Running a one-way ANOVA followed by multiple comparisons tests with XLSTAT

demoANO.xls

Dataset for running a one-way ANOVA

An Excel sheet with both the data and the results can be downloaded by clicking here.

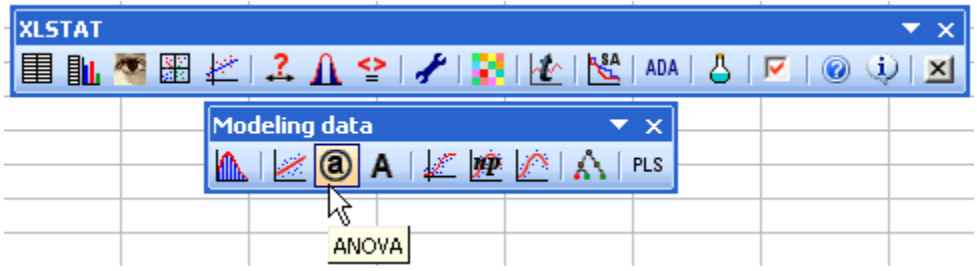
The data correspond to an experiment where 4 new toothpaste formulas were each tested on 6 different patients in order to measure their effect on the whiteness of teeth. All patients had previously used the same toothpaste.

Goal of this tutorial

Using the ANOVA function of XLSTAT we want to find out if the results differ according to the formula used and, if so, which formula is the most effective. The case is a one-way balanced ANOVA because there is only one factor - the formula - and the number of repetitions is the same for each formula.

Setting up the one-way ANOVA

Once XLSTAT is open, select the **XLSTAT / Modeling data / ANOVA command**, or click on the corresponding button of the **Modeling Data** toolbar (see below).

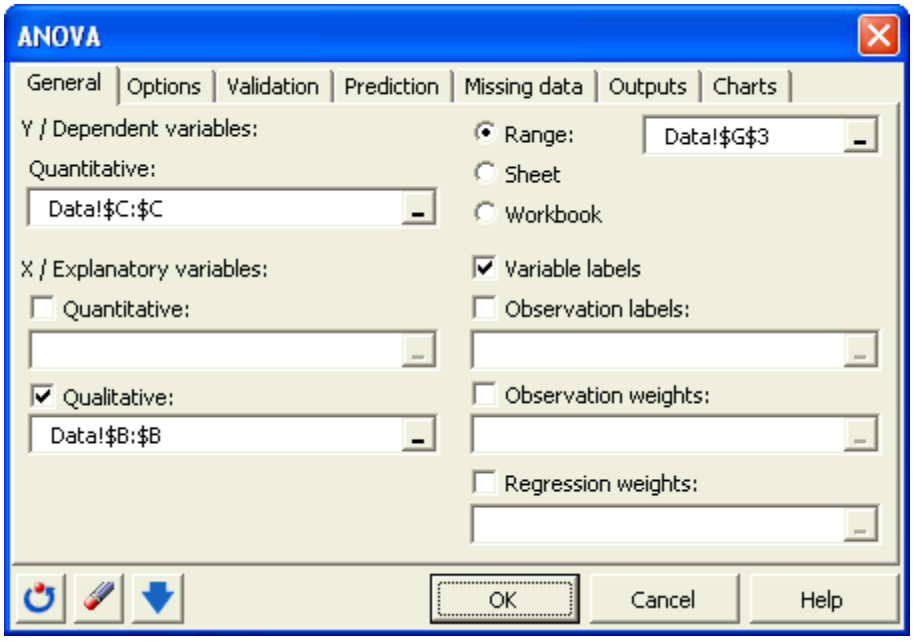


Once you have clicked on the button, the ANOVA dialog box appears.

Select the data on the Excel sheet. To the **Dependent variable** corresponds here "Whiteness" which variability we want to explain by the effect of the "Toothpaste" formula, the latter being the **Qualitative explanatory variable**.

