UNIVERSITY OF THESSALY
SCHOOL OF MEDICINE
POSTGRADUATE PROGRAMME
«RESEARCH METHODOLOGY IN BIOMEDICINE, BIOSTATISTICS AND CLINICAL BIOINFORMATICS»

Two-way ANOVA without interaction with post-hoc test considering the Bonferroni's correction using Python

MASTER THESIS

A. PAPAKONSTANTINOU

COMITTEE: A. KOWALD, CHR. DOXANI, IL. ZINTZARAS

LARISSA, 2017

TWO-WAY ANOVA WITHOUT INTERACTION WITH POST-HOC TEST CONSIDERING THE BONFERRONI'S CORRECTION USING PYTHON

UNIVERSITY OF THESSALY

SCHOOL OF MEDICINE

POSTGRADUATE PROGRAMME

"RESEARCH METHODOLOGY IN BIOMEDICINE, BIOSTATISTICS AND CLINICAL BIOINFORMATICS"

TWO-WAY ANOVA WITHOUT INTERACTION WITH POST-HOC TEST CONSIDERING THE BONFERRONI'S CORRECTION USING PYTHON

MASTER THESIS

A. PAPAKONSTANTINOU

COMMITTEE: A. KOWALD, CHR. DOXANI, IL. ZINTZARAS

LARISSA, 2017

TWO-WAY ANOVA WITHOUT INTERACTION WITH POST-HOC TEST CONSIDERING THE BONFERRONI'S CORRECTION USING PYTHON

Master Thesis

Achilleas Papakonstantinou (Papakonstantinou.Achilleas@gmail.com)

University of Thessaly
School of Medicine
Postgraduate Programme
"Research Methodology in Biomedicine, Biostatistics and Clinical Bioinformatics"

Committee: A. Kowald, Chr. Doxani, Il. Zintzaras

Table of Contents

Table of Contents VIII
A. Abstracts 1
Abstract (Greek) 1
Abstract (English) 2
B. Introduction 3
t test for independent samples 3
Confidence interval (CI) of the difference between two means.... 6
One-Way ANOVA 7
Post hoc tests 9
Bonferroni correction 11
Two-way ANOVA 11
Assumptions 11
C. Methods 13
Importing an existing csv file 13
Saving a csv file 16
D. Results 17
Executing two-way ANOVA with Bonferroni's correction 17
Reporting the results 19
E. Conclusion 20
F. References 21

A. Abstracts

Abstract (Greek)

Το στατιστικό τεστ two-way ANOVA συγκρίνει τη διαφορά των μέσων τιμών ανάμεσα στις ομάδες που έχουν χωριστεί βάσει δυο ανεξάρτητων μεταβλητών. Ο κύριος σκοπός του two-way ANOVA είναι να γίνει κατανοητό αν υπάρχει αλληλεπίδραση ανάμεσα στις δυο ανεξάρτητες μεταβλητές σε σχέση με την εξαρτημένη μεταβλητή. Για παράδειγμα, θα μπορούσε να ελεγχθεί αν υπάρχει αλληλεπίδραση ανάμεσα στο επίπεδο φυσικής δραστηριότητας και φύλου στη συγκέντρωση χοληστερίνης στο αίμα στα παιδιά, όπου η φυσική δραστηριότητα (χαμηλή/μέση/υψηλή) και φύλο (αγόρι/κορίτσι) είναι οι ανεξάρτητες μεταβλητές, και η συγκέντρωση χοληστερίνης είναι η εξαρτημένη μεταβλητή.

Ο όρος αλληλεπίδραση σε ένα τεστ two-way ANOVA πληροφορεί αν η επίδραση μιας από τις ανεξάρτητες μεταβλητές πάνω στην εξαρτημένη μεταβλητή είναι η ίδια για όλες τις τιμές της άλλης ανεξάρτητης μεταβλητής (και το αντίστροφο). Για παράδειγμα, η επίδραση του φύλου (αγόρι/κορίτσι) στη συγκέντρωση χοληστερόλης επηρεάζεται από τη φυσική δραστηριότητα (χαμηλή/μέση/υψηλή);

Η εφαρμογή αυτής της διπλωματικής εργασίας υποθέτει ότι δεν υπάρχει αλληλεπίδραση ανάμεσα στις ανεξάρτητες μεταβλητές. Χρησιμοποιώντας την προγραμματιστική γλώσσα Python, υλοποιείται το τεστ «two-way ANOVA χωρίς αλληλεπίδραση με διόρθωση Bonferroni».

