A Two-Way ANOVA Analysis of Variations in U.S. Health Insurance Coinsurance Across Age Groups and Plan Types with in data analysis process

Abhishek Singh   
Data Science (M.Sc.)  
University of Europe for Applied SciencesPotsdam, Germany  
 abhishek.singh@ue-germany.de

Mati Nakphon  
Data Science (M.Sc.)  
University of Europe for Applied SciencesPotsdam, Germany  
mati.nakphon@ue-germany.deDaniela Oliveira Rocha  
Data Science (M.Sc.)  
University of Europe for Applied SciencesPotsdam, Germany  
daniela.rocha@ue-germany.de

Muthamizhan Alagu  
Data Science (M.Sc.)  
University of Europe for Applied SciencesPotsdam, Germany  
muthamizhan.alagu@ue-germany.de Sarawut Boonyarat  
Data Science (M.Sc.)  
University of Europe for Applied SciencesPotsdam, Germany  
sarawut.boonyarat@ue-germany.de

*Abstract*— The US health system is one of a kind regarding its reliance on private insurance and cost-sharing mechanisms that place financial burdens on patients despite the ACA's mandates. One of the major cost-sharing mechanisms, coinsurance, depends on age and insurance plan types that alter out-of-pocket expenses. Although many studies have been done on the costs of health care, few have addressed how age and types of insurance plans affect coinsurance rates. This paper attempts to fill that gap by using two-way ANOVA to test the effects of age groups and plan types on coinsurance. We do this through the analysis of the healthcare data. Check by two-way ANOVA whether a significant difference exists in the coinsurance rate by various age groups and different plan types to find the interaction of two variables that has put stress on insured persons.

The results show that the plan type significantly affects the coinsurance rate, as evidenced by the p-value less than 0.05, while the age group does not, as the p-value is greater than 0.05, and there is no significant interaction effect. These results underline how plan type features in cost-sharing distributions for informing policy decisions on making healthcare more affordable. This study develops an appreciation for the mechanisms of cost-sharing related to health, with significant emphasis on how plan type would shape the out-of-pocket cost burdens.

Keywords—Two-Way ANOVA, coinsurance rates, insurance plan types, healthcare costs, data analysis, statistical methods, U.S. healthcare system

# **Introduction**

The US health system is one of a kind in the world in its reliance on private insurance and cost-sharing mechanisms, which creates complexities and financial burdens for patients. Unlike most other developed countries with universal health care [1], individuals in the US face a complex system of varying premiums, deductibles, copayments, and coinsurance. These cost-sharing elements mean that healthcare costs in the US topped $3.3 trillion in 2016, or 17.9% of GDP. Even with mandates to buy insurance, such as under the Affordable Care Act, many Americans still face high out-of-pocket expenses even while insured [2].

Coinsurance, in particular, refers to a cost-sharing mechanism whereby policyholders pay a certain percentage of medical costs after their deductible is met; for instance, if the coinsurance is 20%, then the policyholder pays 20% of the allowed costs while insurance covers 80% of the costs [3]. The rate varies widely depending on a host of factors, including the type of insurance plan and demographic characteristics such as age. Understanding these differences is important not only for the policyholders but also for insurers and policy thinkers. The healthcare data categorizes insurance plans mainly into four types: PPO (Preferred Provider Organization), HMO (Health Maintenance Organization), EPO (Exclusive Provider Organization), and POS (Point of Service) [4]. Each of the above plan types may have different coinsurance structures that affect out-of-pocket costs for members.

From an analytical point of view, when we have variables prepared for analyzing the reasons why Americans still need to pay out-of-pocket expenses—such as coinsurance, which is a numerical variable; age, which can be grouped; and plan type, a categorical variable—we should consider statistical methods capable of analyzing these variables. One of the statistical methods is two-way ANOVA, as it's used for analyzing relationships between two independent variables and one dependent variable. Evidence from this statistical method can answer questions we are interested in.

