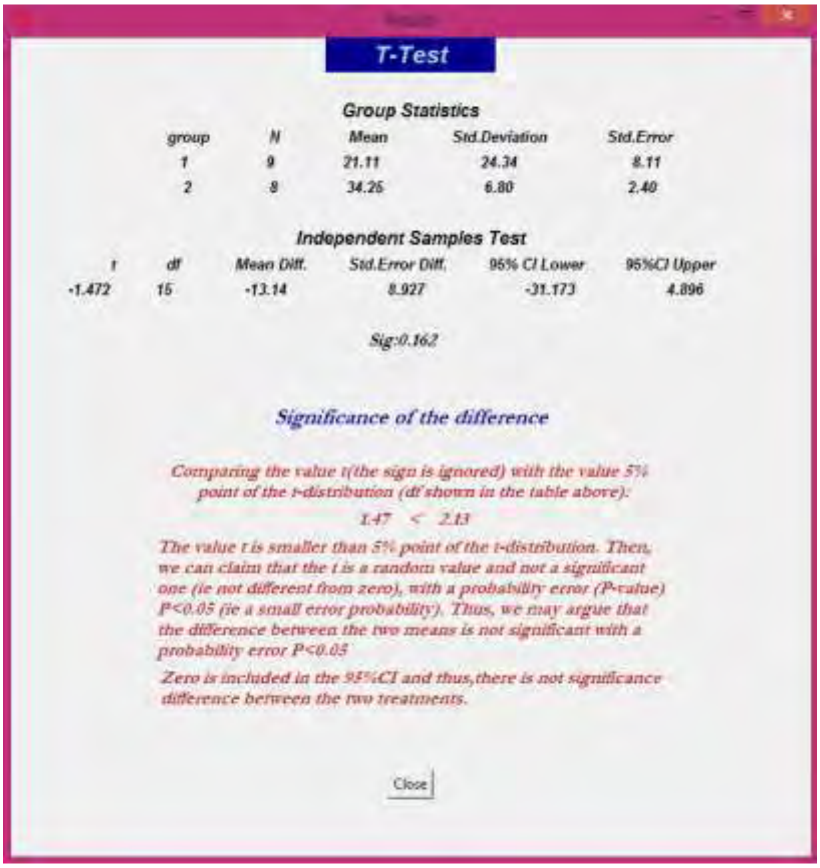


You must select a .xls file

If we press the Button: Independent T-test

We will run the same example of page 8

The results with our application are the same and shown below



T-Test

Group Statistics

group	N	Mean	Std.Deviation	Std.Error
1	9	21.11	24.34	8.11
2	8	34.26	6.80	2.40

Independent Samples Test

t	df	Mean Diff.	Std.Error Diff.	95% CI Lower	95%CI Upper
-1.472	15	-13.14	8.927	-31.173	4.896

Sig:0.162

Significance of the difference

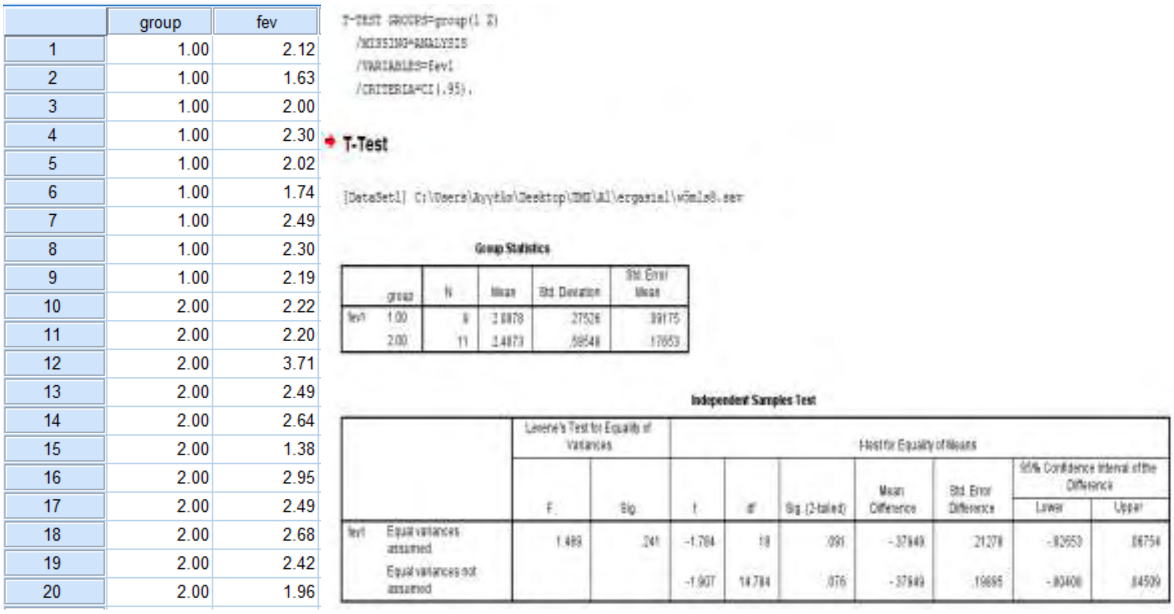
Comparing the value t(the sign is ignored) with the value 5% point of the t-distribution (df shown in the table above):

