**Context**

Data relating to patient health status or the delivery of healthcare is routinely collected from a range of sources and is increasingly being used in clinical decision-making. Organizations are leveraging EHR data, patient registries, and mobile device information to better understand trends and outcomes, leading to improved care delivery.

By leveraging data and technology, the health care sector and its social care partners have an opportunity to improve the efficiency, effectiveness, and sustainability of efforts that address health-related social needs as a regular component of health care delivery.

**Data Set**

Source: https://www.kaggle.com/mirichoi0218/insurance

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column Name | Description | Data Type | Non-null Count | Type of Variable |
| Age | This is an integer indicating the age of the primary beneficiary (excluding those above 64 years, since they are generally covered by the government). | int64 | 1338 | Continuous |
| Sex | This is the policy holder's gender, either male or female. | object | 1338 | Categorical |
| BMI | This is the body mass index (BMI), which provides a sense of how over or under-weight a person is relative to their height. BMI is equal to weight (in kilograms) divided by height (in meters) squared. An ideal BMI is within the range of 18.5 to 24.9. | float64 | 1338 | Continuous |
| Children | This is an integer indicating the number of children / dependents covered by the insurance plan. | Int64 | 1338 | Discrete |
| Smoker | This is yes or no depending on whether the insured regularly smokes tobacco. | object | 1338 | Categorical |
| Region | This is the beneficiary's place of residence in the U.S., divided into four geographic regions - northeast, southeast, southwest, or northwest. | object | 1338 | Categorical |
| Charges | Individual medical costs billed to health insurance | float64 | 1338 | Continuous |

**Questions**

Do smokers incur a disproportional cost compared to non-smokers?

Does gender play a role in overall BMI?

Do people with higher BMI contribute a disproportional cost compared to those who have a lower BMI?

Does age influence medical insurance claim utilization?

Does the occurrence of multiple dependents influence medical insurance claim utilization?

Can patients be identified and categorized as high risk by preforming an analysis on insurance utilization?

**Objective Hypothesis**

1. **Prove (or disprove) that the medical claims made by the people who smoke is greater than those who do not?**

H0:μ1<=μ2 The average charges of smokers is less than or equal to nonsmokers

Ha:μ1>μ2 The average charges of smokers is greater than nonsmokers

*Standard deviation of the population is not known, so a T-stat test will be performed . The > sign in the alternate hypothesis indicates the test is right tailed, that is, all z values that would cause us to reject null hypothesis are in just one tail to the right of the sampling distribution curve.*

1. **Prove (or disprove) with statistical evidence that the BMI of females is different from that of males.**

Where μ1/μ2 is the respective population means for BMI of males and BMI of females

H0:μ1−μ2=0 There is no difference between the BMI of Male and BMI of female.

Ha:μ1−μ2!=0 There is difference between the BMI of Male and BMI of female.

*Standard deviation of the population is not known, so a T-stat test will be performed. Not equal to sign in alternate hypothesis indicate it is a two tailed test.*

1. **Is the proportion of smokers significantly different across different regions?**

H0: Smokers proportions are not significantly different across different regions.

Ha: Smokers proportions are different across different regions.

*Here we are comparing two different categorical variables, smoker, and regions. A Chi-square Test will be performed.*

1. **Is the mean BMI of women with no children, one child, and two children the same?**

H0: μ1 = μ2 = μ3 The mean BMI of women with no children , one child , and two children is same

Ha: At least one of mean BMI women is not same

*A One-way ANOVA test will be performed - Equality of population through variances of samples.*

**Possible Impact of Findings**

Provide frameworks for insurance companies to identify high risk patients and create programs that aim to reduce costs and prolong lives.

Examples of programs that could be implemented:

Smoking cessation – reduction of active smokers

Dietary needs – regimented diet and exercise to lower BMI

Geriatric care – identify common ailments/injuries at critical junctures during a patients lifespan in order to provide preventative measures