The main aim of this case study is to use the data analysis process and statistic with Two-Way ANOVA to determine whether there is a significant difference in coinsurance across age groups and types of insurance plans and the interaction between the two independent variables on the subject's financial costs. Consequent identification of such patterns has much to do with critical determinants like whether costs shared are meaningfully varied across demographics and how the right selection of insurance plans can align cost sharing with particular needs. Analysis-informed insights into designing more tailored, equitable health insurance policies would go a long way toward reducing disparities and creating an inclusive healthcare system.

# **Background**

## The process of data analysis

Epicycles of Data Analysis according to Roger D. Peng and Elizabeth Matsui from Fig 1: It's a way of showing that the process of data analysis could be iterative, going through its main activities in cycles: stating and refining a question, exploring the data, building formal statistical models, interpreting the results, and effectively communicating findings. These activities are not done in a linear sequence but are revisited once new insights have been gained or challenges arise, in a continuous process of refinement. This will ensure, through an iterative cycle, that each stage of the analysis is examined and adjusted in light of evolving understanding for greater accuracy and insight. It is an iterative process that can, according to the scale and complexity of the project, range from a day or even hours for a small analysis to several months for large-scale analyses. The data analyst may work with the precision of an Epicycles framework through complexities to provide reliable and meaningful outcomes. [5]

A diagram of a diagram of a gear system

Description automatically generated with medium confidence

Fig. 1. Epicycles of Analysis.

The data analysis process involves, in sequence, 7 structured steps (Fig. 2) that transform raw information into actionable insights in a reliable and relevant way. Understanding the Problem and Desired Results makes sure that research goals are appropriate for stakeholder expectations and puts a clear focus on the investigation. Setting a clear metric involves establishing criteria such as statistical significance levels, for example, p-values, or identifying key performance indicators that reflect the success of the research in addressing the defined problem. Gathering Data: This involves understanding what particular data will be needed to solve the problem, whether the data is structured or unstructured, the sources that are relevant to the research objectives and are easily accessible, and the selection of appropriate tools that can be used in the collection and analysis to facilitate the process. Data cleaning involves the cleaning of the dataset for missing values, outliers, and inconsistencies for accuracy and reliability. It aims to correct these features to make the data ready for a deep analysis that yields valid and actionable insights. The functions include data manipulation in order to discover trends, find correlations, or identify patterns and variations. Results interpretation involves verification of statistical significance, confidence intervals, and possible limitations of the study. Presenting the findings: This is an area of clearly communicating your findings to the audience by letting them know what assumptions were behind the analysis and why certain hypotheses were set. It will be presented through reports, dashboards, charts, and graphs that will help in bringing forth insights [6]. Will Hillier insists on comprehension of the whole process because a well-structured framework is vital to reliable results. According to Hillier, the core steps include the definition of the question, gathering and cleaning of data, analysis, results dissemination, and learning from any failure. This holistic approach allows for a deeper understanding of the data analysis process and helps to tailor it to specific research needs [7].

A diagram of data analysis process

Description automatically generated

Fig. 2. The Data Analysis Process.

## Two-Way Analysis of Variance (ANOVA)

While one-way ANOVA extends into two-way ANOVA in the inclusion of the two independent variables and one dependent variable, one-way ANOVA deals with the effect resulting from a single independent variable, and two-way ANOVA seeks interactions between independent variables and the impact on the dependent variable. It includes a comparison of the mean difference between groups on at least two independent variables whereby researchers’ study not only the main effect of one independent variable but also the interaction effect between independent variables. Two-way ANOVA serves to assess whether an interaction exists between the two independent variables and the way this interaction may influence the dependent variable. [8]

In a standard two-way ANOVA, there are three main effects that are considered to explain how two independent factors influence a dependent variable.