1.47 < 2.13

The value t is smaller than 5% point of the t-distribution. Then, we can claim that the t is a random value and not a significant one (ie not different from zero), with a probability error (P-value) P<0.05 (ie a small error probability). Thus, we may argue that the difference between the two means is not significant with a probability error P<0.05

Zero is included in the 95%CI and thus,there is not significance difference between the two treatments.

We run and another one example with SPSS and with our application.



	group	fev
1	1.00	2.12
2	1.00	1.63
3	1.00	2.00
4	1.00	2.30
5	1.00	2.02
6	1.00	1.74
7	1.00	2.49
8	1.00	2.30
9	1.00	2.19
10	2.00	2.22
11	2.00	2.20
12	2.00	3.71
13	2.00	2.49
14	2.00	2.64
15	2.00	1.38
16	2.00	2.95
17	2.00	2.49
18	2.00	2.68
19	2.00	2.42
20	2.00	1.96

```
T-TEST GROUPS=group(1 2)
  /MISSING=ANALYSIS
  /VARIABLES=fev1
  /CRITERIA=CI(.95).
```

T-Test

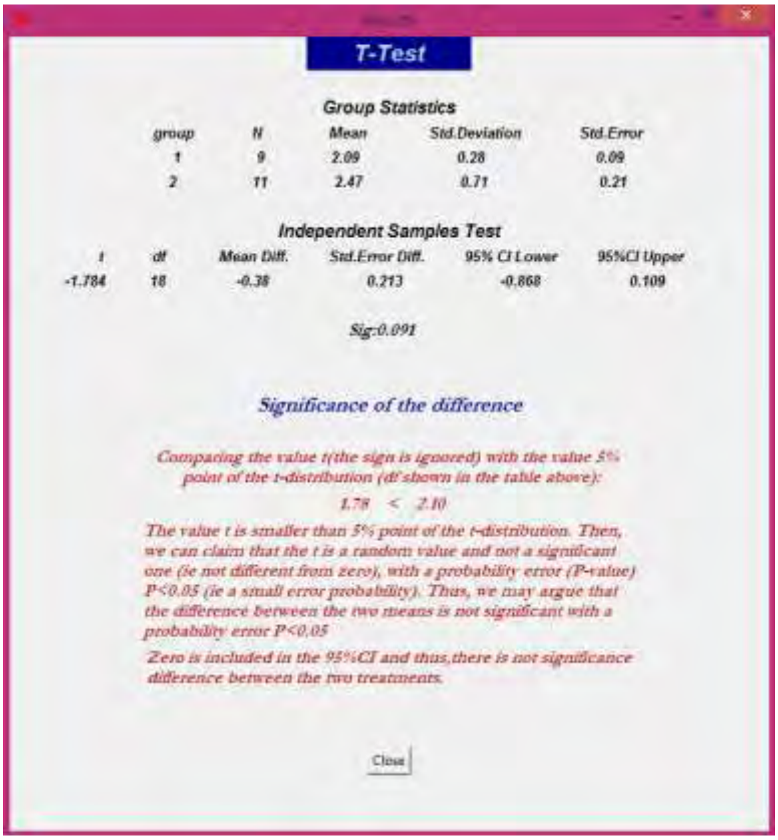
```
[DataSet1] C:\Users\Ayytlo\Desktop\[illegible]\Al\ergasial\w5mls8.sav
```

Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
fev1	1.00	9	2.0878	.27526	.09175
	2.00	11	2.4673	.59548	.17953

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
fev1	Equal variances assumed	1.489	.241	-1.784	18	.091	-.37949	.21278	-.82653	.06754
	Equal variances not assumed			-1.907	14.784	.076	-.37949	.19895	-.80408	.04509



