

LETTERKENNY INSTITUTE OF TECHNOLOGY

ASSIGNMENT COVER SHEET

Lecturer's Name: Dr James Connolly

Assessment Title: _____

Work to be submitted to: Blackboard

Date for submission of work: _____

Place and time for submitting work: _____

To be completed by the Student

Student's Name: Danfeng Wang

Class: _____

Subject/Module: Data Science

Word Count (where applicable): _____

I confirm that the work submitted has been produced solely through my own efforts.

Student's signature: _____ Date: _____

Notes

Penalties: The total marks available for an assessment is reduced by 15% for work submitted up to one week late. The total marks available are reduced by 30% for work up to two weeks late. Assessment work received more than two weeks late will receive a mark of zero. [Incidents of alleged plagiarism and cheating are dealt with in accordance with the Institute's Assessment Regulations.]

Plagiarism: Presenting the ideas etc. of someone else without proper acknowledgement (see section L1 paragraph 8).

Cheating: The use of unauthorised material in a test, exam etc., unauthorised access to test matter, unauthorised collusion, dishonest behaviour in respect of assessments, and deliberate plagiarism (see section L1 paragraph 8).

Continuous Assessment: For students repeating an examination, marks awarded for continuous assessment, shall normally be carried forward from the original examination to the repeat examination.

Table1 shows the variables that will be used to answer the hypothesis test1

Table 1: Description of variables required for test1					
Variable Name	description	Variable Type		Convert to	comments
Gender	PDF 中的描述	Chr	Categorical Dichotomous	factor	
Age		Num			
Ever_married		Chr		Factor	
Bmi		Num		Factor	范围
Date		Chr		Date	
avg_glucose_level		Num			
Hypertension		Int		Factor	
Stroke		Int		Factor	Yes/No

Questions

1. Men and woman differ in getting stroke
H0 - Men and woman do not differ in getting stroke
H1 – Men and woman differ in getting stroke
on the average, the mean age of stroke is 67.73.
H0 – the mean age of man equals to 67.73
H1 – the mean age of man not equals to 67.73
2. smoking has correlation with stroke
3. age has correlation with stroke
4. bmi has correlation with stroke
5. avg_glucose_level has correlation with stroke

Abstract

We calculated the stroke patients by gender, and compare the outcome by using a t test for continuous variable and chi-square test for categorical variables. All hypothesis tests in this reports are two-sides test, and the significant level is 0.05.

这部分最后写，包括 – 检验什么问题，结果是什么，统计方法的简短描述，描述最后的结果包括重要的数据指标比如 p-value 以及含义，表明检验结果说明了什么

The contribution to the stroke of bmi, hypertension, gender are described in this report. These are all risk factors to stroke. Our goal is to find the patterns of health habit and the risk of getting stroke. We test the correlations between blood pressure, diabetes, BMI and stroke separately with chi-square test of independence. The dependent variable (DV) is stroke which is categorical variable.

Data is collected from from 2015 to 2021.

This data set consists of xx rows and xx columns. For patients records from 2015 to 2021, the following variables are reported: gender (Male, Female),It contains categorical variables and continuous variables.

We evaluated the 5000 patients, 50% is woman, 50% is men. Men are younger than woman (数字范围均值+最大值) , 3% of the men has smoke history, 50% of the woman has smoke history. Men had more stroke patients than woman (p-value = 0.4,)

Overall, there is a significant age difference between stroke and non-stroke patients groups. The mean age of stroke group patients is xx which is higher than that of the non-stroke group. So that age is a risk factor contributes to stroke.

Research question

1. Is there relationship between age and stroke?
2. Are the mean ages in stroke and non-stroke patients group different?

Is there relationship with average glucose level and BMI?

Is there relationship between average blood glucose level and stroke?

Is there relationship between obesity and stroke?

Is there relationship between gender and stroke?

1. Whether there is correlation between gender and stroke.
2. Whether there is correlation between smoking and stroke.
3. Whether there is correlation between age and stroke
4. Whether there is correlation between average glucose level and stroke
5. Whether there is correlation between smoking and stroke

Data preparation

Dataset description

The dataset being used in this report is stroke patients' information from 2015 to 2021 (See below table). From this table, we can see the data describes the patient information from aspects of basic information, health status, life styles. The dataset is stored as csv format, before using it, we are going to clean and transform it to make it prepared for the next analysis.

No.	Name	Description
1	Id	Unique id
2	Gender	Male / Female / Other
3	Age	Age of the patient
4	Hypertension	Has hypertension or not (1/0)
5	Heart disease	Has heart disease or not (1/0)
6	Ever married	Yes / No
7	Work type	Children / Gov_jov / never_worked / Private / Sel-employed
8	Residence type	Rural / Urban
9	Avg glucose level	Average blood glucose level
10	BMI	Body mass index
11	Smoking status	Formerly smoked / never smoked / smokes / Unknown
12	Stroke	The patient has had a stroke or not (1/0)
13	Date	Record date

Table 1 Dataset description

Data clean and transformation

The purpose of this stage is to make sure the dataset for analysis is tidy and structured, and prepared for investigate our research questions. The relevant steps will be involved to do this work. After finishing the below steps, descriptive analysis will be applied to give an initial view about the variables correlations.