* Main Effect of the First Factor: This shows the impact that the first independent variable, or, for example, the levels of sunlight exposure—no, low, medium, and high—have on the dependent variable, in this case. This finds whether or not statistically significant differences of the dependent variable between this factor's levels have been created by ignoring the effects from the second factor.
* Main Effect of the Second Factor: Similar to the first, this assesses whether the second independent variable (e.g., watering frequency with levels like daily and weekly) significantly affects the dependent variable, independently of the first factor's influence.
* Interaction Effect: This represents whether, when the two factors are put together, something peculiarly different from what any of the single factors would have suggested results. It shows if one factor's effect is based on another factor's level.

Results from the experiment on sunlight and watering frequency were obtained using Excel. Table 1: The first three sections in the table show summary statistics for each group. For instance, the mean height of plants that were subjected to daily water, but no sunlight was 4.14 inches. The mean height of all plants that received high levels of sunlight was 5.55 inches.

A screenshot of a spreadsheet

Description automatically generated

Fig. 3. The results of a two-way ANOVA conducted in Excel

From this table the last section presents the results of the two-way ANOVA:

* The p-value for watering frequency is 0.975975, which is not statistically significant at an alpha level of 0.05.
* The p-value for sunlight exposure is 3.9E-8 (0.000000039), which is statistically significant at an alpha level of 0.05.
* The p-value for the interaction between watering frequency and sunlight exposure is 0.310898, which is also not statistically significant at an alpha level of 0.05.

These results indicate that sunlight exposure is the only factor with a statistically significant effect on plant height. Since there is no significant interaction effect, the influence of sunlight exposure remains consistent across all levels of watering frequency. In other words, whether a plant is watered daily or weekly does not affect how sunlight exposure influences its growth. [9]

# **The proposed method**

In this section, we will implement the data analysis process, followed the steps outlined below.

## Understanding the problem and desired result

The Centers for Medicare & Medicaid Services have collected coinsurance data across many insurance plan types. This is categorized into options like “no charge”, “20% Coinsurance after deductible”, "60% Coinsurance after deductible” and “No Charge” These coinsurance categories - PPOs, POS plans, HMOs, EPOs, and Indemnity plans - have a set of rules around whether or not one has a provider network, if a referral is required, or the extent to which out-of-network services are covered.

Certain Insurers recommend that young adults go for HMO policies, which usually have inexpensive coinsurance, and a percentage shared in costs. All these can be done so that the premiums remain rather affordable while still offering required coverage. On the reverse side, older adults should consider PPO plans where the coinsurance is greater but offers more flexibility towards the choice of healthcare professionals.

## Setting a clear metric

To clarify, we aim to answer the following question using statistical analysis: How do coinsurance rates vary across age groups? What is the effect of different plan types on coinsurance rates? Is there an interaction between age groups and plan types influencing coinsurance rates?

* **Two-Way ANOVA methods:** To determine that whether age and insurance plan type significantly influence the variation of coinsurance amounts, the most appropriate statistical test to apply is Two-Way ANOVA. In this respect, such a method allows investigating whether two independent variables-age group and plan type-affect the dependent variable of interest (coinsurance amounts) since there could also be an interaction between the two factors.
* It will be good for this condition because Two-way ANOVA provides the main effects of the independent variables alone (age group and plan type) as well as the interaction effects between these factors. This will be important since, for instance, the effect of one factor (for example, age) can change with the levels of another factor such as the type of plan.
* Main effects: Significant differences of coinsurance amount between the different age groups and between the insurance plans will be evaluated.
* The interaction effects will determine whether the relationship between age and coinsurance amounts varies with the type of insurance plan, and vice versa.

If significant differences between main effects or interaction effects are found, this could indicate that age and/or plan type significantly impact the amount of coinsurance, thus supporting the hypothesis that in healthcare, cost-sharing mechanisms differ across demographic and plan factors.

* **Identify the hypothesis**
* Age Group

H0 : Age group does not affect coinsurance

H1 : Age group affects coinsurance.

