**Context**

Information about patients' health and how they receive medical care is regularly gathered from various places. This information is now more often used to help doctors make decisions about treatment. Hospitals and health groups are using electronic health records, lists of patients, and data from mobile devices to learn more about patterns and results, which helps them provide better care.

By using information and technology, both the healthcare system and the social support groups have a chance to make health care better. They can make their work to address health-related social problems a normal part of giving medical care. This can make their efforts to help people with health needs work better, be more useful, and last longer.

**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 people who smoke end up paying more for their health costs than those who don't smoke?

Does being male or female affect how much someone weighs overall?

Do individuals with higher weight cause higher costs compared to those who weigh less?

Does how old someone is affect how much they use their health insurance?

Does having more family members to cover on insurance affect how much insurance is used?

Can we find out if certain patients are more likely to have health risks by looking at how often they use their insurance?

**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 Smokers, on average, are charged less than or equal to nonsmokers

Ha:μ1>μ2 Smokers, on average, are charged greater than smokers

*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 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, regardless of children, is the 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 Findings for users/stakeholders**

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 – less active smokers

Dietary needs – proper diet and exercise to lower BMI

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