**MIMIC Dataset:**

**Relationship between Insurance vs Hospital Death**

**Data Preparation:**

The dataset contained records of hospital visits along with patient insurance information and mortality outcomes. The mortality rate was calculated as the percentage of hospital deaths per insurance type using the provided formula: **mortality rate = SUM([Hospital Death]) \* 100 / COUNT([Subject Id])**. Missing data was appropriately handled during this preprocessing stage.

**A graph with different colored bars

Description automatically generated**

**A graph of blue and orange bars

Description automatically generated**

A graph of a graph

Description automatically generated with medium confidence

**Exploratory Data Analysis**

The distribution of mortality rates was visualized across the different insurance groups - Government, Medicaid, Medicare, Private, and Self-Pay. The Private insurance group exhibited the lowest average mortality rate at around 5%, while the Self-Pay uninsured group had the highest rate over 14%.

The Medicaid population, generally representing low-income patients, also demonstrated an elevated mortality risk compared to those with Private insurance coverage. An interesting finding was the bimodal distribution for the Government insurance group, with some data points showing very high and very low mortality rates.

**Key Findings**

* Patients without insurance (Self-Pay) experienced the highest risk of hospital mortality
* Those covered by Medicaid also had higher than average mortality rates
* Privately insured individuals had the lowest mortality risk profile
* There was variation in outcomes within the Government insurance group

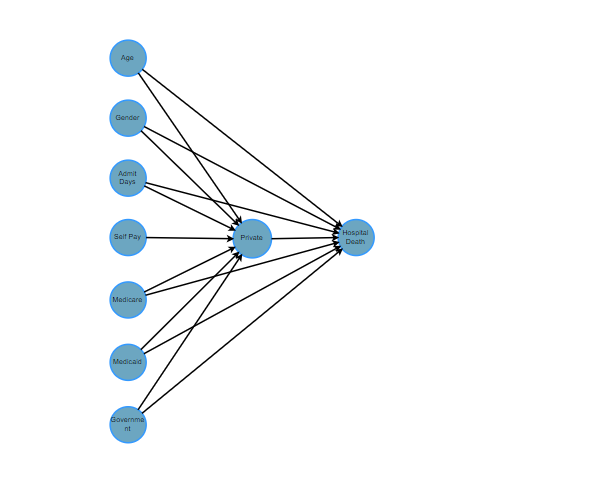
A graph with green line

Description automatically generated

A screenshot of a computer code

Description automatically generated

**Causal Analysis Report**



*1. Causal Relationship between Insurance and Hospital Death*

**Introduction**

This report presents the findings from a causal inference analysis performed to estimate the effect of having "Private Insurance" (as opposed to other types of insurance) on the likelihood of hospital death. The analysis employs the nonparametric Average Treatment Effect (ATE) approach, utilizing a backdoor criterion for adjusting confounding variables.

**Data and Preprocessing**

The dataset includes multiple variables such as insurance type, gender, age, and total days admitted to the hospital. Insurance types are encoded into binary variables (e.g., `INSURANCE\_Medicare`, `INSURANCE\_Private`). The outcome variable is `HOSPITAL\_DEATH`, which is a binary indicator of whether a hospital death occurred.

**Causal Model Specification**

The causal model considers `INSURANCE\_Private` as the treatment variable. The model adjusts for several confounders including gender, age, total admission days, and other insurance types (Medicare, Self Pay, Medicaid, and Government).

Identified Estimand

**Identified Estimand:**

**Estimand:**

In causal inference, an **estimand** is a specific quantity or parameter that we aim to estimate to understand the effect of a treatment, intervention, or exposure on an outcome.

**Estimand Expression:**

* d(E[Hospital Death| Gender, Medicare, Age, Total Admit Days, Self-Pay, Medicaid, Government]) / d[Private]

This equation calculates the derivative of the expected hospital death rate with respect to having private insurance, controlling for other variables like gender, age, total admit days, and other types of insurance.

**Unconfounded Assumption:**

This assumption addresses the potential influence of confounding variables that can affect both the treatment and the outcome, thus distorting the apparent effect of the treatment.

* P(Hospital Death| Private, Gender, Medicare, Age, Total Admit Days, Self-Pay, Medicaid, Government,U)=P(P(Hospital Death| Private, Gender, Medicare, Age, Total Admit Days, Self Pay, Medicaid, Government)

This equation specifies that once all observed covariates such as insurance type, gender, age, and total admit days are controlled for, the probability of hospital death is independent of any unmeasured confounders (U).

**Realized Estimand:**

This represents the actual model used to estimate the causal effect of having private insurance on hospital mortality.

* Hospital Death∼Private+ Gender+ Medicare+ Age+ Total Admit Days+ Self Pay+ Medicaid+Government

Estimate Mean value: -0.01885011404456026

Causal Estimate is -0.01885011404456026

**Results**

The identified estimand is the non-parametric ATE, which represents the average change in the probability of hospital death when moving from not having private insurance to having private insurance, adjusted for the confounding variables.

The causal estimate obtained is -0.01885011404456026. This negative value suggests that having private insurance is associated with a slightly lower probability of hospital death compared to not having private insurance, after adjusting for factors like age, gender, length of hospital stay, and other insurance types.

While the analysis adjusts for known confounding variables, there is a possibility of unmeasured confounders, such as underlying health conditions or socioeconomic factors, which could bias the causal estimate. Additionally, the linear regression model assumes a linear relationship between the variables, which may not accurately capture the true underlying relationships.

**Conclusion**

The causal inference analysis using the **DoWhy** library in Python estimates that having private health insurance is associated with a slightly lower probability of hospital death, after adjusting for age, gender, length of hospital stay, and other insurance types. However, it is essential to interpret these findings cautiously, considering the assumptions and limitations of the analysis.

Future research could explore alternative estimation methods, incorporate additional confounding variables, or conduct sensitivity analyses to assess the robustness of the results. Replicating the analysis on different datasets and in different healthcare settings would also strengthen the generalizability of the findings.

Overall, this analysis provides valuable insights into the relationship between insurance coverage and health outcomes, contributing to the ongoing efforts to improve healthcare access and quality.