* Plan Type

H0 : Plan type does not affect coinsurance.

H1 : Plan type affects coinsurance.

* Interaction Effect

H0 : There is no interaction effect between age group and plan type on coinsurance.

H1: There is an interaction effect between age group and plan type on coinsurance.

* **Two-Way ANOVA Assumptions:** The underlying assumptions and limitations that come with such a statistical analysis have to be considered to ensure that the results obtained from such analyses are valid and reliable [10] [11].
* Normality Assumption: One of the most important assumptions in two-way ANOVA is that of normality. The assumption of normality assumes that the distribution of the residuals should be normal; the values of the dependent variable should follow a bell curve.
* Homogeneity of Variance Assumption: The variation around the mean for each group being compared should be similar among all groups. If your data do not meet this assumption.

## Gathering data

The source data for this analysis is obtained from the U.S. Health Insurance Marketplace, which is available via Kaggle. We consider only three files from the source that are relevant for our analysis:

* **BenefitsCostSharing.csv:** contains information on coinsurance by detail, including different coinsurance structures, such as "no charge", "0%", "10%, 65%, 100% Coinsurance after deductible" etc. This will help examine how coinsurance values may change across insurance plans.
* **Rate.csv**: The file contains, for every insurance plan, age groupings for which the records will be useful in making sense of how age may influence the coinsurance structure.
* **PlanAttributes.csv**:It lists the kinds of plan types available, for instance, PPO, POS, HMO, EPO, and Indemnity, that one can find within the marketplace. This gives us scope to study, afterwards, how coinsurance values could be varied with respect to this type.

## Cleaning Data

Data cleaning is an important step for any analysis of data because it ensures the quality, consistency, and reliability of the dataset before any further analysis can be conducted.