Keywords

Two-way ANOVA, Bonferroni, Python

Abstract (English)

The statistical test two-way ANOVA compares the mean differences between groups that have been split on two independent variables. The primary purpose of a two-way ANOVA is to understand if there is an interaction between the two independent variables on the dependent variable. For example, there could be determined whether there is an interaction between physical activity level and gender on blood cholesterol concentration in children, where physical activity (low/moderate/high) and gender (male/female) are the independent variables, and cholesterol concentration is the dependent variable.

The interaction term in a two-way ANOVA informs whether the effect of one of the independent variables on the dependent variable is the same for all values of the other independent variable (and vice versa). For example, is the effect of gender (male/female) on cholesterol concentration influenced by physical activity (low/moderate/high)?

The application of this thesis assumes there is no interaction between the two independent variables. Using the programming language Python, the test "two-way ANOVA without interactions with Bonferroni correction" is implemented.

Keywords

Two-way ANOVA, Bonferroni, Python

B. Introduction

t test for independent samples

The independent-samples t-test (or independent t-test, for short) compares the means between two unrelated groups on the same continuous, dependent variable. For example, an independent t-test can be used to understand whether glucose levels differed based on treatments A and B for two separate groups (i.e., the dependent variable would be "glucose levels" and your independent variable would be "treatments", which has two groups: "A" and "B") [1].

This statistical test will compare the two groups with the following formula :

$$t = \frac{\bar{x}_A - \bar{x}_B}{SE}$$

where $SE = \sqrt{s_{pooled}\left(\frac{1}{n_T} + \frac{1}{n_p}\right)}$

and $s_{pooled} = \frac{(n_T-1){s_T}^2+(n_p-1){s_p}^2}{(n_T-1)-(n_p-1)}$

How confident are we that a given value t (e.g. $t = -4.07$) is significant? Alternatively, what is the error probability (i.e. the p-value or the probability of false-positive result) for claiming that t is significant?

We can answer the question by comparing the absolute value of $t = 4.07$ with the value of the 5% point (0.05) of the t-distribution with degrees of freedom as follows:

$$df = (n_T - 1) + (n_p - 1)$$

df \ Pr	0.50	0.20	0.10	0.05	0.02	0.010	0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

In Medicine we use the aforementioned 5% value. Other disciplines use different points, for example psychology uses $P_r = 10\%$.

For $df = 22 \Rightarrow P_r = 2.07$

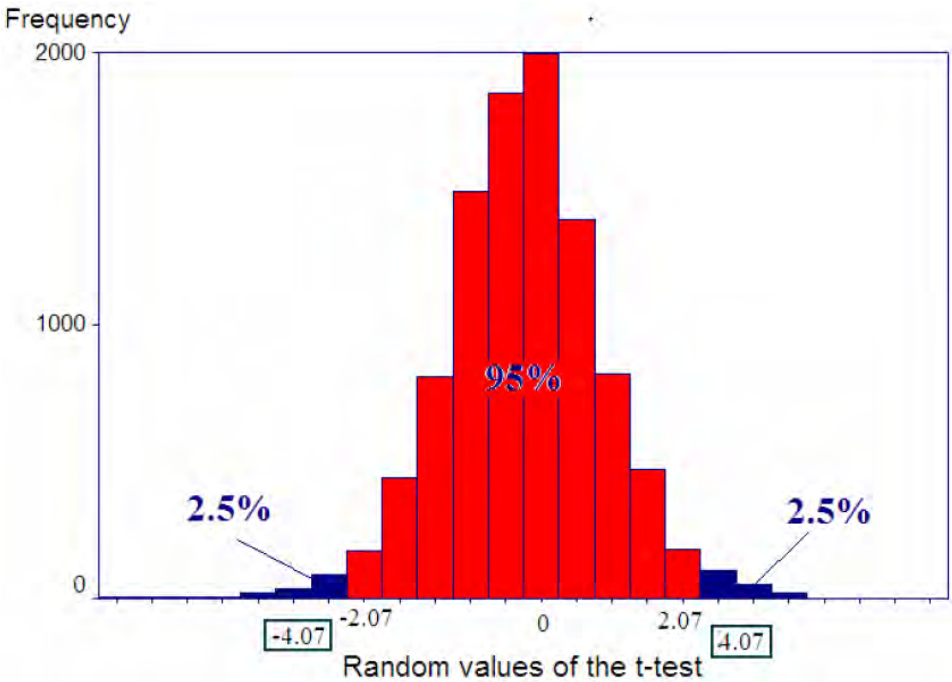
If the absolute t value is above a certain threshold then we can claim the difference between means (taking into account the variability of the data and the size of the trial) is significant (i.e. different from zero).

