

STROKE PREDICTION

GROUP MEMBERS

1. ISHIKA SHETTY: SS2021469
2. PRIYANKA KHOT: SS2021449
3. PRANJAL SHINDE: SS2021470
4. DHIRAJ POOJARY: SS2021480
5. PRIYANKA MANDYA: SS2021462
6. SHREYASH MHASKAR: SS2021456

OBJECTIVES:

To find the factors which lead to stroke.

RESEARCH METHODOLOGY:

We have analyzed these predictions using chi-square test for independence of attributes, significance of two population means i.e. t-test, two-way ANOVA and regression analysis.

Data source: - [kaggle](#).

The sample was of 4909 people

TESTING:

CHI SQUARE

HYPERTENSION & STROKE

<u>HYPOTHESIS:</u> H ₀ : - variables are independent H ₁ : - variables are dependent	<u>DECISION CRITERIA:</u> $\chi^2_{cal} > \chi^2_{tab}, P < \alpha$ $\alpha=0.05$ i.e Reject H ₀ $\chi^2_{tab}= 3.84$
<u>TEST STATISTICS:</u> $\chi^2_{cal} =97.275$ df=1 Pvalue <2.2e-16	<u>RESULT:</u> $\chi^2_{cal} > \chi^2_{tab}$ So REJECT H ₀ Therefore, hypertension and stroke are dependent.

GENDER & STROKE

<u>HYPOTHESIS :</u> H ₀ :- variables are independent H ₁ :- variables are dependent	<u>DECISION CRITERIA :</u> $\chi^2_{cal} > \chi^2_{tab} , P > \alpha$ $\alpha=0.05$ i.e Accept H ₀ $\chi^2_{tab}= 3.84$
<u>TEST STATISTICS :</u> $\chi^2_{cal} =0.16657$ df=1 Pvalue=0.6832	<u>RESULT :</u> $\chi^2_{cal} \nless \chi^2_{tab}$ So ACCEPT H ₀ Therefore, gender and stroke are independent.

t-TEST:

GENDER & BMI MEAN

<u>HYPOTHESIS:</u> H ₀ : -mean of variables are same H ₁ : -not H ₀	<u>DECISION CRITERIA:</u> $t_{cal} \not> t_{tab}$, $p > \alpha$ $\alpha=0.05$ i.e Accept H ₀ t _{tab} =1.960
<u>TEST STATISTICS:</u> t _{cal} =1.6304 df=4499 Pvalue=0.1031	<u>RESULT:</u> $t_{cal} < t_{tab}$ So ACCEPT H ₀ Therefore, the mean BMI of genders are same.

GENDER AND AVERAGE GLUCOSE LEVEL MEAN

<u>HYPOTHESIS:</u> H ₀ : - mean of variables are same H ₁ : - not H ₀	<u>DECISION CRITERIA:</u> $t_{cal} < t_{tab}$, $p > \alpha$ $\alpha=0.05$ i.e Accept H ₀ t _{tab} =1.960
<u>TEST STATISTICS:</u> t _{cal} =0 df=4996 Pvalue =1	<u>RESULT:</u> $t_{cal} < t_{tab}$ So ACCEPT H ₀ Therefore, the mean of average glucose level are same.

TWO WAY ANOVA

FACTORS

2 independent variables are: working class, smoking status
Dependent continuous variable is average glucose level.

Hypothesis:

Null hypothesis 1: all working class have same average glucose level v/s alternatively: not
null hypothesis 1

Null hypothesis 2: average glucose level is same for all smoking status v/s alternatively:
not null hypothesis 2

Decision criteria:

As $F_{cal} > F_{tab}$ i.e we REJECT H_0

RESULT:

- ✧ Self employed people have the highest average glucose rate.
- ✧ And the people who depend on their children for income have the minimum average glucose rate.

REGRESSION ANALYSIS:

Two variables are Age & Avg Glucose Level

$$r=0.2358382$$

Therefore, Age & Avg glucose level are positively correlated.

Two variables are Age & BMI

$$r=0.333398$$

Therefore, Age & BMI are positively correlated.

Two variables are BMI & Avg Glucose Level

$$r=0.1755022$$

Therefore, BMI & Avg Glucose Level are positively correlated.

CONCLUSION:

- 1) The insights from chi square test of independence of attributes are: -
 - i) Hypertension and stroke are dependent as it is observed that $\chi^2_{cal} > \chi^2_{tab}$. Therefore, hypertension can be one of the causes for stroke.
 - ii) The next two attributes considered for chi square test of independence are gender and stroke and it is observed that they are independent as $\chi^2_{cal} \not> \chi^2_{tab}$. Therefore, two attributes are independent.

2) The results obtained from significance of two population means i.e is t-test are as follows:

- i) Two populations considered are BMI and gender and after testing it was observed that mean BMI in male and female are same as $t_{cal} \not> t_{tab}$.
- ii) The second two populations considered for the t test were against gender and average glucose levels and the result it was found to be that mean average glucose levels in males and females are same as the $t_{cal} < t_{tab}$.

3)The next technique used is two-way ANOVA with factor A as working class and factor B smoking status and the result observed was that different working classes have different level of glucose and people with different smoking status have different level of glucose.

4) The regression analysis was carried out on age average glucose level and BMI the result observed from the correlation coefficient was that age average glucose level and BMI are positively correlated but share a weak correlation.

5) Therefore, with our analysis of the data we came to a conclusion that smoking, hypertension working class, residence type and heart disease are one of the reasons of the stroke.