T-Test

Group Statistics

group	N	Mean	Std.Deviation	Std.Error
1	9	2.09	0.28	0.09
2	11	2.47	0.71	0.21

Independent Samples Test

t	df	Mean Diff.	Std.Error Diff.	95% CI Lower	95%CI Upper
-1.784	18	-0.38	0.213	-0.868	0.109

Sig:0.091

Significance of the difference

Comparing the value t(the sign is ignored) with the value 5% point of the t-distribution (df shown in the table above):

1.78 < 2.10

The value t is smaller than 5% point of the t-distribution. Then, we can claim that the t is a random value and not a significant one (ie not different from zero), with a probability error (P-value) P<0.05 (ie a small error probability). Thus, we may argue that the difference between the two means is not significant with a probability error P<0.05

Zero is included in the 95%CI and thus,there is not significance difference between the two treatments.

If we press the Button: One-way Anova

We will execute the same example of page 14

The results with our application are the same and shown below

Data

Groups	SumOfSquares	df	MaenSquare	F	sig.
BetweenGroups	7729.330	2	3864.665	6.192	0.007
WithinGroups	13730.900	22	624.132		
Total	21460.230	24			

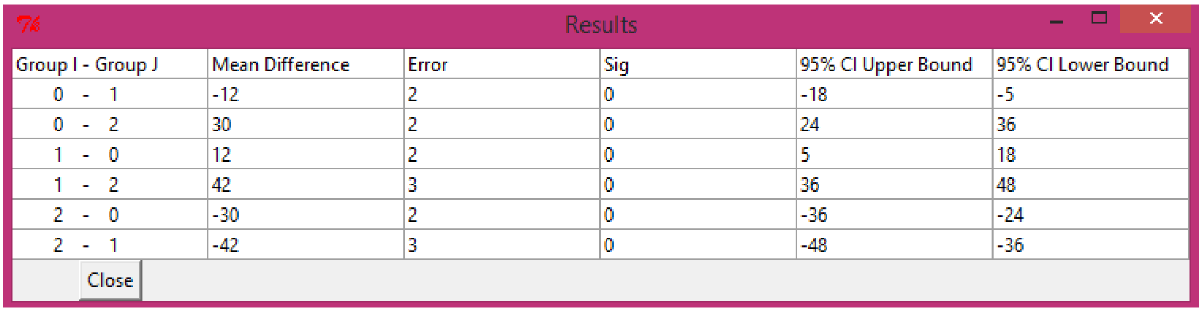
Conclusion:

6.192 > 3.44

The value f is larger than 5% point of the f-distribution. Then, we can claim that the f is not a random value but a significant one (ie not different from zero), with a probability error (P-value) P<0.05 (ie a small error probability). Thus, we may argue that the difference between the groups is significant with a probability error P<0.05

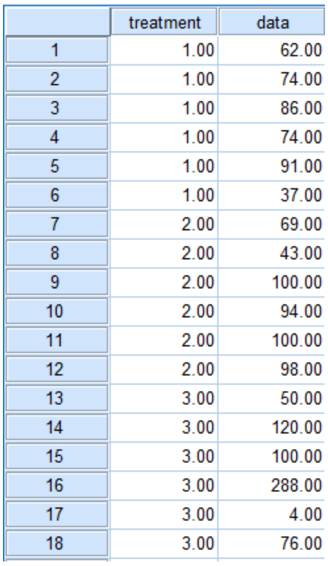
The p-value is less than the significance level of 0.05, so we reject the null hypothesis. The null hypothesis is that the groups were the same, but we reject that, so at least there is one team of groups that differs.

There is enough evidence to support the claim that there is a difference in the groups. The ANOVA doesn't tell which group is different, you would need to look at confidence intervals and run post hoc tests to determine that.

