

Experimental design and biometric research. Toward innovations

Sylwester Białowas
Editor



eISBN 978-83-8211-079-1

<https://doi.org/10.18559/978-83-8211-079-1>



© Copyright by Poznań University of Economics and Business
Poznań 2021



This textbook is available under the Creative Commons 4.0 license – Attribution-Noncommercial-No Derivative Works



INDEPENDENT SAMPLES— SINGLE HYPOTHESIS TESTING



Sylwester Białowąs

Poznań University of Economics and Business



Blaženka Knežević

Faculty of Economics and Business, University of Zagreb



Adrianna Szyszka

Poznań University of Economics and Business



Berislav Žmuk

Faculty of Economics and Business, University of Zagreb

Abstract: In this chapter, the “between-subject” is dealt with, as well as the single hypothesis approach. Both parametrical and non-parametrical versions of the tests are described. All tests are introduced, and the full, step-by-step SPSS guidance is presented. The sections regarding effect size and about writing the report are also included.

Keywords: Kruskal-Wallis H test, Mann-Whitney U test, one-way ANOVA, *t*-test.

As described in the first part of the book, the analysis of the experiment results will follow one of two approaches: between-subject and within-subject. This division is reflected in the analytical part. The first two chapters (1 and 2) are devoted to the between-subject approach, first if only one hypothesis is verified, and the second chapter—when more hypotheses are verified. In the last sub-chapter of this part (3), the authors deal with the within-subject approach.

1.1. Independent samples *t*-test

General information

The independent samples *t*-test is one of the most popular statistical tests. It is used to compare the means of two groups (e.g. age, height, balance in a savings account, bio food expenses, exam scores). It is a basic test in experimental designs when one group is a control group (e.g. receives placebo or usual treatment), while the other one is administered what we want to test. In the *t*-test, the means and standard deviations of two groups are computed and it is checked whether there is a statistically significant difference between means. The compared groups may be selected by the researchers while assigning participants to different conditions or may occur naturally (Verma & Abdel-Salam, 2019). It should be borne in mind that the differences between groups may be caused not only by the manipulation of the researcher but also by different aspects that influence variance, such as individual differences or IQ (Field, 2013).

Hypothesis

In order to compare the scores for two groups, the null and alternate hypotheses should be stated. The null hypothesis is that the mean scores in the two groups are equal. This indicates that the observed difference is due to chance alone. The alternate hypothesis is that the means in two groups differ from each other (Lind, Marchal, & Wathen, 2006).

$$H_0: m_1 = m_2$$

$$H_1: m_1 \neq m_2$$

Assumptions

The following assumptions are associated with the independent samples *t*-test:

- the level of measurement should be interval or ratio (what in SPSS is indicated as the scale level of measurement);

- the samples must be disjoint, which means there should be no relationship between the subjects in the groups, the samples should be unrelated to each other;
- samples should be randomly selected, which means that the data constitute a representative portion of the total population and every individual has the same chance to be selected into the sample (Verma & Abdel-Salam, 2019; Waters, 2011);
- the data should follow normal distribution and the dataset should not include outliers. The researcher should check if there are any extreme (unusually high or low) values in the dataset (Verma & Abdel-Salam, 2019);
- the sample should be reasonably large. Although we can technically carry out the t -test with a group of any size, the results of the t -test are considered stronger with larger samples. It is often recommended that each sample should have about 30 observations, but groups do not necessarily have to include the same number of participants (Lind et al., 2006; Waters, 2011).

Example

Dataset: dwell time of studying information about managing electricity expenses in two groups.

The community managing the apartment blocks has chosen two random groups, each consisting of 105 families living in medium-size flats. (Note, the groups don't have to be equal, they can have different number of cases). Both groups got one page with information on sustainable household management. Electricity management comprised 30% of the page. One of the groups received additional information about the future increase in the price of electricity. The other group was the control (without this info). Using eye-tracking gear, the dwell time in the area of interest (AOI) covering the info about electricity expenses was recorded for every participant.

Data info:

- variable 1: group—nominal, 1—group given the special information, 2—control group;
- variable 2: dwell time—scale, recorded time in seconds spent in the AOI (part about electricity management).

Hypotheses:

H0: There is no difference in dwell time between the groups.

H1: There is a difference in dwell time among both groups.

Testing the assumptions

In the presented example, both groups contain 105 observations, thus the assumption of the group size is met. The size of each group can be read from every SPSS output (e.g. first line of the Kolmogorov-Smirnov test).

Splitting the file—this will cause SPSS to show all the results divided according to the selected groups. In this case, the file division will be carried out according to the variable “group”, therefore, all the results will be shown separately for the groups—“control” and “informed”. This command is valid until it is revoked. For revoking, please open the dialogue box again and click ‘analyze all cases, do not create groups’.

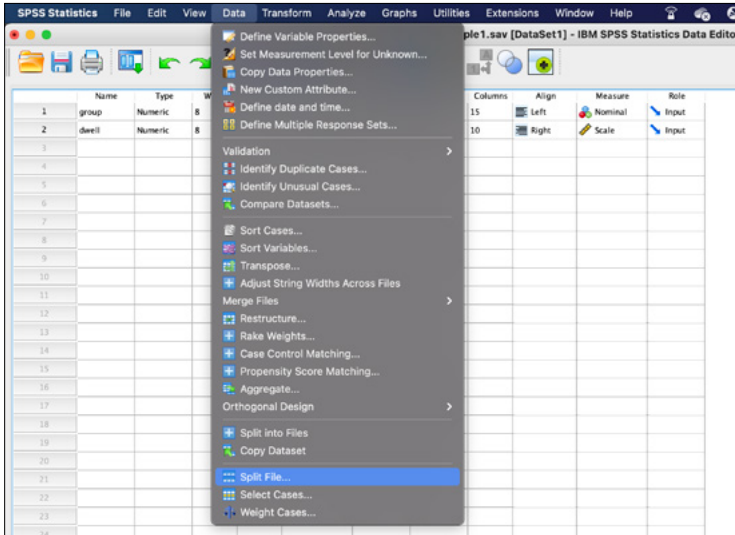


Figure 1. Splitting the file—path

Source: The authors' own elaboration, IBM SPSS screenshot.

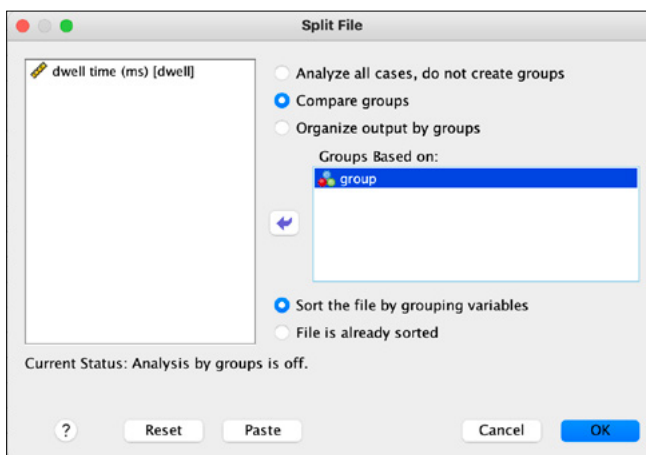


Figure 2. Splitting the file—dialogue box

Source: The authors' own elaboration, IBM SPSS screenshot.

Normality of distribution

The commonly used test for evaluating the normality is the Kolmogorov-Smirnov test. This test allows to compare the set of scores obtained in the study to the normally distributed scores.

Hypotheses for the Kolmogorov-Smirnov test

Null hypothesis (H0): The data follow normal distribution.

Alternate hypothesis (H1): The data significantly differ from normal distribution.

Performing the Kolmogorov-Smirnov test will produce a table with the output for both groups separately (group splitting is still valid).

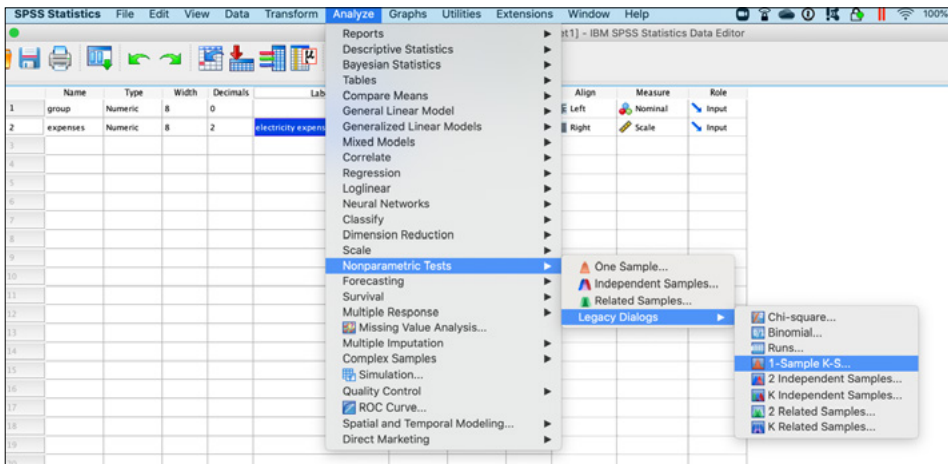


Figure 3. Kolmogorov-Smirnov test for normality of distribution—path

Source: The authors' own elaboration, IBM SPSS screenshot.

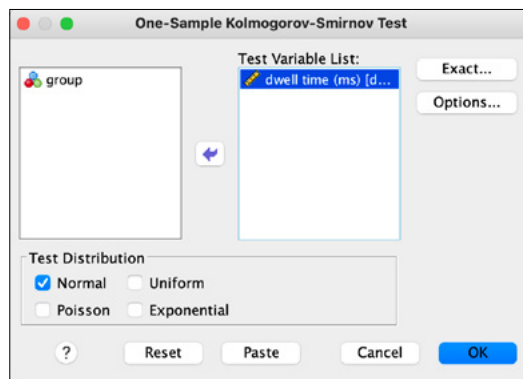


Figure 4. Kolmogorov-Smirnov test for normality of distribution—dialogue box

Source: The authors' own elaboration, IBM SPSS screenshot.

One-Sample Kolmogorov-Smirnov Test			
group			dwelt time (ms)
1 informed	N		105
	Normal Parameters ^{a,b}	Mean	6250.686
		Std. Deviation	577.2935
	Most Extreme Differences	Absolute	.058
		Positive	.058
		Negative	-.046
	Test Statistic		.058
Asymp. Sig. (2-tailed)		.200 ^{c,d}	
2 control	N		105
	Normal Parameters ^{a,b}	Mean	5859.429
		Std. Deviation	522.4182
	Most Extreme Differences	Absolute	.068
		Positive	.051
		Negative	-.068
	Test Statistic		.068
Asymp. Sig. (2-tailed)		.200 ^{c,d}	

Figure 5. Kolmogorov-Smirnov test for normality of distribution—results

Source: The authors' own elaboration, IBM SPSS screenshot.

