Causes of Stroke

https://www.medicinenet.com/stroke_symptoms_and_treatment/article.htm

https://www.hopkinsmedicine.org/health/conditions-and-diseases/stroke/risk-factors-for-stroke

1. Background

1.1 Background

According to the World Health Organization (WHO), stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths. Generally, a stroke is a medical emergency. The affected individuals, family, friends, or bystanders need to access emergency care immediately since they will get different permanent sequelae. There are two main causes of stroke including a blocked artery(ischemic stroke) and the leaking or bursting of a blood vessel (hemorrhagic stroke). Some people may have only a temporary disruption of blood flow to the brain, known as a transient ischemic attack (TIA), which doesn't cause lasting symptoms. Although anyone can have a stroke at any age, the chance of getting stroke increases if people have certain risk factors including high blood pressure, heart disease, smoking, high cholesterol level, and others.

1.2 Definition

Stroke is a medical condition in which poor blood flow to the brain causes cell death.

Body Mass Index (BMI) is a person's weight in kilograms divided by the square of height in meters. A high BMI can be an indicator of high body fatness. A BMI of between 18.5 and 24.9 is ideal. A BMI of between 25 and 29.9 is overweight.

Normal blood glucose level: To be considered "normal," fasting glucose must be under 100 mg/dl. Moreover, nondiabetic people should have a glucose level of no higher than 140 mg/dl after meals, and glucose should return to pre-meal levels within 2-3 hours.

2. Objectives

The objective of this project is to find out whether a patient is likely to get a stroke based on the input parameters using data visualization methods. I select seven factors and some characteristic of patients that may cause stroke and test whether they are the main causes of stroke:

1) Relationship between gender and stroke
2) Relationship between marriage and stroke
3) Relationship between residence type and stroke
4) Relationship between work type and stroke
5) Relationship between age and stroke
6) Relationship between glucose level and stroke
7) Relationship between BMI and stroke
8) Characteristics of the patients who have stroke

3. Main Results (Data Visualization)

Here is the main result of the dataset on the problem. With Python, I have plot some figures to

see the following factor:

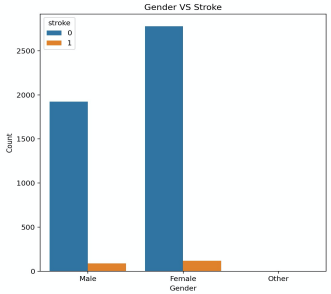


Fig (1)

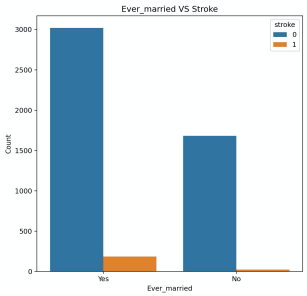


Fig (2)

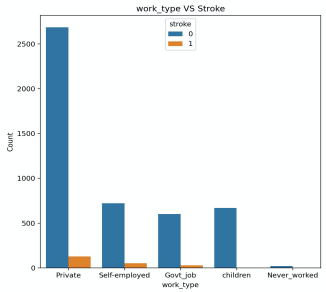


Fig (3)

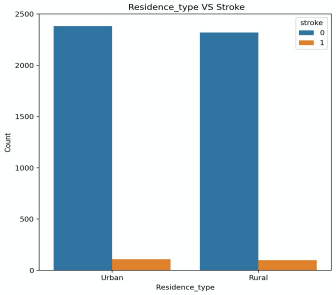
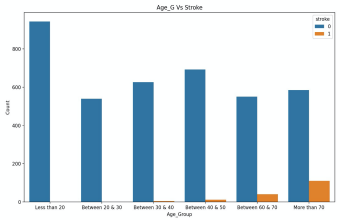


Fig (4)



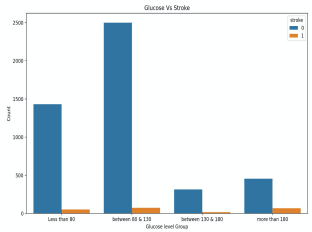


Fig (5)