* **Observation Total Rows and Columns:** In this work, we chose data from the year 2014 and three CSV files: rate, benefit cost sharing, and plan attributes. These three datasets provide detailed information with regard to health insurance rates, cost-sharing structures, and plan attributes. Filtering for Business Year = 2014, the datasets contained 3,796,388 records from rate, 1,164,869 records from Benefit Cost Sharing, and 18,719 records from Plan Attributes, respectively. This huge amount of information was then cleaned to maintain consistency and relevance for the analysis.
* **Data Types and Possible Issues:** The datasets are a mix of numeric, categorical, and string data types; hence, these need to be validated and transformed with care for proper analysis. Some numeric fields had text or special characters in them; hence, these needed some preprocessing steps like removal or conversion into appropriate formats. The missing values (NaN) were treated using imputation techniques or, where domain expertise was required for interpretation, the records were removed. Standardization efforts included specific adjustments to ensure data consistency. For example, records that had "family option" in the rate.csv file were filtered out, and the age group "0-20" was replaced with a representative value of 20. In the BenefitCostSharing.csv dataset, all the records with "no charge" in the coinsurance column were converted to numeric values of 0. These preprocessing steps thus yielded a clean and structured dataset, which provided a sound basis for subsequent statistical analyses.
* **Categorical Variables:** When analyzing the health plans, age served as an independent variable in the two-way ANOVA test. Since the age variable in our dataset is continuous, we applied the binning technique to segregate ages into discrete groups. The reason for that is the analysis of age groups interacting with the health plans should help derive conclusions regarding variations in health-related outcomes for these subjects. We decided to split the age category into
* 0-20: This age group includes children who usually are under their parents or guardians for health insurance covers.
* 21-25: This category involves a young adult, covered till the age of 26 under parental insurance, who will eventually seek their cover.
* 26-40: The individuals that fall within this category range from early to mid-career level. They increasingly seek private insurance plans and employer-sponsored insurance.
* 41-60: Is representative of the pre-retirement age, usually associated with higher attention to preventative care and management of chronic conditions.
* 61-64: The near-retirement category is covered through employer plans and private health insurance before this group turns 65 to get onto Medicare.
* 65+: Seniors who are qualified for Medicare represent the final stage in the transition of working-age into retirement.
* **Processing Complex Coinsurance Description:**
* This is a dataset whereby most of the times in this column, there exist a lot of descriptions such as "50% coinsurance after deductible." These should be preprocessed and standardised into their numeric equivalents. We came up with an approach that extracts the numeric part from the descriptions, like 50% | Coinsurance after deductible, and then transforms the extracted values to the numeric data type.
* Preprocessing of missing values in the column for coinsurance: In this case, those values were dropped from the dataset because interpretation required domain expertise that was not available. This is a very important step to make sure that only clean and actionable data flow through into further analyses.
* **Mean Coinsurance Calculation:** One health plan, as defined by StandardComponentId, can have several benefits associated with it. Each of these benefits is classified under a separate BenefitName but falls under the same StandardComponentId, representing the overall plan or coverage component. To have a better view of the cost-sharing structure of each plan, we summary statistics of the coinsurance by StandardComponentId, computing the mean coinsurance across each plan. This allowed for a fairer comparison among the different plans and ages.
* **Merge dataset:** In the analysis of the relationship between age, plan type, and coinsurance, it was necessary to integrate information found in three separate datasets: rate, PlanAttributes, and BenefitCostSharing. Each of these three datasets contained variables critical to the analysis but did not have direct links for easy integration. The process of merging started by joining the rate.csv dataset, which includes age, with the PlanAttributes dataset, which includes plan type, on the common key called PlanId. Moreover, this joined data was merged further on the StandardComponentId key with BenefitCostSharing, which contains coinsurance. Once the data was integrated, a new column was added called AgeGroup; to do this, some binning techniques were applied so as to categorize the ages coherently following a certain structure. Therefore, this integrated data set, with representations of age, plan type, and coinsurance in it, will form the backbone of the two-way ANOVA test that might be conducted later in the steps.
* **Random Selection of Observations:** The random sampling was done in order to have representative analyses and prevent any over presentation that could lead to bias. Precisely, we select 1,000 observations each from the six age groups to make it 6,000 observations, to ensure a fair representation of the subjects of each category and thus lead to a much stronger finding of the result from the data analysis. Random sampling is important in such studies to reduce the effects of sampling bias and ensure that the findings are generalizable.

# **expriments**

## Analyzing and mining data

After cleaning the data and preparing the data so that it passed all assumptions for statistical testing, Fig 4 provides evidence of meeting the assumptions from a two-way ANOVA with the variables 'Agegroup,' 'Plantype,' and 'Coinsurance'.

A table with numbers and letters

Description automatically generated

Fig. 4. Appropriate data for statistical testing.

We check three assumptions before use two-way ANOVA: normality of the dependent variable and homogeneity of variable.

#### Normality of the dependent variable: The testing of whether coinsurance rates vary across different age groups and plan types would involve comparing the quantiles of coinsurance rates for each age group and plan type with that of a normal distribution curve using a Q-Q plot. By doing this we see that the two sets of quantiles fall roughly along a straight line, which implies that the distribution of coinsurance rates is roughly normal (Fig. 5).

A graph with a red line

Description automatically generated

Fig. 5. The distribution with Q-Q plot.

Moreover, from the histogram plots, it was observed that for some subgroups such as, age group 0-20 and having a PPO plan, for an age group of 21-25 and also having a PPO plan, dependent variable 'coinsurance' had approximately taken a normal distribution depicting a bell shape (Fig. 6).

A graph of a graph

Description automatically generatedA graph of a graph with blue lines

Description automatically generated

Fig. 6. The distribution with Histogram plot.

