

## BASIC GUIDELINE TO CONDUCT ONE-WAY ANOVA USING XLSTAT IN EXCEL: APPROACH TO CARIES DATA

*Mohamad Arif Awang Nawi*

School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM),  
16150 Kubang Kerian, Kelantan, Malaysia

*Corresponding author: Mohamad Arif Awang Nawi, School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Kelantan, Malaysia, Email: [mohamadarif@usm.my](mailto:mohamadarif@usm.my)*

<https://doi.org/10.32827/ijphcs.6.6.202>

### ABSTRACT

**Background:** The One-Way ANOVA is commonly used to test statistical differences among the means of two or more groups or intervention or change score. There are two assumption are needed for this analysis such as normality distribution and randomness & independence. In ANOVA, multiple comparison methods are designed to investigate differences between specific pairs of means. This provides the information that is of most use to the researcher. The objective in this study is to provide basic guideline how to conduct one-way ANOVA analysis using XLSTAT Excel.

**Materials and Methods:** In this paper will explain how to conduct one-way ANOVA using XLSTAT software, an add on of Microsoft Office Excel. Firstly, checking normality and secondly ANOVA analysis. Secondly, set up and interpret a one-way Analysis of Variance (ANOVA) followed by Bonferroni multiple comparisons in Excel using the XLSTAT software.

**Result:** We want to compare mean difference between smoker status relate with number of caries. In order to make a confident and reliable decision, we will need evidence to support our approach. This is where the concept of ANOVA comes into play. From Shapiro-Wilk test, as the computed p-value is greater than the significance level  $\alpha=0.05$ , the caries data was extracted follows a Normal distribution. The Bonferroni test is applied to all pairwise differences between means. Based on the p-values ( $P > Diff$ ), all pairs of smoking status appear to be significantly different effect on number of caries (heavy smoker vs non-smoker), (Heavy smoker vs ex- smoker) and (Ex-smoker vs non-smoker).

**Conclusion:** Heavy smoker vs non-smoker gives the higher mean difference = 9.8986 which can conclude that heavy smoker will increase the number of caries among patients compare to non-smoker patients. Besides that, heavy smoker vs ex-smoker is statistically significant difference effect on number of caries where heavy smoker is effect on number of caries compare with ex-smoker = 6.8986. XLSTAT is a powerful yet flexible Excel data analysis add-on that allows users to analyze, customize and share results within Microsoft Excel.

**Keywords:** One-Way ANOVA, XLSTAT software, Compare Means, Caries Data

## 1.0 Introduction

A one-way ANOVA is important statistical method to compare two means from two independent (unrelated) groups using the F-distribution. The null hypothesis for the test is that the two means are equal. Therefore, a significant result means that the two or more means are unequal. This method can be used only for numerical data. The ANOVA, developed by Ronald Fisher in 1918, extends the t and the z test which have the problem of only allowing the nominal level variable to have two categories. This test is also called the Fisher analysis of variance (Aczel, 1989). There are two assumption are needed for this analysis. First assumption is randomness and independence. Both conditions must be met because the validity of any experiment depends on random sampling and/or the randomization process. The second assumption is normality distribution. Distribution of each group should be normal. The Kolmogorov-Smirnov or the Shapiro-Wilk test may be used to confirm normality of the group (Wan Muhamad Amir W Ahmad et al. 2019).

When to use a one-way ANOVA? The One-Way ANOVA is commonly used to test statistical differences among the means of two or more groups or intervention or change score. Based on figure 1 below, researcher want to study the effect of smoking status on caries patients and form three groups: Non-smoker, Ex-smoker and heavy smoker.