Step1 – load the dataset from csv, and store each variable into suitable types. Below table is the type conversion plan. The hypertension, heart disease variables' values are 0 or 1. 0 means patient has no such disease, and 1 has the opposite meaning. In order to know clearly what the data represents, we convert them to *No* and *Yes* accordingly when converting them to factor type.

Table 2 Type conversion

No.	Name	Target type and format
1	Gender	Factor
2	Hypertension	Factor with Yes/No labels
3	Heart_disease	Factor with Yes/No labels
4	Ever_married	Factor
5	Work_type	Factor
6	Residence_type	Factor
7	Smoking_status	Factor
8	Stroke	Factor with Yes/No labels
9	Date	Date with format yyyy-mm-dd

The other four columns, which has not been listed in the above table includes id, age, avg_glucose_level, bmi, have been loaded as numeric.

After conversion, we check the structure and sample data to make sure all of the variables have been loaded into the corresponding types and the total amount of the data is 5110 rows with 13 columns.

Step2 – process missing value. Before dealing with the missing data, we calculate and plot the missing values of the dataset first. It can be seen that only bmi variable has 201 rows of missing values. We remove it.

Step3 – process outliers. From the dataset introduction and looking at the summary of the data, we have known that there are patients with *Other* gender, and smoking status with *Unknown* type. First, we use summary function to give an overview of the data, and filter out these outliers by subset function. Re-check the data summary to make sure the outliers have been removed. Some factor values are no longer useful because of the filter operation, we remove them by reconverting them to char and to factor type again.

Step4 – get subset. The analysis of this report focus on the risk factors of stroke, *id* and *Date* is useless in the data investigation. We remove both columns in the final dataset.

Descriptive statistics

Now the dataset has been prepared. We plot the correlation between the variables. The variables, by their content nature, cover life-style related variables (gender, age, marital status, work type, residence type), and health status related variables (hypertension, heart disease, average glucose value).

Assumption of age dependent variable

We assume age independent variable is normally distributed, and the age of all patients on both stroke group and non-stroke group is normally distributed. In addition, we assume stroke patients are older than that of the non-stroke patients.

