INTRODUCTION:

We have selected the brain stroke dataset, which is available on Kaggle and contains information on patients who have experienced a stroke. We aim to investigate the potential links between work status, hypertension, and glucose levels and the incidence of brain stroke. Additionally, we also want to examine whether age and average glucose level could potentially act as mediators in this relationship and provide insights on how these factors may be interrelated. The objective of our study is to identify any important factors that may contribute to the likelihood of brain stroke occurrence in the population.

Importing and viewing file:

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.2.3
```

```
brain_stroke <- read_excel("C:/Users/madhu/Downloads/brain_stroke.xlsx")
View(brain_stroke)
```

```
# viewing number of rows and columns in the data set
dim(brain_stroke)
```

```
## [1] 4981   11
```

```
head(brain_stroke)
```

```
## # A tibble: 6 x 11
##   gender   age hypertension heart_disease ever_married work_type  Residence_type
##   <chr>  <dbl>        <dbl>         <dbl> <chr>        <chr>      <chr>
## 1 Male      67            0             1 Yes          Private    Urban
## 2 Male      80            0             1 Yes          Private    Rural
## 3 Female    49            0             0 Yes          Private    Urban
## 4 Female    79            1             0 Yes          Self-empl~ Rural
## 5 Male      81            0             0 Yes          Private    Urban
## 6 Male      74            1             1 Yes          Private    Rural
## # i 4 more variables: avg_glucose_level <dbl>, bmi <dbl>, smoking_status <chr>,
## #   stroke <dbl>
```

```
#viewing data Categories of variables
str(brain_stroke)
```

```
## tibble [4,981 x 11] (S3: tbl_df/tbl/data.frame)
##  $ gender           : chr [1:4981] "Male" "Male" "Female" "Female" ...
##  $ age              : num [1:4981] 67 80 49 79 81 74 69 78 81 61 ...
##  $ hypertension     : num [1:4981] 0 0 0 1 0 1 0 0 1 0 ...
##  $ heart_disease    : num [1:4981] 1 1 0 0 0 1 0 0 0 1 ...
##  $ ever_married     : chr [1:4981] "Yes" "Yes" "Yes" "Yes" ...
##  $ work_type        : chr [1:4981] "Private" "Private" "Private" "Self-employed" ...
##  $ Residence_type   : chr [1:4981] "Urban" "Rural" "Urban" "Rural" ...
##  $ avg_glucose_level: num [1:4981] 229 106 171 174 186 ...
##  $ bmi              : num [1:4981] 36.6 32.5 34.4 24 29 27.4 22.8 24.2 29.7 36.8 ...
##  $ smoking_status   : chr [1:4981] "formerly smoked" "never smoked" "smokes" "never smoked" ...
##  $ stroke           : num [1:4981] 1 1 1 1 1 1 1 1 1 1 ...
```

Summary statistics:

```
summary(brain_stroke)
```

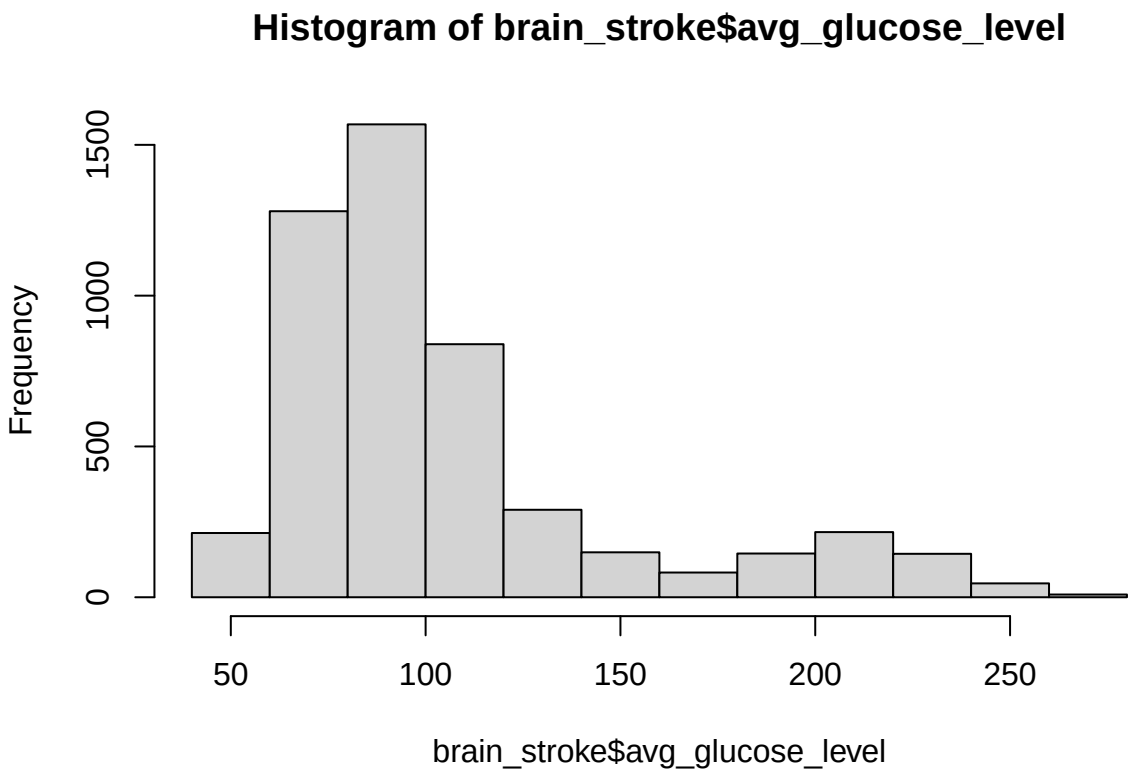
```
##     gender                age          hypertension     heart_disease
##  Length:4981        Min.   : 0.08   Min.   :0.00000   Min.   :0.00000
##  Class :character   1st Qu.:25.00   1st Qu.:0.00000   1st Qu.:0.00000
##  Mode  :character   Median :45.00   Median :0.00000   Median :0.00000
##                     Mean   :43.42   Mean   :0.09617   Mean   :0.05521
##                     3rd Qu.:61.00   3rd Qu.:0.00000   3rd Qu.:0.00000
##                     Max.   :82.00   Max.   :1.00000   Max.   :1.00000
##  ever_married        work_type         Residence_type     avg_glucose_level
##  Length:4981        Length:4981        Length:4981        Min.   : 55.12
##  Class :character   Class :character   Class :character   1st Qu.: 77.23
##  Mode  :character   Mode  :character   Mode  :character   Median : 91.85
##                                                           Mean   :105.94
##                                                           3rd Qu.:113.86
##                                                           Max.   :271.74
##       bmi        smoking_status         stroke
##  Min.   :14.0   Length:4981        Min.   :0.00000
##  1st Qu.:23.7   Class :character   1st Qu.:0.00000
##  Median :28.1   Mode  :character   Median :0.00000
##  Mean   :28.5                      Mean   :0.04979
##  3rd Qu.:32.6                      3rd Qu.:0.00000
##  Max.   :48.9                      Max.   :1.00000
```

Exploratory data analysis:

Plotting histograms to check normality of the data

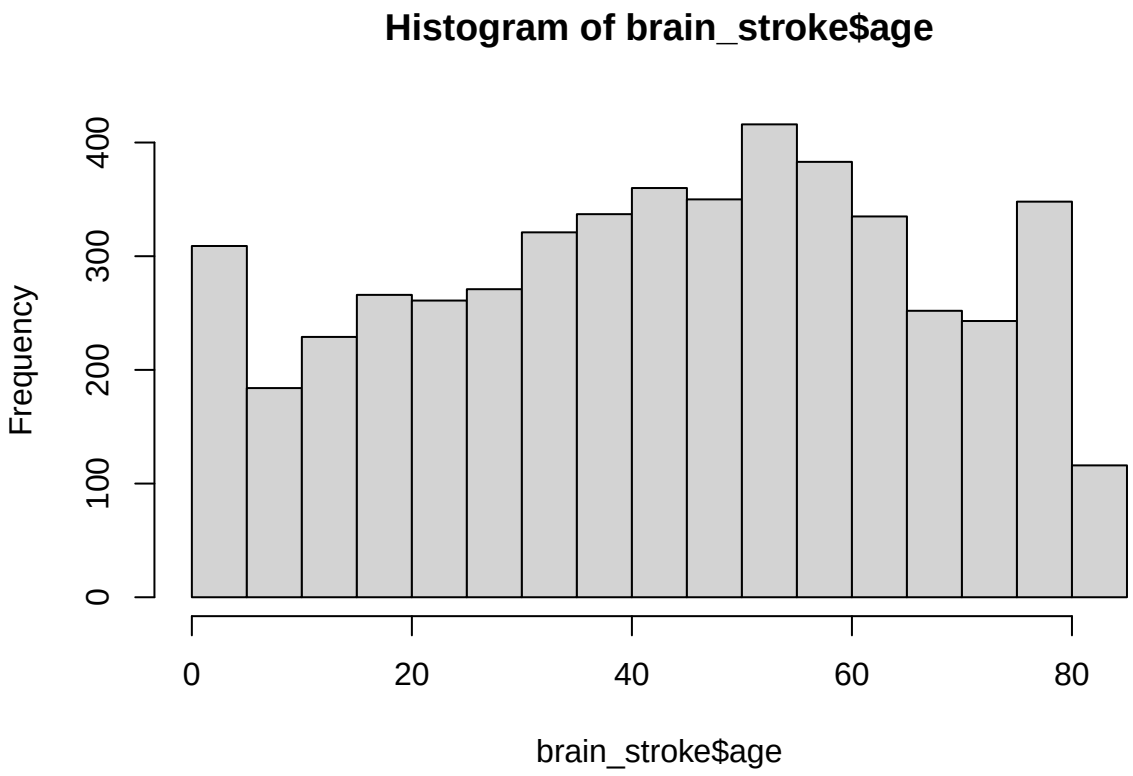
1) For average glucose levels