#### Homogeneity of Variance: the homogeneity of variance assumption will be tested by applying Levene's Test. This is a test to see whether the variances in different groups are equal; this is one of the important assumptions when doing most statistical analyses like Two-Way ANOVA. Levene's test operates under the principle of first calculating the absolute deviation of each data point from its group mean, then testing for equality among the groups regarding these absolute deviations. The null hypothesis for the Levene's Test is the equality-homogeneity of the variances across all groups. The alternative hypothesis states that at least one group is different in its variance from other groups. In interpreting this result, we check out the p-value: If the p-value > 0.05, we fail to reject the null hypothesis, and this would mean there is no significant difference in the variances, hence equal variances across groups. If the p-value is ≤ 0.05, we reject the null hypothesis. This means that there is a significant difference in variances between groups.

The result suggest that p-value from Levene's Test libraly with python from our analysis is 0.804442810550771 (Fig. 7), larger than the common significance level of 0.05. Hence, we fail to reject the null hypothesis of variance. Therefore, we are assured of homogeneity of variance, meaning in perspective, the variances in coinsurance rates between and within different age groups and plan types are approximately equal, thus satisfying one of the key assumptions necessary to carry out a Two-Way ANOVA.

#### A screenshot of a computer code Description automatically generated

Fig. 7. The result of homogeneity of variance.

## Interpreting results

To determine if age groups and plan type each have a statistically significant effect on coinsurance rate, and if the effects on age groups on coinsurance rate depend on whether the plan type is in the PPO, POS, HMO, EPO, and Indemnity, a two-way ANOVA was conducted. According to Table 1, the result using the Pingouin-a library in Python for conducting a two-way ANOVA.

* The p-value of the variable Age Group had a value of 0.4267589, or 4.267589e-01, which is beyond the level of significance. Because the p-value is > 0.05, we fail to reject the null hypothesis. Based on this view, it can be postulated that age group insignificantly contributes to a significant effect in the variation between subjects concerning coinsurance rates.
* The p-value for the variable Plan Type was 1.171894e-286, which is practically 0, less than 0.05. Since p < 0.05, we reject the null hypothesis. This means that there is a significant difference between plan types concerning the coinsurance rate.
* The interaction effect of the age group and plan type had a p-value of 0.9705367 (9.705367e-01), which is greater than the 0.05 level of significance. Since p-value > 0.05, we fail to reject the null hypothesis. This would insinuate that the interaction of the age group and plan type does not have a significant effect on the coinsurance rates.

Table 1Two-Way ANOVA Table generated by Python.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | SS | DF | MS | F | p-value  (p-unc) | Partial η² (np²) |
| Age Group | 0.086137 | 5.0 | 0.017227 | 0.982485 | 4.267589e-01 | 0.000822 |
| Plan Type | 26.112719 | 4.0 | 6.528180 | 372.305836 | 1.171894e-286 | 0.199649 |
| Age Group  x  Plan Type | 0.172878 | 20.0 | 0.008644 | 0.492966 | 9.705367e-01 | 0.001649 |
| Residual | 104.680693 | 5970.0 | 0.017534 |  |  |  |

As the result in Table 2, which was generated using Excel, there was no significant interaction between age groups and plan type on coinsurance rates (p = .968). However, there was a statistically significant main effect of plan type on coinsurance rates (p < .05) as well.

Table 2 Two-Way ANOVA Table generated by Excel.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Source of Variation | SS | DF | MS | F | p-value | F crit |
| Sample | 0.01396914 | 5 | 0.00279383 | 0.33103005 | 0.89449631 | 2.21439767 |
| Columns | 77.8093165 | 4 | 19.4523291 | 2304.83304 | 0 | 2.37222858 |
| Interaction | 0.08434595 | 20 | 0.0042173 | 0.49969164 | 0.96825949 | 1.57087302 |
| Within | 252.940796 | 29970 | 0.0084398 |
| Total | 330.848427 | 29999 |  |  |  |  |

## Presenting finding

The findings showed that insurance plan type plays a significantly contributing role in determining the co-insurance rate, therefore resulting in a statistically significant effect: p < 0.05. That means the class or type of insurance—which might be PPO, POS, HMO, EPO, or indemnity—has a significantly related effect on the degree to which financial burdens have been shared between patients and payers. This result highlighted that better plan design can help with designing more equitable and affordable options for a diverse group of individuals. It therefore indicates that the plan type remains a key factor in either assessing or reforming the policies that policymakers and insurers face regarding insurance offerings.