Table 3 age distribution by stroke

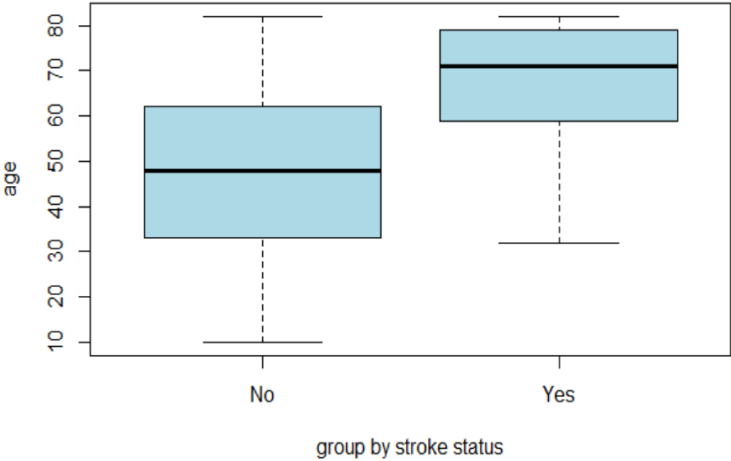
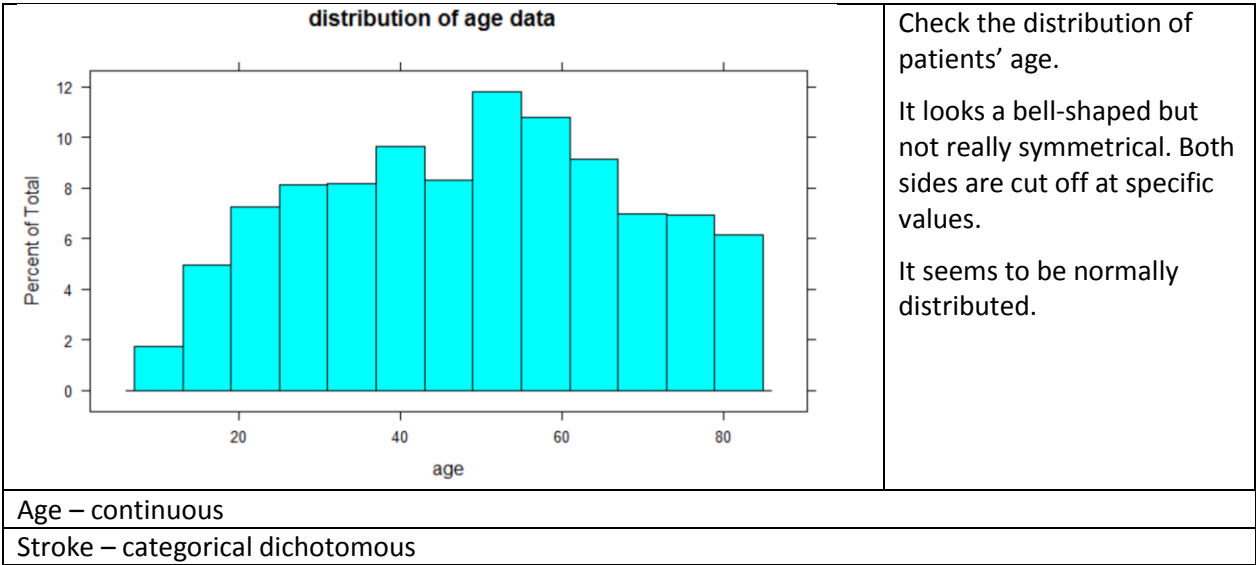
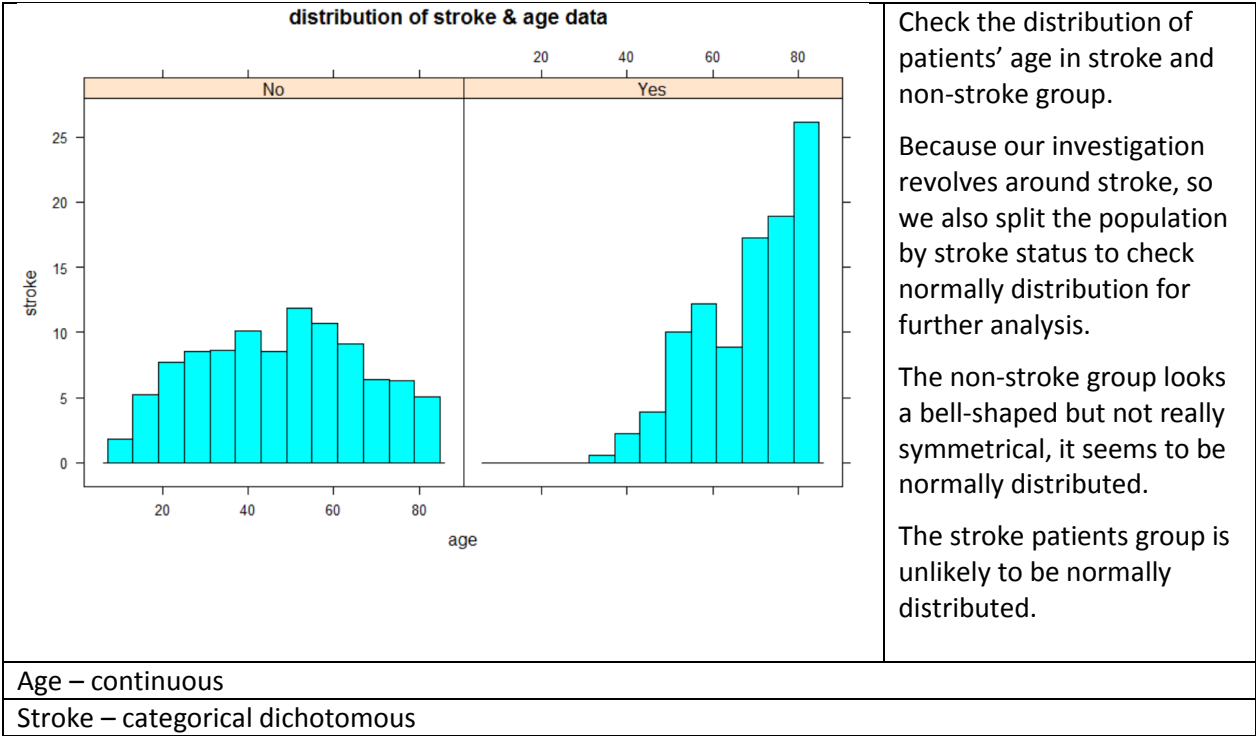
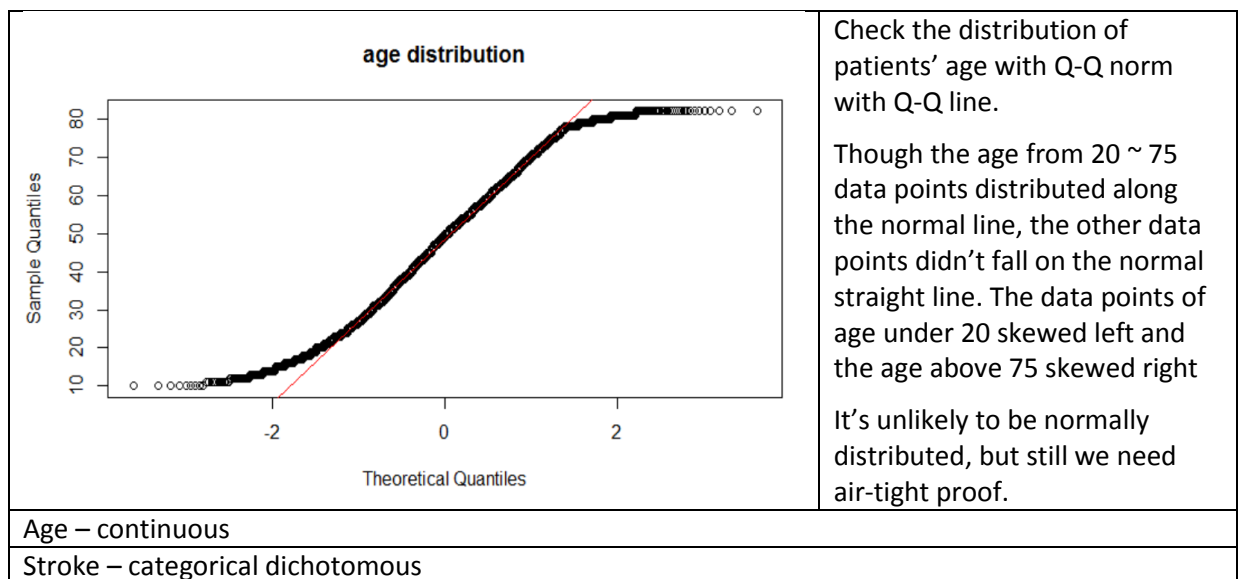
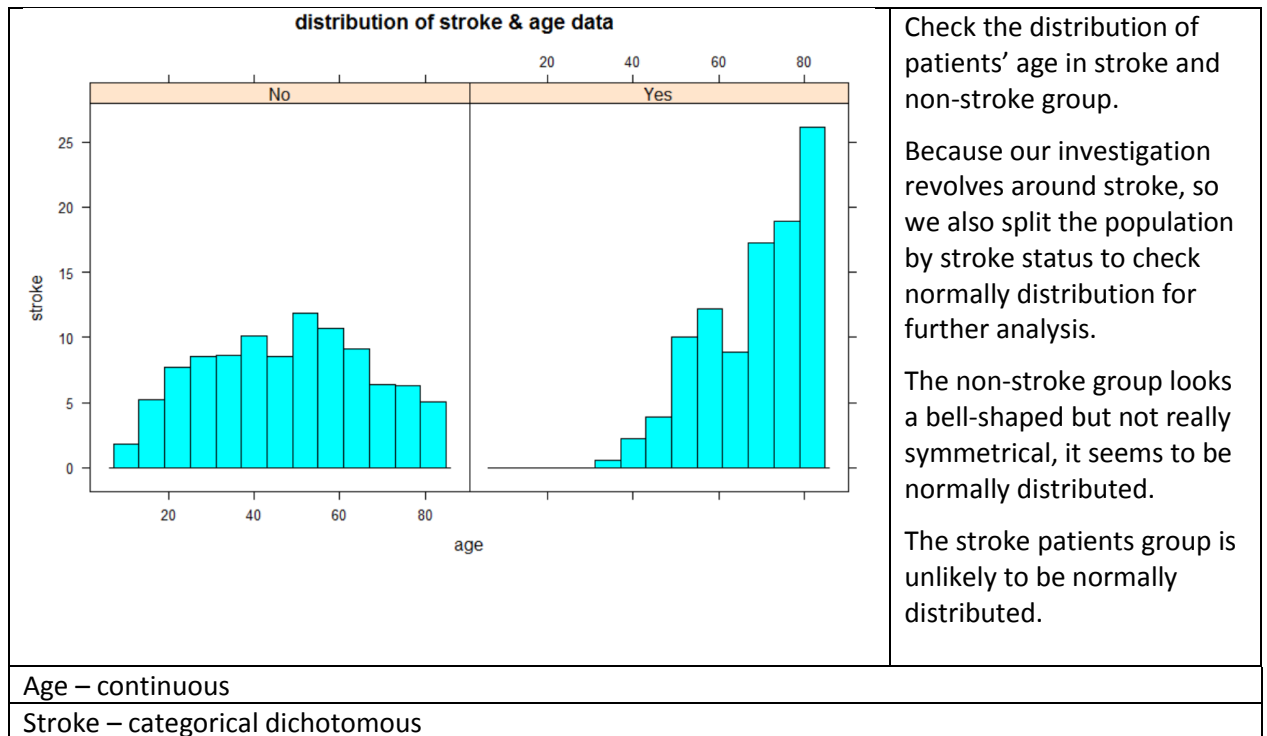
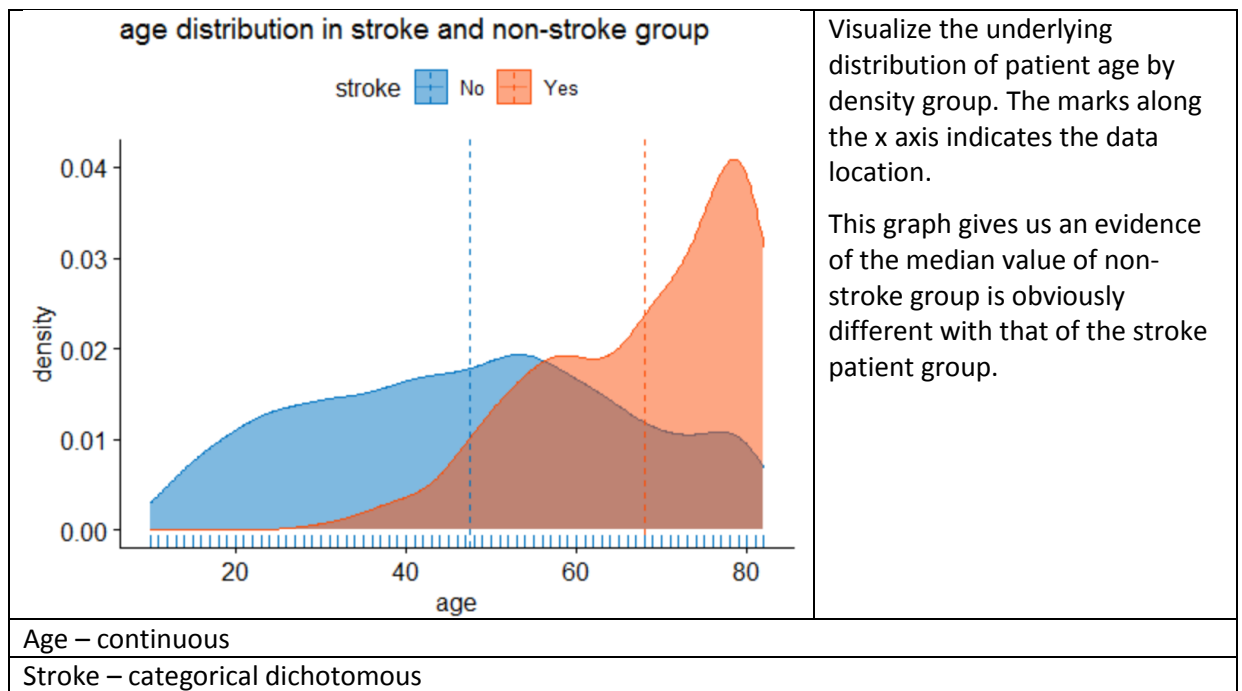
<p style="text-align: center;">stroke status vs age</p>  <p style="text-align: center;">group by stroke status</p>	<p>Check the distribution of patients' age.</p> <p>The median age of stoke group are higher than that of the non-stroke group. We calculate the accurate mean value of both group, the result shows that non-stroke patients is approximately at 47.57 of mean age, while the mean age of stroke group is 68.05.</p> <p>It looks a bell-shaped but not really symmetrical. Both sides are cut off at specific values.</p> <p>It seems to be normally distributed.</p>
Age – continuous	
Stroke – categorical dichotomous	