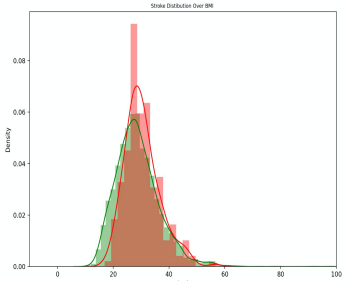


Fig (7)

Fig (6)

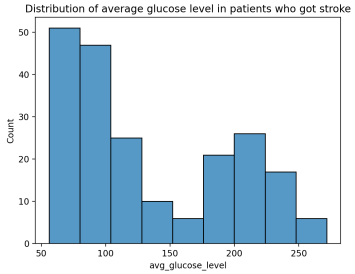


Fig (8)

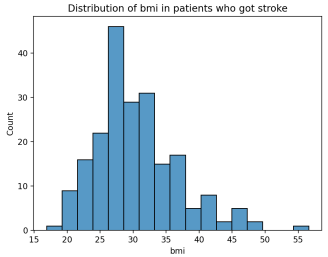


Fig (9)

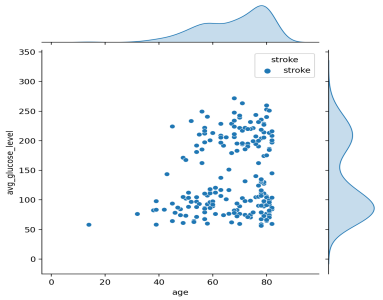


Fig (10)

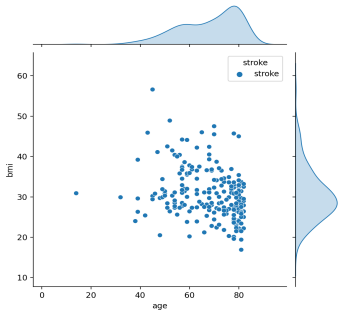


Fig (11)

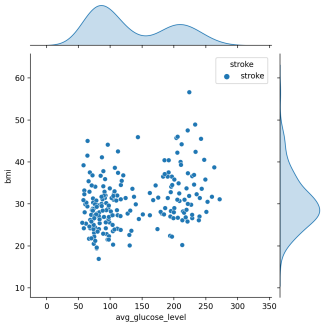


Fig (12)

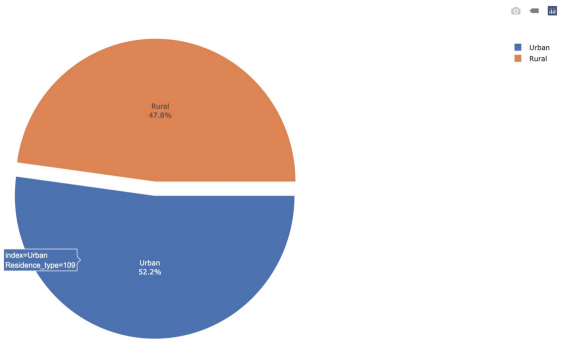


Fig (13)

4. Interpretation, Comparison, and Discussions

Here are the findings after plotting the graphs:

1) Compared to male participants, the number of female participants suffering from stroke is significantly higher.

2) Compared to participants who have never been married, the number of participants who have been married and suffering from stroke is obviously higher.

3) In terms of the number of strokes, there is no significant gap between participants who live in an urban place and participants living in a rural place.

4) The number of participants whose work type is Private and who have never got a stroke is particularly high.

5) At the age of 0-30, almost none of the participants are suffering from a stroke. At the age of 30-70, the number of strokes has continued to rise with the growth of age. And among them, the number of stroke patients over the age of 70 is the most.

6) Participants with a Body Mass Index(BMI) of over 25 are 2-3 times more likely to suffer from stroke compared to those with a normal weight (with BMI less than 25)

7) People with higher average blood glucose levels are more likely to get stroke. However, the graph cannot show it clearly due to the limitations.