```
hist(brain_stroke$avg_glucose_level)
```



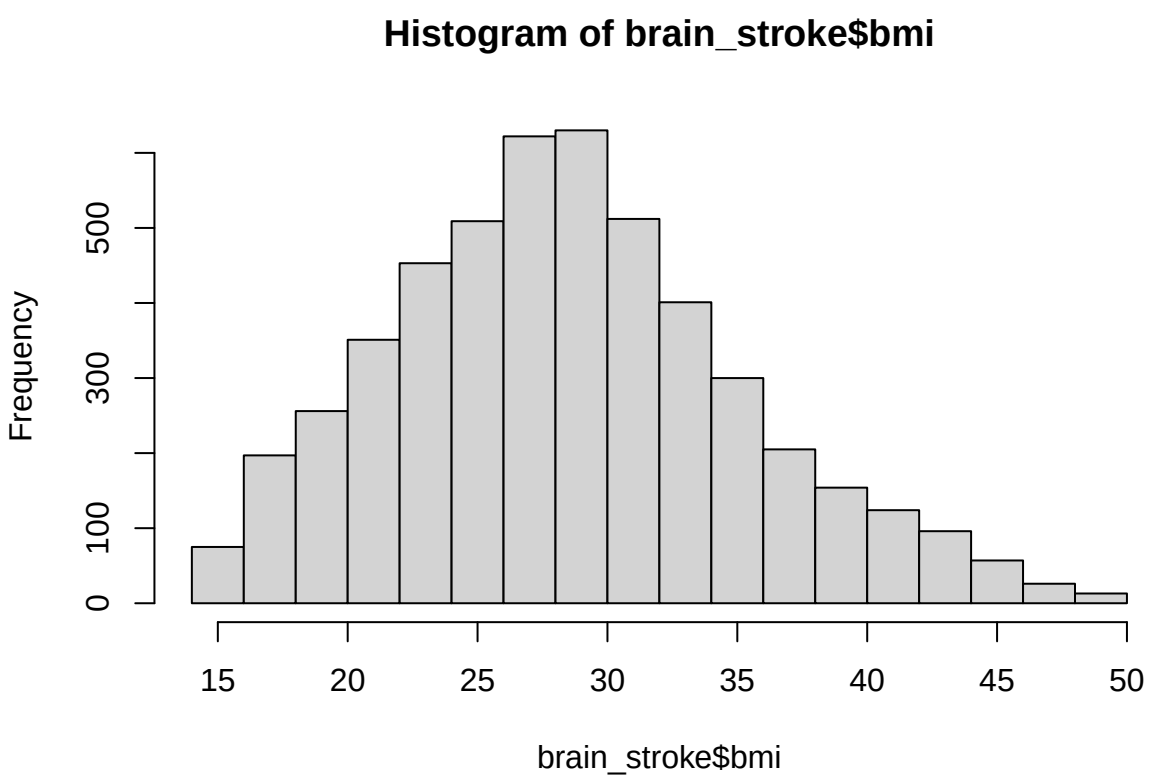
2) For age

```
hist(brain_stroke$age)
```



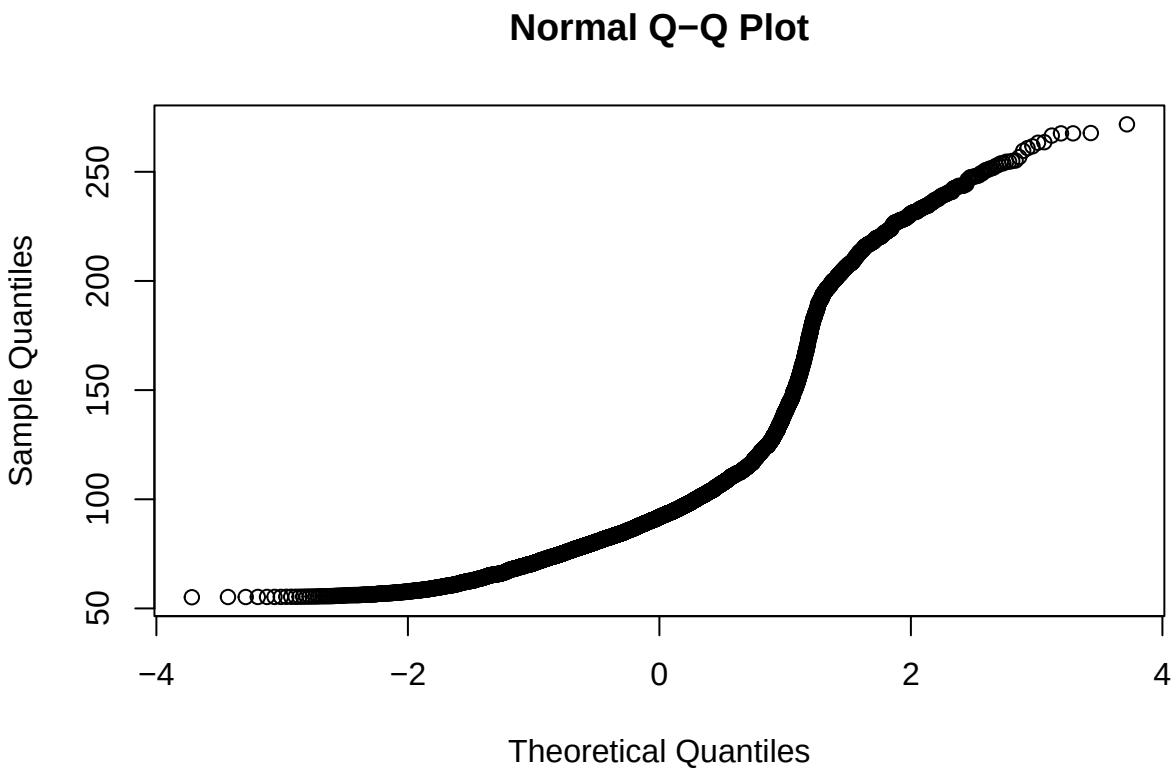
3) For BMI

```
hist(brain_stroke$bmi)
```

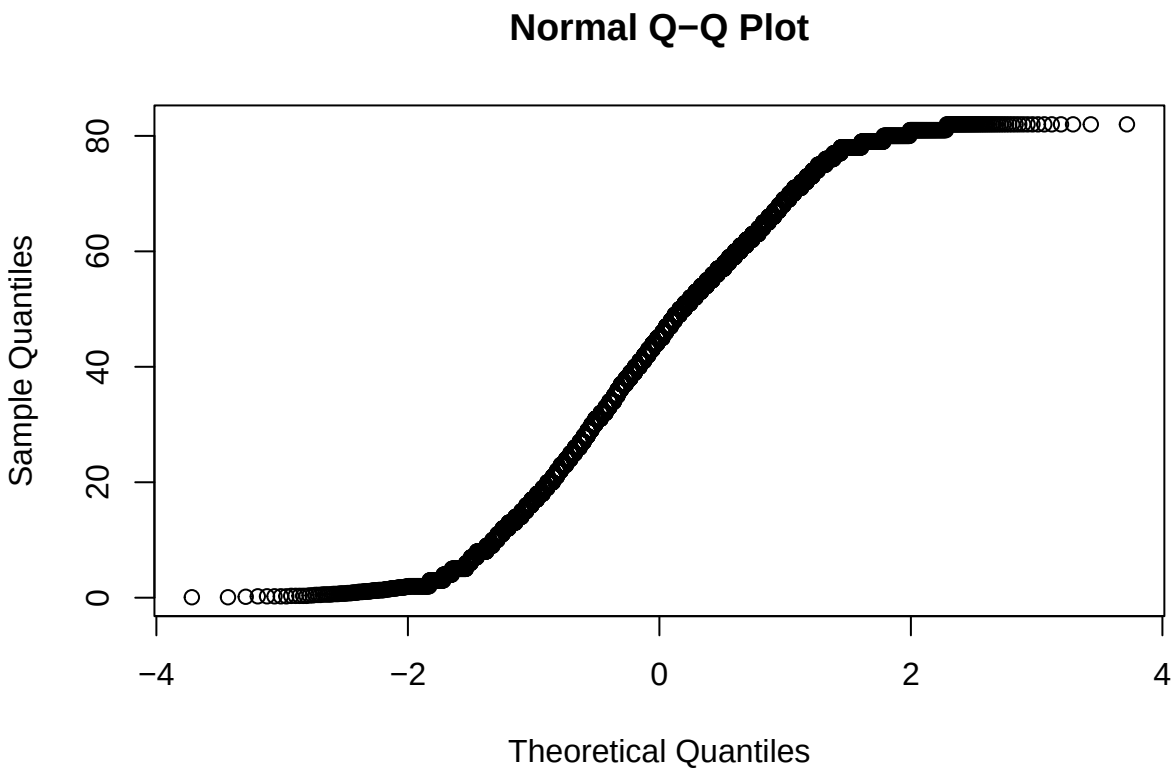


Q-Q PLOTS FOR NORMALITY

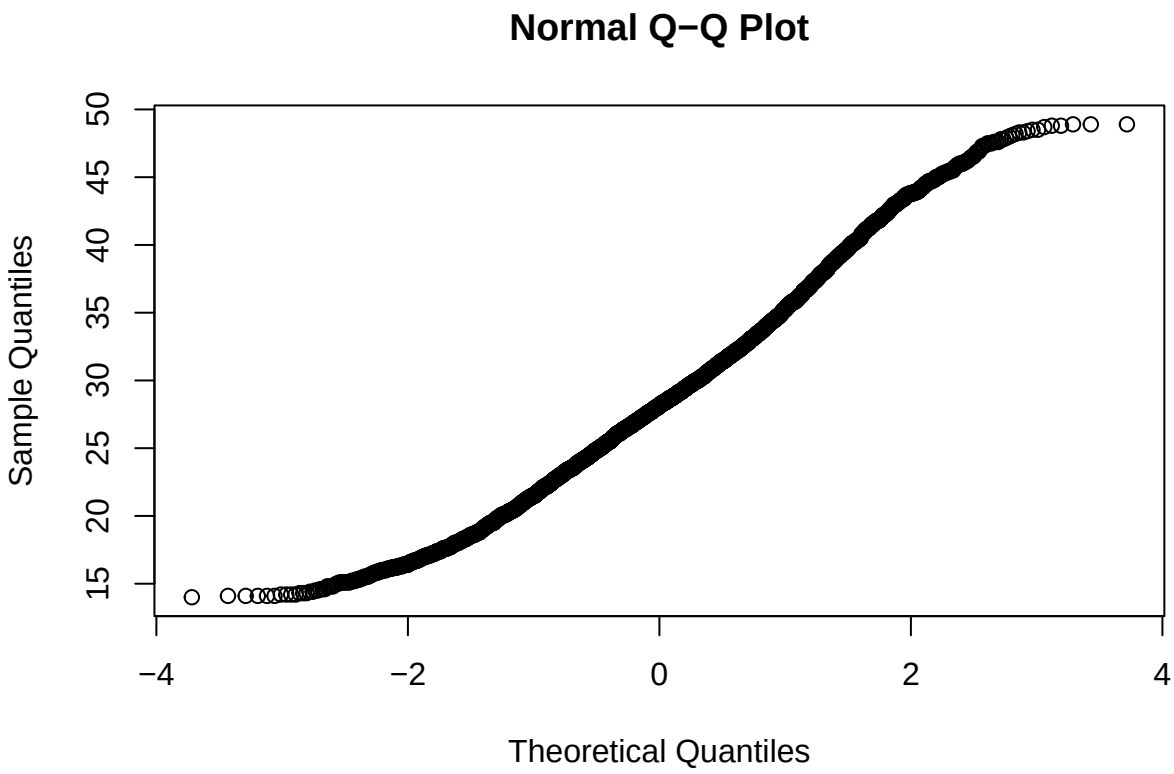
```
qqnorm(brain_stroke$avg_glucose_level)
```