Table 4 age normality check by stroke







We did Shapiro-Wilk test in order to get a precise answer of whether the age is normally distributed in the entire population. The null hypothesis of Shapiro-Wilk test is that *a variable is normally distributed in some population*. If $p\text{-value} < 0.05$, the null hypothesis is rejected.

According to the test result, the p-value of non-stroke group, stroke group and entire population is 6.236922e-23, 4.133363e-09, 8.500619e-25 respectively, which are all far smaller than 0.05. In this case, we conclude that age is not normally distributed in our dataset.

Assumption of average glucose variable

We assume that *average glucose level* variable is normally distributed. and the *average glucose level* of all patients on both stroke group and non-stroke group is normally distributed.

We use boxplots of ggplot2 to see the outliers for stroke patient group and non-stroke-patient group. The middle line of the boxplots indicates both stroke and non-stroke patients group have similar median values (90-110). Non-stroke patients group has many outliers, while the stroke patients group has no outliers.

Side-by-side boxplots are provided by ggplot2. The boxplots below seem to indicate one outlier for treatment group C and D. Furthermore, both the mean (circle with +) and median (middle line) values are at the 75th percentile. This indicates that the data is highly skewed by the effects of the outlier(s).

用 descriptive statistics techniques 分析。描述过程，用数据和图表 discuss。直到 dataset prepared

Hypothesis testing

Question 1

Is there relationship between age and stroke?

H0 – age has no correlation with stroke

H1 – age has a correlation with stroke

Many diseases have significant different in age, gender, and other group identity. We use this hypothesis to test whether there is correlation between patients age and stroke. The dependent variable of this test is categorical dichotomous which should apply Chi-squared test regardless the independent variable is normally distributed or not.

Question 2

Is the age in stroke patient group and non-stroke patient group different?

H0 – age in stroke group and non-stroke group are identical

H1 – age in stroke group and non-stroke group are non-identical

From the boxplot and density plot of the above analysis, we see the distribution of patients age in stroke and non-stroke group look different, and the stroke group has higher mean value than that of the non-stroke patient group. We make this hypothesis to test whether there is a difference influence between population's age and stroke

Both stroke patients and non-stroke patients are independent, which are the patients who has stroke or not respectively, and do not affect each other.

With Man-Whitney test, we can decide the age distribution in both groups are different without having them being normally distributed.

Question 3

Is there relationship between average blood glucose level and BMI?

H0 – average glucose level has no correlation with BMI.

H1 – average glucose level has a correlation with BMI.

Average glucose level is the key indicator of diabetes. Diabetes used to be believed a risk factor of other diseases such as heart diseases¹. We use the above hypothesis to prove this relationship.

Question 3

Is there relationship between obesity and stroke?