Moreover, I have visualized the numerical data with the count of the patients who had strokes. From the related figures, they show:

(1) The older patients who had strokes before have higher average blood glucose level and BMI.

(2) There is no significant relationship between the BMI and average blood glucose level of the patients since the graph is dispersed.

It seems the main cause of stroke for most patients is due to the age factor; however, it may not be the main reason, instead, high average blood glucose level and BMI are the major causes of stroke. Although it can not be shown clearly due to some limitations of the data source, nemours studies have been proved that glucose levels would progressively increase with age (Ko et al., 2006). Also, older people tend to have higher BMI due to loss of muscles (Katie, 2020).

Although most of the results correspond with those previous studies that some features will significantly lead to stroke, this study challenges the previous studies that females are easier to get strokes (UT Southwestern Medical Center, 2016). Nonetheless, it may be due to the potential limitations of this study, and this seems to be the reason that Ihave some contrasts to previous studies.

As mentioned above, there are some limitations in the dataset, which might affect the results. First, the number of female participants and male participants is unevenly distributed. The number of female participants is significantly more than that of male participants (about 1000 people). In this case, it seems to be unfair to directly compare the relationship between gender and stroke. Second, the dataset has only 4.3% of stroke cases, which is a case of highly imbalanced data. Therefore, it would make the dataset highly biased and inaccurate.

5 . Conclusions

After the statistical analysis, the conclusions corresponding to the hypothesis are deducted as followings:

1) There is a positive relationship between age and strokes older people are relatively easier to get a stroke.
2) There is a relationship between average glucose level and stroke it is easier to get strokes when the average glucose level is higher than 100 mg/dL.
3) There is a relationship between BMI and stroke it is easier to get strokes when BMI is higher than 24.9.
4) There is a relationship between age and BMI that more older people used to have relatively higher BMI.
5) There is a relationship between age and average glucose level that more older people used to have relatively higher average glucose levels.
6) There is a relationship between marriage and stroke. People who have never been married are more likely to get strokes. It can be explained by most people who have ever been married would be older compared to those who haven't been married.
7) There is a relationship between gender and stroke more males get a stroke compared to females.
8) There is no significant effect of residence type and work type on stroke.

As stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths, it is important to understand how to prevent stroke. Through analysis of data, this study established the relationship between different features and stroke. I found that (1) there is no significant effect of residence type and work type on stroke; (2) age could be the effect on getting stroke, as older people are easier to have relatively higher BMI and average glucose levels while both of them would lead to stroke. Therefore, future research into stroke might focus on how to prevent older people/elders from getting stroke.

Reference

Reeves, M. J., Bushnell, C. D., Howard, G., Gargano, J. W., Duncan, P. W., Lynch, G., Khatiwoda, A., & Lisabeth, L. (2008). Sex differences in stroke: epidemiology, clinical presentation, medical care, and outcomes. *The Lancet. Neurology*, *7*(10), 915 – 926. https://doi.org/10.1016/S1474-4422(08)70193-5

Kelly-Hayes M. (2010). Influence of age and health behaviors on stroke risk: lessons from longitudinal studies. *Journal of the American Geriatrics Society*, *58 Suppl 2*(Suppl 2), S325 – S328. https://doi.org/10.1111/j.1532-5415.2010.02915.x

Jenna, F. (2019, April 10). What is the link between diabetes and stroke? *Medical News Today*. https://www.medicalnewstoday.com/articles/324924

Ko, G. T., Wai, H. P., & Tang, J. S. (2006). Effects of age on plasma glucose levels in non-diabetic Hong Kong Chinese. *Croatian medical journal*, *47*(5), 709 – 713.