```
qqnorm(brain_stroke$age)
```




```
qqnorm(brain_stroke$bmi)
```



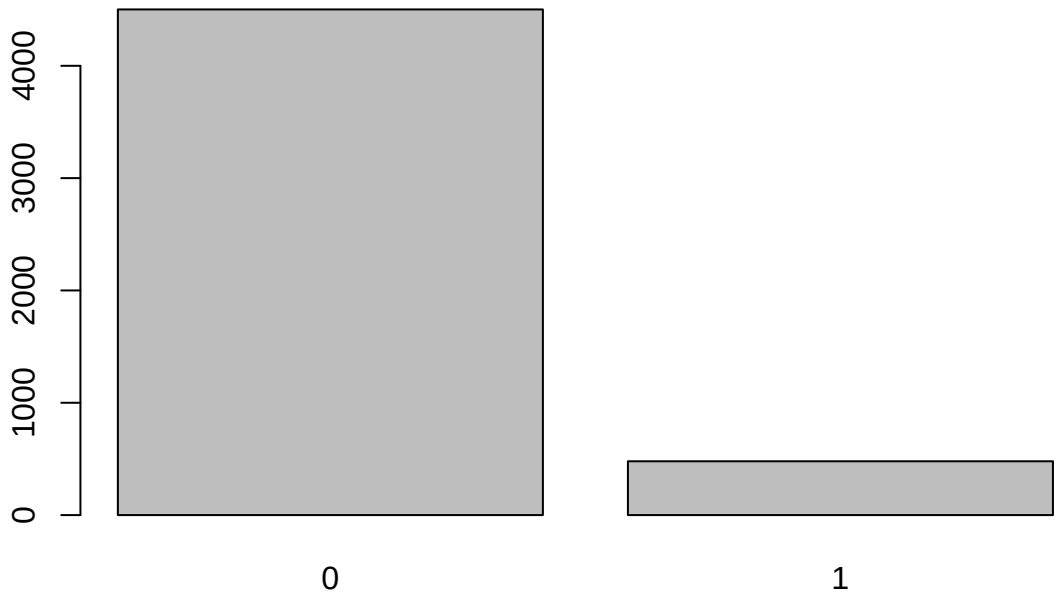
Plotting bar plots

1) For hypertension

```
# One way table to summarize hypertension data
table(brain_stroke$hypertension)
```

```
##
##    0    1
## 4502  479
```

```
# plotting bar plot for hypertension
barplot(table(brain_stroke$hypertension))
```

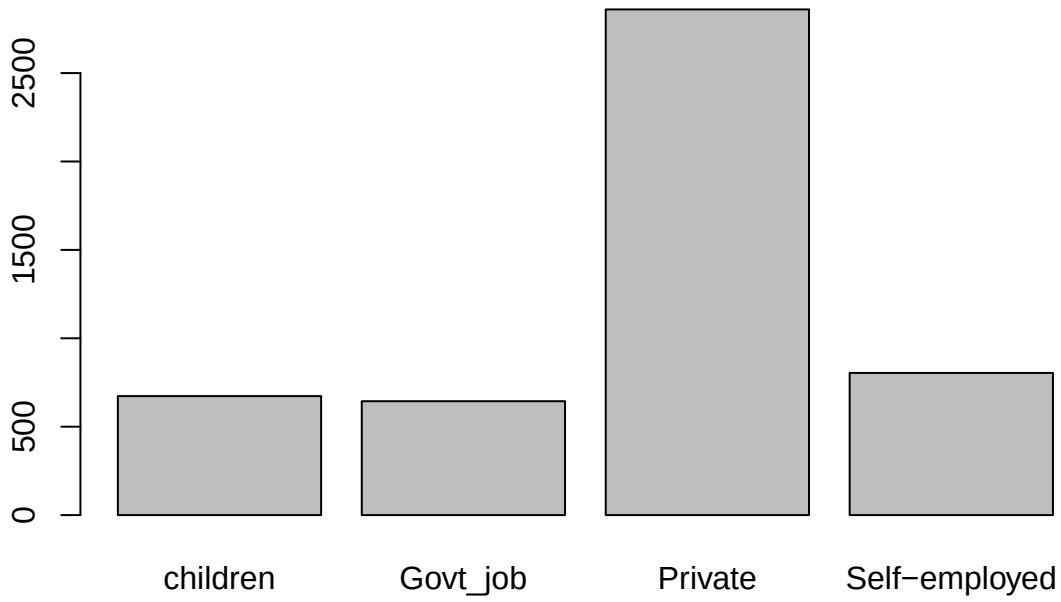


2) For work_type

```
# One way table to summarize work_type data
table(brain_stroke$work_type)
```

```
##
##      children      Govt_job       Private Self-employed
##           673           644          2860           804
```

```
# plotting bar plot for work_type
barplot(table(brain_stroke$work_type))
```

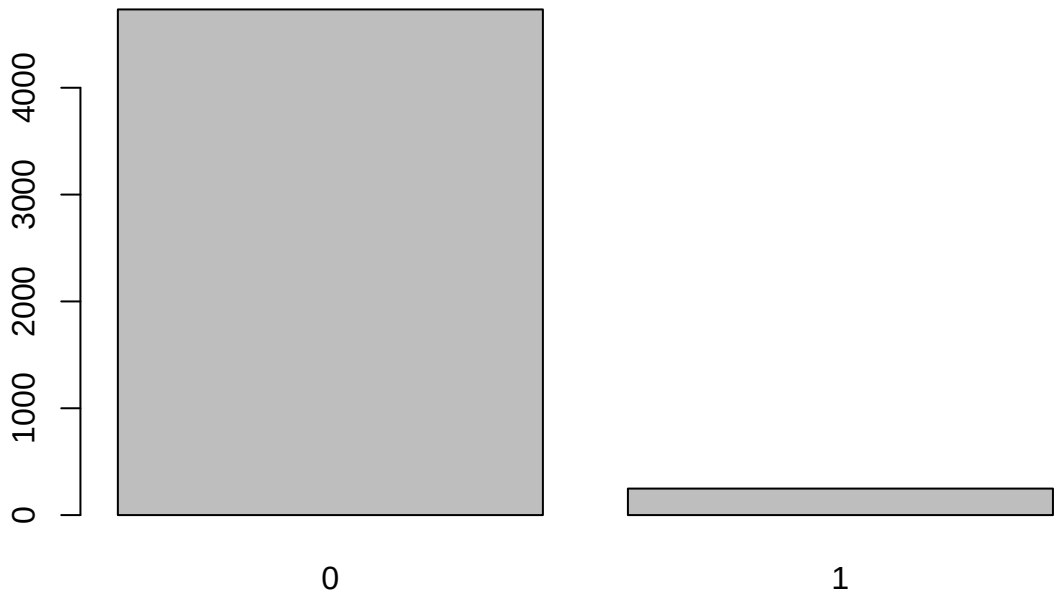


3) For stroke