The hypothesis is determined by interpreting the p -value. If the test is significant ($p < .05$), this means that the data do not follow normal distribution. If the test is non-significant ($p > .05$), the distribution of the obtained scores is normal (Field, 2013; Verma & Abdel-Salam, 2019). In this case, for both groups $p = .200$, which indicates that the assumption of normality is fulfilled.

The next step is performing the t -test itself. Firstly, the splitting of the groups needs to be revoked by clicking in the command “analyze all cases, do not create groups” in the dialogue box (see Figure 2).

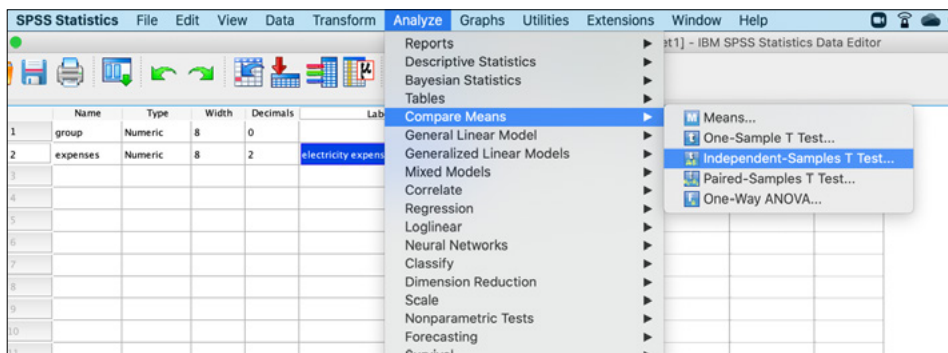


Figure 6. Independent samples t -test—path

Source: The authors' own elaboration, IBM SPSS screenshot.

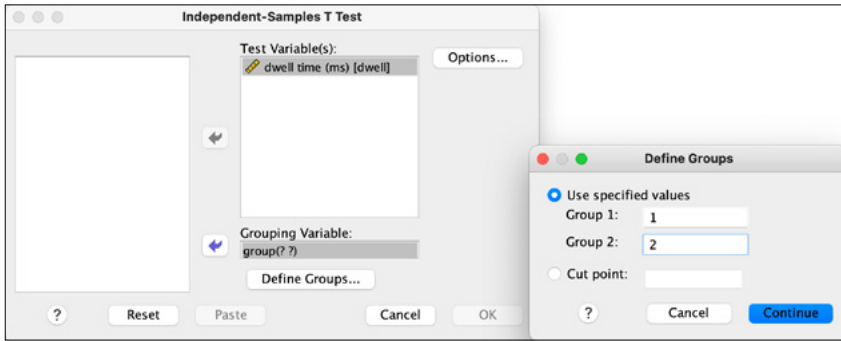


Figure 7. Independent samples t -test—dialogue box

Source: The authors' own elaboration, IBM SPSS screenshot.

Group Statistics					
	group	N	Mean	Std. Deviation	Std. Error Mean
dwell time (ms)	1 informed	105	6250.686	577.2935	56.3381
	2 control	105	5859.429	522.4182	50.9828

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	
dwell time (ms)	Equal variances assumed	1.663	.199	5.149	208	.000	391.2571	75.9817	241.4641	541.0502	
	Equal variances not assumed			5.149	205.959	.000	391.2571	75.9817	241.4554	541.0589	

Figure 8. Independent samples t -test—results

Source: The authors' own elaboration, IBM SPSS screenshot.

Results

The results are interpreted from the lower table (Independent Samples Test). First, the Levene's test of homogeneity in the second column is read (Sig.):

- if $p > .05$, the results are interpreted from the upper row (equal variances assumed);
- if $p < .05$, the results are interpreted from the lower row (equal variances not assumed).

Now, a decision can be made about the significance of the t -test. In this case, it equals $p = .199$. This value is greater than the critical value of $p = .05$, indicating that the results will be read from the upper row (equal variances assumed).

In the lower table, it can be checked if the difference is statistically significant by interpreting the p -value from the 5th column (Sig. 2-tailed). It can be found that $p < .001$, which is lower than the critical value of $p = .05$. This means that the null

hypothesis can be rejected and the results interpreted as the statistically significant difference between the groups.

In the upper table of the outcome (group statistics), it can be noted that the mean for the informed group is 6250.7, while for the control group it totals 5859.4.

The independent t-test hypotheses resolution:

$p < .05$ —there is a significant difference between the groups; reject H_0 ;

$p > .05$ —there is no significant difference between the groups; do not reject H_0 .

Effect size

In order to examine whether the observed difference is significant, the effect size can be calculated. This enables determining the size of the observed effect in a standardised way, which makes the results easy to compare (e.g. with different studies). For the independent samples t -test, a popular measure is Cohen's d :

$$d = \frac{|\bar{x}_1 - \bar{x}_2|}{s_{control}}$$

\bar{x}_1, \bar{x}_2 —means of both groups;

$s_{control}$ —standard deviation of the control group (Cohen, 1988; Field, 2013).

Cohen's d has the following interpretation:

Below 0.2—no effect;

< 0.2–0.5)—small effect;

< 0.5–0.8)—moderate effect;

0.8 and above—large effect.

$$d = \frac{|62.51 - 58.59|}{5.77} = 0.68$$

In this case, a moderate effect can be observed ($d = 0.68$).

Summary

The community managing the apartment blocks has randomly chosen two groups, each consisting of 105 families living in medium-size flats. Both groups received one page with information on sustainable household management and 30% of the page was devoted to electricity management. One of the groups got additional information about future increases in the price of electricity. The other group was the control (without this info). Using eye-tracking gear, dwell time in the AOI, covering the info about electricity expenses, has been recorded for every participant.

Data info:

- variable 1: group—nominal, 1—group given the special information, 2—control group;
- variable 2: dwell time—scale, recorded time in seconds spent in the area of interest (part about electricity management).

The dwell time for the electricity expenses regarding AOI in the group without extra information was at an average of 5859.4 milliseconds ($SD = 522.4$), whereas in the informed group—at an average of 6250.7 milliseconds ($SD = 577.3$). The t -test revealed that the difference of 391.3 milliseconds is statistically significant ($p < .001$, $\alpha = .05$), suggesting that the informed participants focused their attention on the electricity part longer than the control group. The effect size for this analysis ($d = 0.68$) was found to be moderate.

More info about the t -test

The independent samples t -test enable comparison of the scores in two separate groups (populations), and test if there are differences between them. Precisely, the t -test is commonly used in statistics to examine whether the means of two populations are the same (Field, 2013; Verma & Abdel-Salam, 2019; Waters, 2011). Before performing the t -test, the above-mentioned assumptions should be fulfilled (see the ‘Assumptions’ part in this chapter). However, there is another assumption that has not been mentioned in this chapter so far—the homogeneity of variance regarding the data. This means that the samples should be selected from populations that have equal variance with reference to some criterion. The reason for not mentioning the homogeneity of variance is because performing the t -test in SPSS enables interpretation of the results even if this assumption is violated. Specifically, together with the t -test output, this generates Levene’s test and calculates the results for both equal and unequal variances (in case of lack of homogeneity between groups, the results can be read from the lower row). What also should be emphasized is that violating the homogeneity of variance assumption applies only if the sizes of tested groups are unequal (Field, 2013). However, the other assumptions may sometimes be violated as well and there are certain ways to deal with some of them. If the assumption of normal distribution is not fulfilled, there are techniques to convert the data distribution into at least quasi-normal (e.g. log, root, or Box-Cox transformation). If this is not possible, non-parametric tests should be used (Verma & Abdel-Salam, 2019).

It should be noted that the Kolmogorov-Smirnov test is not the only way of checking the normality of distribution. Another popular test that is usually used for this purpose is the Shapiro-Wilk test (considered as better for smaller samples). The hypotheses and interpretation of test statistics are analogical. However, among researchers, there are discussions about the necessity of testing normality before using the independent sample t -test. That is because if tests depend on the hypothesis testing, this may consequently show significant effects for large samples, even in case

of irrelevant influences. On the other hand, for smaller samples, the test results may not indicate that the assumption is violated (Field, 2013). For more experienced investigators, the normality of distribution can be assessed using histograms or by assessing skewness and kurtosis.

For effect size calculation, the simplified formula of Cohen's d was proposed with standard deviation of the control group in the denominator. This approach is justified because it can be assumed that the treatment in the study may affect not only the mean, but also dispersion in the dataset. However, there are other possibilities of standard deviation calculations used as a standardiser. The commonly accepted formula has pooled standard deviation in the denominator that is given by the following equation:

$$s^* = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

This formula finds its application especially when there is a remarkable difference between standard deviations of both groups. Its advantage also depends on including the sizes of samples (Borenstein, Hedges, Higgins, & Rothstein, 2009; Cumming, 2012; Dean & Illowsky, 2013; Field, 2013).

Furthermore, when calculating effect size, there is inconsistency in terminology. The formula with no pooling (with control standard deviation in the denominator) is also referred to as Glass' d or Glass' Δ . The researchers sometimes refer to Hedge's g as the measure with pooled standard deviation as a standardiser. Recently, using the d for all different formulas prevails. However, it is crucial to explain how the effect size was calculated (Cumming, 2012).

References

- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Effect sizes based on means. In *Introduction to meta-analysis*. John Wiley & Sons.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge Taylor & Francis Group.
- Dean, S., & Illowsky, B. (2013). *Introductory statistics*. OpenStax College.
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). Sage edge.
- Lind, D. A., Marchal, W. G., & Wathen, S. A. (2006). *Basic statistics for business & economics* (5th ed.). McGraw-Hill.
- Verma, J. P., & Abdel-Salam, G. A.-S. (2019). *Testing statistical assumptions in research*. John Wiley & Sons, Inc.
- Waters, D. (2011). *Quantitative methods for business* (5th ed.). Pearson Education Limited.

1.2. Mann-Whitney U test

General information

The Mann-Whitney U test is a nonparametric test that is an alternative for the independent sample t -test. Generally, this test can be carried out when the assumptions for using the t -test are not met. The Mann-Whitney U test is used particularly in two cases—when at least one variable is ordinal or when the continuous data do not follow normal distribution. The Mann-Whitney U test assesses whether samples are drawn from the same population.

Assumptions

The following assumptions associated with the Mann-Whitney U test can be put forward:

- the measurement level of the dependent variable should be at least ordinal;
- the samples must be disjoint—there should be no relationship between the subjects in both groups, the samples should be unrelated to each other (Verma & Abdel-Salam, 2019).