H0 – obesity has no correlation with stroke

H1 – obesity has a correlation with stroke

Question 5

Is there relationship between gender and stroke?

H0 – gender has no correlation with stroke.

H1 – gender has relationship with stroke.

Describe the questions in hypothesis test.

Describe the relationship between variables.

Explain how each of the hypothesis test will be valuable when answering each question. 参照问题定义

Statistical methods

Using the hypothesis test to describe examine each test.

Describe in detail each statistical test by examining the following characteristics

- The structure of the data variables I will examine
 - How selected variables enable me to answer the question
 - Any assumptions you are making about your data variables such as **normality**
-

¹ Whiteley, L., Padmanabhan, S., Hole, D. and Isles, C., 2005. Should diabetes be considered a coronary heart disease risk equivalent?: results from 25 years of follow-up in the Renfrew and Paisley survey. *Diabetes Care*, 28(7), pp.1588-1593.

Test 1

The below frequency table summarizes two variables from the **stroke** dataset: gender and stroke.

Variable Name	Description	Type
Gender	Gender of the patient	Categorical independent variable with 2 levels labelled with Male, Female
Stroke	Whether the patient has had a stroke in the past or not	Categorical dichotomous

H0 – there is no correlation between gender and stroke

H1 – there is correlation between gender and stroke.

There exists gender bias with some diseases. Before making hypothesis test of the correlation of gender and stroke status, we see the data summation first from these two dimensions. Below table shows the number of patients by gender by stroke. Percentage of female patients had stroke in this dataset, is approximately 4.14% (120/2897), and the according percentage of man is about 4.25% (209/4908).

Number of patients by gender by stroke			
Gender	Non-stroke	Stroke	Total
Female	2777	120	2897
Male	1922	89	2011
Total	4699	209	4908

Judging from the percentage difference between male who got stroke and that of the female group, it appears that there is a slight difference between gender and stroke. However, in order to prove it, we need to do hypothesis test.

We are going to test whether there is correlation between gender and the stroke based on this dataset. Gender is a typical categorical variable, the value of gender consists of Male and Female. Because we found there are outliers in the gender variable in data preparation stage, we have removed the row with gender values are neither Male nor Female, and then convert gender from char to factor. Stroke in this dataset means whether the patient ever got stroke or not, we convert the variable into factor and label the original 0 and 1 value with No and Ye respectively.

Both gender and stroke are categorical variables. Gender is independent and stroke we take it as dependent variable. When both the dependent variable and independent variable are categorical, we use Chi-square test to do hypothesis testing.

The null hypothesis in chi-square is that there is no relationship between the independent variable and the dependent variable. So our test is defined as the above Table shows.

We choose Alpha = 0.05 in this test which means when these two variables have a 0.05 or less probability, there is no relationship between the two variables.

p-value of this test is 0.6805 which indicates weak evidence against the null hypothesis, so that there is no correlation between gender and stroke based on this patients dataset.

Test 2

Whether there is correlation between smoking and stroke.

H0 – the probability of stroke patients who are smokers equal to that of the non smokers.

H1 – the probability of stroke patients who are smokers not equal to that of the non smokers.

Variable Name	Description	Type	Target Type
Smoke	Whether is a smoker	Categorical dichotomous	Factor
Stroke	Whether the patient has had a stroke in the past or not	Categorical dichotomous	Factor

Number of patients by smoking status by stroke			
Smoking	Non-stroke	Stroke	Total
No	1768	84	1852
Yes	1477	96	1573
Total	3245	180	3425

Test 3

Whether there is correlation between age and stroke.

H0 – there is no correlation between age and stroke.

H1 – there is correlation between age and stroke

The dependent variable is age, which is continuous variable. Stroke is categorical dichotomous variable. The distribution of stroke is not normally distributed. The dependent variable is not normally distributed in this dataset, so that we use Kruskai-Wallis test of Non-parametric test.

These data samples are independent, and the samples do not affect each other. Using this test, we can decide if the population distribution is identical without assuming them to follow the normal distribution.

The p-value of the test is 2.2e-16 which is obviously less than the significant level 0.05, we can conclude that there are significant age differences between the patients of stroke and non-stroke groups.

Variable Name	Description	Type	Target Type
---------------	-------------	------	-------------

Age		Continuous	
Stroke	Whether the patient has had a stroke in the past or not	Categorical dichotomous	Factor

Question 4

Variable Name	Description	Type	Target Type
BMI	Body mass index	Continuous	
Stroke	Whether the patient has had a stroke in the past or not	Categorical dichotomous	Factor

Question 5

Variable Name	Description	Type	Target Type
BMI	Body mass index	Continuous	
Stroke	Whether the patient has had a stroke in the past or not	Categorical dichotomous	Factor

Result

用 R 对 variables 应用统计学方法回答问题。解释每个测试的输出。解释如何决定每个假设的

XX percent of the sample are woman, the mean age of the sample is xxx years, and woman is significantly older than men (55 years' vs 50 years). The two groups did not differ in health.

Recognition of stroke

Blood pressure, smoking, and obesity are risk factors of stroke. BMI in a good level (according to WHO) has lower risk in getting stroke.

