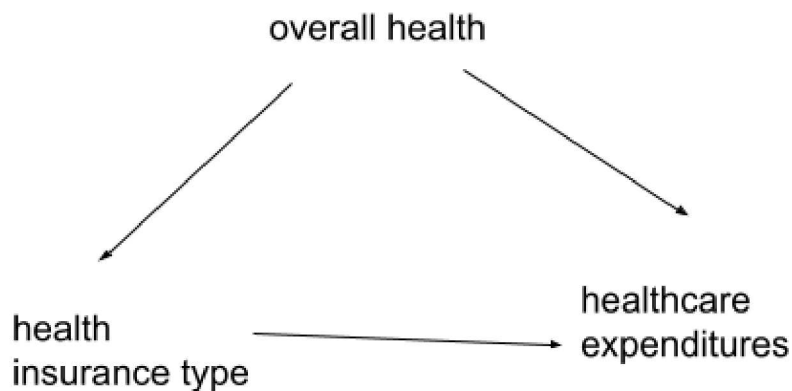


Interim Submission

Literature review

A causal graph is a diagram that looks like this. Variables relevant to a causal investigation are connected by arrows. They'll be immensely valuable to us because, as we'll see, they are actually depictions of statistical models. For this reason, they'll help us harness our subject matter knowledge to design for exchangeability across a wide range of study designs, and they'll help us analyze data appropriately to answer causal questions.



Formally, graphs are mathematical objects that are collections of nodes or vertices, indicated by the circles, and edges linking the nodes, indicated by the arrows. In the case of these arrows, the edges are called directed edges because the connections between nodes have a direction. A graph whose edges are all directed is called a directed graph. If the edges connecting the nodes did not have directions, the edges would be called undirected edges. A graph whose edges are all undirected is called an undirected graph.

Causal information is deemed highly valuable and desirable along many dimensions of the human endeavor, including in science, engineering, business, and

law. The causal inference framework in Pearl's framework and related methods have emerged as vital within epidemiology. Cause and effect relationships play a central role in how we perceive and make sense of the world around us, how we act upon it, and ultimately, how we understand ourselves. the Causal Hierarchy Theorem (CHT), demonstrates that the three layers of the hierarchy almost always separate in a measure-theoretic sense.

In a more systematic fashion, causality plays a central role in how we probe the physical world around us and ultimately understand Nature.

For any proper framework causal inference gives insight like ;

1. The causal mechanisms underlying the phenomenon under investigation should be accounted for – indeed, formalized – in the analysis.
2. This collection of mechanisms (even if mostly unobservable) should be formally tied to its output: the generated phenomena and corresponding datasets.

Structural Causal Models (SCMs) describe the collection of mechanisms underpinning a phenomenon of interest. Each SCM naturally defines a qualitative hierarchy of concepts, which is called the Pearl Causal Hierarchy, or PCH. Following Pearl's presentation, we label the layers (or rungs, or levels) of the hierarchy associational, interventional, and counterfactual. The concepts of each layer can be described in a formal language and correspond to distinct notions within human cognition. Each of these allows one to articulate with mathematical precision qualitatively different types of questions regarding the Observed variables of the underlying system.

Overview of the data source and formats

We extract the data from [kaggle](#) directly upload it to our local machine. The data has 569 rows and 33 columns. It contains two class Class distributions: 357 benign (not cancer) and 212 malignant (cancer). Attribute Information ID number and

Diagnosis(M = malignant, B = benign). It has Ten real-valued features computed for each cell nucleus.

Techniques used to perform causal inference

First, we Perform feature extraction by finding out the relationships of each variable with targeted variables. and the relationship between each other variables. And after that, we conduct the scaling and regression method.

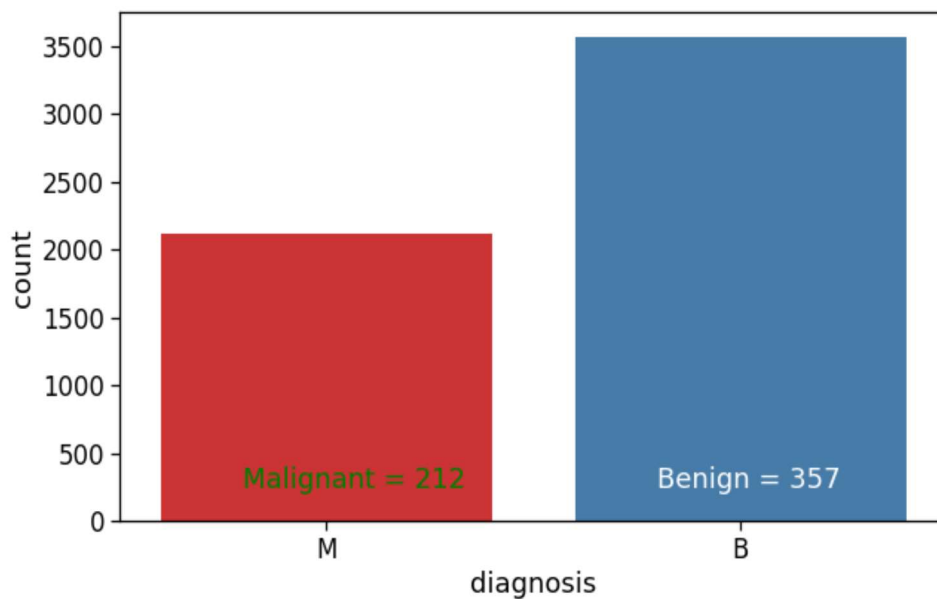
Then we make a causal graph. To implement causal graphs we use CausalGraphicalModel is a python module for describing and manipulating [Causal Graphical Models](#) and [Structural Causal Models](#).

Causal Graphical Models As inference tools, the graphs enable researchers to estimate effect sizes from non-experimental data.

Structural Causal models have a 4E framework that is 1) Equation (model or equation specification), 2) Estimation of free parameters, 3) Evaluation of models and model fit, 4) Explanation and communication, as well as execution of results.

Insights derived from analysis

The plot below shows the frequency of cancer stages.



In order to check the correlation between the features, we plotted a correlation matrix. It is effective in summarizing a large amount of data where the goal is to see patterns.