Example

Dataset: The company managing sharing bicycles decided to check the impact of the station location on the use of bicycles. Two comparable high-schools were chosen. In the case of one of them (control group), the location of the station was 200 m from the entrance and in the other (test group), the station was located just in front of the entrance.

After two months of experiment, two random samples of students from each school were selected. Respondents declared the frequency of using the shared bicycles.

Data info:

- variable 1: group—nominal (1—distant location, 2—close location);
- variable 2: freq.—ordinal (declared frequency of using the shared bicycles; 1—more than once a day; 2—every day; 3—2–4 times a week; 4—once a week; 5—once a month; 6—less than once a month; 7—never).

Hypotheses:

H0: There is no difference in the frequency of using shared bicycles between the groups.

H1: The frequency of using shared bicycles differs in both groups.

Testing the assumptions

There are two unrelated groups and the frequency is measured on an ordinal scale.

In the first step, the medians are computed for both groups. In order to obtain the output, the file is split (procedure described in 1.1.) and descriptive statistics are run.

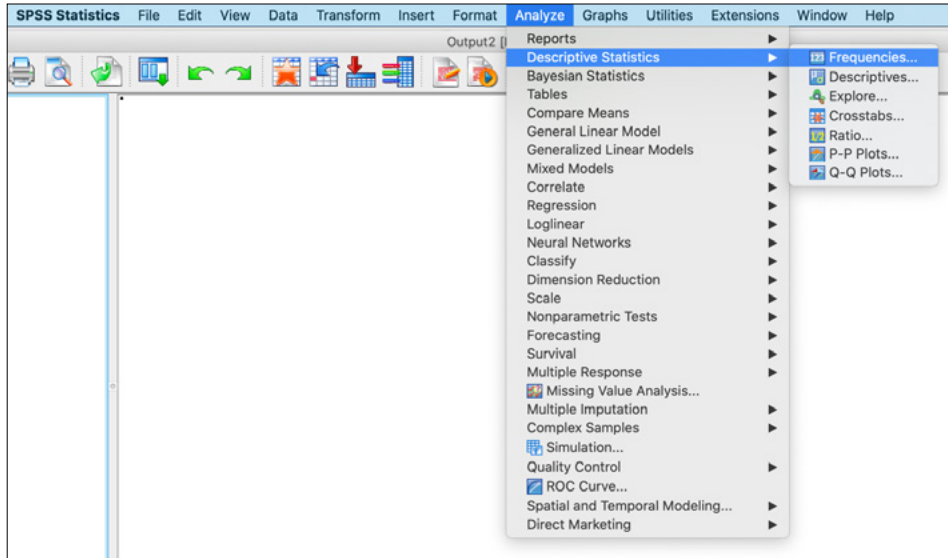


Figure 9. Descriptive statistics—path

Source: The authors' own elaboration, IBM SPSS screenshot.

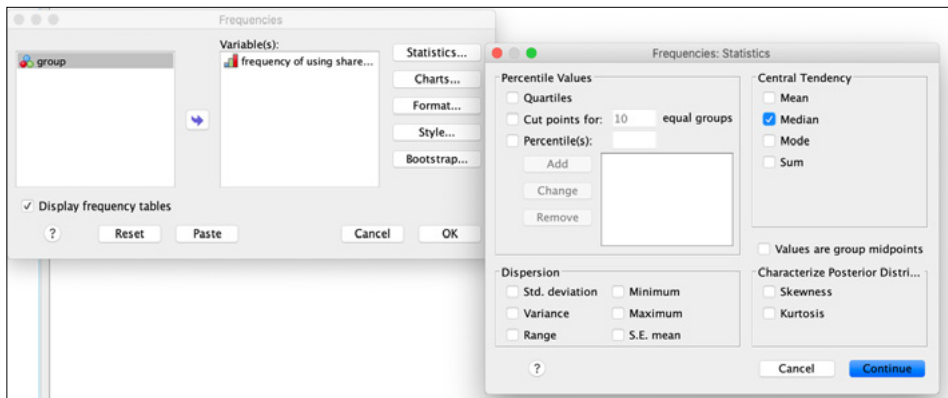


Figure 10. Descriptive statistics—dialogue box

Source: The authors' own elaboration, IBM SPSS screenshot.

Statistics			
frequency of using shared bicycles			
1 control	N	Valid	45
		Missing	0
	Median		5.00
2 test	N	Valid	39
		Missing	0
	Median		3.00

Figure 11. Descriptive statistics—results

Source: The authors' own elaboration, IBM SPSS screenshot.

The medians are 3 and 5 for the test and control groups, respectively. The number of observations can be seen as well. In the next step, the Mann-Whitney U test is performed. It will be compared whether the difference between groups is statistically significant. Before running the test, it must be remembered to split the groups by using the command “analyse all cases, do not create groups” in the dialog box (procedure described in 1.1).

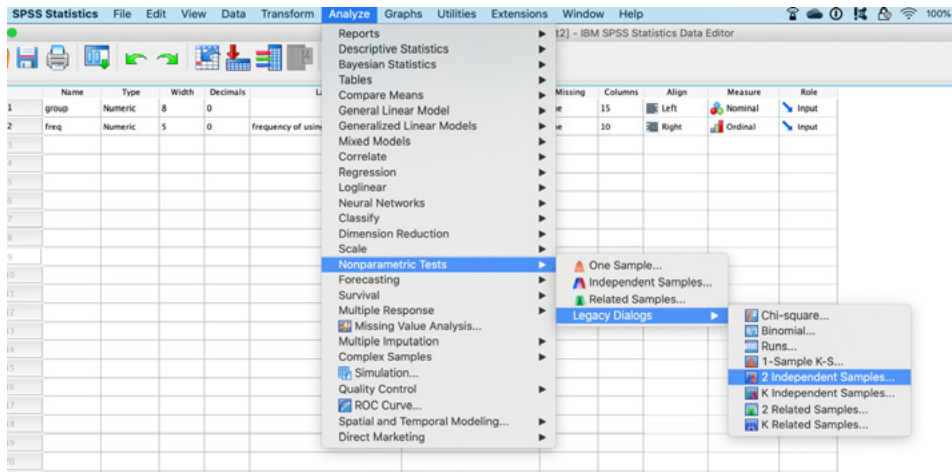


Figure 12. The Mann-Whitney U test—path

Source: The authors' own elaboration, IBM SPSS screenshot.

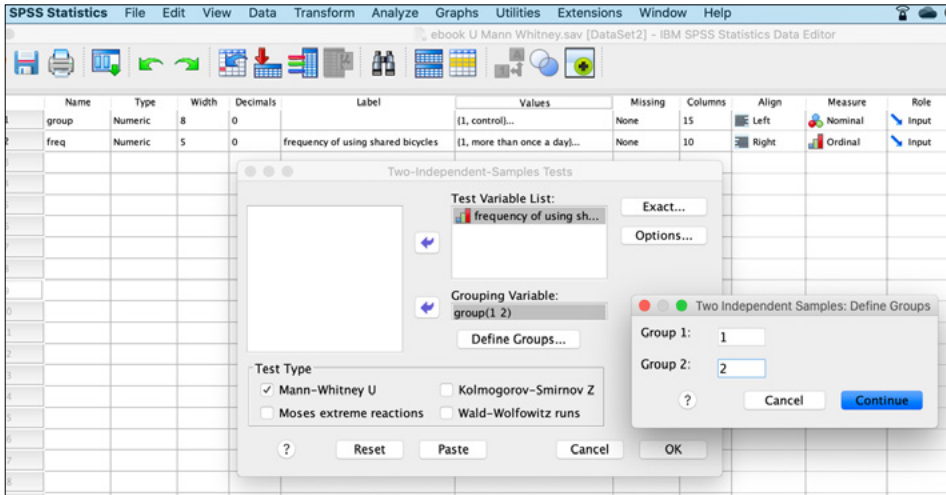


Figure 13. Mann-Whitney U test—dialogue box

Source: The authors' own elaboration, IBM SPSS screenshot.

Mann-Whitney Test				
Ranks				
	group	N	Mean Rank	Sum of Ranks
frequency of using shared bicycles	1 control	45	48.37	2176.50
	2 test	39	35.73	1393.50
	Total	84		

Test Statistics ^a	
	frequency of using shared bicycles
Mann-Whitney U	613.500
Wilcoxon W	1393.500
Z	-2.407
Asymp. Sig. (2-tailed)	.016

a. Grouping Variable: group

Figure 14. Mann-Whitney U test—results

Source: The authors' own elaboration, IBM SPSS screenshot.

Results

The results are interpreted from the last row in the lower table (test statistics). The significance equals $p = .016$, which is lower than the critical level of $p = .05$. This indicates that there is a significant difference in frequencies of using shared bicycles between the groups.

In Figure 11 (descriptive statistics results), it can be noted that the median for the control group is five (once a month) and for the test group it totaled three (2–4 times a week).

Mann-Whitney U test hypotheses resolution:

$p < .05$ —there is a significant difference between the groups; reject H_0 ;

$p > .05$ —there is no significant difference between the groups; do not reject H_0 .

Effect size

The effect size measure for the Mann-Whitney U test is the r (*do not confuse with Pearson's r*), which is calculated using the statistic Z value and n_1, n_2 being the total number of observations in both groups:

$$r = \frac{|Z|}{\sqrt{n_1 + n_2}}$$

The r has the following interpretation:

Below .1—no effect;

< .1-.3)—small effect;

< .3-.5)—moderate effect;

.5 and above—large effect (Pallant, 2011; Field, 2013).

$$r = \frac{|-2.407|}{\sqrt{84}} = .26$$

In this case, a small effect ($r = .26$) can be observed.

Summary

Dataset: The company managing sharing bicycles decided to check the impact of the station location on use of the bicycles. Two comparable high-schools were chosen. In the case of one of them (control group), the location of the station was set 200 m from the entrance and in the other (test group), the station was located just in front of the entrance.

After two months of the experiment, two random student samples from each school took part in the study. Respondents declared the frequency of using the shared bicycles.

Data info:

- variable 1: group—nominal (1—distant location, 2—close location);
- variable 2: freq.—ordinal (declared frequency of using the shared bicycles; 1—more than once a day; 2—every day; 3—2–4 times a week; 4—once a week; 5—once a month; 6—less than once a month; 7—never).

Using the bicycles in the test group was more frequent ($Mdn = 3$; once a month) than in the control ($Mdn = 5$; 2–3 times a week). The Mann-Whitney U test allows to indicate that this difference is statistically significant: $U(N_{control} = 45, N_{test} = 39) = 613.50, Z = -2.41, p = .016$.

It can be assumed that the location of the station has significant impact on the frequency of using the bicycles. This effect is considered small ($r = .26$).