Example: Testing the difference between means by comparing the effectiveness of two treatments.

To assess the effectiveness of Tysabri, 24 patients with RRMS were randomized to receive either Tysabri or placebo. The primary end-point was the ARR. The results were as follows:

Tysabri (n=12)	Placebo (n=12)
0.38	0.80
0.37	0.70
0.24	0.39
0.08	0.06
0.04	0.49
0.03	0.07
0.07	0.63
0.19	0.83
0.35	0.62
0.38	0.95
0.27	0.92
0.20	0.81

If we simulate the study ten thousand times by randomly permutating all the numbers, and after every permutation we calculate the t-test, we expect 5% of the random t-tests to be greater of 2.07 or less than -2.07.



The value $t = 4.07$ is larger than 2.07. Thus, we can claim that $t = 4.07$ is not a random value, but a significant one (i.e. different from zero), with probability error (p-value) $P < 0.05$ (i.e. a small error probability).

This means that the difference between the two means is significant with a probability error $P < 0.05$.

Using a statistical software we can calculate the exact $P = 0.001$.

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 CI is defined as:

$$(D - t \cdot SE, D + t \cdot SE)$$

More specifically, the 95% CI is defined as:

$$(D - t_{0.05} \cdot SE, D + t_{0.05} \cdot SE)$$

If zero is not included in the 95% CI then there is statistically significant difference between the two treatments.

One-Way ANOVA

The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of two or more independent (unrelated) groups (although it is used more often when there are a minimum of three, rather than two groups). For example, a one-way ANOVA could be used to understand whether glucose levels differed based on treatment received amongst patients, dividing them into three independent groups (e.g., A, B and C). Also, it is important to realize that the one-way ANOVA is an omnibus statistic test and cannot tell which specific groups were statistically significantly different from each other; it only tells that at least two groups were different. Since there may be three, four, five or more groups in a study design, determining which of these groups differ from each other is important [2]. This can be done using a post hoc test; this thesis will use the Bonferroni's correction.

Bottom line: There is a dependent continuous variable and a categorical variable with two or more categories. When there are only two categories, then the one-way ANOVA is equivalent to the t-test for independent samples.

<u>Example</u>: In a study the liver weight (x) expressed as a percentage of the body weight of mice belonging in $k = 4$ groups that were fed with 4 different diets. In order to examine if there is difference between the four groups, the mean values of the four groups will be compared using one-way ANOVA.

a	*b*	*c*	*d*	
3.42	3.17	3.34	3.64	
3.96	3.63	3.72	3.93	
3.87	3.38	3.81	3.77	
4.19	3.47	3.66	4.18	
3.58	3.39	3.55	4.21	
3.76	3.41	3.51	3.88	
Mean value $\bar{x}_a$=3.80	$\bar{x}_b$=3.41	$\bar{x}_c$=3.60	$\bar{x}_d$=3.94	$\bar{x}$=3.69 (total)

There are two types of variances in this study. One is the sum of squares between groups and the other is the sum of squares within groups. The latter is considered as standard error or random variance.

The comparison of the variance between the four groups with the random variance is a result of the F-test as follows:

Source of variation	df	SS	MS
Between groups	$4-1=3$	0.954	0.318
Within groups (error)	$23-3=20$	0.876	$s^2=0.044$
Total	$24-1=23$	1.83	

The total variance of the dataset is:

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_j - \bar{x}\right)^2}{n-1} = \frac{1.83}{24-1}$$

The Mean Square is:

$$MS = \frac{SS}{df}$$

The F test compares the mean square between the groups with the error MS as follows:

$$F = \frac{Between\ groups\ MS}{Error\ MS} = \frac{0.318}{0.044} = 7.23$$

If the mean variance between the four groups is greater than the random mean variance, then the differences between the four groups is not random (i.e. there is statistically significant difference).

As with t-test, there is a table for the F-distribution:

Degrees of freedom in denominator	Degrees of freedom in numerator												
	1	2	3	4	5	6	7	8	9	10	20	40	∞
1	161	200	216	225	230	234	237	239	241	242	248	251	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.66	8.59	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.80	5.72	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.56	4.46	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.87	3.77	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.44	3.34	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.15	3.04	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	2.94	2.83	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.77	2.66	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.65	2.53	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.54	2.43	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.46	2.34	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.39	2.27	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.33	2.20	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.28	2.15	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.23	2.10	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.19	2.06	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.16	2.03	1.88
20	4.35	3.49	3.10	2.87	2.71	2.62	2.51	2.45	2.39	2.35	2.12	1.99	1.84
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.01	1.87	1.71
30	4.17	2.21	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	1.93	1.79	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.84	1.69	1.51
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.66	1.50	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.57	1.39	1.00

Post hoc tests

With two groups the interpretation of a significant difference is reasonably straightforward, but how the significant variation among the means of three or more groups is interpreted? Further analysis is required to find out how the means differ, for example whether one group differs from all the others. Note that only differences between individual groups should be investigated when the overall comparison of groups in the analysis of variance is significant, unless certain comparisons were intended in advance of the analysis [3].