```
# One way table to summarize stroke data
table(brain_stroke$stroke)
```

```
##
##    0    1
## 4733  248
```

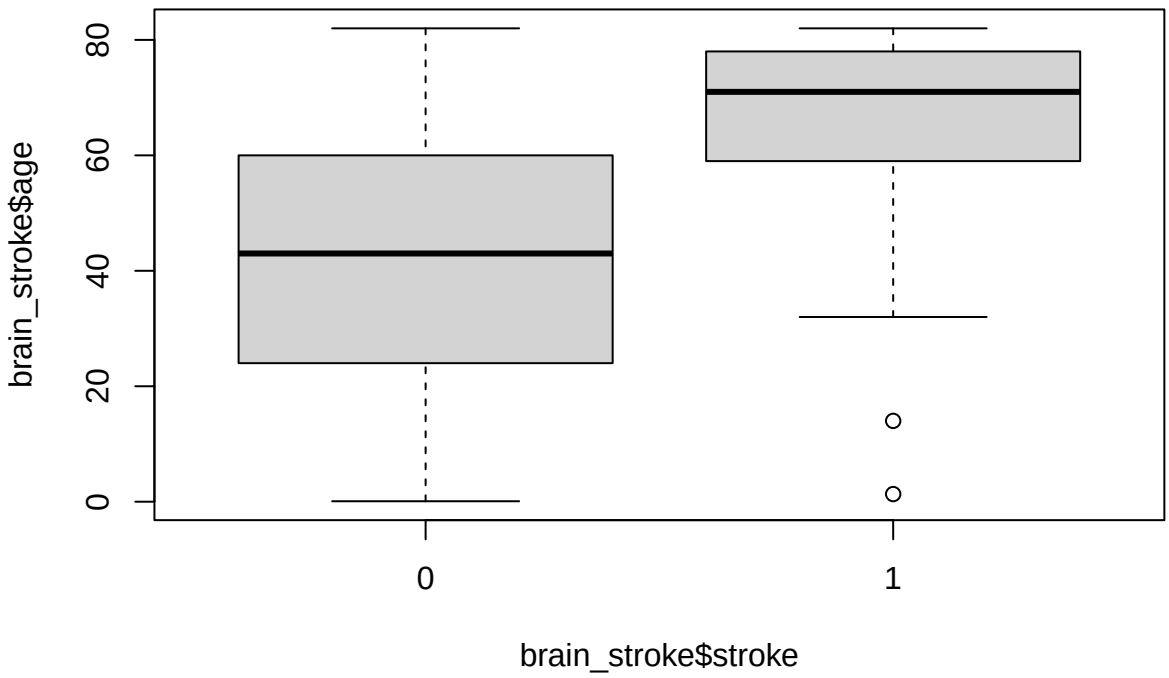
```
# plotting bar plot for stroke
barplot(table(brain_stroke$stroke))
```



Plotting box plot to check for outliers in data:

1) For age

```
boxplot(brain_stroke$age~brain_stroke$stroke)
```



2) For average glucose levels

```
boxplot(brain_stroke$avg_glucose_level~brain_stroke$stroke)
```

Importing and viewing the new data set after categorizing and coding variables:

```
library(readxl)
brain_stroke1 <- read_excel("C:/Users/madhu/Downloads/brain_stroke1.xlsx")
View(brain_stroke1)
```

Sampling:

```
#installing package sampling to perform stratified random sampling
#install.packages("sampling")
library("sampling")
```

```
## Warning: package 'sampling' was built under R version 4.2.3
```

```
set.seed(1)
strat.samp <- strata(brain_stroke1, stratanames =
c("gender", "heart_disease", "bmi_status"), size = rmultinom(30, 30, 30),
method = "srswr")
strat.samp
```

Performing random stratification of sample considering gender, heart_disease, and bmi as confounding variables.

```
##        gender heart_disease bmi_status ID_unit       Prob Stratum
## 20       Male             1          3      20 0.31433319       1
## 45       Male             1          3      45 0.31433319       1
## 81       Male             1          3      81 0.31433319       1
## 164      Male             1          3     164 0.31433319       1
## 271      Male             1          3     271 0.31433319       1
## 565      Male             1          3     565 0.31433319       1
## 756      Male             1          3     756 0.31433319       1
## 756.1    Male             1          3     756 0.31433319       1
## 821      Male             1          3     821 0.31433319       1
## 842      Male             1          3     842 0.31433319       1
## 1214     Male             1          3    1214 0.31433319       1
## 1745     Male             1          3    1745 0.31433319       1
## 1830     Male             1          3    1830 0.31433319       1
## 1838     Male             1          3    1838 0.31433319       1
## 2021     Male             1          3    2021 0.31433319       1
## 2181     Male             1          3    2181 0.31433319       1
## 2335     Male             1          3    2335 0.31433319       1
## 2400     Male             1          3    2400 0.31433319       1
## 2725     Male             1          3    2725 0.31433319       1
## 2772     Male             1          3    2772 0.31433319       1
## 3508     Male             1          3    3508 0.31433319       1
## 3803     Male             1          3    3803 0.31433319       1
## 4040     Male             1          3    4040 0.31433319       1
## 4393     Male             1          3    4393 0.31433319       1
## 4783     Male             1          3    4783 0.31433319       1
## 4815     Male             1          3    4815 0.31433319       1
## 4901     Male             1          3    4901 0.31433319       1
## 4901.1   Male             1          3    4901 0.31433319       1
## 4903     Male             1          3    4903 0.31433319       1
## 4940     Male             1          3    4940 0.31433319       1
## 133    Female             0          3     133 0.03568919       2
## 190    Female             0          3     190 0.03568919       2
## 242    Female             0          3     242 0.03568919       2
## 360    Female             0          3     360 0.03568919       2
## 1273   Female             0          3    1273 0.03568919       2
## 1381   Female             0          3    1381 0.03568919       2
## 1461   Female             0          3    1461 0.03568919       2
## 1511   Female             0          3    1511 0.03568919       2
## 1538   Female             0          3    1538 0.03568919       2
## 1571   Female             0          3    1571 0.03568919       2
## 1575   Female             0          3    1575 0.03568919       2
## 1682   Female             0          3    1682 0.03568919       2
## 1869   Female             0          3    1869 0.03568919       2
## 2037   Female             0          3    2037 0.03568919       2
## 2404   Female             0          3    2404 0.03568919       2
## 2665   Female             0          3    2665 0.03568919       2
## 2831   Female             0          3    2831 0.03568919       2
## 2933   Female             0          3    2933 0.03568919       2
## 3314   Female             0          3    3314 0.03568919       2
## 3328   Female             0          3    3328 0.03568919       2
```

```
## 3854   Female            0          3    3854 0.03568919       2
## 4181   Female            0          3    4181 0.03568919       2
## 4261   Female            0          3    4261 0.03568919       2
## 4313   Female            0          3    4313 0.03568919       2
## 4313.1 Female            0          3    4313 0.03568919       2
## 4412   Female            0          3    4412 0.03568919       2
## 4476   Female            0          3    4476 0.03568919       2
## 4587   Female            0          3    4587 0.03568919       2
## 4675   Female            0          3    4675 0.03568919       2
## 4795   Female            0          3    4795 0.03568919       2
## 64     Female            0          1      64 0.03737875       3
## 355    Female            0          1     355 0.03737875       3
## 356    Female            0          1     356 0.03737875       3
## 634    Female            0          1     634 0.03737875       3
## 648    Female            0          1     648 0.03737875       3
## 670    Female            0          1     670 0.03737875       3
## 732    Female            0          1     732 0.03737875       3
## 770    Female            0          1     770 0.03737875       3
## 770.1  Female            0          1     770 0.03737875       3
## 921    Female            0          1     921 0.03737875       3
## 1804   Female            0          1    1804 0.03737875       3
## 2177   Female            0          1    2177 0.03737875       3
## 2255   Female            0          1    2255 0.03737875       3
## 2363   Female            0          1    2363 0.03737875       3
## 2741   Female            0          1    2741 0.03737875       3
## 2756   Female            0          1    2756 0.03737875       3
## 2959   Female            0          1    2959 0.03737875       3
## 3292   Female            0          1    3292 0.03737875       3
## 3414   Female            0          1    3414 0.03737875       3
## 3512   Female            0          1    3512 0.03737875       3
## 3872   Female            0          1    3872 0.03737875       3
## 3887   Female            0          1    3887 0.03737875       3
## 3935   Female            0          1    3935 0.03737875       3
## 3983   Female            0          1    3983 0.03737875       3
## 4053   Female            0          1    4053 0.03737875       3
## 4171   Female            0          1    4171 0.03737875       3
## 4308   Female            0          1    4308 0.03737875       3
## 4639   Female            0          1    4639 0.03737875       3
## 4656   Female            0          1    4656 0.03737875       3
## 4776   Female            0          1    4776 0.03737875       3
## 63       Male            0          2      63 0.04978200       4
## 75       Male            0          2      75 0.04978200       4
## 122      Male            0          2     122 0.04978200       4
## 183      Male            0          2     183 0.04978200       4
## 237      Male            0          2     237 0.04978200       4
## 265      Male            0          2     265 0.04978200       4
## 547      Male            0          2     547 0.04978200       4
## 580      Male            0          2     580 0.04978200       4
## 677      Male            0          2     677 0.04978200       4
## 695      Male            0          2     695 0.04978200       4
## 903      Male            0          2     903 0.04978200       4
## 1179     Male            0          2    1179 0.04978200       4
## 1261     Male            0          2    1261 0.04978200       4
## 1331     Male            0          2    1331 0.04978200       4
```