References

- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). Sage edge.
- Pallant, J. (2011). *SPSS survival manual: a step by step guide to data analysis using SPSS* (4th ed.). Allen & Unwin.
- Verma, J. P., & Abdel-Salam, G. A.-S. (2019). *Testing statistical assumptions in research*. John Wiley & Sons, Inc.

1.3. One-way analysis of variance (ANOVA)

Theoretical background

One-way analysis of variance (ANOVA) is used to determine if there is a significant difference between means of several subpopulations (groups) dependent on one factor. In ANOVA, independent variables are organised in categorical groups (Dean & Illowsky, 2013; Field, 2013; Fraser, 2016). For example, if the difference in one's average daily income in January, February, March and April is to be tested, then there will be four groups of data (according to particular month), and daily income expressed in some currency will be the dependent variable. If it is to be tested whether there is a difference in sales when merchandise is displayed in a window, in the centre of the shop or at some point behind sales person, there will be three groups: "window", "centre", "behind", and for one particular product, sales will be measured in some period according to those positions. The value of the daily sales will be the dependent variable. Also, ANOVA is useful when wanting to observe if there is a significant difference in consumer behaviour regarding various socio-demographic characteristics. In addition, ANOVA can be useful when wanting to analyse effectiveness of sales force in different locations.

One-way ANOVA is usually utilised when comparing three or more categorical independent groups to establish whether there is a statistically significant difference between them (Field, 2013; Barrow, 2017). One-way ANOVA can be used in the case of just two categorical independent groups, but in that case, the independent sample t -test is more frequently used. It is recommended each category (group) contain at least two units or two measurements in order to be able to calculate variance.

Hypothesis

The null hypothesis is that the means of all groups are equal, i.e. that the observed difference between groups is due to chance. On the other hand, the alternate hypothesis is that there is at least one pair of groups where the mean between them is significantly different.

Assumptions

The following assumptions can be associated with one-way ANOVA independent samples (Dean & Illowsky, 2013; Field, 2013; Randolph & Myers, 2013):

- the independent variable should consist of two or more categorical, independent groups. Typically, one-way ANOVA is used when there are three or more groups, but it can also be used for only two groups (even though the independent sample *t*-test is more commonly used in that case);
- the samples are disjoint, there is no relationship between the observations in each group or between the groups themselves. For example, one participant has to be exclusively in one group;
- the dependent variable should be measured at the interval or ratio level (i.e. they have to be continuous);
- the dependent variable should be approximately normally distributed for each category of the independent variable and there should be no significant outliers;
- homogeneity of variance is required. Therefore, it is recommended to perform Levene's test for homogeneity before application of one-way ANOVA.

Example

Dataset: quantity of food waste measured in grams per month, per person observed in four groups of consumers, according to age groups: 18–25; 26–40; 41–60; above the age of 60. Food waste as a problem is growing in the modern world. There are some studies in which it is shown that age might be the crucial factor when explaining difference in consumer behaviour regarding food waste. Thus, it is enquired whether there is a difference between generations of consumers regarding food waste on a monthly basis. Therefore, research was carried out in which the respondents were asked to assess the quantity of wasted food on a personal level within one month in grams. The survey was carried out using a random sample of 200 respondents.

Data info:

- variable 1: groups—nominal (1—age 18–25, 2—age 26–40, 3—age 41–60, 4—above the age of 60);
- variable 2: food waste quantity—numeric (grams of wasted food in grams per person in a month).

Hypotheses:

H0: There is no difference in mean food waste quantities between the groups.

H1: There is at least one group at which mean food waste quantity is different than in the other groups.

Testing the hypotheses in SPSS

In this example, 200 respondents were studied, and for each respondent, two types of data were collected: (1) age, (2) level of food waste in grams per month, per person. Three questionnaires (observations) were not valid, thus the dataset was based on 197 valid questionnaires (or observations).

SPSS does not require grouping the collected survey data, but data is entered as an observation per row. In Figure 15, in the first row—in the “Generation” column, data on the generation of the respondent is entered and in column “Foodwaste” data on food waste for this respondent is entered, therefore, in row 40, it can be observed that the respondent’s age is 18–25 (generation group numbered as 1) and respondent wastes 195 grams of food per month (see Figure 15).

Prior to analysis, the type of loaded data has to be checked, and it is recommended that numerical denomination of categories is used for the independent variable (Field, 2013). That means instead of the text “Group 1 (18–25)”, 1 should be used to denominate this particular generation of consumers, similarly—“Group 2 (26–40)” will be coded as 2, etc. It is important to emphasize that with introducing numeric codes for the variable does not strengthen its measurement level. It is still categorical (in our case: ordinal). In Figure 15, see column = “Generation”.

Row	Generation	Foodwaste
40	1	195
41	1	129
42	1	230
43	1	161
44	1	350
45	1	285
46	1	152
47	1	230
48	1	235
49	1	158
50	1	234
51	1	176
52	1	154
53	2	292
54	2	302
55	2	309
56	2	363
57	2	324
58	2	454
59	2	364
60	2	339
61	2	502
62	2	453
63	2	430
64	2	407
65	2	503
66	2	331

Figure 15. Excerpt from dataset in SPSS

Source: The authors’ own elaboration.

Before conducting ANOVA, it is recommended to calculate means of groups (Note: This step can be skipped, because the latter in the one-way ANOVA procedure, the option to display descriptive statistics can be chosen, which will show the summarised descriptive statistics data for each group).

In Figure 16, the screenshot shows the command for calculation of means, while in Figure 17, it is demonstrated how to set options in order to calculate means for various age groups (generations) of consumers based on the dependent variable “Food waste in grams” from the dataset.

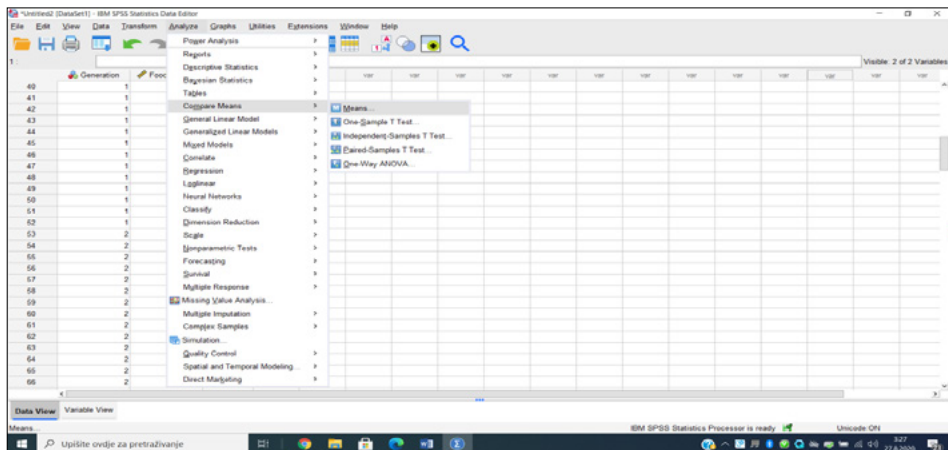


Figure 16. SPSS Command to calculate means

Source: The authors' own elaboration.

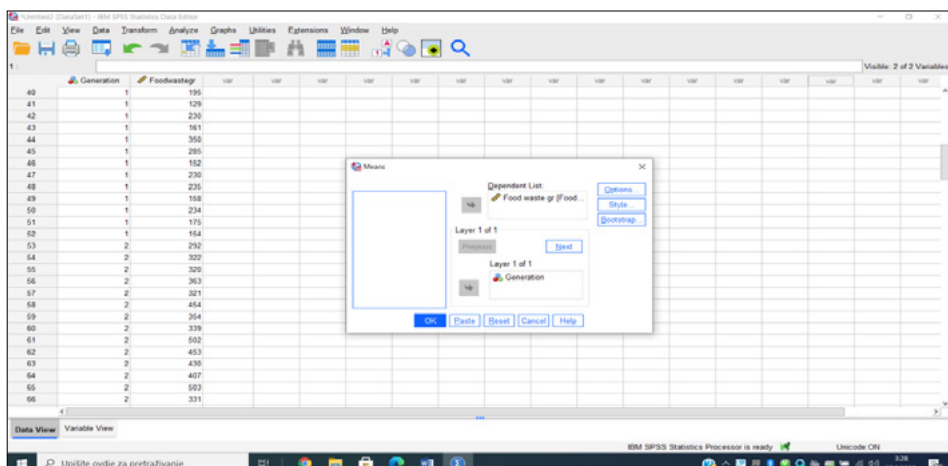


Figure 17. Setting variables to calculate means for groups in the dataset

Source: The authors' own elaboration.

Output of means calculation is given in Figure 18. The first part of the output is the summary based on the total sample from which it can be read how many cases from the dataset are included or excluded from means calculation. In this case, all observations were correct, therefore, all data is included in the calculation of means. The second part of the report are means, number of cases (observations) and standard errors for each group in the sample, i.e. for each generation of consumers. For instance, for generation 4 or “Group 4 (60+)”, it can be observed that average monthly food waste per person is 290.52 grams, the result based on 50 cases (observations) with the standard deviation of 89.30. Compared to the total sample, this generation has a lower average of food waste. Namely, the average monthly food waste, taking all 197 respondents into account, is 376.85 grams per person.

Case Processing Summary						
	Cases					
	Included		Excluded		Total	
	N	Percentage	N	Percentage	N	Percentage
Food waste (gr) * Generation	197	100.0%	0	0.0%	197	100.0%

Report			
Food waste (gr)			
Generation	Mean	N	Std. Deviation
1	261.88	52	80.236
2	363.18	50	99.787
3	620.82	45	12.660
4	290.52	50	89.300
Total	376.85	197	169.944

Figure 18. SPSS means report

Source: The authors' own elaboration.

After that, the procedure for one-way ANOVA will be started. Selection of SPSS required command is shown in Figure 19, while in Figure 20, the dialogue for tuning up settings in the presented example is given.

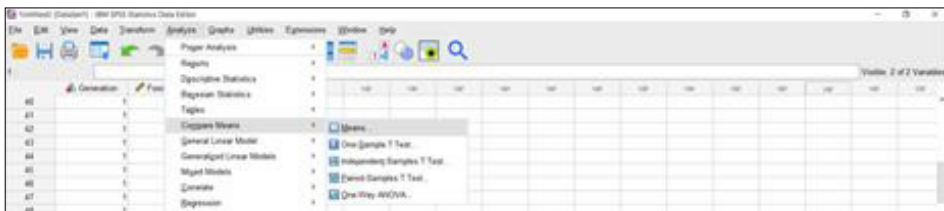


Figure 19. SPSS Command for one-way ANOVA

Source: Authors' own elaboration.