If each pair of means is compared in turn then there is a 5% chance of a false positive result when there is no real difference (a Type I error). So, if there are for example four groups and all six paired tests are performed then the probability of at least one false positive result is very much greater than 5%. Several methods have been proposed to deal with this problem, such as Bonferroni, Newman-Keuls, Duncan and Scheffe. Each method is aimed at controlling the overall Type I error rate at no more than 5% (or some other specified level).

The disadvantage of all of these methods is that they are 'conservative', in that they err on the side of safety (non-significance). It is possible to find that, although the F test in the analysis of variance is statistically significant, no pair of means is significantly different.

There is no simple nor totally satisfactory solution to these problems, but the following strategy should be used:

1) Decision in advance of the analysis which groups are particularly interested in comparing (the fewer the better)
2) Perform modified t tests to compare the pairs of groups of interest, using the Bonferroni (or some other) method to adjust the P values.

The modified t test is based on the pooled estimate of variance from all the groups (which is the residual variance in the ANOVA table), not just the pair being considered. So t is calculated as

$$t = \frac{\bar{x}_1 - \bar{x}_2}{se(\bar{x}_1 - \bar{x}_2)}$$

where $se(\bar{x}_1 - \bar{x}_2) = s_{res} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

Bonferroni correction

Because the comparisons between the groups are not independent, when multiple comparisons (numbered k) are performed the significance level (P) changes to $P' = kP$.

For example, if six comparisons are made between the groups, then the adjusted P value will be $P' = kP = 6P$. The adjusted P value can increase too much if too many variables are added to the test. That's why a careful consideration of the proper pair testing is needed beforehand.

Two-way ANOVA

The two-way ANOVA is similar to the one-way ANOVA except that the comparison is made between the mean differences of groups of a dependent continuous variable that has been split on two independent categorical variables instead of one.

In the previous example with the mice, if there was one more categorical variable (e.g. gender) a two-way ANOVA would be required. The liver weight as the dependent continuous variable, and the diets and gender as the two categorical variables.

Assumptions

When a two-way ANOVA is chosen for data analysis, part of the process involves checking that the data can actually be analyzed using a two-way ANOVA. This is needed because it is only appropriate to use a two-way ANOVA if the data "pass" six assumptions that are required for a two-way ANOVA to give a valid result [4].

Sometimes when analyzing data, one or more of these assumptions is violated. This is not uncommon when working with real-world. There are the following assumptions:

1. The dependent variable should be measured at the continuous level (i.e. they are interval or ratio variables). Examples of continuous variables include revision time (measured in hours), intelligence (measured using IQ score), exam performance (measured from 0 to 100), weight (measured in kg), and so forth.
2. The two independent variables should each consist of two or more categorical, independent groups. Example independent variables that meet this criterion include gender (2 groups: male or female), ethnicity (3 groups: Caucasian, African American and Hispanic), profession (5 groups: surgeon, doctor, nurse, dentist, therapist), and so forth.
3. There should be independence of observations, which means that there is no relationship between the observations in each group or between the groups themselves. For example, there must be different participants in each group with no participant being in more than one group.
4. There should be no significant outliers. Outliers are data points within the data that do not follow the usual pattern (e.g., in a study of 100 students' IQ scores, where the mean score was 108 with only a small variation between students, one student had a score of 156, which is very unusual, and may even put her in the top 1% of IQ scores globally). The problem with outliers is that they can have a negative effect on the two-way ANOVA, reducing the accuracy of the results.
5. The dependent variable should be approximately normally distributed for each combination of the groups of the two independent variables. However, two-way ANOVA only requires *approximately* normal data because it is quite robust to violations of normality, meaning the assumption can be a little violated and still provide valid results.
6. There needs to be homogeneity of variances for each combination of the groups of the two independent variables.