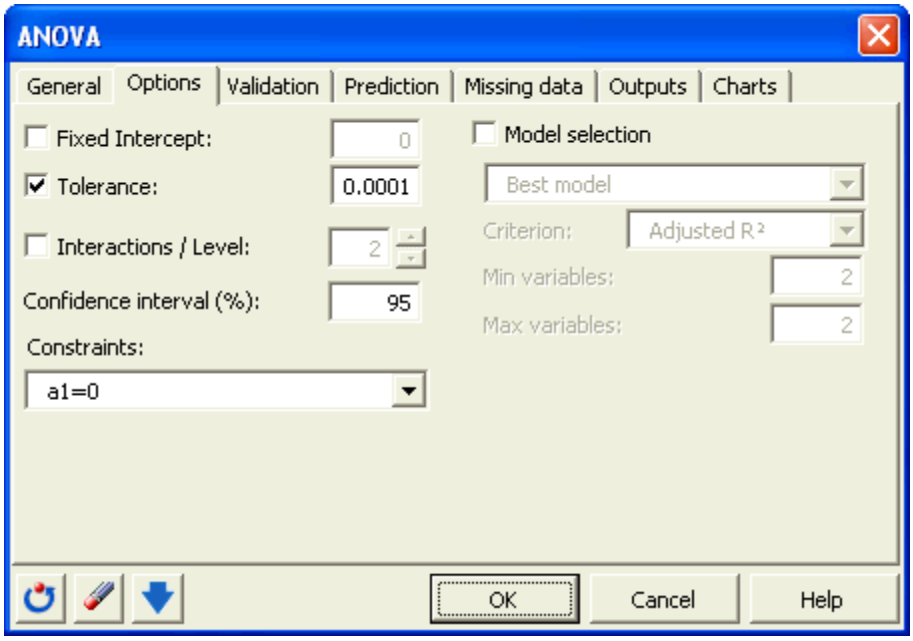
As we selected the column title for both variables, we also clicked on the option **Variable labels**.

In this example we want to display the results on the same sheet where the data are stored, so we chose the **Range** option and selected the cell that corresponds to the top left corner of the results report to be displayed.



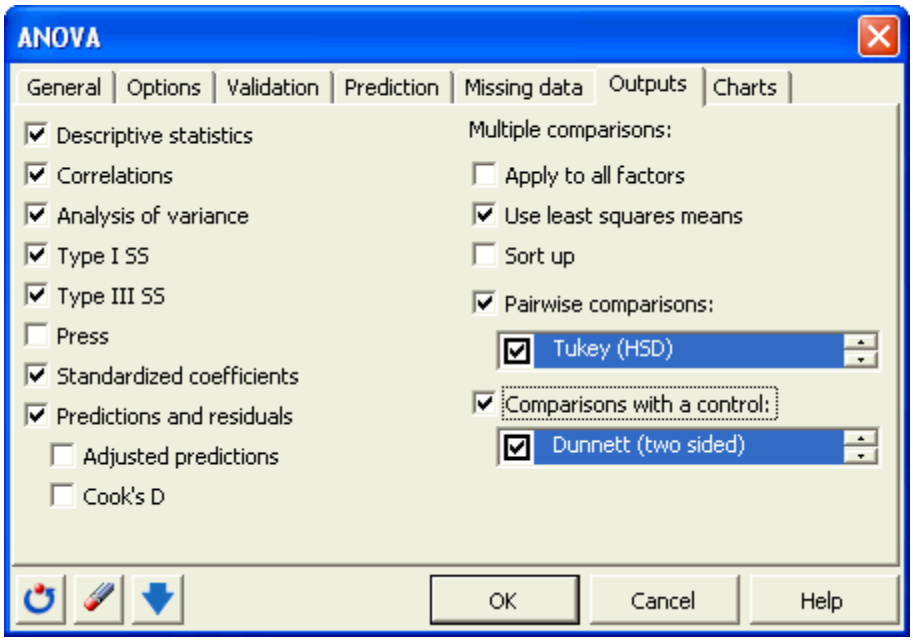
In the **Options** tab, we left the constraint option at **a1=0**, meaning that we want the model to be built using the assumption that the T1 toothpaste has the basic effect on whiteness: we know the average for T1 is the lowest and this guarantees that the other effects will be positive.