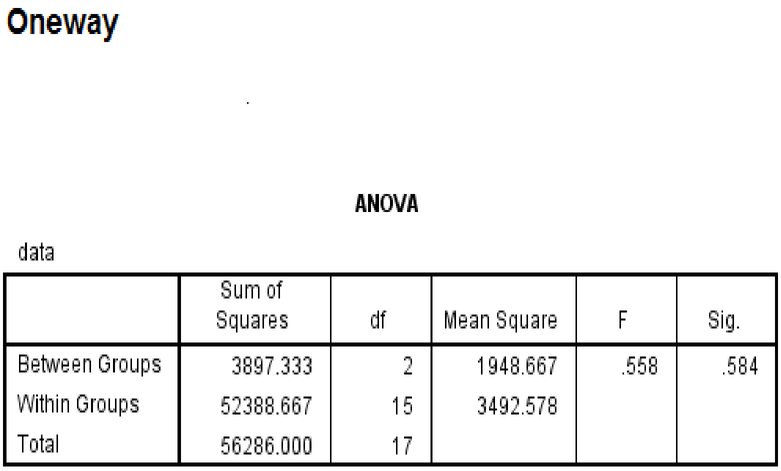


Group I - Group J	Mean Difference	Error	Sig	95% CI Upper Bound	95% CI Lower Bound
0 - 1	-12	2	0	-18	-5
0 - 2	30	2	0	24	36
1 - 0	12	2	0	5	18
1 - 2	42	3	0	36	48
2 - 0	-30	2	0	-36	-24
2 - 1	-42	3	0	-48	-36

We run and another one example with SPSS and with our application.



	treatment	data
1	1.00	62.00
2	1.00	74.00
3	1.00	86.00
4	1.00	74.00
5	1.00	91.00
6	1.00	37.00
7	2.00	69.00
8	2.00	43.00
9	2.00	100.00
10	2.00	94.00
11	2.00	100.00
12	2.00	98.00
13	3.00	50.00
14	3.00	120.00
15	3.00	100.00
16	3.00	288.00
17	3.00	4.00
18	3.00	76.00



Oneway

ANOVA

data

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3897.333	2	1948.667	.558	.584
Within Groups	52388.667	15	3492.578		
Total	56286.000	17			

The P-value between groups (Sig.) P = 0.584, so there is no significant difference between groups (P <0.05). From the 95% CI the difference between groups are compared in pairs is containing zero.

ANOVA

Data

Groups	SumOfSquares	df	MaenSquare	F	sig.
BetweenGroups	3897.190	2	1948.595	0.558	0.584
WithinGroups	52388.667	15	3492.578		
Total	56285.857	17			

Conclusion:

0.558 < 3.68

The value f is smaller than 5% point of the f-distribution. Then, we can claim that the f is a random value and not a significant one (ie not different from zero), with a probability error (P-value) P<0.05 (ie a small error probability). Thus, we may argue that the non difference between the groups is not significant with a probability error P<0.05

The p-value is larger than the significance level of 0.05, so we can't reject the null hypothesis. The null hypothesis is that the groups were the same.

Chapter 4 – Conclusion

The independent samples t-test is used to test the hypothesis that the difference between the means of two samples is equal to 0 (this hypothesis is therefore called the null hypothesis). The program displays the difference between the two means, and the 95% Confidence Interval (CI) of this difference. Next follow the test statistic t, the Degrees of Freedom (DF) and the two-tailed probability P. When the P-value is less than the conventional 0.05, the null hypothesis is rejected and the conclusion is that the two means do indeed differ significantly.

The purpose of ANOVA is almost the same as the t tests. The goal is to determine whether the mean differences that are obtained for the sample data are sufficiently large to justify a conclusion that there are mean differences between the populations from which the samples were obtained

ANOVA allows researcher to evaluate all of the mean differences in a single hypothesis test using a single alpha-level(in our examples a=0.05) and, thereby, keeps the risk of a Type I error under control no matter how many means are being compared ⍰.However, what if we just compared each of the groups in a pairwise manner like the Bonferroni's correction, using a 'testwise' alpha-level = alpha-level / (number of tests)

Chapter 5 - Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Dr. habil. Axel Kowald for the continuous support of my study and related research, for his patience, motivation, and immense knowledge. His guidance assisted me in writing of this thesis.

Besides my supervisor, I would like to thank the rest of my thesis committee: Prof. Elias Zintzaras and Prof. George Rachiotis, for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives and gave me the opportunity to join their team as PhD student.

I also thank my fellow labmates for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we had this year.

Last but not the least, I would like to thank my family: my parents and my brother and my best friends, for supporting me spiritually throughout writing this thesis and my life in general.

Chapter 6 –References

http://biomath.med.uth.gr/

https://statistics.laerd.com

http://oak.ucc.nau.edu

http://uk.sagepub.com/sites/default/files/upm-binaries/33663_Chapter4.pdf

http://pip.ucalgary.ca/psyc-312/multiple-group-experiments-and-analysis-of-variance/one-way-analysis-of-variance/anova2.html

https://people.richland.edu/james/lecture/m113/anova.html

http://www.java2s.com

https://stat.ethz.ch/education/semesters/as2010/anova/ANOVA_how_to_do.pdf

https://onlinecourses.science.psu.edu/stat800/book/export/html/58