ME 308 ASSIGNMENT

ANOVA | Analysis of Variance

Abstract: This assignment outlines the steps taken to enhance understanding and appreciate the use cases of ANOVA. The document details the entire process, beginning with data generation based on the conditions outlined in the problem statement, then applying ANOVA procedure as explained the lecture to the data for three values of risk factors, followed by directly generating ANOVA table using MS Excel Data Analysis tool and solver. Finally, conclusions are made using the F statistic value as obtained from each iteration. The process chosen for analogy is ethanol blending of petrol.

Introduction: The ***practical and logical*** use case of Single Factor ANOVA is vast and is applicable to a diverse spectrum of fields. It is used to compare populations using their sample mean values. To be specific one of the areas where this can be used is to test the effect of ethanol blended petrol (with different conc. of ethanol) on the emission score of the vehicle measuring the level of pollutants (E.g. NO_x gases). This will help us realize whether ethanol blending really creates a difference in terms of reducing pollutants in the exhaust and contribution to cleaner air. Hence, here treatments would be varying levels of ethanol concentration and replications would repeated tests on the engine to obtain the emission score.

Now let us look at the ANOVA analysis performed on the generated data. The solution is explained in various phases:

1. *Data Generation:* Data generation was done using a python script using NumPy & SciPy with DOB, number of replications, tolerance and number of decimal places as input parameters. The generated data is shown.

Treatment_1	Treatment_2	Treatment_3	Treatment_4
6.345	10.215	14.11	18.591
6.013	9.534	13.751	18.104
6.424	10.358	13.603	17.849
6.882	9.831	14.256	18.107
5.963	9.998	14.469	16.731
5.963	9.83	14.418	17.901
6.911	11.156	13.521	18.283
6.487	10.148	13.79	19.025
5.84	9.584	14.114	17.704
6.369	10.6	14.44	17.511
5.843	9.496	13.704	17.714
5.842	10.268	13.852	18.653
6.212	9.097	13.386	18.264
5.085	9.438	13.34	17.696
5.184	10.262	14.358	18.387
5.792	10.554	14.633	18.111
5.556	10.248	13.91	18.688
6.25	10.093	14.455	17.582
5.611	9.993	14.129	17.83
5.347	9.357	13.619	17.787
6.852	9.766	14.129	17.078
5.967	9.906	14.725	18.243
6.121	10.726	13.928	18.22
5.341	10.341	14.739	18.05
5.801	9.203	12.619	17.891

2. *Manual ANOVA:* According to the ANOVA scheme the table was prepared with all the parameters for four iterations of risk factor i.e, $\alpha \in \{0.01, 0.025, 0.05, 0.1\}$. The tables are mentioned below:

The hypothesis testing:

H_0: $\mu_1=\mu_2=\mu_3=\mu_4$

H_1: At least one of μ_i is different from others.

It has been observed that in all cases **F-Statistic > $F_{\alpha,\, p-1,\, p(r-1)}$** which leads to rejection of the null hypothesis in every case.

This leads to the conclusion that there is significant difference between the means of these treatments.

α	Source of Variation	df	SS	MS	F-Statistic	F-Crit	Reject H0?
0.01	Treatments	3	1999.972	666.657	2666.591	3.9924	Yes
	Error	96	24.00035	0.25			
	Total	99	2023.972				

α	Source of Variation	df	SS	MS	F-Statistic	F-Crit	Reject H0?
0.025	Treatments	3	1999.972	666.657	2666.591	3.2553	Yes
	Error	96	24.00035	0.25			
	Total	99	2023.972				

α	Source of Variation	df	SS	MS	F-Statistic	F-Crit	Reject H0?
0.05	Treatments	3	1999.972	666.657	2666.591	2.6994	Yes
	Error	96	24.00035	0.25			
	Total	99	2023.972				

α	Source of Variation	df	SS	MS	F-Statistic	F-Crit	Reject H0?
0.1	Treatments	3	1999.972	666.657	2666.591	2.1417	Yes
	Error	96	24.00035	0.25			
	Total	99	2023.972				

3. *ANOVA using MS Excel Tool:* The generated data was prepared for ANOVA using Excel data analysis tool. Using the tool, the data to be analysed was selected and ANOVA was executed for various values of the risk factor ($\alpha \in \{0.01, 0.025. 0.05. 0.1\}$). The tables generated are mentioned below:

$\alpha = 0.01$

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1999.972	3	666.6573	2666.590597	2.8E-92	3.992403
Within Groups	24.00035	96	0.250004			
Total	2023.972	99				

$\alpha = 0.025$

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1999.972	3	666.6573	2666.591	2.8E-92	3.255332
Within Groups	24.00035	96	0.250004			
Total	2023.972	99				

$\alpha = 0.05$

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1999.972	3	666.6573	2666.591	2.8E-92	2.699393
Within Groups	24.00035	96	0.250004			
Total	2023.972	99				

$\alpha = 0.1$

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1999.97	3	666.657	2666.59	2.8E-92	2.14173
Within Groups	24.0003	96	0.25			
Total	2023.97	99				

It is evident from the table that the ANOVA table generated by MS Excel is almost similar to the manually calculated table and hence leads to the same conclusion about the populations" behavior. **F-Statistic > $F_{\alpha, p-1, p(r-1)}$** for all cases. Now that H_0 is rejected we can rank sample means by preference.

Conclusion: The values in the data indicated the concentration of NO_x pollutants with decreasing levels of ethanol in the petrol samples. ANOVA was used to statistically assess whether variations in ethanol blending lead to significant differences in NOx emissions. By comparing the variance within groups (samples with the same ethanol content) to the variance between groups (samples with different ethanol contents), ANOVA tests the null hypothesis (H_0) that all group means of emission levels are equal. Since the computed F-statistic was significantly greater than the critical value (F-crit) at all significance levels (α=0.01, 0.025, 0.05, 0.1), we reject H_0. This confirms that ethanol blending has a statistically significant impact on NOx emissions.

Ethanol, as an oxygenated fuel, enhances combustion efficiency and lowers the adiabatic flame temperature (T_{peak}), reducing NOx formation, which follows the Zeldovich mechanism:

$$NO \propto \exp(-E_a / RT_{peak})$$

With ethanol reducing T_{peak}, NOx emissions decrease exponentially. The strong ANOVA results validate this relationship, reinforcing ethanol's role in eliminating NOx pollution and improving air quality as decreasing levels of ethanol increased the values of emission to very high levels.

References:

1. The code for data generation: https://github.com/Hrishikesh-H/ME308_ANOVA_Assigment.git
2. https://youtu.be/ZvfO7-J5u34
3. Lecture Slides / Notes.
4. Zeldovich mechanism - Glossary of Meteorology

Submitted By:

Hrishikesh Hiremath

Roll No: 20003037