```
## 1438     Male             0          2    1438 0.04978200       4
## 1471     Male             0          2    1471 0.04978200       4
## 1565     Male             0          2    1565 0.04978200       4
## 1672     Male             0          2    1672 0.04978200       4
## 1958     Male             0          2    1958 0.04978200       4
## 2250     Male             0          2    2250 0.04978200       4
## 2308     Male             0          2    2308 0.04978200       4
## 2535     Male             0          2    2535 0.04978200       4
## 2621     Male             0          2    2621 0.04978200       4
## 3240     Male             0          2    3240 0.04978200       4
## 3445     Male             0          2    3445 0.04978200       4
## 3688     Male             0          2    3688 0.04978200       4
## 3826     Male             0          2    3826 0.04978200       4
## 3827     Male             0          2    3827 0.04978200       4
## 4943     Male             0          2    4943 0.04978200       4
## 4977     Male             0          2    4977 0.04978200       4
## 181      Male             1          2     181 0.37194378       5
## 182      Male             1          2     182 0.37194378       5
## 502      Male             1          2     502 0.37194378       5
## 1373     Male             1          2    1373 0.37194378       5
## 1373.1   Male             1          2    1373 0.37194378       5
## 1444     Male             1          2    1444 0.37194378       5
## 1444.1   Male             1          2    1444 0.37194378       5
## 2052     Male             1          2    2052 0.37194378       5
## 2112     Male             1          2    2112 0.37194378       5
## 2337     Male             1          2    2337 0.37194378       5
## 2337.1   Male             1          2    2337 0.37194378       5
## 2343     Male             1          2    2343 0.37194378       5
## 2614     Male             1          2    2614 0.37194378       5
## 2838     Male             1          2    2838 0.37194378       5
## 2838.1   Male             1          2    2838 0.37194378       5
## 2838.2   Male             1          2    2838 0.37194378       5
## 2854     Male             1          2    2854 0.37194378       5
## 3088     Male             1          2    3088 0.37194378       5
## 3117     Male             1          2    3117 0.37194378       5
## 3119     Male             1          2    3119 0.37194378       5
## 3370     Male             1          2    3370 0.37194378       5
## 3370.1   Male             1          2    3370 0.37194378       5
## 3385     Male             1          2    3385 0.37194378       5
## 3719     Male             1          2    3719 0.37194378       5
## 4249     Male             1          2    4249 0.37194378       5
## 4821     Male             1          2    4821 0.37194378       5
## 4821.1   Male             1          2    4821 0.37194378       5
## 4891     Male             1          2    4891 0.37194378       5
## 4899     Male             1          2    4899 0.37194378       5
## 4900     Male             1          2    4900 0.37194378       5
## 302    Female             0          2     302 0.03696429       6
## 393    Female             0          2     393 0.03696429       6
## 442    Female             0          2     442 0.03696429       6
## 446    Female             0          2     446 0.03696429       6
## 606    Female             0          2     606 0.03696429       6
## 980    Female             0          2     980 0.03696429       6
## 1163   Female             0          2    1163 0.03696429       6
## 1245   Female             0          2    1245 0.03696429       6
```

```
## 1285   Female            0          2    1285 0.03696429       6
## 1365   Female            0          2    1365 0.03696429       6
## 1592   Female            0          2    1592 0.03696429       6
## 1700   Female            0          2    1700 0.03696429       6
## 1706   Female            0          2    1706 0.03696429       6
## 1740   Female            0          2    1740 0.03696429       6
## 1751   Female            0          2    1751 0.03696429       6
## 1820   Female            0          2    1820 0.03696429       6
## 1820.1 Female            0          2    1820 0.03696429       6
## 2216   Female            0          2    2216 0.03696429       6
## 2668   Female            0          2    2668 0.03696429       6
## 3125   Female            0          2    3125 0.03696429       6
## 3188   Female            0          2    3188 0.03696429       6
## 3375   Female            0          2    3375 0.03696429       6
## 3389   Female            0          2    3389 0.03696429       6
## 3555   Female            0          2    3555 0.03696429       6
## 3625   Female            0          2    3625 0.03696429       6
## 3877   Female            0          2    3877 0.03696429       6
## 4043   Female            0          2    4043 0.03696429       6
## 4668   Female            0          2    4668 0.03696429       6
## 4692   Female            0          2    4692 0.03696429       6
## 4812   Female            0          2    4812 0.03696429       6
## 130    Female            1          3     130 0.49042879       7
## 178    Female            1          3     178 0.49042879       7
## 178.1  Female            1          3     178 0.49042879       7
## 187    Female            1          3     187 0.49042879       7
## 304    Female            1          3     304 0.49042879       7
## 526    Female            1          3     526 0.49042879       7
## 674    Female            1          3     674 0.49042879       7
## 885    Female            1          3     885 0.49042879       7
## 974    Female            1          3     974 0.49042879       7
## 1410   Female            1          3    1410 0.49042879       7
## 1485   Female            1          3    1485 0.49042879       7
## 1613   Female            1          3    1613 0.49042879       7
## 2598   Female            1          3    2598 0.49042879       7
## 2671   Female            1          3    2671 0.49042879       7
## 2705   Female            1          3    2705 0.49042879       7
## 2721   Female            1          3    2721 0.49042879       7
## 2721.1 Female            1          3    2721 0.49042879       7
## 2836   Female            1          3    2836 0.49042879       7
## 2849   Female            1          3    2849 0.49042879       7
## 3313   Female            1          3    3313 0.49042879       7
## 3336   Female            1          3    3336 0.49042879       7
## 3336.1 Female            1          3    3336 0.49042879       7
## 3774   Female            1          3    3774 0.49042879       7
## 3774.1 Female            1          3    3774 0.49042879       7
## 3876   Female            1          3    3876 0.49042879       7
## 4818   Female            1          3    4818 0.49042879       7
## 4859   Female            1          3    4859 0.49042879       7
## 4859.1 Female            1          3    4859 0.49042879       7
## 4866   Female            1          3    4866 0.49042879       7
## 4956   Female            1          3    4956 0.49042879       7
## 12     Female            1          2      12 0.53211570       8
## 30     Female            1          2      30 0.53211570       8
```

```
## 125    Female            1          2     125 0.53211570       8
## 214    Female            1          2     214 0.53211570       8
## 225    Female            1          2     225 0.53211570       8
## 1216   Female            1          2    1216 0.53211570       8
## 1628   Female            1          2    1628 0.53211570       8
## 1642   Female            1          2    1642 0.53211570       8
## 1642.1 Female            1          2    1642 0.53211570       8
## 1642.2 Female            1          2    1642 0.53211570       8
## 1968   Female            1          2    1968 0.53211570       8
## 1968.1 Female            1          2    1968 0.53211570       8
## 2036   Female            1          2    2036 0.53211570       8
## 2036.1 Female            1          2    2036 0.53211570       8
## 2080   Female            1          2    2080 0.53211570       8
## 2080.1 Female            1          2    2080 0.53211570       8
## 2080.2 Female            1          2    2080 0.53211570       8
## 2315   Female            1          2    2315 0.53211570       8
## 2392   Female            1          2    2392 0.53211570       8
## 2527   Female            1          2    2527 0.53211570       8
## 2597   Female            1          2    2597 0.53211570       8
## 2771   Female            1          2    2771 0.53211570       8
## 2910   Female            1          2    2910 0.53211570       8
## 3716   Female            1          2    3716 0.53211570       8
## 3921   Female            1          2    3921 0.53211570       8
## 3921.1 Female            1          2    3921 0.53211570       8
## 4078   Female            1          2    4078 0.53211570       8
## 4553   Female            1          2    4553 0.53211570       8
## 4788   Female            1          2    4788 0.53211570       8
## 4970   Female            1          2    4970 0.53211570       8
## 18     Female            0          4      18 0.13052039       9
## 226    Female            0          4     226 0.13052039       9
## 354    Female            0          4     354 0.13052039       9
## 590    Female            0          4     590 0.13052039       9
## 691    Female            0          4     691 0.13052039       9
## 701    Female            0          4     701 0.13052039       9
## 889    Female            0          4     889 0.13052039       9
## 889.1  Female            0          4     889 0.13052039       9
## 929    Female            0          4     929 0.13052039       9
## 1412   Female            0          4    1412 0.13052039       9
## 1495   Female            0          4    1495 0.13052039       9
## 1524   Female            0          4    1524 0.13052039       9
## 1680   Female            0          4    1680 0.13052039       9
## 2082   Female            0          4    2082 0.13052039       9
## 2276   Female            0          4    2276 0.13052039       9
## 2348   Female            0          4    2348 0.13052039       9
## 2388   Female            0          4    2388 0.13052039       9
## 2900   Female            0          4    2900 0.13052039       9
## 2988   Female            0          4    2988 0.13052039       9
## 3134   Female            0          4    3134 0.13052039       9
## 3410   Female            0          4    3410 0.13052039       9
## 3539   Female            0          4    3539 0.13052039       9
## 3596   Female            0          4    3596 0.13052039       9
## 3900   Female            0          4    3900 0.13052039       9
## 4252   Female            0          4    4252 0.13052039       9
## 4330   Female            0          4    4330 0.13052039       9
```