C. Methods

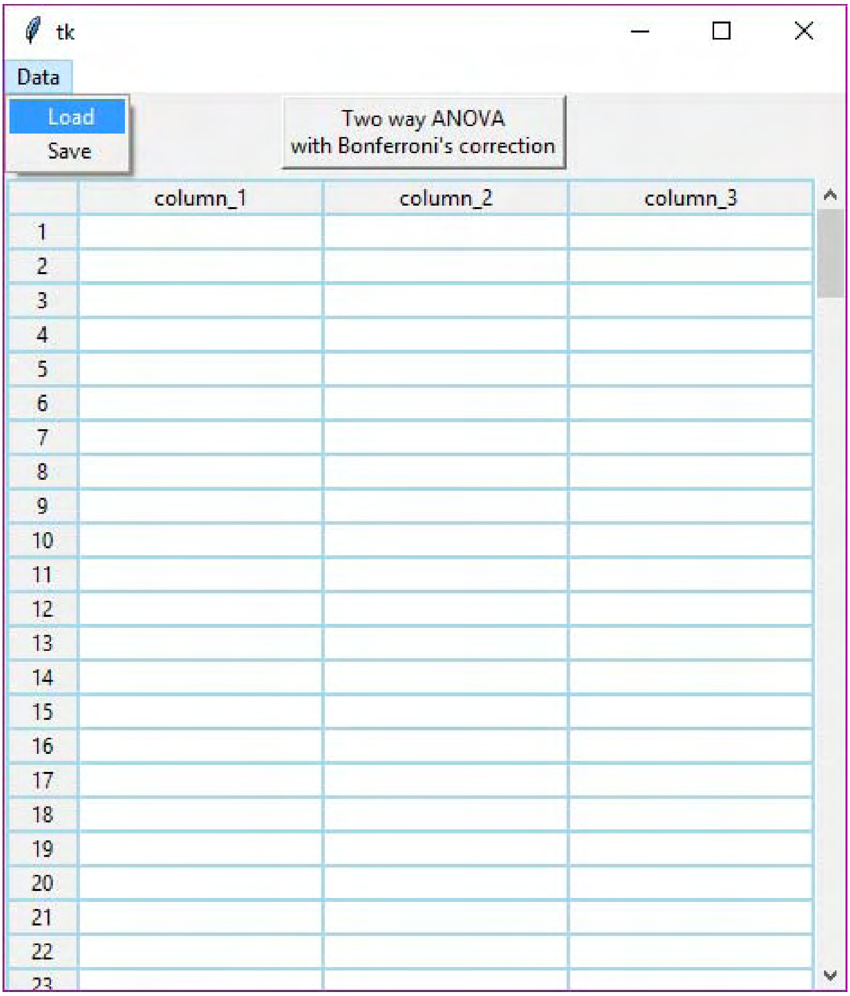
When working with large datasets it is very hard, frustrating and prone to mistakes to make by hand all the calculations required for a statistical test.

In order to avoid this, for my thesis a software in the programming language Python has been developed which calculates the two-way ANOVA when there is no interaction between the groups. Optionally, the Bonferroni's correction can be calculated when one of the categorical variables has at least 3 categories.

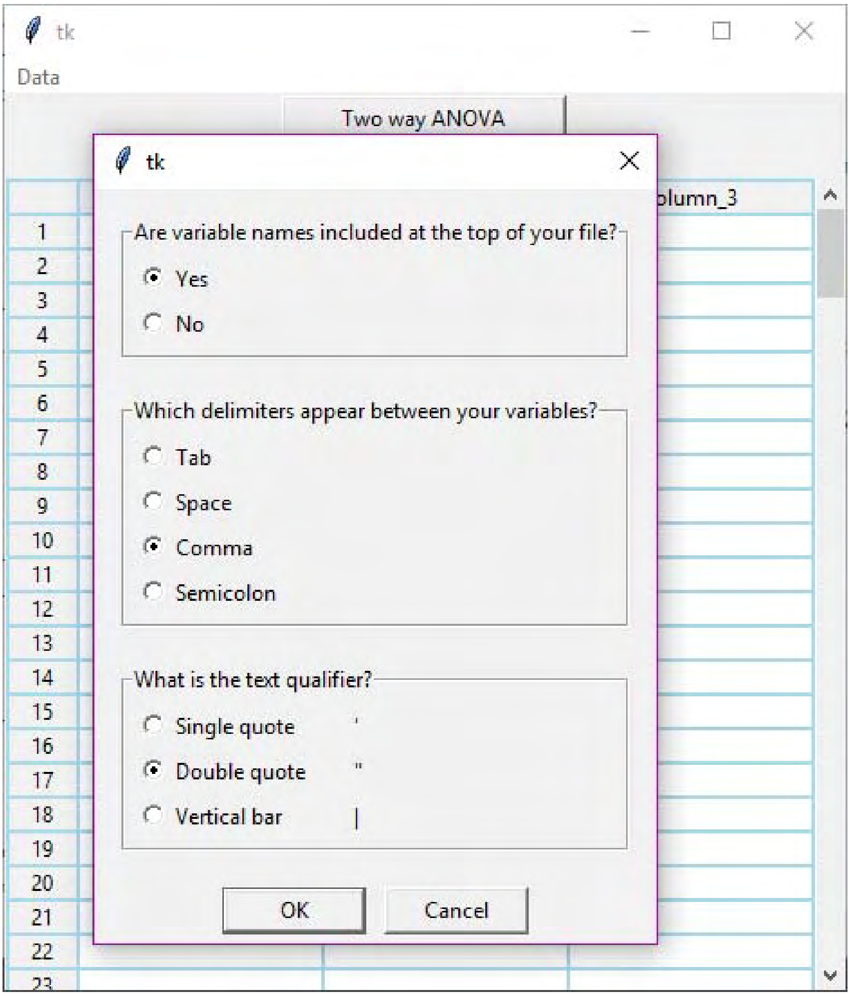
Using Python, the powerful Python statistical libraries could be accessed and a GUI could be designed for easier user experience. The application is compatible with csv files which store a table row in each line of the file and the values are separated by a specific character for example comma.