Kurth, T., Gaziano, J. M., Berger, K., Kase, C. S., Rexrode, K. M., Cook, N. R., Buring, J. E., & Manson, J. E. (2002). Body mass index and the risk of stroke in men. *Archives of internal medicine*, *162*(22), 2557 – 2562. https://doi.org/10.1001/archinte.162.22.2557

UT Southwestern Medical Center. (2016, May 9). *Why are women at higher risk for stroke than men?* https://utswmed.org/medblog/stroke-symptoms-women-risk/

Yvette, B. (2020, January 18). How much should I weigh for my height and age? *Medical News Today*. https://www.medicalnewstoday.com/articles/323446

Casey, M., & Chimene, R. (n.d.). What should your glucose levels be? Here's the ultimate guide to healthy blood sugar ranges. Levels. https://www.levelshealth.com/blog/what-should-my-glucose-levels-be-ultimate-guide

Katie, D. (2020, May 21). *BMI in the Elderly: What You Need to Know*. The Geriatric Dietitian. https://www.thegeriatricdietitian.com/bmi-in-the-elderly/#Sarcopenia_in_Older_Adults

The National Heart, Lung, and Blood Institute. (n.d.). *Stroke*. Retrieved April 23, 2021, from https://www.nhlbi.nih.gov/health-topics/stroke

Appendix

Sample Statistics used in hypothesis Testing:

Cleaned dataset:

https://drive.google.com/file/d/1fDjCE3gftK3WyeshsFR8ZqIfsOAiFrBO/view?usp=sharing

Code Example

```
import pandas as pd
import numpy as np
import plotly.express as px
import seaborn as sns
import matplotlib.pyplot as plt
import csv

df=pd.read_csv("CleanedHealth.csv")
df.info()

df=df.dropna(how='any')
df.drop("id", axis=1, inplace=True)

segment = []
plt.figure(figsize=(10,5))
plt.title("Stroke count")
g = plt.pie(df.stroke.value_counts(), labels=df.stroke.value_counts().index,autopct='%1.1f%%', startangle=180)

plt.show()
for i in df['age'].values:
if (i < 20):
segment.append('Less than 20')
if (i >= 20) & (i<30):
segment.append('Between 20 & 30')
if (i >= 30) & (i<40):
segment.append('Between 30 & 40')
if (i >= 40) & (i<50):
segment.append('Between 40 & 50')
if (i >= 50) & (i<60):
segment.append('Between 50 & 60')
if (i >= 60) & (i<70):
segment.append('Between 60 & 70')
if (i >= 70):
segment.append('More than 70')

df['age_segment']=segment

segment2 = []

for i in df['avg_glucose_level'].values:
if (i < 80):
segment2.append('Less than 80')
if (i >= 80) & (i<130):
segment2.append('between 80 & 130')
if (i >= 130) & (i<180):
segment2.append('between 130 & 180')
if (i >= 180):
segment2.append('more than 180')

df['glucose_seg']= segment2


df['glucose_seg'] = pd.Categorical(df.glucose_seg,
categories=["Less than 80","between 80 & 130","between 130 & 180","more than 180"],
ordered=True)
segment5 = []

for i in df['stroke'].values:
if (i == 1):
segment5.append('stroke')
if (i == 0):
segment5.append('no stroke')

df['stroke']=segment5
dfst=df[df['stroke']=='stroke']

dfnst=df[df['stroke']=='no stroke']

#below are some graph code examples
plt.figure(figsize=(12,7))
sns.countplot(x='work_type', hue='stroke', data=df)

plt.xlabel('worktype')
plt.ylabel('Count')
plt.title( "work_type Vs Stroke")
plt.show()

plt.figure(figsize=(12,10))

sns.jointplot(x=dfst['avg_glucose_level'],y=dfst['bmi'] ,hue=dfst['stroke'])

plt.show()
sns.jointplot(x=dfst['age'],y=dfst['bmi'],hue=dfst['stroke'])
plt.show()

sns.jointplot(x=dfst['age'],y=dfst['avg_glucose_level'],hue=dfst['stroke'])
plt.show()
restgrp=dfst['Residence_type'].value_counts().reset_index()

fig = px.pie(restgrp, values="Residence_type", names="index",template="seaborn")

fig.update_traces(rotation=90, pull=0.05, textinfo="percent+label")
```

```
fig.show()

sns.histplot(dfst['bmi'])
plt.title( "Distribution of bmi in patients who got stroke")
plt.show()
```