```
## 4545   Female            0          4    4545 0.13052039       9
## 4573   Female            0          4    4573 0.13052039       9
## 4598   Female            0          4    4598 0.13052039       9
## 4730   Female            0          4    4730 0.13052039       9
## 486      Male            0          1     486 0.07014564      10
## 617      Male            0          1     617 0.07014564      10
## 679      Male            0          1     679 0.07014564      10
## 705      Male            0          1     705 0.07014564      10
## 757      Male            0          1     757 0.07014564      10
## 869      Male            0          1     869 0.07014564      10
## 909      Male            0          1     909 0.07014564      10
## 1012     Male            0          1    1012 0.07014564      10
## 1042     Male            0          1    1042 0.07014564      10
## 1139     Male            0          1    1139 0.07014564      10
## 1156     Male            0          1    1156 0.07014564      10
## 1250     Male            0          1    1250 0.07014564      10
## 1320     Male            0          1    1320 0.07014564      10
## 1624     Male            0          1    1624 0.07014564      10
## 1624.1   Male            0          1    1624 0.07014564      10
## 2066     Male            0          1    2066 0.07014564      10
## 2118     Male            0          1    2118 0.07014564      10
## 2726     Male            0          1    2726 0.07014564      10
## 2729     Male            0          1    2729 0.07014564      10
## 2925     Male            0          1    2925 0.07014564      10
## 3170     Male            0          1    3170 0.07014564      10
## 3189     Male            0          1    3189 0.07014564      10
## 3469     Male            0          1    3469 0.07014564      10
## 3699     Male            0          1    3699 0.07014564      10
## 3714     Male            0          1    3714 0.07014564      10
## 3779     Male            0          1    3779 0.07014564      10
## 3801     Male            0          1    3801 0.07014564      10
## 4066     Male            0          1    4066 0.07014564      10
## 4719     Male            0          1    4719 0.07014564      10
## 4738     Male            0          1    4738 0.07014564      10
## 65       Male            0          4      65 0.27200659      11
## 66       Male            0          4      66 0.27200659      11
## 66.1     Male            0          4      66 0.27200659      11
## 66.2     Male            0          4      66 0.27200659      11
## 72       Male            0          4      72 0.27200659      11
## 531      Male            0          4     531 0.27200659      11
## 675      Male            0          4     675 0.27200659      11
## 720      Male            0          4     720 0.27200659      11
## 1108     Male            0          4    1108 0.27200659      11
## 1211     Male            0          4    1211 0.27200659      11
## 1359     Male            0          4    1359 0.27200659      11
## 1424     Male            0          4    1424 0.27200659      11
## 1456     Male            0          4    1456 0.27200659      11
## 1456.1   Male            0          4    1456 0.27200659      11
## 1792     Male            0          4    1792 0.27200659      11
## 1903     Male            0          4    1903 0.27200659      11
## 2223     Male            0          4    2223 0.27200659      11
## 2456     Male            0          4    2456 0.27200659      11
## 2456.1   Male            0          4    2456 0.27200659      11
## 2681     Male            0          4    2681 0.27200659      11
```

```
## 2902     Male            0         4    2902 0.27200659      11
## 2902.1   Male            0         4    2902 0.27200659      11
## 3122     Male            0         4    3122 0.27200659      11
## 3645     Male            0         4    3645 0.27200659      11
## 3645.1   Male            0         4    3645 0.27200659      11
## 3645.2   Male            0         4    3645 0.27200659      11
## 3754     Male            0         4    3754 0.27200659      11
## 4020     Male            0         4    4020 0.27200659      11
## 4383     Male            0         4    4383 0.27200659      11
## 4755     Male            0         4    4755 0.27200659      11
## 103      Male            0         3     103 0.04500349      12
## 175      Male            0         3     175 0.04500349      12
## 456      Male            0         3     456 0.04500349      12
## 804      Male            0         3     804 0.04500349      12
## 946      Male            0         3     946 0.04500349      12
## 1384     Male            0         3    1384 0.04500349      12
## 1489     Male            0         3    1489 0.04500349      12
## 1514     Male            0         3    1514 0.04500349      12
## 1711     Male            0         3    1711 0.04500349      12
## 1713     Male            0         3    1713 0.04500349      12
## 1790     Male            0         3    1790 0.04500349      12
## 1859     Male            0         3    1859 0.04500349      12
## 1881     Male            0         3    1881 0.04500349      12
## 1884     Male            0         3    1884 0.04500349      12
## 1939     Male            0         3    1939 0.04500349      12
## 2225     Male            0         3    2225 0.04500349      12
## 2591     Male            0         3    2591 0.04500349      12
## 2631     Male            0         3    2631 0.04500349      12
## 2661     Male            0         3    2661 0.04500349      12
## 2754     Male            0         3    2754 0.04500349      12
## 2760     Male            0         3    2760 0.04500349      12
## 3040     Male            0         3    3040 0.04500349      12
## 3162     Male            0         3    3162 0.04500349      12
## 3306     Male            0         3    3306 0.04500349      12
## 3638     Male            0         3    3638 0.04500349      12
## 3794     Male            0         3    3794 0.04500349      12
## 3815     Male            0         3    3815 0.04500349      12
## 4029     Male            0         3    4029 0.04500349      12
## 4275     Male            0         3    4275 0.04500349      12
## 4959     Male            0         3    4959 0.04500349      12
## 87     Female            1         1      87 0.76862255      13
## 292    Female            1         1     292 0.76862255      13
## 357    Female            1         1     357 0.76862255      13
## 581    Female            1         1     581 0.76862255      13
## 581.1  Female            1         1     581 0.76862255      13
## 713    Female            1         1     713 0.76862255      13
## 1100   Female            1         1    1100 0.76862255      13
## 2183   Female            1         1    2183 0.76862255      13
## 2239   Female            1         1    2239 0.76862255      13
## 2239.1 Female            1         1    2239 0.76862255      13
## 2239.2 Female            1         1    2239 0.76862255      13
## 2239.3 Female            1         1    2239 0.76862255      13
## 2386   Female            1         1    2386 0.76862255      13
## 3104   Female            1         1    3104 0.76862255      13
```

```
## 3104.1 Female             1          1    3104 0.76862255      13
## 3104.2 Female             1          1    3104 0.76862255      13
## 3104.3 Female             1          1    3104 0.76862255      13
## 3104.4 Female             1          1    3104 0.76862255      13
## 3156   Female             1          1    3156 0.76862255      13
## 3617   Female             1          1    3617 0.76862255      13
## 3617.1 Female             1          1    3617 0.76862255      13
## 3617.2 Female             1          1    3617 0.76862255      13
## 3945   Female             1          1    3945 0.76862255      13
## 3945.1 Female             1          1    3945 0.76862255      13
## 3945.2 Female             1          1    3945 0.76862255      13
## 3945.3 Female             1          1    3945 0.76862255      13
## 4093   Female             1          1    4093 0.76862255      13
## 4239   Female             1          1    4239 0.76862255      13
## 4239.1 Female             1          1    4239 0.76862255      13
## 4685   Female             1          1    4685 0.76862255      13
## 93       Male             1          1      93 0.92649055      14
## 93.1     Male             1          1      93 0.92649055      14
## 93.2     Male             1          1      93 0.92649055      14
## 93.3     Male             1          1      93 0.92649055      14
## 94       Male             1          1      94 0.92649055      14
## 217      Male             1          1     217 0.92649055      14
## 217.1    Male             1          1     217 0.92649055      14
## 217.2    Male             1          1     217 0.92649055      14
## 810      Male             1          1     810 0.92649055      14
## 810.1    Male             1          1     810 0.92649055      14
## 810.2    Male             1          1     810 0.92649055      14
## 810.3    Male             1          1     810 0.92649055      14
## 810.4    Male             1          1     810 0.92649055      14
## 1233     Male             1          1    1233 0.92649055      14
## 1233.1   Male             1          1    1233 0.92649055      14
## 1233.2   Male             1          1    1233 0.92649055      14
## 1233.3   Male             1          1    1233 0.92649055      14
## 1947     Male             1          1    1947 0.92649055      14
## 2445     Male             1          1    2445 0.92649055      14
## 2445.1   Male             1          1    2445 0.92649055      14
## 2498     Male             1          1    2498 0.92649055      14
## 2498.1   Male             1          1    2498 0.92649055      14
## 2498.2   Male             1          1    2498 0.92649055      14
## 2498.3   Male             1          1    2498 0.92649055      14
## 3424     Male             1          1    3424 0.92649055      14
## 3424.1   Male             1          1    3424 0.92649055      14
## 3424.2   Male             1          1    3424 0.92649055      14
## 3424.3   Male             1          1    3424 0.92649055      14
## 3490     Male             1          1    3490 0.92649055      14
## 3614     Male             1          1    3614 0.92649055      14
## 180    Female             1          4     180 0.99578728      15
## 180.1  Female             1          4     180 0.99578728      15
## 180.2  Female             1          4     180 0.99578728      15
## 180.3  Female             1          4     180 0.99578728      15
## 180.4  Female             1          4     180 0.99578728      15
## 180.5  Female             1          4     180 0.99578728      15
## 204    Female             1          4     204 0.99578728      15
## 204.1  Female             1          4     204 0.99578728      15
```