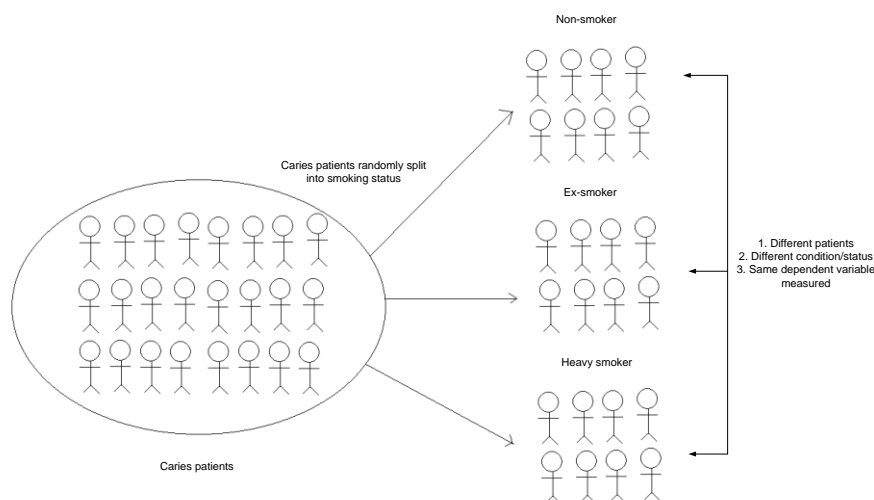


Figure 1: Example of ANOVA

A one-way ANOVA will tell you that at least two groups were different from each other. But it won't tell researcher what groups were different. If F-statistic test is significant, researcher may need to run post hoc test to tell know exactly which groups had a difference in means. Post hoc tests are tests of the statistical significance of differences between group means calculated after having done ANOVA that shows an overall difference. Multiple comparison methods are designed to investigate differences between specific pairs of means. This provides the information that is of most use to the researcher. One possible approach to the multiple comparison problem is to make each comparison independently using a suitable statistical procedure (Rafter et al. 2002). The objective in this study is to provide basic guideline how to conduct one-way ANOVA analysis using XLSTAT Excel.

## 2.0 Materials and Methods

There are many software's that can implement ANOVA of different kinds. Some of them are SPSS, MINTAB, MATLAB, STATISTICA, E-View, SAS, R - software and XLSTAT in EXCEL. In this paper will explain how to conduct one-way ANOVA using XLSTAT (Addinsoft, 2019) software, an add on of Microsoft Office Excel. Firstly, checking normality and secondly ANOVA analysis.

### 2.1 Checking normality of One-way ANOVA in XLSTAT Excel

Normality tests involve the null hypothesis that the variable from which the sample is drawn follows a normal distribution. Thus, a low p-value indicates a low risk of being wrong when stating that the data are not normal.

The data represent number of caries among 101 patients.

1. Select the XLSTAT → Describing data → click at Normality tests.
2. The dialog box of Normality tests appears. Select the samples in the Data field.
3. Click at Shapiro-Wilk test.
4. Go to Charts: Click the Q-Q plot option to allow us to visually check the normality of the samples.
5. Click at OK button and the results are displayed on a new sheet.

### 2.2 One-way ANOVA

Analysis of variance (ANOVA) is a tool used to partition the observed variance in a variable into components attributable to different sources of variation. Analysis of variance (ANOVA) uses the same conceptual framework as linear regression. The main difference comes from the nature of the explanatory variables: instead of quantitative, here they are qualitative. In ANOVA, explanatory variables are often called factors.

If  $p$  is the number of factors, the ANOVA model is written as follows:

$$y_i = \beta_0 + \sum_{j=1}^q \beta_{k(i,j)} + \epsilon_i$$

where  $y_i$  is the value observed for the dependent variable for observation  $i$ ,  $k(i,j)$  is the index of the category (or level) of factor  $j$  for observation  $i$  and  $\epsilon_i$  is the error of the model.

In this paper shows how to set up and interpret a one-way Analysis of Variance (ANOVA) followed by Bonferroni multiple comparisons in Excel using the XLSTAT software. Step by step to produced One-way ANOVA in XLSTAT Excel as shown below:

1. Open XLSTAT and select the XLSTAT → Modeling data → click at ANOVA command.
2. ANOVA dialog box appears. Data format: Click at Colum.
3. Y/Dependent variables: Quantitative: Select the data on the Excel sheet "Caries".
4. X/Explanatory variables: Click at Qualitative: Select the data on the Excel sheet "Smoking status".
5. Click at Outputs tab, → Click at Pariwise comparisons option → Click at Bonferroni test.
6. Click at OK button.

### 3.0 Result

Consider a scenario where we have three smoker status to apply on caries patients. Once we have the test results, one approach is to assume that the smoker status which took the higher number of caries among patients. We want to compare mean difference between smoker status relate with number of caries. In order to make a confident and reliable decision, we will need evidence to support our approach. This is where the concept of ANOVA comes into play.

**Table 1:** Description Data

Dependent variable	Independent variable
Number of Caries	Smoking status 1 = Non-Smoker 2 = Ex-smoker 3 = Heavy smoker

#### 3.1 Normality Distribution in ANOVA

Test interpretation:

H<sub>0</sub>: The variable from which the sample was extracted follows a Normal distribution.

H<sub>a</sub>: The variable from which the sample was extracted does not follow a Normal distribution.