Importing an existing csv file

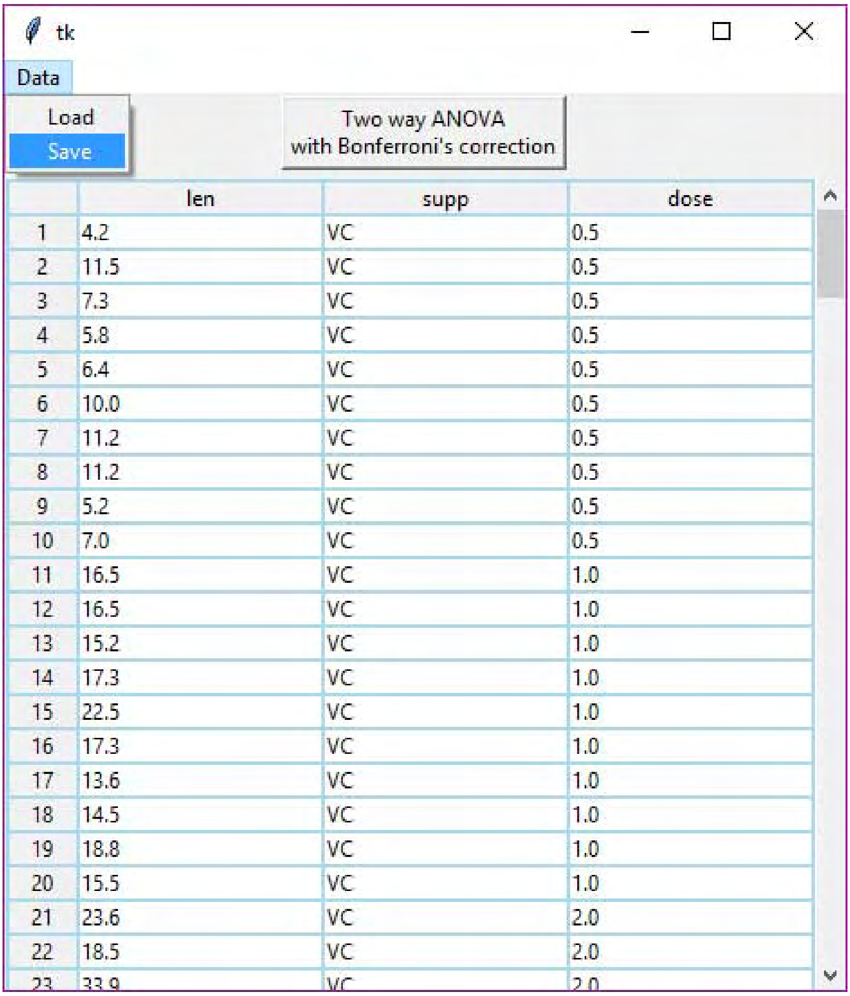
Through clicking **Data** -> **Load**, an open file dialog pops up and the desirable csv file can be imported.