The definition of obesity ($BMI > 30 \text{ kg/m}^2$)² is a risk factor of diabetes. It has been widely used in white population. In this dataset, the personal information does not include race, we cannot have an accurate level to define the obesity status which might lead to bias of our investigation.

Baseline clinical characteristics for men and women in each of the four groups are shown in [Table 1](#). In all, 112 men and 116 women had diabetes. The average age of men and women with diabetes only and CHD only was 55–56 years. Men with

² Chiu, M., Austin, P.C., Manuel, D.G., Shah, B.R. and Tu, J.V., 2011. Deriving ethnic-specific BMI cutoff points for assessing diabetes risk. *Diabetes care*, 34(8), pp.1741-1748.

diabetes only were less likely to be current smokers ($P < 0.001$) than subjects with CHD only. The other differences in risk profiles in these two subgroups were small and not statistically significant.

In all, 4,267 (60%) of the men and 3,746 (45%) of the women died during the 25 years of follow-up. There were 192 (84%) deaths in 228 subjects with diabetes and 2,016 (65%) deaths in 3,092 subjects with CHD. The highest mortality was in the group with both diabetes and CHD (37 of 38 or 100.2 deaths/1,000 person-years in men; 37 of 39 or 93.6 deaths/1,000 person-years in women). The lowest mortality was in the group with neither diabetes nor CHD (29.2 deaths/1,000 person-years in men, 19.4 deaths/1,000 person-years in women).

Men and women with diabetes only and with CHD only formed an intermediate risk group. Men with diabetes only had marginally higher mortality than men with CHD only (54.0 vs. 50.5 deaths per 1,000 person-years), whereas women with diabetes only appeared to have a considerably higher risk of vascular death than women with CHD only (46.7 vs. 29.2 deaths per 1,000 person-years) ([Fig. 1](#)). Similar trends were observed for each group of causes of death. Specifically, men with diabetes only had marginally higher CHD and other vascular mortality than men with CHD only, whereas women with diabetes only had higher CHD and other vascular mortality than women with CHD only ([Fig. 1](#)).

Overall survival over the course of 25 years is shown in [Fig. 2](#). This confirms the similarity in outcome between men with diabetes only and men with CHD only (log-rank $\chi^2 = 0.19$, $P = 0.664$) as well as the difference in outcome between women with diabetes only and women with CHD only (log-rank $\chi^2 = 8.54$, $P = 0.004$). Survival was least in men and women with both diabetes and CHD and greatest in those with neither ([Fig. 2](#)).

The similarity in men and the difference in women persisted after adjustments for age, smoking, hypertension, serum cholesterol, BMI, and social class ([Table 2](#)).

Adjusted hazard ratios (HRs) for CHD and all-cause mortality in men with diabetes only compared with men with CHD only were 1.17 (95% CI 0.78–1.74; $P = 0.450$) and 1.20 (0.92–1.56; $P = 0.172$), respectively. Corresponding HRs for women were 1.97 (1.27–3.08; $P = 0.003$) for CHD mortality and 1.80 (1.37–2.35; $P < 0.001$) for all-cause mortality. [Table 2](#) also shows HRs for the other covariates in the Cox model.

Increasing age, cigarette smoking, hypertension, and hyperlipidemia were all associated with CHD mortality. There were trends toward increased CHD mortality with increasing BMI, but these did not achieve statistical significance. Low social class predicted CHD mortality in women but not men. Increasing age, cigarette smoking, hypertension, and low social class were each associated with all-cause mortality in both sexes. Hyperlipidemia did not predict all-cause mortality in men or women, whereas BMI ≥ 35.0 kg/m² did ([Table 2](#)).

Conclusion

The p-value of the chi-squared tests is less than 0.05, we infer dependence. If it not less than 0.05, we failed to prove the dependence. The p-value of test 1 is roughly 0.00005, which is quite tiny. In other words, we have evidence that xxxxx and stroke are not independent. The

proportion of stroke patients in this dataset who is obesity is roughly half that of the non-stroke patient group.

The main finding of our study is that, after adjusting for known cardiovascular risk factors, middle-aged men and women with diabetes but no clinical evidence of CHD experience a lifetime vascular risk that is at least as high as (and for women, possibly higher than) that for subjects of similar age who have CHD but no diabetes. In this respect our data reinforce those of Haffner et al. (3) in their study of Finnish men and women. These authors compared the risk of fatal CHD in 890 subjects with type 2 diabetes without prior myocardial infarction and 69 with prior myocardial infarction but no diabetes and found an HR of 1.2 (95% CI 0.6–2.4). When comparing the two studies, however, one should note that Haffner et al. (3) followed their subjects for 7 years rather than 25 years, had only eight deaths in their patients with CHD but no diabetes, and were therefore unable to examine outcome in men and women separately.