As the computed p-value is greater than the significance level  $\alpha=0.05$ , one cannot reject the null hypothesis H<sub>0</sub>. The risk to reject the null hypothesis H<sub>0</sub> while it is true is 7.41%.

The first result displayed is the Q-Q plot for the sample (Caries). The Q-Q plot allows to compare the cumulative distribution function of the sample to the cumulative distribution function of normal distribution with the same mean and standard deviation. In the case of sample following a normal distribution, we should observe an alignment with the first bisecting line. In the other cases some deviations from the bisecting line should be observed.

Table 2: Shapiro-Wilk test (Caries)

W	0.9770
p-value (Two-tailed)	0.0741
alpha	0.05

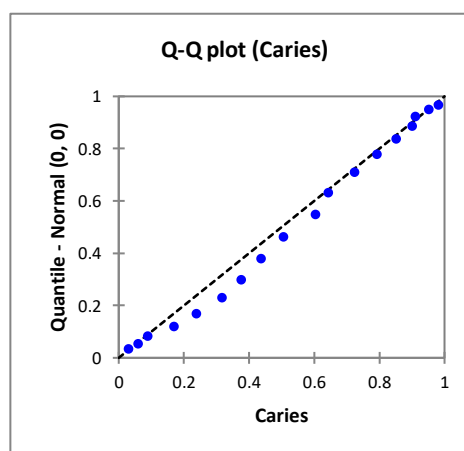


Figure 2: Q-Q Plot of Caries

### 3.2 One-Way ANOVA

Table 3: Goodness of fit statistics:

Observations	101.0000
Sum of weights	101.0000
DF	98.0000
R <sup>2</sup>	0.6304
Adjusted R <sup>2</sup>	0.6229
MSE	8.1866
RMSE	2.8612
DW	1.7677

The first results displayed by XLSTAT are the goodness of fit coefficients according Table 3 above, including the R<sup>2</sup> (coefficient of determination), the adjusted R<sup>2</sup> and several other statistics. The coefficient of determination (here 0.63) gives a fair idea of how much of the variability of the modelled variable (here the number of caries) is being explained by the explanatory variables (here the smoking status); in our case, we have 63% of the variability explained. The other 37% are hidden in other variables which are not available, and which the model hides in "random errors".

Table 4: Analysis of variance

Source	DF	Sum of squares	Mean squares	F	Pr > F
Model	2	1368.6606	684.3303	83.5912	< 0.0001
Error	98	802.2899	8.1866		
Corrected Total	100	2170.9505			

*Computed against model  $Y=Mean(Y)$*

The test used here is the Fisher's F test. Given that the probability corresponding to the F value, in this case, is 0.001 (Table 4), it means that we would take a 0.1% risk to conclude that the null hypothesis (no effect of the smoking status) is wrong. So, we can conclude with confidence that there is an effect of the smoking status on the number of caries among patients (based on Figure 3).

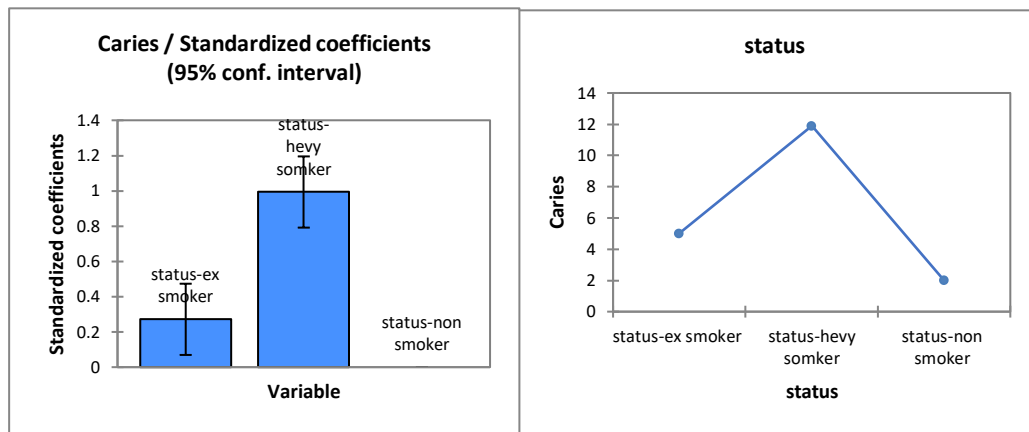


Figure 3: Standardized Coefficients of Caries

Table 5: Analysis of the differences between the categories with a confidence interval of 95%