In the follow-up dialog, the appropriate parameters of the csv file can be chosen.


tk
Data
Two way ANOVA
tk
olumn_3
Are variable names included at the top of your file?
Yes
No
Which delimiters appear between your variables?
Tab
Space
Comma
Semicolon
What is the text qualifier?
Single quote '
Double quote "
Vertical bar |
OK
Cancel

Saving a csv file



	len	supp	dose
1	4.2	VC	0.5
2	11.5	VC	0.5
3	7.3	VC	0.5
4	5.8	VC	0.5
5	6.4	VC	0.5
6	10.0	VC	0.5
7	11.2	VC	0.5
8	11.2	VC	0.5
9	5.2	VC	0.5
10	7.0	VC	0.5
11	16.5	VC	1.0
12	16.5	VC	1.0
13	15.2	VC	1.0
14	17.3	VC	1.0
15	22.5	VC	1.0
16	17.3	VC	1.0
17	13.6	VC	1.0
18	14.5	VC	1.0
19	18.8	VC	1.0
20	15.5	VC	1.0
21	23.6	VC	2.0
22	18.5	VC	2.0
23	33.9	VC	2.0

A csv file can be saved by clicking Data -> Save. The .csv extension is recommended. The resulting csv file by default will separate the values with comma, quote the text with double quotes and have the variable names above the values like so:

```
"","column_1","column_2","column_3"
1,1.0,2.0,3.0
2,4.0,5.0,6.0
3,7.0,8.0,9.0
4,10.0,11.0,12.0
```

D. Results

Executing two-way ANOVA with Bonferroni's correction

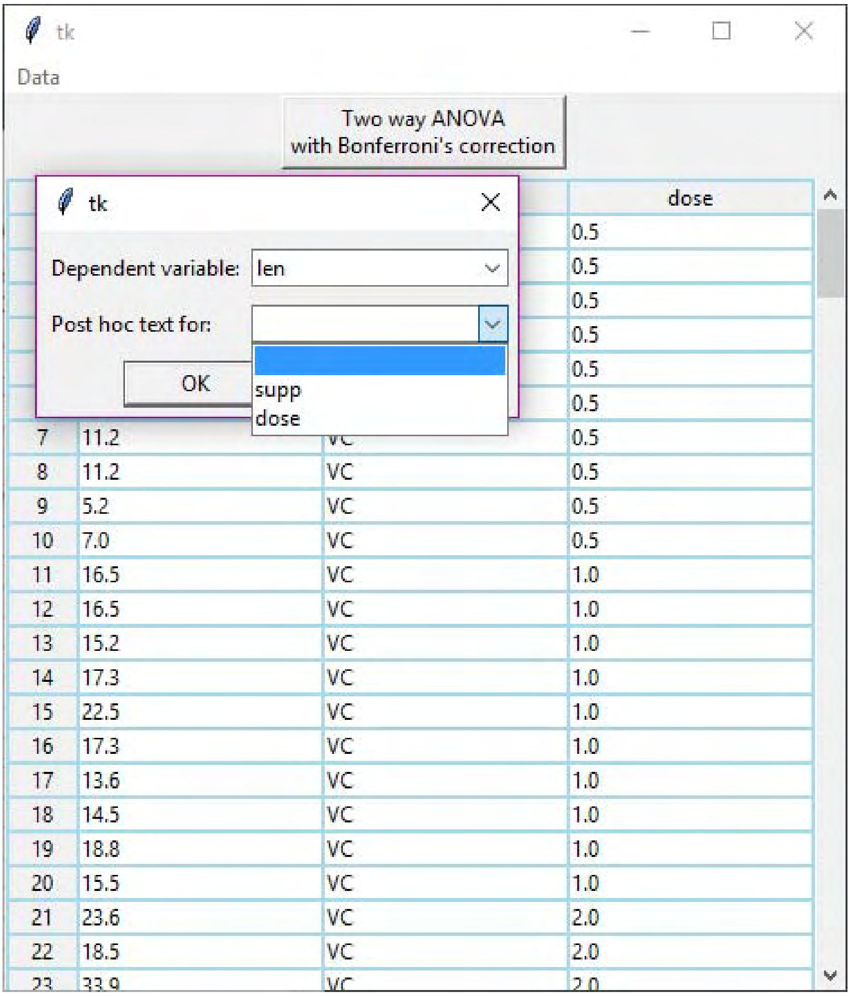
A sample dataset was used with name "ToothGrowth.csv" that can be downloaded from https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html .

The dependent continuous variable is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice - a form of vitamin C (ascorbic acid) and coded as VC).