In contrast, no significant effect of age group was observed on coinsurance rates, p > 0.05. This indicates that, as a singular factor, age does not have any real impact on the economic burden of co-insurance. Although this may somewhat run counter to intuition, as health needs and costs across one's lifetime often follow a curvilinear trajectory, it demonstrates that one age group is not uniquely burdened over another by the structuring of co-insurance.

Besides, there was no significant interaction effect of age group and plan type on coinsurance rates either, p > 0.05. This result would suggest that the impact of plan type on coinsurance rates is consistent across age groups, and there is no meaningful synergy or dependency between the two variables.

# **conclusions**

The study concludes that, data analysis process and a two-way ANOVA statistical method that could provide identification of significant differences either between main effects or interaction effects. This would afford an insight into how the variation of cost-sharing mechanisms goes on across demographic groups and insurance plans by analyzing how more than one factor, such as age and insurance plan type, affects the coinsurance rates. The findings from this study in determining the coinsurance rate from the type of insurance plan are significant with a statistically significant effect: p-value < 0.05. This implies that different insurance plans, including PPO, POS, HMO, EPO, and indemnity, vary in their impacts on how the financial burdens are distributed between the patients and the insurers. This, therefore, supports the hypothesis that cost-sharing mechanisms of healthcare vary across plan factors.

In the future, we will conduct further work beyond the two-way ANOVA by considering a multiple regression model to investigate the relationship between a variety of independent variables and their interaction on coinsurance rates. This would provide a better understanding of how these factors interact and cause variation in coinsurance rates, offering a full investigation into the various ways in which multiple factors combine to drive the economic burdens faced by patients within the healthcare system.

##### References

1. "Fundamentals of Statistics," John Wiley & Sons, available at: <http://samples.jblearning.com/9781284262988/9781284265064_FM.pdf>.
2. "Understanding Two-Way ANOVA," YouTube video, available at: <https://www.youtube.com/watch?v=DublqkOSBBA>.
3. "Co-insurance," Healthcare.gov, available at: <https://www.healthcare.gov/glossary/co-insurance/>.
4. "Understanding HMO, PPO, EPO, POS," UnitedHealthcare, available at: <https://www.uhc.com/understanding-health-insurance/types-of-health-insurance/understanding-hmo-ppo-epo-pos>.
5. R. D. Peng and E. Matsui, The Art of Data Science: A Guide for Anyone Who Works with Data, 1st ed., Skybrude Consulting, LLC, 2016, pp. 4-8.
6. Prof. Dr. Talha Ali Khan, "Data Analytics Lectures 1-4," Lecture Notes, pp. 105-112.
7. CareerFoundry, "The Data Analysis Process: Step by Step," CareerFoundry Blog. [Online]. Available: <https://careerfoundry.com/en/blog/data-analytics/the-data-analysis-process-step-by-step/>.
8. Prof. Dr. Talha Ali Khan, "Data Science and Statistics Lectures 8-12," Lecture Notes, pp. 181-194.
9. Statology, "Two-Way ANOVA in Excel," [Online]. Available: <https://www.statology.org/two-way-anova-excel/>.
10. "Assumptions and Limitations: Understanding the Scope of Two-Way ANOVA," FasterCapital, available at: <https://fastercapital.com/content/Assumptions-and-limitations--Understanding-the-Scope-of-Two-Way-ANOVA.html>.
11. Scribbr, "Two-Way ANOVA," available at: <https://www.scribbr.com/statistics/two-way-anova/>.