Contrast	Difference	Standardize d difference	Critical value	Pr > Diff	Significant
Heavy smoker vs non-smoker	9.8986	9.7615	2.4358	< 0.0001	Yes
Heavy smoker vs ex-smoker	6.8986	10.0138	2.4358	< 0.0001	Yes
Ex-smoker vs non-smoker	3.0000	2.6667	2.4358	0.0090	Yes

*Bonferroni Test**Modified significance level: 0.05*

Now we obtain the answer to our initial question: is there a significant difference between the smoking status, and how should this difference be classified? As shown in the next table, the Bonferroni test is applied to all pairwise differences between means. The risk of 5% we have chosen is used to determine the critical value  $q$ , which is compared to the standardized difference between the means. Based on the  $p$ -values below ( $Pr > Diff$ ), all pairs of smoking status appear to be significantly different effect on number of caries (heavy smoker vs non-smoker), (Heavy smoker vs ex-smoker) and (Ex-smoker vs non-smoker).

#### 4.0 Conclusion

The conclusion is that the smoking status show significantly different effects on number of caries. Based on analysis of the differences using Bonferroni test, heavy smoker vs non-smoker gives the higher mean difference = 9.8986 which can conclude that heavy smoker will increase the number of caries among patients compare to non-smoker patients. Beside that, heavy smoker vs ex-smoker is statistically significant difference effect on number of caries where heavy smoker is effect on number of caries compare with ex-smoker = 6.8986.

Ashkanane et al. 2017 and Mohammed et al. 2018 have examined the effect of nicotine and cigarette extracts on oral bacteria. In a recent review<sup>28</sup> on the effects of nicotine on oral microorganisms and human tissues there is indirect evidence of a link between smoking and caries. A few researchers have reported a higher caries experience happening (DMFT) and is suspected to be linked with cigarette smoking (Aguilar et al. 2008). Cigarette smoking is known for its negative effects towards oral health. Smoking is a precedent to gum diseases which are clinically presented as gingival swelling and inflammation, loss of gingival attachment, gingival recession, and deeper periodontal pockets (Benedetti et al. 2013). This showed that smokers have poorer oral health compared to non-smokers in terms of caries experience.

XLSTAT is a powerful yet flexible Excel data analysis add-on that allows users to analyze, customize and share results within Microsoft Excel. With over 220 standards to advanced statistical features available, XLSTAT is the preferred tool for statistical analysis in businesses and universities, large and small, and for 100,000+ users in over 120 countries across the world.

## Declaration

Author(s) declare that there is no conflict of interest with the publication of this article.

## Authors' contribution

Author: Initiation of idea, literature searching, drafting the manuscript, final manuscript review and editing

## References

Aczel, A.D. 1989. Complete Business Statistics, Irwin.

Aguilar-Zinser, V., Irigoyen, M. E., Rivera, G., Maupomé, G., Sánchez-Pérez, L., Velázquez, C. 2008. Cigarette smoking and dental caries among professional truck drivers in Mexico. *Caries Res.* 42(4), 255–262. doi: 10.1159/000135670 21.

Ashkanane A, Gomez GF, Levon J, Windsor LJ, Eckert GJ, Gregory RL. 2017. Nicotine Upregulates Coaggregation of *Candida albicans* and *Streptococcus mutans*. *Journal of Prosthodontics Official Journal of the American College of Prosthodontists*.

Benedetti, G., Campus, G., Strohmenger, L., Lingström, P., Tobacco and dental caries: A systematic review. 2013. *Acta Odontol. Scand.* 363–371.



Mohammed AY, Gomez GF, Eckert GJ , et al. 2018. The Impact of Nicotine and Cigarette Smoke Condensate on Metabolic Activity and Biofilm Formation of *Candida albicans* on Acrylic Denture Material. *Journal of Prosthodontics*. doi:10.1111/jopr.12945.

Rafter, J.A., Abell, M.L., Braselton, J.P. 2002. "Multiple Comparison Methods for Means," *SIAM Review*, 44 (2). 259-278. 2002.

Wan Muhamad Amir W Ahmad, Bassaruddin Ahmad, Sarimah Abdullah & Nor Azlida Aleng@Mohamad. 2017. *Statistical Analysis SPSS for Health Sciences*. Pulau Pinang: Penerbit USM.

XLSTAT (One-Way ANOVA & Multiple Comparisons In Excel Tutorial)-Addinsoft. 2019. Available from: [https://help.xlstat.com/s/article/one-way-anova--multiple-comparisons-in-excel-tutorial?language=en\\_US](https://help.xlstat.com/s/article/one-way-anova--multiple-comparisons-in-excel-tutorial?language=en_US). Accessed 2019 November 15.