```
## 204.2  Female            1          4     204 0.99578728      15
## 204.3  Female            1          4     204 0.99578728      15
## 204.4  Female            1          4     204 0.99578728      15
## 1598   Female            1          4    1598 0.99578728      15
## 1598.1 Female            1          4    1598 0.99578728      15
## 1598.2 Female            1          4    1598 0.99578728      15
## 1598.3 Female            1          4    1598 0.99578728      15
## 2156   Female            1          4    2156 0.99578728      15
## 2156.1 Female            1          4    2156 0.99578728      15
## 2156.2 Female            1          4    2156 0.99578728      15
## 2156.3 Female            1          4    2156 0.99578728      15
## 2156.4 Female            1          4    2156 0.99578728      15
## 2156.5 Female            1          4    2156 0.99578728      15
## 2156.6 Female            1          4    2156 0.99578728      15
## 4063   Female            1          4    4063 0.99578728      15
## 4063.1 Female            1          4    4063 0.99578728      15
## 4063.2 Female            1          4    4063 0.99578728      15
## 4539   Female            1          4    4539 0.99578728      15
## 4539.1 Female            1          4    4539 0.99578728      15
## 4539.2 Female            1          4    4539 0.99578728      15
## 4539.3 Female            1          4    4539 0.99578728      15
## 4539.4 Female            1          4    4539 0.99578728      15
## 191    Female            0          0     191 0.16309278      16
## 321    Female            0          0     321 0.16309278      16
## 873    Female            0          0     873 0.16309278      16
## 873.1  Female            0          0     873 0.16309278      16
## 1004   Female            0          0    1004 0.16309278      16
## 1067   Female            0          0    1067 0.16309278      16
## 1082   Female            0          0    1082 0.16309278      16
## 1284   Female            0          0    1284 0.16309278      16
## 1316   Female            0          0    1316 0.16309278      16
## 1316.1 Female            0          0    1316 0.16309278      16
## 1506   Female            0          0    1506 0.16309278      16
## 1534   Female            0          0    1534 0.16309278      16
## 1584   Female            0          0    1584 0.16309278      16
## 2138   Female            0          0    2138 0.16309278      16
## 2298   Female            0          0    2298 0.16309278      16
## 2344   Female            0          0    2344 0.16309278      16
## 2410   Female            0          0    2410 0.16309278      16
## 2438   Female            0          0    2438 0.16309278      16
## 2680   Female            0          0    2680 0.16309278      16
## 2843   Female            0          0    2843 0.16309278      16
## 3073   Female            0          0    3073 0.16309278      16
## 3356   Female            0          0    3356 0.16309278      16
## 3640   Female            0          0    3640 0.16309278      16
## 3940   Female            0          0    3940 0.16309278      16
## 4082   Female            0          0    4082 0.16309278      16
## 4161   Female            0          0    4161 0.16309278      16
## 4165   Female            0          0    4165 0.16309278      16
## 4277   Female            0          0    4277 0.16309278      16
## 4445   Female            0          0    4445 0.16309278      16
## 4556   Female            0          0    4556 0.16309278      16
## 277      Male            0          0     277 0.16857648      17
## 332      Male            0          0     332 0.16857648      17
```

```
## 336      Male             0          0     336 0.16857648      17
## 986      Male             0          0     986 0.16857648      17
## 1187     Male             0          0    1187 0.16857648      17
## 1192     Male             0          0    1192 0.16857648      17
## 1464     Male             0          0    1464 0.16857648      17
## 1464.1   Male             0          0    1464 0.16857648      17
## 1475     Male             0          0    1475 0.16857648      17
## 1475.1   Male             0          0    1475 0.16857648      17
## 1640     Male             0          0    1640 0.16857648      17
## 1640.1   Male             0          0    1640 0.16857648      17
## 2168     Male             0          0    2168 0.16857648      17
## 2534     Male             0          0    2534 0.16857648      17
## 2654     Male             0          0    2654 0.16857648      17
## 3149     Male             0          0    3149 0.16857648      17
## 3177     Male             0          0    3177 0.16857648      17
## 3194     Male             0          0    3194 0.16857648      17
## 3273     Male             0          0    3273 0.16857648      17
## 3369     Male             0          0    3369 0.16857648      17
## 3499     Male             0          0    3499 0.16857648      17
## 3603     Male             0          0    3603 0.16857648      17
## 3629     Male             0          0    3629 0.16857648      17
## 3742     Male             0          0    3742 0.16857648      17
## 4076     Male             0          0    4076 0.16857648      17
## 4101     Male             0          0    4101 0.16857648      17
## 4199     Male             0          0    4199 0.16857648      17
## 4474     Male             0          0    4474 0.16857648      17
## 4487     Male             0          0    4487 0.16857648      17
## 4625     Male             0          0    4625 0.16857648      17
## 334      Male             1          4     334 0.99578728      18
## 334.1    Male             1          4     334 0.99578728      18
## 334.2    Male             1          4     334 0.99578728      18
## 1183     Male             1          4    1183 0.99578728      18
## 1183.1   Male             1          4    1183 0.99578728      18
## 1183.2   Male             1          4    1183 0.99578728      18
## 1183.3   Male             1          4    1183 0.99578728      18
## 1183.4   Male             1          4    1183 0.99578728      18
## 1801     Male             1          4    1801 0.99578728      18
## 1801.1   Male             1          4    1801 0.99578728      18
## 1926     Male             1          4    1926 0.99578728      18
## 1926.1   Male             1          4    1926 0.99578728      18
## 1926.2   Male             1          4    1926 0.99578728      18
## 1926.3   Male             1          4    1926 0.99578728      18
## 1926.4   Male             1          4    1926 0.99578728      18
## 1926.5   Male             1          4    1926 0.99578728      18
## 1959     Male             1          4    1959 0.99578728      18
## 1959.1   Male             1          4    1959 0.99578728      18
## 1959.2   Male             1          4    1959 0.99578728      18
## 1959.3   Male             1          4    1959 0.99578728      18
## 1959.4   Male             1          4    1959 0.99578728      18
## 1959.5   Male             1          4    1959 0.99578728      18
## 2440     Male             1          4    2440 0.99578728      18
## 2440.1   Male             1          4    2440 0.99578728      18
## 2440.2   Male             1          4    2440 0.99578728      18
## 2440.3   Male             1          4    2440 0.99578728      18
```

```
## 2440.4   Male             1          4    2440 0.99578728      18
## 2440.5   Male             1          4    2440 0.99578728      18
## 2440.6   Male             1          4    2440 0.99578728      18
## 2440.7   Male             1          4    2440 0.99578728      18
```

```
# Taking a stratified random sample as bs.sample
num_samples=5
for (i in 1:num_samples) {bs.samp <- brain_stroke1[strat.samp$ID_unit, ]}
bs.samp
```

```
## # A tibble: 540 x 16
##    gender   age hypertension heart_disease ever_married work_type Residence_type
##    <chr>  <dbl>        <dbl>         <dbl> <chr>        <chr>     <chr>
##  1 Male      82            0             1 Yes          Private   Rural
##  2 Male      63            0             1 Yes          Private   Rural
##  3 Male      58            0             1 Yes          Private   Rural
##  4 Male      80            1             1 Yes          Private   Urban
##  5 Male      57            0             1 Yes          Private   Rural
##  6 Male      52            0             1 No           Private   Rural
##  7 Male      61            1             1 Yes          Govt_job  Rural
##  8 Male      61            1             1 Yes          Govt_job  Rural
##  9 Male      71            0             1 Yes          Private   Urban
## 10 Male      71            0             1 Yes          Self-emp~ Rural
## # i 530 more rows
## # i 9 more variables: avg_glucose_level <dbl>, bmi <dbl>, smoking_status <chr>,
## #   stroke <dbl>, work <dbl>, age_groups <dbl>, blood_glucose <dbl>,
## #   bmi_status <dbl>, blood_glucose2 <dbl>
```

```
#set.seed(1)
#sample.size = sample(1:nrow(brain_stroke1), 500, replace = FALSE)
#bs.samp = brain_stroke1[sample.size, ]
#bs.samp
```

Simple random sampling:

Correlation test:

```
cor1 = cor.test(bs.samp$stroke, bs.samp$age, method = "spearman")
```

```
## Warning in cor.test.default(bs.samp$stroke, bs.samp$age, method = "spearman"):
## Cannot compute exact p-value with ties
```

```
cor1
```

```
##
##  Spearman's rank correlation rho
##
```

```
## data:  bs.samp$stroke and bs.samp$age
## S = 19434369, p-value = 9.315e-10
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.2594713
```

```
cor2 = cor.test(bs.samp$stroke, bs.samp$work, method = "spearman")
```

```
## Warning in cor.test.default(bs.samp$stroke, bs.samp$work, method = "spearman"):
## Cannot compute exact p-value with ties
```

```
cor2
```

```
##
##  Spearman's rank correlation rho
##
## data:  bs.samp$stroke and bs.samp$work
## S = 24062997, p-value = 0.05361
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## 0.08310168
```

```
cor3 = cor.test(bs.samp$stroke, bs.samp$blood_glucose, method = "spearman")
```

```
## Warning in cor.test.default(bs.samp$stroke, bs.samp$blood_glucose, method =
## "spearman"): Cannot compute exact p-value with ties
```

```
cor3
```

```
##
##  Spearman's rank correlation rho
##
## data:  bs.samp$stroke and bs.samp$blood_glucose
## S = 18217504, p-value = 3.72e-13
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.3058388
```

```
cor4 = cor.test(bs.samp$stroke, bs.samp$blood_glucose2, method = "spearman")
```

```
## Warning in cor.test.default(bs.samp$stroke, bs.samp$blood_glucose2, method =
## "spearman"): Cannot compute exact p-value with ties
```

