Technical Report

(Medical Charges Analysis: The Impact of Age and Smoking Status)

Problem Statement: Healthcare costs are influenced by various factors. This analysis focuses on the relationship between medical charges and two potential predictors: an individual's age and their smoking status.

Dataset:

Here is the sample dataset:

Smoker(Yes/No)	Age	Charges
Yes	48	478556
No	36	78455
Yes	54	8881
Yes	22	5300
No	15	4502
No	63	4021

Data consist of following variables:

Age: Numerical (age of individual)
Smoker: Categorical (smoker/non-smoker)
Charges: Numerical (medical charges)

Where **'Age'** and **'Smoker'** are input factors and **'Charges'** is the response variable.

Hypotheses:

- ***HO1:*** Age is unrelated to differences in medical charges.
- ***HA1:*** Age is associated with medical cost differences.

- ***HO2:*** Smoking status (smoker vs. non-smoker) is not associated with differences in medical charges.
- ***HA2:*** Smoking status is associated with differences in medical charges.
- ***HO3:*** There is no interaction effect between age and smoking status on medical charges.
- ***HA3:*** There is an interaction effect between age and smoking status on medical charges (meaning the impact of one factor on charges depends on the level of the other factor).

Two-Way ANOVA

Purpose: Two-way ANOVA assesses the influence of two categorical independent variables (age and smoking status) and their interaction on a continuous dependent variable (charges).

Model:

$$\textbf{charges} \sim \textbf{C(age)} * \textbf{C(smoker)}$$

where C( ) indicates treatment of variables as categorical.

The final ANOVA Table:

variable	sum of squares	degrees of freedom	F-statistic	p-value
age	6.530507e+09	46	1.946522	3.690067e-03
smoker	1.980275e+10	1	271.516245	4.717964e-32
age:smoker	3.740502e+09	46	1.114916	3.198056e-01
Residual	8.387405e+09	115	NaN	NaN

Interpretation:

Age: Significant p-value(< 0.05), ***Hence we reject the Null Hypothesis,*** and it suggests that ***there's an association between age and charges.***

Smoker: Significant p-value(< 0.05)**, *Hence we reject the Null Hypothesis*** and it suggests that ***there's an association between smoking status and charges.***

Interaction: P-value (> 0.05) so ***we fail to reject the null hypothesis. There's not enough evidence to suggest an interaction effect.***

Conclusion

1. Age has a significant impact on medical charges.
2. Smoking status also plays a significant role in medical charges.
3. No significant interaction effect was found between age and smoking status.