

To: Professor Yuce
Applied Mathematics
New York City College of Technology

From: Brian Holliday
Applied Mathematics
New York City College of Technology

Subject: Lab 21: One Way ANOVA

Figure:

Figure 1: GLM Procedure Smokers Data

Figure 2: Distribution of FEF

Figure 3: t-test LSD for FEF

Figure 4: GLE Procedure: Contrast and Parameter.

Date: 4/23/20

Summary:

The question that ANOVA is concerned with answering is whether there is a difference in the means of two populations. To do this ANOVA uses the F-statistic where:

$$F\text{-statistics} = \text{Variation among sample means} / \text{Variation within groups}$$

If the variation among the sample means is higher than the variation within groups, then the value of the F-statistic will be low. If the variation within groups is high, then the F-statistic will be low. If the F-statistic is high then we have strong evidence for the null hypothesis, which is means are equal. If the variation within groups is high, then then we probably will not reject the null hypothesis. We looked at data about smokers using ANOVA test in order to see if FEF or the forced mid-expiratory flow is different among six groups, nonsmokers, passive smokers, non-inhalers, light smokers, moderate smokers and heavy smokers.

When we ran the ANOVA test on different groups:

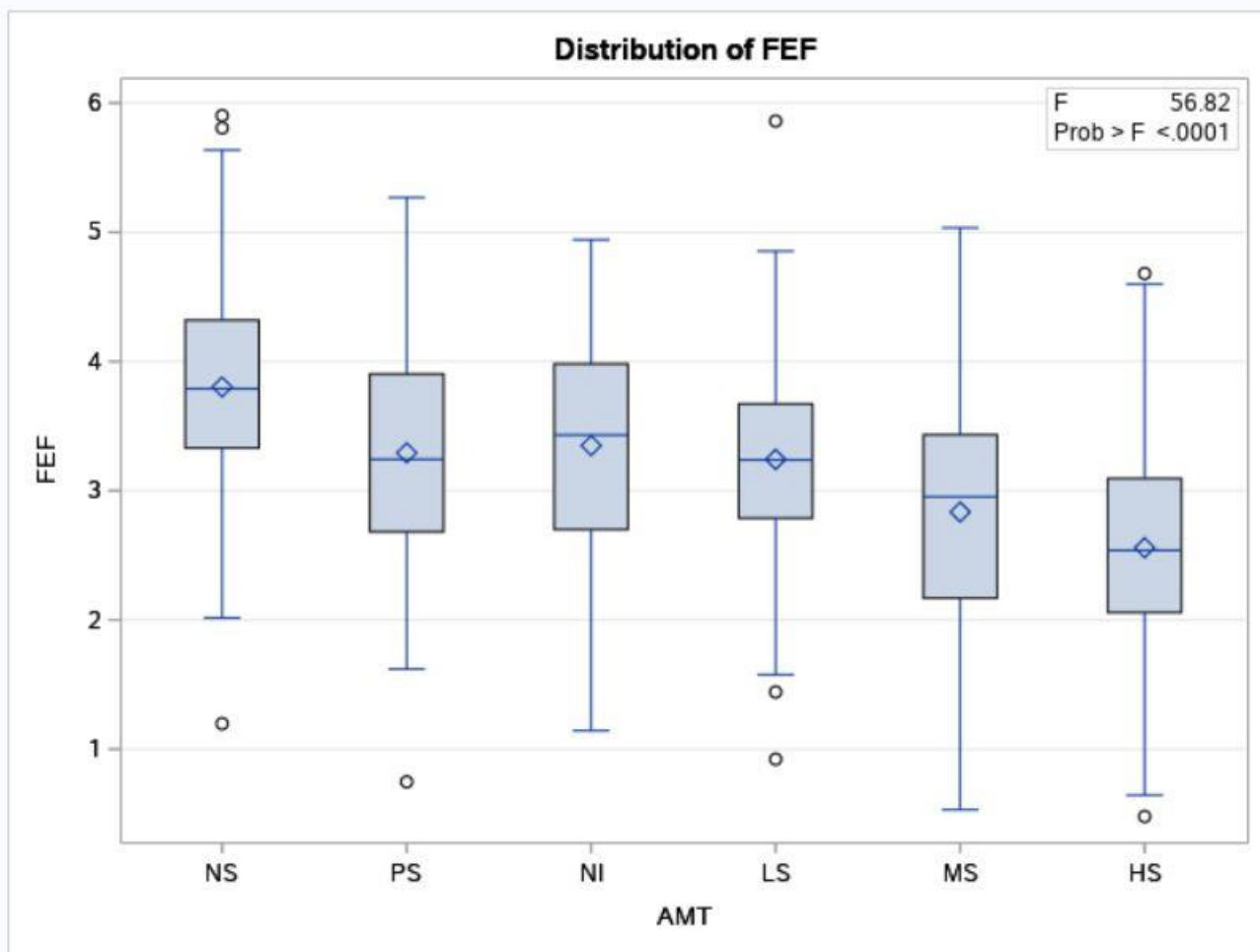
Figure 1

The GLM Procedure					
Dependent Variable: FEF					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	182.6239362	36.5247872	56.82	<.0001
Error	1044	671.0601090	0.6427779		
Corrected Total	1049	853.6840452			

R-Square	Coeff Var	Root MSE	FEF Mean
0.213925	25.40532	0.801734	3.155773

We could see from running the procedure that p-value for our model is <0.001, which is well below our 5 percent threshold. Since this is the case, we reject the null hypothesis. In this case the null hypothesis is there is not equality among the different groups. We strongly support that smoking influences FEF. We also have a linear model as part of our procedure, but our groups are not lined in way that would make the linear model feasible. Therefore, we have a low R² value.

Figure 2