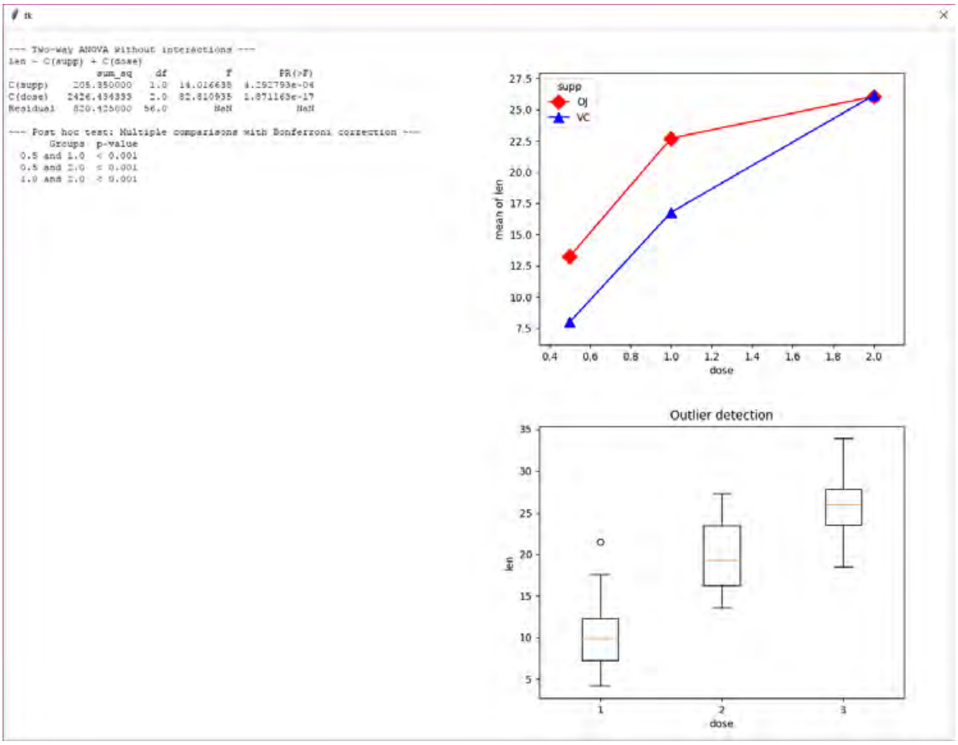
It is a data frame with 60 observations on 3 variables.

[,1]	*len*	numeric	Tooth length
[,2]	*supp*	categorical	Supplement type (VC or OJ).
[,3]	*dose*	numeric	Dose in milligrams/day

Through clicking the button "Two way ANOVA with Bonferroni's correction", a dependent continuous variable "len" was chosen and a post hoc text for a categorical variable "dose".



The results appear in a separate window:



The first plot is an interaction plot which as the name suggests shows if the data have interaction. The lines should be parallel for the interaction not to exist.

The second plot is the outlier detection. There should be no significant outliers outside the boxes.

The dataset should pass the above two tests for the results to be considered valid, although some interaction and outliers will still give valid results.

Reporting the results

A two-way ANOVA was conducted that examined the effect of supplement type and dose in milligrams/day on tooth growth of guinea pigs.

The analysis shows that there is statistically significant correlation between the type of supplement ($p < 0.001$) and dose ($p < 0.001$) and tooth growth. Using the post hoc test (Bonferroni's correction) all three doses (0.5, 1.0 and 2.0) are found to have statistically significant difference ($p < 0.001$).

E. Conclusion

For this thesis the version 3.6.2 of Python with Anaconda 4.3.21 and the open source integrated development environment JetBrains PyCharm was used.

In the file *"ach_multiple_comparisons.py"* the multiple comparisons and the calculation of the Bonferroni correction is performed. The graphical user interface was designed with the Python library *tkinter*.

The application overall is lightweight in contrast with many commercially available software. There is also a graphical user interface to make it easier to use.

F. References

[1] Lund Research Ltd, "Independent t-test in SPSS Statistics - Procedure, output and interpretation of the output using a relevant example," 2013. [Online]. Available: https://statistics.laerd.com/spss-tutorials/independent-t-test-using-spss-statistics.php. [Accessed Sep. 2017].

[2] Lund Research Ltd, "One-way ANOVA in SPSS Statistics - Step-by-step procedure including testing of assumptions," 2013. [Online]. Available: https://statistics.laerd.com/spss-tutorials/one-way-anova-using-spss-statistics.php. [Accessed Sep. 2017].

[3] D. G. Altman, Practical Statistics for Medical Research, 1990.

[4] Lund Research Ltd, "Two-way ANOVA in SPSS Statistics - Step-by-step procedure including testing of assumptions," 2013. [Online]. Available: https://statistics.laerd.com/spss-tutorials/two-way-anova-using-spss-statistics.php. [Accessed Sep. 2017].