Independent samples—single hypothesis testing

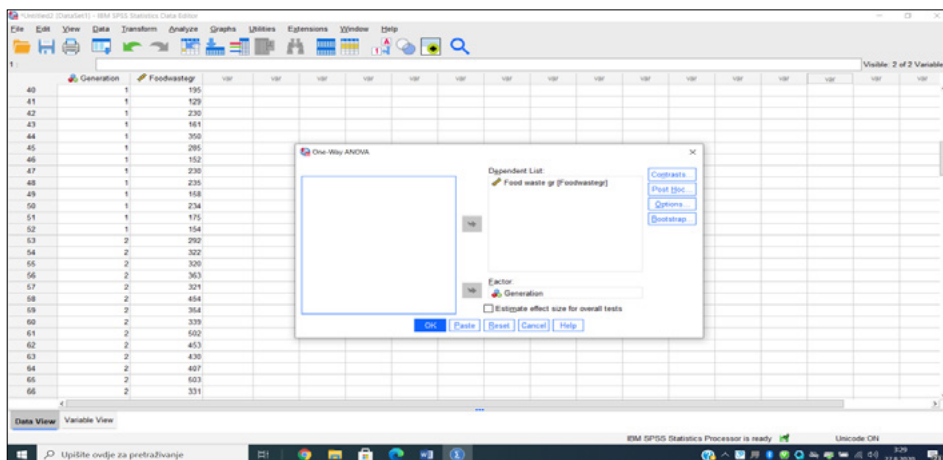


Figure 20. Settings of one-way ANOVA for groups in the dataset

Source: The authors' own elaboration.

Together with basic settings, SPSS can be set to perform post hoc analysis in the same run. Therefore, the 'Post Hoc' button should be clicked, and the dialogue box shown at Figure 21 will appear. Usually, it is enough to do Tukey's post hoc analysis at the confidence level of .05 (necessary settings are shown in Figure 21). When everything is set up, the analysis will be run.

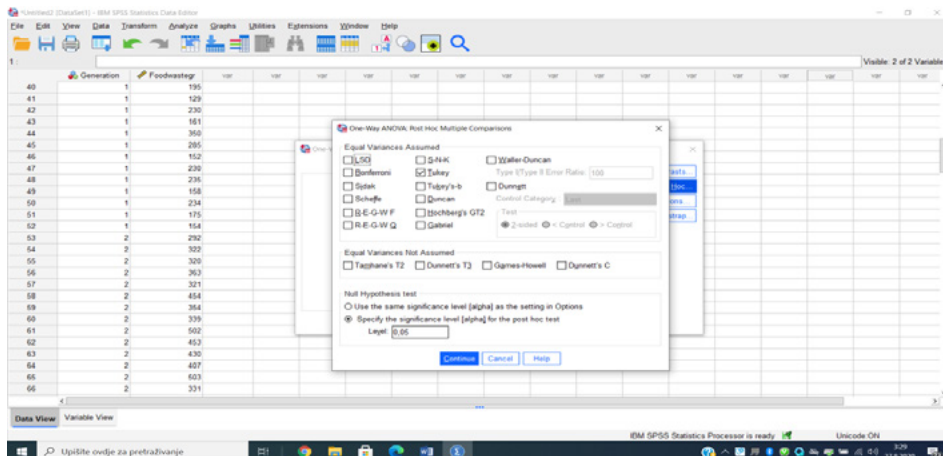


Figure 21. Settings of Tukey's post hoc analysis

Source: The authors' own elaboration.

In Figure 22, the output of one-way ANOVA is shown. It has to be emphasized that the significance value on the output is written as equal to 0.000 ($p = .000$), but

that does not mean that the significance value equals zero. That is just the way SPSS tells us that the significance value is below .001. Thus, in accordance with the output, it can be concluded that the significance value is very small and, for sure, lower than .05. Therefore, at the significance level of .05, the null hypothesis of the test that there is no difference in mean food waste quantities between the groups can be rejected. So, it can furthermore be concluded that there is at least one group in which mean food waste quantity is different than in the other age groups (generations).

Post hoc analysis provides scrutinized insight into differences between pairs of groups. As a result, the significance value (see Sig. column in Post hoc analysis) can be observed for each age group compared to other age groups. In the presented example, it can be seen that all significance values are less than .05, except for the value use to compare age groups (generation) 1 and 4 ($p = .469$). Therefore, for instance, it can be assumed that the average quantity of food waste per month, per person from generation 1 is statistically different compared to generations 2 and 3, respectively. However, at the significance level of .05, the hypothesis cannot be rejected that there is no statistically significant difference between generations 1 and 4 regarding the average quantity of food waste per month, per person.

ANOVA					
Food waste gr					
	Sum of squares	df	Mean square	F	Sig.
Between groups	3747782.985	3	1249260.995	126.044	.000
Within groups	1912881.745	193	9911.304		
Total	5660664.731	196			

Post hoc tests

Multiple comparisons						
Dependent variable: Food waste gr						
Tukey HSD						
(I) Generation	(J) Generation	Mean difference (I - J)	Std. Error	Sig.	95% Confidence interval	
					Lower bound	Upper bound
1	2	-101.295*	19.719	.000	-152.40	-50.19
	3	-358.938*	20.270	.000	-411.47	-306.41
	4	-28.635	19.719	.469	-79.74	22.47
2	1	101.295*	19.719	.000	50.19	152.40
	3	-257.642*	20.457	.000	-310.66	-204.63
	4	72.660*	19.911	.002	21.06	124.26
3	1	358.938*	20.270	.000	306.41	411.47
	2	257.642*	20.457	.000	204.63	310.66
	4	330.302*	20.457	.000	277.29	383.32
4	1	28.635	19.719	.469	-22.47	79.74
	2	-72.660*	19.911	.002	-124.26	-21.06
	3	-330.302*	20.457	.000	-383.32	-277.29

Figure 22. Output of one-way ANOVA in SPSS

Source: The authors' own elaboration.

Testing hypotheses in Excel

In order to perform analysis of the same dataset in Excel, collected data has to be prepared for analysis, i.e. collected data has to be classified into columns that represent groups (Balakirshnan, Render, & Stair, 2007; Winston, 2016; Fraser, 2016). In our case columns will represent groups by age—generations of consumers. Therefore, in this case, the collected data will be classified into four columns and each column will be labelled according to consumer generation (in Figure 23, see title of columns in row 3). Then, all observed values will be entered for each generation of consumers. For instance, if a certain respondent is from generation 2 (age 26-40) and wastes 407 grams of food per month, his/her data is entered into the second column—‘Group 2 (26-40)’ (in Figure 23, see row 15). In the SPSS dataset, data on this respondent was entered as a simple observation in a single row as 2 and 407 (see Figure 15, row 64).

	Group 1 (18-25)	Group 2 (26-40)	Group 3 (41-60)	Group 4 (60+)
4	269	292	740	178
5	244	322	600	368
6	215	320	622	277
7	390	363	546	253
8	163	321	631	205
9	293	454	507	174
10	338	354	672	348
11	330	339	617	292
12	700	502	483	331
13	233	453	651	199
14	223	430	502	267
15	396	407	757	438
16	388	503	827	426
17	384	331	561	351
18	185	496	813	153
19	398	230	557	377
20	354	289	679	402
21	351	268	822	340
22	165	382	375	220
23	366	519	540	189
24	272	225	511	237
25	212	531	593	423
26	328	250	748	240
27	300	408	520	417
28	206	373	551	236

Figure 23. Excerpt from dataset for one-way ANOVA of food waste according to age

Source: The authors' own elaboration.

Then, ‘Data tab’ has to be selected and ‘Data Analysis’ (within Analysis group of commands) clicked. (Note that Data Analysis pack is not default package, you have to install it in your Excel). From among the list of methods, ‘Anova: Single Factor’ is chosen (see Figure 24).

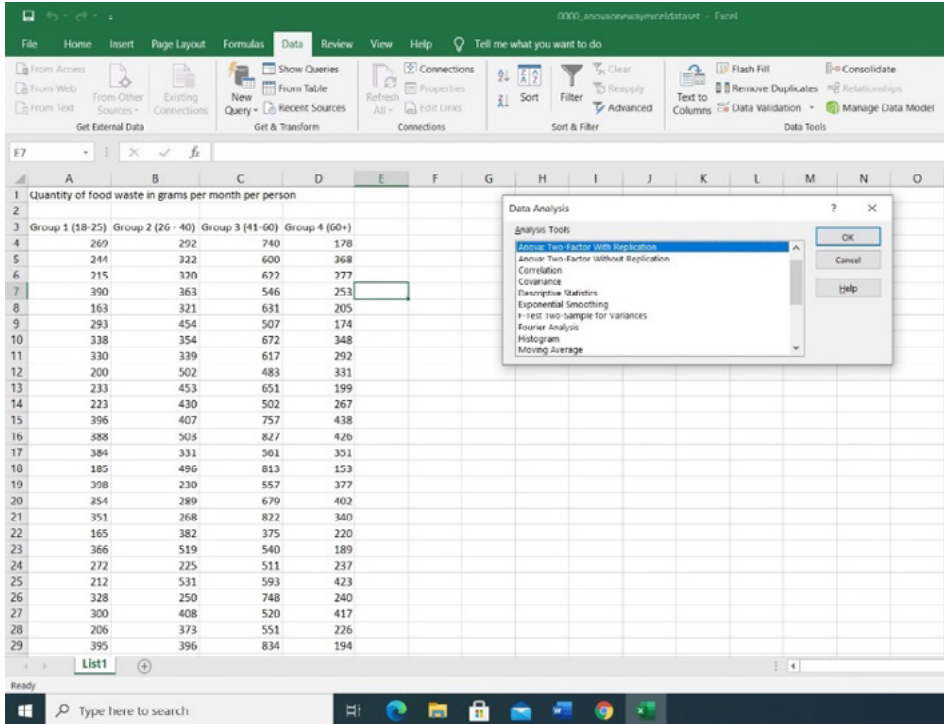


Figure 24. Data analysis tab in Excel—selection of ANOVA method: Single Factor

Source: Authors' own elaboration.

In the dialogue box of Anova: Single Factor—configuration has to be carried out as follows (see Figure 25):

- input range of the dataset including labels, in this example—A3:D55;
- position of data labels, in this example—First Row (there are names of the observed groups);
- way of organising groups of data, in this case, data is organised in columns, therefore, 'Columns' is chosen;
- output range—data can be chosen to be shown at some position in the active worksheet. Then, the exact cell, from which our results are going to be presented (such as F3), has to be specified; but in this case, we rather specified 'New worksheet' was indicated as the location for results. A name for the output can be specified (in this example—'Anova1');
- finally, the level of significance, i.e. alpha value. The default value, already set to .05, can be used.