Not all authors have reached the same conclusions. Although all recognize that type 2 diabetes is associated with high vascular and all-cause mortality, the belief that diabetes is a CHD risk equivalent is not supported as strongly as we expected, given the views and recommendations of the guideline writers. The Haffner study (3) and an analysis of registry data of patients hospitalized for unstable angina and non-Q-wave myocardial infarction in six different countries (15) both showed that people with diabetes but no previous cardiovascular disease had the same event rates as people with previous vascular disease only. By contrast, the cross-sectional and cohort studies from Tayside (4), the U.S. Physicians Health Study (16), the U.S. Nurses' Health Study (17), and an analysis of data from the Multiple Risk Factor Intervention Trial (MRFIT) (18) all concluded that patients with diabetes only carry lower vascular risks than those with CHD only. So why are there discrepancies in the results?

The studies differ in several respects including age range, sex, criteria for diagnosis of diabetes and CHD, and duration of diabetes. None of these seems likely to account for the discrepancies in results in our view, except for the duration of diabetes, which varied substantially. Studies comparing incident and prevalent cases of diabetes showed higher mortality risks for prevalent diabetes, presumably by virtue of their longer duration of follow-up (3,17). Hu et al. (17) explored this possibility further and were able to show a linear relationship between duration of diabetes and risk. Diabetes duration of <5 years carried a lesser risk of fatal CHD than previous myocardial infarction. Risks were equivalent after 10 years of diabetes and began to exceed those of previous myocardial infarction after 15 years of diabetes. Although we do not know when subjects in the Renfrew and Paisley Survey first developed their diabetes, the 25-year follow-up guaranteed a long duration of diabetes and may therefore be the main reason why our results support the view that diabetes is a CHD risk equivalent.

Studies such as ours have strengths and limitations. We do not know if there were differences between respondents and nonrespondents because we did not have permission to track the nonrespondents, although we believe that a 79% response rate means that subjects in the Renfrew and Paisley Survey were likely to have been representative of the general population from which they were drawn. The inclusion

of both sexes, the long duration of follow-up, and the large number of deaths are also strengths, as is the adjustment for the effects on outcome of six possible confounding variables including age, smoking habit, blood pressure, cholesterol, BMI, and social class.

Set against these strengths are a number of limitations. The diagnosis of diabetes was for the most part based on the response to the question, "Do you have or have you ever had diabetes?" Self-reporting of medical conditions such as diabetes is considered to be reliable (19), and other studies we reviewed also based their analyses on self-reported illness (16,17). We did not include any new cases of diabetes or CHD after screening. This is not so much a limitation as an acknowledgment that our findings reflect the risks associated with prevalent rather than incident diabetes. We were unable to distinguish type 1 from type 2 diabetes but suspect, given the age range of our cohort, that the majority had type 2 diabetes. Of greater potential concern is that we may have overestimated the risk of diabetes and underestimated the risk of CHD. Self-reporting might exaggerate the risks of diabetes by excluding milder cases. However, the prevalence of diabetes at recruitment (1972–1976) in the subgroup of subjects who had a random glucose measurement (74 of 4,702 or 1.5%) was not dissimilar from the self-reported prevalence of diabetes (190 of 14,039 or 1.3%). At the time of data collection, the random glucose cutoff for diagnosis of diabetes was generally accepted to be 14 mmol/l. A single random glucose of >11.1 mmol/l (the cutoff point used to extend the diabetes population in our analysis) does not on its own fully validate a diagnosis of diabetes but may be associated with a higher CHD risk than abnormal fasting glucose levels (20). The prevalence of type 2 diabetes has since increased to ~3% (21) as a result of increasing levels of obesity; however, this should not alter the ability of a diagnosis of diabetes made 25 years prior to predict outcome. We may in addition also have underestimated the risk of CHD, particularly in women, by relying on the Rose Angina Questionnaire. It is well known that a higher proportion of women with a typical history of angina do not have CHD (22). Even if we did underestimate CHD risk, however, it is unlikely that we did so to the extent that the CHD death rate in women with CHD only (9.8/1,000 person-years) would then have exceeded that in women with diabetes only (16.9/1,000 person-years). In conclusion, although there remain some differences in the detail, U.S. (23,24), European (25), and U.K. (26,27) guidelines now all recommend that type 2 diabetes be considered a CHD risk equivalent and that people with diabetes should be treated as if they already have vascular disease. The move to more aggressive therapies is supported by the results of the Heart Protection Study (5) and the Collaborative Atorvastatin Diabetes Study (6), which provide clear evidence of benefit from cholesterol lowering drug therapy in patients with diabetes only and no evidence of vascular disease. Two other statin trials containing subgroups of patients with type 2 diabetes failed to show benefit, although the Anglo Scandinavian Cardiac Outcomes Trial simply was not powered to do so (8) and the difference in serum cholesterol between treated and usual care groups in the Antihypertensive and Lipid Lowering Treatment to Prevent Heart Attack Trial was likely to have been insufficient for a difference in outcome to emerge (7). Our study shows that middle-aged men and women with diabetes but no clinical evidence of CHD experience a lifetime vascular risk that is at least as high as (and for women, possibly higher than) that for subjects of similar age who have CHD but no diabetes. These data support the view that

patients with type 2 diabetes who have not yet had a myocardial infarction or developed angina should be given the same advice and treatment as is currently recommended for patients with overt vascular disease to prevent vascular events.

讨论每个问题的分析，统计结果表明。对结果明细进行批判。提供深度讨论这个结果表明。