Applying a constraint to the ANOVA model is necessary for theoretical reasons, but it has no effect on the results (goodness of fit, predictions). The only difference it makes is in the way you write the model.

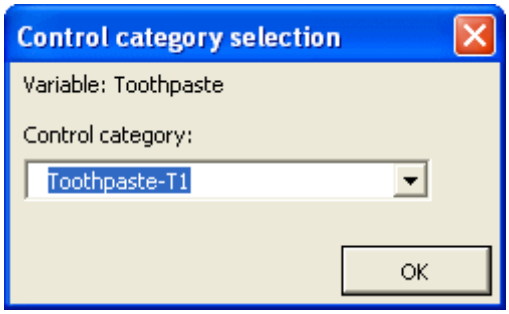


In the **Outputs** tab we activated the **Pariwise comparisons** option to be able to run a **Tukey's test** and a **REGWQ test**, we activated the **Comparisons with a control** option to run two sided **Dunnett's test**.

To understand more about the relatively complex subject of multiple comparison tests, see the work by Jason C. Hsu.



The computations begin once you have clicked on the **OK** button, then stop to ask you which group is the control group for the Dunnett test. We selected T1 as the control group.



Once the user has clicked on the **OK** button, the computations resume and the results are displayed.

Interpreting the one-way ANOVA results

The first results displayed by XLSTAT are the goodness of fit coefficients, including the R'² (coefficient of determination), the adjusted R'² and several other statistics.

Goodness of fit statistics:	
Observatio	24,000
Sum of we	24,000
DF	20,000
R²	0,556
Adjusted R	0,489
MSE	12,083
RMSE	3,476
MAPE	10,478
DW	1,750
Cp	4,000
AIC	63,428
SBC	68,140
PC	0,622

The coefficient of determination (here 0.56) gives a fair idea of how much of the variability of the modeled variable (here the whiteness) is being explained by the explanatory variables (here the type of toothpaste); in our case we have 56% of the variability explained. The other 44% are hidden in other variables which are not available, and which the model hides in "random errors".

The analysis of variance table is a very important result to look at (see below). This is where we determine whether the explanatory variable (the toothpaste formula) brings significant information (null hypothesis H0) to the model or not. In other words, it is way of asking yourself whether it is valid to take the mean to describe the whole population, or if the information provided by the categories (here the toothpaste type) is of value or not.

Analysis of variance:

Source	DF	Sum of squares	Mean square	F	Pr > F
Model	3	302,167	100,722	8,336	0,001
Error	20	241,667	12,083		
Corrected Total	23	543,833			

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, it means that we would take a 0.1% risk to conclude that the null hypothesis (no effect of the toothpaste formulas) is wrong.

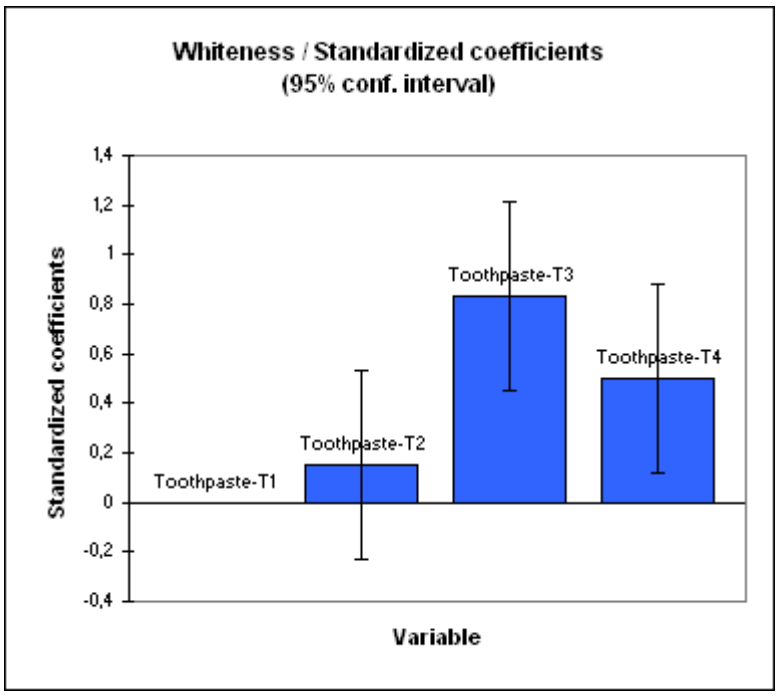
So we can conclude with confidence that there is an effect of the toothpaste formulas on the whiteness of the patients' teeth. Note that the R'^2 is not very good (0.56), meaning that some of the information offering a complementary explanation of the variations of the whiteness is missing, which is no real surprise.