Independent samples—single hypothesis testing

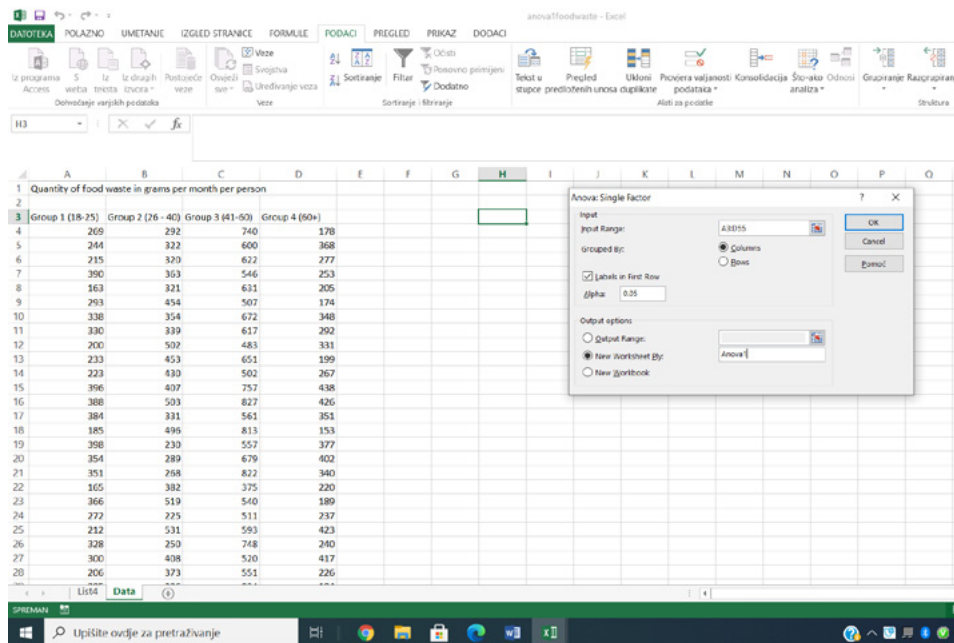


Figure 25. Dialog box—ANOVA: Single Factor

Source: The authors' own elaboration.

In Figure 26, the results of data analysis are shown, and the results can be interpreted. First of all, basic descriptive statistical data on each age group is obtained (see SUMMARY). From this part, it can be read how many respondents are in which group, then, what the average food waste in each group is, as well as the variance within each group. For instance, the lowest average of 261.88 grams of food waste per person, per month is shown in 'Group 1' (aged 18–25). The highest average value is in 'Group 3' (aged 41–60) and amounts to 620.82 grams a month, per person. In addition, ANOVA results are shown. In this table, the most important reading is p -value, because using this value, it can be decided not to reject or to reject the null hypothesis. In this case, the p -value is 3.08×10^{-45} , or if rounded and truncated to four decimal points, the p -value is: .0000. However, the more precise would be if it were said that the p -value is lower than .0001 (p -value $< .0001$). In this way, it can be concluded that the significance value is much lower than that of .05. Consequently, that result means that the null hypothesis H_0 can be rejected and that there is no difference in mean food waste quantities between groups. In other words, at a significance level of .05, it may be concluded that there is at least one group in which mean food waste quantity is statistically different than in the other age groups (generations).

The screenshot shows an Excel spreadsheet with the following data:

Groups	Count	Sum	Average	Variance
Group 1 (18-25)	52	13618	261.8846	6437.869
Group 2 (26 - 40)	50	18159	363.18	9957.416
Group 3 (41-60)	45	27937	620.8222	16042.88
Group 4 (60+)	50	14526	290.52	7974.5

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	3747783	3	1249261	126.0441	3.08E-45	2.651396
Within Groups	1912882	193	9911.304			
Total	5660665	196				

Figure 26. One-way ANOVA results

Source: The authors' own elaboration.

If ANOVA shows that there is a statistically significant difference between observed groups, post hoc analysis has to be carried out by comparing pairs of groups in order to explain which groups differ in comparison to the other groups. For this purpose, several t -tests can be performed in Excel.

In the presented example, the t -test will be performed between Group 1 and Group 2 as an example. This kind of comparison is then done to compare Group 1 to Group 3, Group 1 to Group 4, Group 2 to Group 3 and Group 2 to Group 4. The t -tests have to be repeated accordingly to investigate differences between all possible pairs of groups in the dataset.

Steps for performing the t -test in Excel are the following: first, it must be specified which type of t -test it to be performed. This is done via the 'Data analysis' tab (see Figure 27).

During this step, the t -test: 'Two Samples Assuming Equal Variance' is chosen.

Independent samples—single hypothesis testing

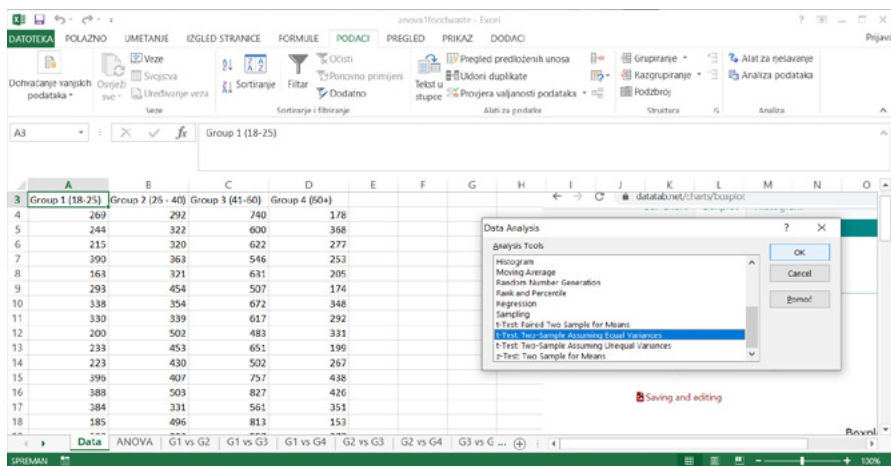


Figure 27. Data analysis tab in Excel—selection of t -test type

Source: The authors' own elaboration.

Then, in the t -test dialogue box (see Figure 28), it has to be specified which data is to be compared. The first pair of data comprises Group 1 (18–25) and Group 2 (26–40). Therefore, the range of data for Group 1 in 'Variable 1 Range' is specified, and the same is done for 'Variable 2 Range', giving the range of data from Group 2. Moreover, the data has data labels in the first row of selected data range, thus, 'Labels' have to be checked. Finally, the location for the output or results are specified. In this case, it was decided to have a new worksheet named 'G1 vs. G2'.

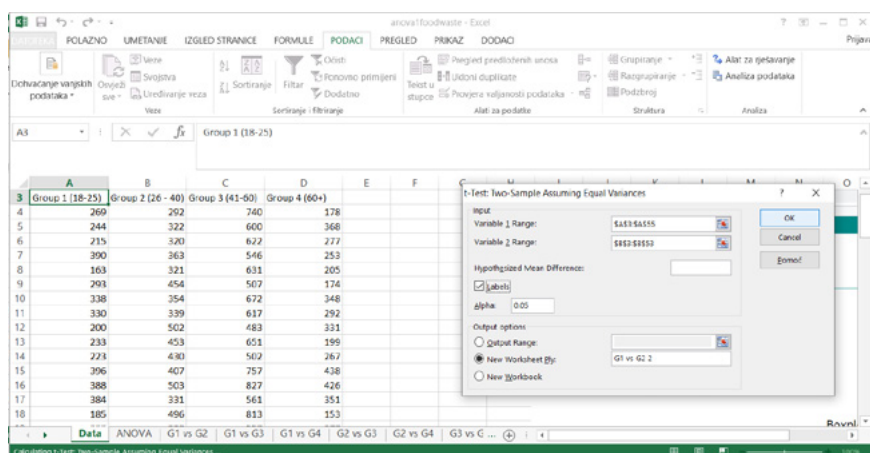


Figure 28. Dialogue box for t -test: Two Sample Assuming Equal Variances

Source: The authors' own elaboration.

After clicking ‘OK’, the output of the t -test is shown (see Figure 29) and interpretation can be carried out on the basis of the analysed pair of variables (in this test—Group 1 and Group 2).

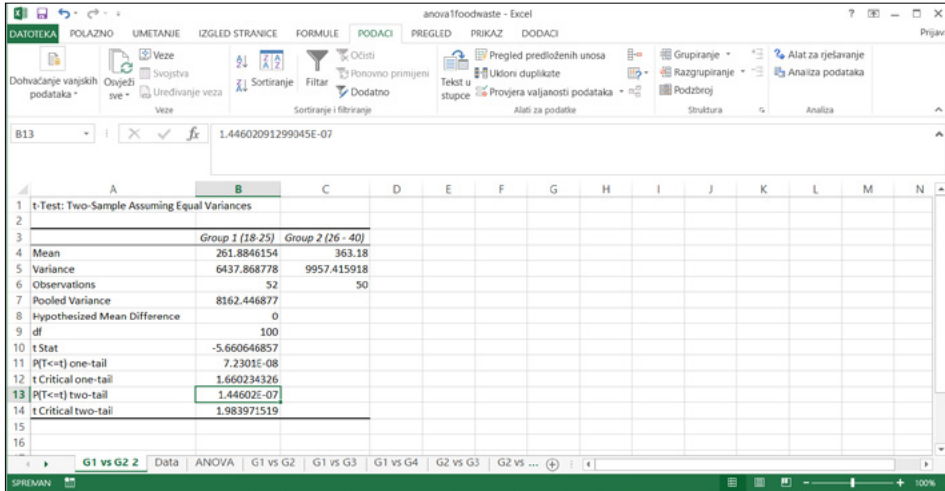


Figure 29. T -test results for pair of groups

Source: The authors' own elaboration.

If the post hoc t -test results are to be interpreted, the p -value for two-tail comparison is used (see Figure 21). Based on the level of significance of .05, it can be concluded that the given p -value (in this case: 1.4607×10^{-7} or truncated to four decimal digits: 0.0000), is lower than .05 and that there is a statistically significant difference in means between Group 1 (18–25) and Group 2 (26–40). However, by doing so, an erroneous conclusion could be drawn. Therefore, as suggested in literature on the subject, before final conclusions, the significance level of .05 has to be adjusted according to number of groups of data in the ANOVA. As in this case there are 4 groups of data (according to the age of respondents), the relevant value for comparison would be .05 divided by 4, i.e. .0125. Thus, in order to carry out the correct interpretation and reach adequate conclusions, the given p -value of the t -test has to be compared for each pair of groups in the dataset to .0125, not to .05! In this case, .0000 is lower than .0125, and it may be concluded that there is a statistically significant difference between the means for Group 1 (18–25) and Group 2 (26–40).