We can see the distribution of FEF among the different categories. The boxplot does confirm our hypothesis from the models that the means are different. We can look at the mean of NS (Non-Smoker) and HS (Heavy Smoker) to see that there is a significant difference in the FEF between the two groups.

Figure 3

The GLM Procedure

t Tests (LSD) for FEF

Note: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	1044
Error Mean Square	0.642778
Critical Value of t	1.96224

Comparisons significant at the 0.05 level are indicated by ***.

AMT Comparison	Difference Between Means	95% Confidence Limits		
NS - NI	0.45373	0.20498	0.70247	***
NS - PS	0.50974	0.35242	0.66706	***
NS - LS	0.56070	0.40339	0.71802	***
NS - MS	0.96702	0.80970	1.12433	***
NS - HS	1.24406	1.08674	1.40138	***
NI - NS	-0.45373	-0.70247	-0.20498	***
NI - PS	0.05601	-0.19273	0.30476	
NI - LS	0.10698	-0.14177	0.35572	
NI - MS	0.51329	0.26455	0.76203	***
NI - HS	0.79033	0.54159	1.03907	***
PS - NS	-0.50974	-0.66706	-0.35242	***
PS - NI	-0.05601	-0.30476	0.19273	
PS - LS	0.05096	-0.10636	0.20828	
PS - MS	0.45727	0.29996	0.61459	***
PS - HS	0.73432	0.57700	0.89163	***

We can also view the t-test to look for the difference between each individual group. Here we can also confirm if that difference is significant as they are noted with the three stars.

Figure 4

The GLM Procedure

Dependent Variable: FEF

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
ns vs ps	1	25.9835335	25.9835335	40.42	<.0001
ns vs ave inhale	1	125.1839038	125.1839038	194.75	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > t
ns vs ave inh	0.98179227	0.07035190	13.96	<.0001
smo mn	2.73514990	0.03711309	73.70	<.0001
nonsmo_mn_vs_smo_mn	0.81240958	0.05462898	14.87	<.0001

Lastly, we can test individual groups within our 6 categories against each other, with either a contrast or parameter test. In the contrast you can put a weight on the groups to have them factor in with the model, but they must add up to zero. A parameter is different in that the groups do not have to add up to zero, but you must include in the estimate specification the term intercept T where T is the total of the linear coefficients for the group terms. We can also see that for our models, all models reject the null hypothesis.