The following table gives details on the model. This table is helpful when predictions are needed. In this particular case it is not very useful. We can already notice that the toothpaste T2 has an effect which 95% confidence range includes 0, indicating that there is no evidence that T2 is very different from T1.

Model parameters:

Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)
Intercept	19,000	1,419	13,389	< 0,0001	16,040	21,960
Toothpaste	0,000	0,000				
Toothpaste	1,667	2,007	0,830	0,416	-2,520	5,853
Toothpaste	9,167	2,007	4,568	0,000	4,980	13,353
Toothpaste	5,500	2,007	2,741	0,013	1,314	9,686

The bar chart of the standardized coefficients allow to visually compare the relative impact of the categories, and to see if the confidence intervals include 0 or not.



The next table shows the residuals. We can look at the reduced residuals (standardized residuals) more specifically, residuals which, given the assumptions of the ANOVA model, should be normally distributed. This means, among other things, that 95% of the residuals should be in the interval [-1.96, 1.96].

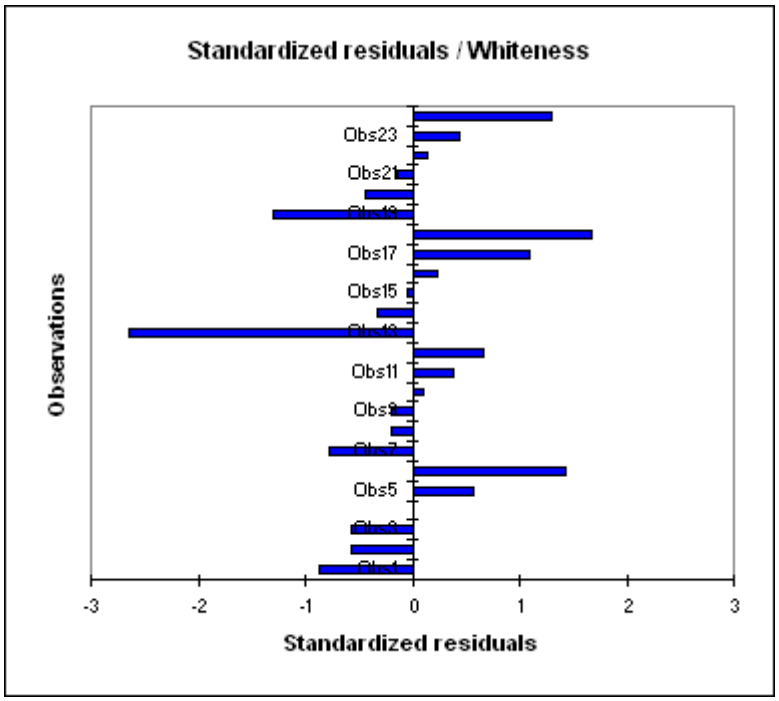
All values outside this interval are potential outliers, or might suggest that the normality assumption is wrong. It seems here that there is one strong outlier (13th observation) with a residual equal to -2.8279.

To explain the difference, one should first verify that the right toothpaste was given to the 13th patient, and secondly, one should try to understand why the response to the formula wasn't the same as for the other patients.

The histogram of the residuals allows you to quickly visualize the residuals that are out of the expected range.