After this, the t -test is iteratively repeated for all pairs of data in a similar way. In Table 1, the totalled t -test p -values relevant for each pair of groups is shown. The p -value is given in a default calculated format (scientific) and then in numeric

format truncated to 4 decimal digits. From the table, it may be concluded that at the level of .05, there is a statistically significant difference between all observed groups, except for Group 1 (18-25) and Group 4 (60+). For that pair of groups, the calculated p -value is higher than the adjusted significance level of .0125. Therefore, we cannot reject the hypothesis that there is no difference between the means of those groups.

Table 1. *T*-test relevant p -values for given example

Pair of groups		t-test p -value (scientific)	t-test p -value (numeric)	Decision (according to adjusted significance 0.0125)
Group 1 (18-25)	Group 2 (26 - 40)	1.44602E-07	0.0000	reject H0 (there is difference)
Group 1 (18-25)	Group 3 (41 - 60)	2.23E-30	0.0000	reject H0 (there is difference)
Group 1 (18-25)	Group 4 (60+)	0.091316825	0.0913	not reject H0 (there is no difference)
Group 2 (26 - 40)	Group 3 (41-60)	1.16784E-18	0.0000	reject H0 (there is difference)
Group 2 (26 - 40)	Group 4 (60+)	0.000220728	0.0002	reject H0 (there is difference)
Group 3 (41-60)	Group 4 (60+)	3.46163E-26	0.0000	reject H0 (there is difference)

Source: The authors' own elaboration.

It must be borne in mind that such post hoc analysis is performed only in the case when ANOVA indicates that there is a difference between means of several groups of data in the dataset in order to interpret data more accurately and precisely (Fraser, 2016; Winston, 2016).

Summary of the example

Dataset: the food waste quantity in city A is inspected. In the conducted survey, a total of 200 respondents participated. However, three questionnaires have been declared invalid. Consequently, in the analysis, 197 data units about monthly food waste quantity of the respondents are used. In order to get better insight into monthly food waste quantity, the respondents have been divided into four categories according to age.

Data info:

- variable 1: groups—nominal (1—age 18–25, 2—age 26–40, 3—age 41–60, 4—above the age of 60);
- variable 2: food waste quantity—numeric (wasted food in grams per person, per month).

The one-way ANOVA approach was used to inspect whether the average monthly food waste quantity can be considered the same across all four age groups. However, the results of one-way ANOVA have shown that there was a statistically significant difference between age groups ($F(3,193) = 126.044, p < .001$). Tukey's post hoc test revealed that the average monthly food waste quantity for people aged 18–25 was

statistically significantly lower than the average monthly food waste quantity for those aged 26–40 ($p < .001$), while the average monthly food waste quantity for individuals aged 41–60 ($p < .001$). However, there was no statistically significant difference in the average monthly food waste quantity for people aged 18–25 or for those above the age of 60 ($p = .469$).

More info about one-way ANOVA

One-way ANOVA is used to inspect whether there are any statistically significant differences between the means of two or more independent groups. Despite the fact that one-way ANOVA can be used for comparing means between two independent groups, it is more often applied in cases where there are three or more independent groups, whereas in the cases of two independent groups, the t -test for independent samples is applied.

In order for one-way ANOVA to be used, six assumptions have to be fulfilled. Three of them can be checked without any computer software use: independent variable should consist of two or more categorical independent groups; independence of observations; dependent variable should be measured at the interval or ratio level. Those assumptions are straightforward and they can be verified very quickly. The other three assumptions should be checked using a computer program.

The fourth assumption is that dependent variable should be approximately normally distributed. The normality of data can be tested, for example, by use of the Shapiro-Wilk test for normality of distribution, and Kolmogorov-Smirnov test. The normality of data can be inspected graphically as well by using, for example, the normal Q-Q plot. In case of not normal distribution, the data should be converted into that normal by applying certain techniques. Technically spoken, there should be at least two units in each group to apply one-way ANOVA. However, the more units there are in each group, the larger the sample size. Consequently, it is more likely that the normality assumption will be fulfilled.

Because outliers have huge impact on the mean values, their presence has certain influence on the results of one-way ANOVA. Therefore, outlier analysis should be performed before conducting one-way ANOVA. The most straightforward approach to detect outliers is to standardise all values and then to check whether any of them deviate from the mean value more than three standard deviations. The outliers can be detected by using different graphical approaches as well. It has to be emphasized that outliers may have different sources. They can appear due to certain characteristics of the observed unit, but can also be the product of technical error (for example, data is mistyped).

The final assumption of one-way ANOVA application is homogeneity of variance between the groups. This assumption can be checked by Levene's test for homogeneity of variance. The null hypothesis of the test contains the assumption

that the observed groups all have equal population variances. In the given case this assumption is not met, thus, Welch's ANOVA should be used instead of this classic one-way ANOVA approach.

References

- Balakirshnan, N., Render, B., & Stair, R. M. (2007). *Managerial decision modeling with spreadsheets*. Pearson Prentice Hall.
- Barrow, M. (2017). *Statistics for economics, accounting and business studies*. Pearson.
- Dean, S., & Illowsky, B. (2013). *Introductory statistics*. OpenStax College.
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). Sage edge.
- Fraser, C. (2016). *Business statistics for competitive advantage with Excel 2016*. Springer.
- Randolph, K. A., & Myers, L. L. (2013). *Basic statistics in multivariate analysis*. Oxford: University Press.
- Winston, W. L. (2016). *Microsoft Excel 2016: Data analysis and business modelling*. Microsoft Press.

1.4. Kruskal-Wallis H test

General information

The Kruskal-Wallis H test is a commonly used nonparametric alternative to one-way ANOVA. It can be used when one-way ANOVA assumptions are violated—for example, when the dependent variable is measured on an ordinal scale. The test is similar to the Mann-Whitney U test, but it is used to compare scores in three or more groups. Since the Kruskal-Wallis H test does not require normality of data distribution, it does not allow comparison of means but ranks. The procedure includes ordering the observations from lowest to highest, and giving them ranks (Pallant, 2011; Verma & Abdel-Salam, 2019).

Hypotheses:

H0: There is no difference between the scores.

H1: There is at least one difference between the scores.

Assumptions

The following assumptions are associated with the Kruskal-Wallis H test:

- the measurement level of the dependent variable should be at least ordinal;
- there should be one independent variable divided into three or more groups;
- groups do not have common elements.

Example

Dataset: The company managing sharing bicycles decided to check the impact of the station location on the use of the bicycles. Three comparable high-schools were cho-

sen and for each of them, a different proximity of the station was set. The first school had a distant location, 200 m from the entrance; the second school had a middle location (100 m); while the third had the station set exactly in front of the entrance.

After two months of experiment, three random samples of students from each school have been selected (39, 44 and 45 students). Respondents declared the frequency of using the shared bicycles.

Data info:

- variable 1: group—nominal (1—close location (N = 39), 2—middle location (N = 44), 3—distant location (N = 45));
- variable 2: freq.—ordinal (declared frequency of using the shared bicycles; 1—more than once a day; 2—every day; 3—2–4 times a week; 4—once a week; 5—once a month; 6—less than once a month; 7—never).

Hypotheses:

H0: There is no difference in the frequency of using shared bicycles between the groups.

H1: The frequency of using shared bicycles differs among the groups, at least one group is different from the other.

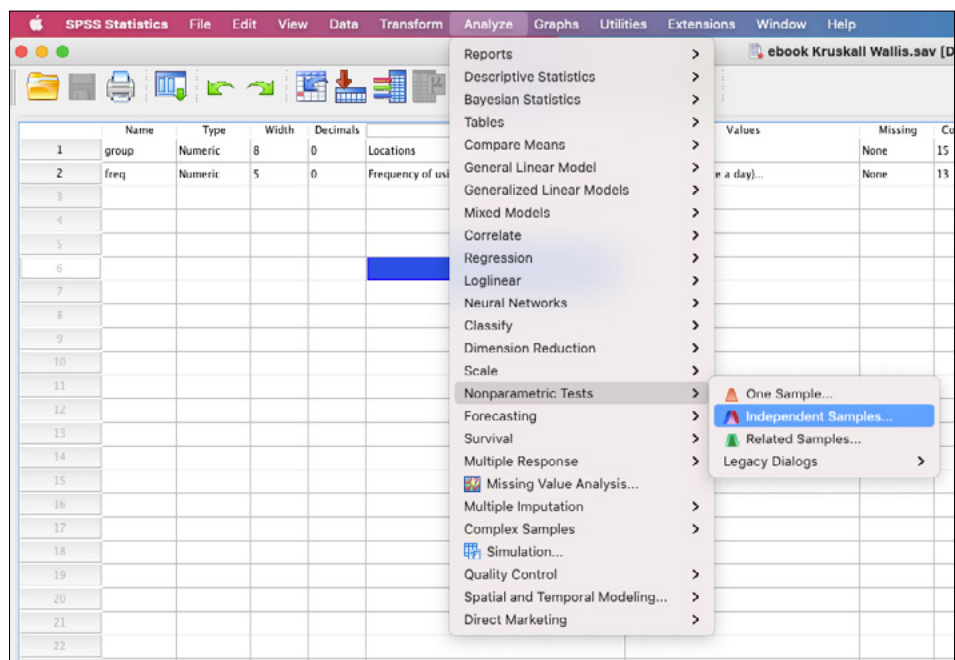


Figure 30. Kruskal-Wallis H test—path

Source: The authors' own elaboration, IBM SPSS screenshot.

Independent samples—single hypothesis testing

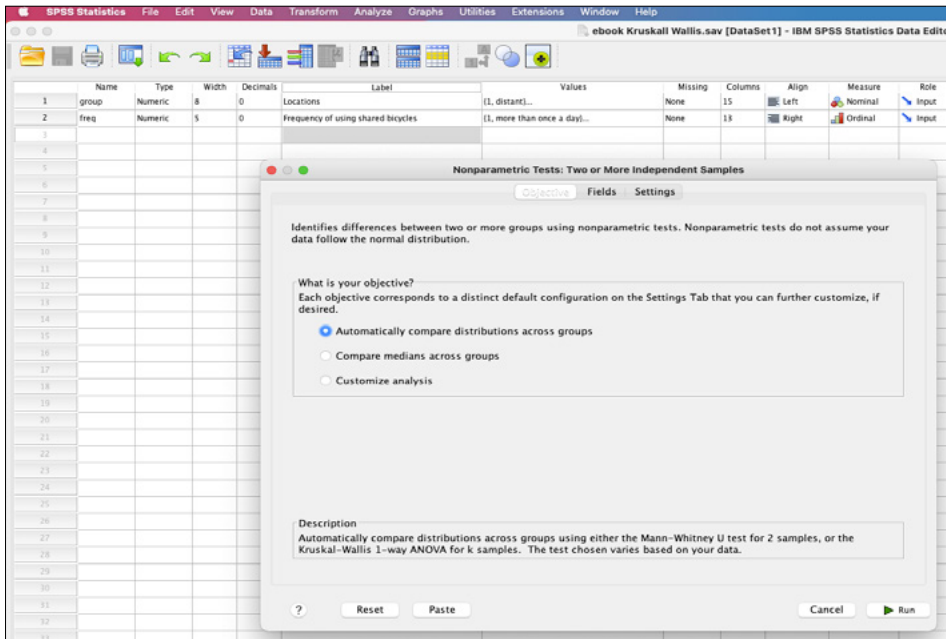


Figure 31. Kruskal-Wallis H test—dialogue box (1)

Source: The authors' own elaboration, IBM SPSS screenshot.

In the first dialogue box, three tabs can be seen—‘Objective’, ‘Fields’ and ‘Settings’. The objective of the analysis is defined by choosing the default option—‘Automatically compare distributions across groups.’

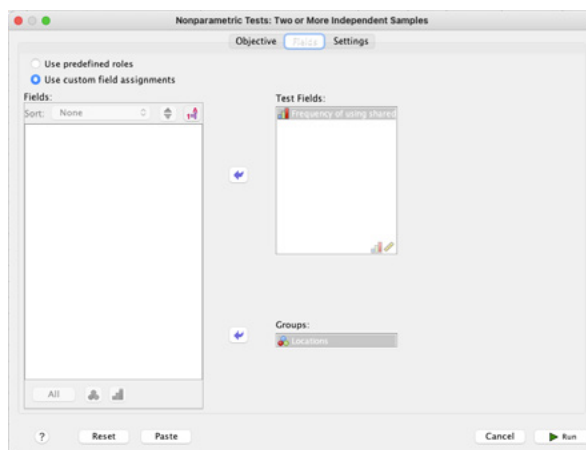


Figure 32. Kruskal-Wallis H test—dialogue box (2)

Source: The authors' own elaboration, IBM SPSS screenshot.

In the next step, we move to the tab ‘Fields’ where the analysed variable (‘Test fields’) and grouping variable (‘Groups’) are chosen.

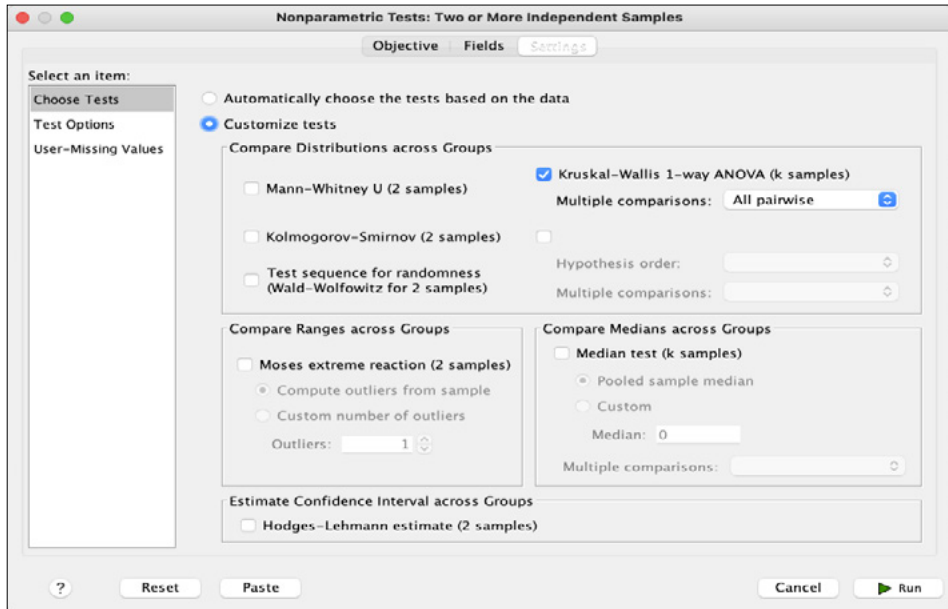


Figure 33. Kruskal-Wallis H test—dialogue box (3)

Source: The authors’ own elaboration, IBM SPSS screenshot.

In the last step, we choose ‘Customize tests’ and select ‘Kruskal-Wallis 1-way ANOVA’ (k samples) with multiple comparisons: ‘All pairwise’.

➔ **Nonparametric Tests**

Hypothesis Test Summary				
	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Frequency of using shared bicycles is the same across categories of Locations.	Independent-Samples Kruskal-Wallis Test	.038	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .050.

Independent-Samples Kruskal-Wallis Test

Frequency of using shared bicycles across Locations

Independent-Samples Kruskal-Wallis Test Summary

Total N	128
Test Statistic	6.539 ^a
Degree Of Freedom	2
Asymptotic Sig. (2-sided test)	.038

a. The test statistic is adjusted for ties.

Figure 34. Kruskal-Wallis H test—results

Source: The authors’ own elaboration, IBM SPSS screenshot.

Results

The hypothesis is decided upon by interpreting the ‘Asymptotic Sig. (2-sided test)’ value in the lower table. In this case, it equals $p = .038$. This value is lower than the critical value of $p = .05$, which indicates that there is at least one significant difference in scores across the various groups. The first dialogue box presents only a general result of the Kruskal-Wallis H test—which of the groups is significantly different from the other ones is still unknown.

Sample 1–Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig. ^a
1.00 close–2.00 middle	-13.939	8.037	-1.734	.083	.249
1.00 close–3.00 distant	-20.103	7.995	-2.514	.012	.036
2.00 middle–3.00 distant	-6.165	7.748	-.796	.426	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.
a. Significance values have been adjusted by the Bonferroni correction for multiple ...

Figure 35. Kruskal-Wallis H test—pairwise comparisons (1)

Source: The authors’ own elaboration, IBM SPSS screenshot.

In order to identify the differences between the groups, pairwise comparisons are examined. ‘Adj. Sig.’ value for the last column is interpreted. In the presented example, the p -value is lower than the critical value of $p = .05$ when comparing only the close and distant locations ($p = .036$). This means that there is a significant difference in the frequency of using shared bikes between these groups. The p -values for other comparisons: $p = .249$ and $p = 1.000$, mean that there is no significant difference in the frequency of using bikes.

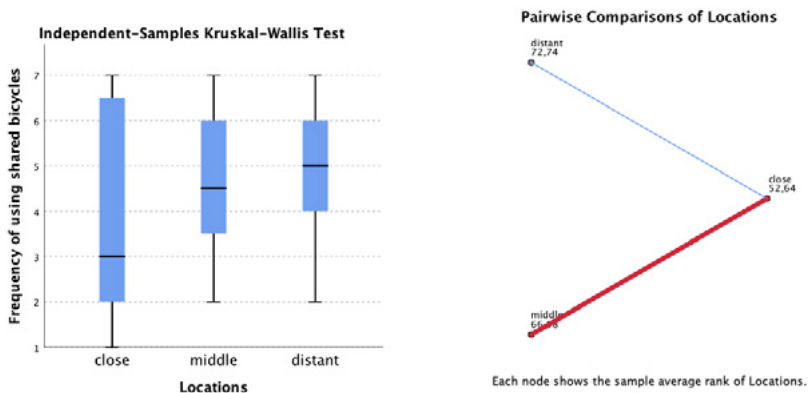


Figure 36. Kruskal-Wallis H test—pairwise comparisons (2)

Source: The authors’ own elaboration, IBM SPSS screenshot.

The results can be compared by looking at the box-and-whiskers graph and at the visual representation of pairwise comparisons. The blue line in the graph on the right indicates a significant difference in the frequency of using shared bikes between close and distant locations. The red line indicates an insignificant difference.

Results and post hoc tests

Kruskal-Wallis H test hypotheses resolution:

$p < .05$ —there is at least one significant difference in scores across different groups; reject H_0 ;

$p > .05$ —there is no significant difference in scores across different groups; do not reject H_0 .

Effect size

The effect size measure for Kruskal-Wallis H test is calculated following the procedure for the Mann-Whitney U test (Pallant, 2011).

The effect size measure (r) is based on the statistic Z and N values which is total number of observations in both groups:

$$r = \frac{|Z|}{\sqrt{N}}$$

The effect size can only be calculated for significant differences between groups. The Z value for each comparison is expressed as ‘Std. Test Statistic’ in the ‘Pairwise Comparisons of Locations’ table.

The r has the following interpretation:

Below .1—no effect;

< .1-.3)—small effect;

< .3-.5)—moderate effect;

.5 and more—large effect.

$$r = \frac{|-2.514|}{\sqrt{84}} = 0.27$$

In this case, a small effect size ($r = .27$) can be observed.

Summary

Dataset: The company managing sharing bicycles decided to check the impact of the station location on the use of the bicycles. Three comparable high-schools were chosen, and for each of them a different proximity of the station was set. The first school had a distant location, 200 m from the entrance, the second one had moderate location (100 m), while the third school was set exactly in front of the entrance.

After two months of the experiment, three random samples of students from each school were selected (39, 44 and 45 students). Respondents declared the frequency of using the shared bicycles.

Data info:

- variable 1: group—nominal (1—close location (N = 39), 2—middle location (N = 44), 3—distant location (N = 45));
- variable 2: freq.—ordinal (declared frequency of using the shared bicycles; 1—more than once a day; 2—every day; 3—2–4 times a week; 4—once a week; 5—once a month; 6—less than once a month; 7—never)

The Kruskal-Wallis H test allowed to reveal that the frequency of using shared bikes differed statistically significantly across different locations. Pairwise comparisons indicated that there is a difference in the frequency of using shared bikes between the students from school with close location of the station and with distant location of the station ($G_c, n = 39, G_d, n = 45, Z = -2.514; p = .036$). Students from schools close to the station used bikes more often ($Mdn = 3$) than students from those with distant locations ($Mdn = 5$). However, this effect was rather small ($r = .027$). The analysis did not show any significant differences between other groups.

More information

The result of Kruskal-Wallis H test does not inform us about the between-group comparisons. In order to compare separate groups pairwise, Bonferroni adjustment needs to be applied. This involves multiplying the significance by the number of tests (significance level equal to $p = .012$ after multiplication is shown as adjusted significance (('Adj. Sig.') $p = .036$). The same result may be obtained by dividing the alpha level of .05 by the number of tests that are intended to be used, and by implementing the initial significance level ('Sig.'). While interpreting the group comparisons the revised alpha level should be used as the criteria for determining significance. However, the described procedure shows the results applying Bonferroni adjustment.

References

- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). Sage edge.
- Pallant, J. (2011). *SPSS Survival Manual: a step by step guide to data analysis using SPSS* (4th ed.). Allen & Unwin.
- Verma, J. P., & Abdel-Salam, G. A.-S. (2019). *Testing statistical assumptions in research*. John Wiley & Sons, Inc.