Predictions and residuals:

Observation	Weight	Whiteness	Fitted(Whitene	Residual	Std. residua
Obs1	1	16,000	19,000	-3,000	-0,863
Obs2	1	17,000	19,000	-2,000	-0,575
Obs3	1	17,000	19,000	-2,000	-0,575
Obs4	1	19,000	19,000	0,000	0,000
Obs5	1	21,000	19,000	2,000	0,575
Obs6	1	24,000	19,000	5,000	1,438
Obs7	1	18,000	20,667	-2,667	-0,767
Obs8	1	20,000	20,667	-0,667	-0,192
Obs9	1	20,000	20,667	-0,667	-0,192
Obs10	1	21,000	20,667	0,333	0,096
Obs11	1	22,000	20,667	1,333	0,384
Obs12	1	23,000	20,667	2,333	0,671
Obs13	1	19,000	28,167	-9,167	-2,637
Obs14	1	27,000	28,167	-1,167	-0,336
Obs15	1	28,000	28,167	-0,167	-0,048
Obs16	1	29,000	28,167	0,833	0,240



Now we obtain the answer to our initial question: is there a significant difference between the treatments, and how should this difference be classified?

As shown on the next table, the Tukey's HSD (Honestly Significantly Different) 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.

Only two pairs appear to be significantly different (T1, T3) and (T2,T3). The means and the categories are then classified based on this analysis. We can see here that there is no transitivity (means not significantly different, and <> means significantly different):
T4 T3 T4 T2 but T2 <> T3

Toothpaste / Tukey (HSD) / Analysis of the differences between the categories with a confidence interval of 95%:

Contrast	Difference	ardized diff	ritical valu	Pr > Diff	Significant
T3 vs T1	9,167	4,568	2,799	0,001	Yes
T3 vs T2	7,500	3,737	2,799	0,007	Yes
T3 vs T4	3,667	1,827	2,799	0,290	No
T4 vs T1	5,500	2,741	2,799	0,056	No
T4 vs T2	3,833	1,910	2,799	0,256	No
T2 vs T1	1,667	0,830	2,799	0,839	No
Tukey's d critical value:			3,958		

Category	LS means	Groups		
T3	28,167	A		
T4	24,500	A	B	
T2	20,667		B	
T1	19,000		B	

The REQWQ procedure gives different results (see below), which shows that one needs to be very cautious when using comparison methods.

Three pair of categories are different in this case (T1 and T4 appear to be significantly different with this method). The groupings give now three superimposed groups of categories.

Toothpaste / REGWQ / Analysis of the differences between the categories with a confidence interval of 95%:

Contrast	Difference	ardized diff	ritical valu	Pr > Diff	ha (Modifie	Significant
T3 vs T1	9,167	4,568	2,799	0,001	0,050	Yes
T3 vs T2	7,500	3,737	2,530	0,004	0,050	Yes
T3 vs T4	3,667	1,827	2,530	0,083	0,050	No
T4 vs T1	5,500	2,741	2,530	0,032	0,050	Yes
T4 vs T2	3,833	1,910	2,417	0,071	0,025	No
T2 vs T1	1,667	0,830	2,417	0,416	0,025	No

Category	LS means	Groups		
T3	28,167	A		
T4	24,500	A	B	
T2	20,667		B	C
T1	19,000			C

Next, we performed a Dunnett's test to compare each category with the control category T1. The Dunnett's test agrees with the REQWQ procedure that the T1 and T4 categories are significantly different.

Toothpaste / Dunnett (two sided) / Analysis of the differences between categories and the control category Toothpaste-T1

Category	Difference	ardized diff	ritical valu	ical differer	Significant
T3 vs T1	-9,167	-4,568	2,540	5,098	Yes
T4 vs T1	-5,500	-2,741	2,540	5,098	Yes
T2 vs T1	-1,667	-0,830	2,540	5,098	No

Conclusion for this one-way ANOVA

The conclusion is that the 4 toothpaste formulas show significantly different effects on whiteness. As the T1 toothpaste is already on the market, it is the toothpastes T3 or T4, which show a significant increase in whiteness, which should be selected as newcomers to the market.