```
cor4
```

```
##
##  Spearman's rank correlation rho
```

```
##
## data:  bs.samp$stroke and bs.samp$blood_glucose2
## S = 18586642, p-value = 4.661e-12
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.2917731
```

Interpretation:

1. For the correlation between **bs.samp$stroke** and **bs.samp$age**, the Spearman's rank correlation coefficient was estimated as 0.259 and the p-value was estimated as 9.315e-10. This suggests that there is a strong positive correlation between stroke and age in the sample.
2. For the correlation between **bs.samp$stroke** and **bs.samp$work**, the Spearman's rank correlation coefficient was estimated as 0.0831 and the p-value was estimated as 0.05361. This suggests that there may be a weak positive correlation between stroke and work in the sample, but this correlation is not statistically significant at the conventional 5% level.
3. For the correlation between **bs.samp$stroke** and **bs.samp$blood_glucose**, the Spearman's rank correlation coefficient was estimated as 0.306 and the p-value was estimated as 3.72e-13. This suggests that there is a strong positive correlation between stroke and blood glucose in the sample.
4. For the correlation between **bs.samp$stroke** and **bs.samp$blood_glucose2**, the Spearman's rank correlation coefficient was estimated as 0.292 and the p-value was estimated as 4.661e-12. This suggests that there is a strong positive correlation between stroke and blood glucose level measured after a meal in the sample.
5. Overall, these results suggest that age, blood glucose, and blood glucose level after a meal are strongly positively correlated with stroke, while work may have a weaker positive association with stroke that is not statistically significant in this sample. However, these results are based on a single sample, and the associations may differ in other populations.

Unpaired two sample t-test:

un-paired two Sample t-test for comparing the mean stroke status of data elements with two categories i.e., average glucose levels

```
ndiabetic = subset(bs.samp, blood_glucose == 0)
diabetic = subset(bs.samp, blood_glucose == 1)
t.test(ndiabetic$stroke, diabetic$stroke, na.rm = TRUE,
       conf.level = 0.95, alternative = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  ndiabetic$stroke and diabetic$stroke
## t = -4.8835, df = 99.291, p-value = 3.993e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.3389517 -0.1430964
## sample estimates:
##  mean of x  mean of y
## 0.05567929 0.29670330
```

Interpretation:

The Welch two-sample t-test was conducted to determine if there was a significant difference in the mean values of the variable 'stroke' between two groups, 'ndiabetic' and 'diabetic'. The test statistic (t) was calculated to be -4.8835 with a degrees of freedom (df) of 99.291, and a very small p-value of 3.993e-06. This indicates strong evidence against the null hypothesis of no difference between the means of the two groups. Therefore, we reject the null hypothesis and conclude that there is a significant difference between the mean values of 'stroke' for the two groups.

The alternative hypothesis suggests that the true difference in means is not equal to 0. The 95% confidence interval of the difference between means is (-0.3389517, -0.1430964), which does not include 0. This indicates that the difference between the means is statistically significant.

The sample estimates for the mean value of 'stroke' for group 'ndiabetic' and 'diabetic' are 0.05567929 and 0.29670330, respectively. This suggests that patients with diabetes have a higher mean value for 'stroke' compared to those without diabetes.

Chi-square test:

1. For age groups:

```
# Creating a two-way contingency variable in R
table = table(brain_stroke1$stroke, brain_stroke1$age_groups)
table
```

```
##
##        0    1    2    3    4    5
##   0    2   44  527  248 3051  861
##   1    0    0    1    1   87  159
```

```
# Chi-square test
chisq.test(table)
```

```
## Warning in chisq.test(table): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  table
## X-squared = 314, df = 5, p-value < 2.2e-16
```

Interpretation:

The result of the Pearson's Chi-squared test indicates that there is a significant association between the variables in the contingency table.

The test statistic (X-squared) is 314, which is large and suggests that the observed counts in the table are different from what would be expected if there was no association between the variables.

The degrees of freedom (df) are 5, which corresponds to the number of categories in one variable minus 1, times the number of categories in the other variable minus 1.

The p-value is less than 2.2e-16, which is essentially zero. This indicates that the probability of observing such an extreme test statistic by chance, assuming no association between the variables, is extremely low.

Therefore, we can reject the null hypothesis of no association and conclude that there is a significant association between the variables in the contingency table.

2. For work:

```
# Creating a two-way contingency variable in R
table = table(brain_stroke1$stroke, brain_stroke1$work_type)
table
```

```
##
##     children Govt_job Private Self-employed
##   0      671      611    2712           739
##   1        2       33     148            65
```

```
# Chi-square test
chisq.test(table)
```

```
##
##  Pearson's Chi-squared test
##
## data:  table
## X-squared = 47.832, df = 3, p-value = 2.312e-10
```

Interpretation:

The Pearson's chi-squared test shows a statistically significant result with a chi-squared value of 47.832 and 3 degrees of freedom (df), which corresponds to a very small p-value of 2.312e-10. This indicates that there is strong evidence to reject the null hypothesis that the observed frequencies in the contingency table are equal to the expected frequencies, and that there is a significant association between the two categorical variables being analyzed.

3. For average blood glucose levels in 2 categories i.e., non-diabetic and diabetic:

```
# Creating a two-way contingency variable in R
table = table(brain_stroke1$stroke, brain_stroke1$blood_glucose)
table
```

```
##
##        0    1
##   0 4373  360
##   1  193   55
```

```
# Chi-square test
chisq.test(table)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table
## X-squared = 63.617, df = 1, p-value = 1.511e-15
```

Interpretation:

The result of the Pearson's Chi-squared test with Yates' continuity correction indicates that there is a significant association between the variables in the contingency table.

The test statistic (X-squared) is 63.617, which is large and suggests that the observed counts in the table are different from what would be expected if there was no association between the variables.

The degrees of freedom (df) are 1, which corresponds to the number of categories in one variable minus 1, times the number of categories in the other variable minus 1, minus 1 for the correction.

The p-value is 1.511e-15, which is essentially zero. This indicates that the probability of observing such an extreme test statistic by chance, assuming no association between the variables, is extremely low.

Therefore, we can reject the null hypothesis of no association and conclude that there is a significant association between the variables in the contingency table, even after applying Yates' continuity correction to adjust for the small expected cell frequencies.

4. For average blood glucose in 4 categories i.e., low, normal, pre-diabetic and diabetic:

```
#Creating a two-way contingency variable in R
table = table(brain_stroke1$stroke, brain_stroke1$blood_glucose2)
table
```

```
##
##        0    1    2    3
##   0  706 3179  488  360
##   1   27  122   44   55
```

```
# Chi-square test
chisq.test(table)
```

```
##
##  Pearson's Chi-squared test
##
## data:  table
## X-squared = 86.324, df = 3, p-value < 2.2e-16
```

Interpretation:

The result of the Pearson's Chi-squared test indicates that there is a significant association between the variables in the contingency table.

The test statistic (X-squared) is 86.324, which is large and suggests that the observed counts in the table are different from what would be expected if there was no association between the variables.

The degrees of freedom (df) are 3, which corresponds to the number of categories in one variable minus 1, times the number of categories in the other variable minus 1.

The p-value is less than 2.2e-16, which is essentially zero. This indicates that the probability of observing such an extreme test statistic by chance, assuming no association between the variables, is extremely low.

Therefore, we can reject the null hypothesis of no association and conclude that there is a significant association between the variables in the contingency table.

Logistic Regression:

1. For age groups

```
model <- glm(stroke ~ age_groups, data = bs.samp,
             family = "binomial"(link = "logit"))
summary(model)
```

```
##
## Call:
## glm(formula = stroke ~ age_groups, family = binomial(link = "logit"),
##     data = bs.samp)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.6352  -0.6352  -0.3584  -0.1975   2.3567
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -7.5720     1.2672  -5.975 2.30e-09 ***
## age_groups    1.2148     0.2724   4.459 8.22e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 342.22  on 539  degrees of freedom
## Residual deviance: 311.91  on 538  degrees of freedom
## AIC: 315.91
##
## Number of Fisher Scoring iterations: 6
```

Interpretation:

The coefficients table shows that age_groups is a significant predictor of the likelihood of having a stroke ($z = 4.459$, $p < 0.001$), with an estimated coefficient of 1.2148. The negative intercept coefficient (-7.5720) indicates that the probability of having a stroke decreases as age increases.

The deviance residuals describe the difference between the predicted and observed values for the dependent variable. The values range from -0.6352 to 2.3567, with higher values indicating a larger difference between predicted and observed values.

The AIC (Akaike Information Criterion) is a measure of the model's goodness-of-fit, with lower values indicating a better fit. The AIC value for this model is 315.91.

In conclusion, the logistic regression model suggests that age_groups is a significant predictor of the likelihood of having a stroke, with older age groups having a higher probability of having a stroke.

2. For work type

```
model <- glm(stroke ~ work, data = bs.samp,
             family = "binomial"(link = "logit"))
summary(model)
```

```
##
## Call:
## glm(formula = stroke ~ work, family = binomial(link = "logit"),
##     data = bs.samp)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.5645  -0.4525  -0.4525  -0.3609   2.3512
##
```

```
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.1705     0.4237  -7.482 7.31e-14 ***
## work          0.4715     0.1885   2.501   0.0124 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 342.22  on 539  degrees of freedom
## Residual deviance: 335.12  on 538  degrees of freedom
## AIC: 339.12
##
## Number of Fisher Scoring iterations: 5
```

Interpretation:

The coefficients table shows that work is a significant predictor of the likelihood of having a stroke (z = 2.501, p = 0.0124), with an estimated coefficient of 0.4715. The positive coefficient indicates that being employed is associated with a higher probability of having a stroke, compared to being unemployed or retired.

The deviance residuals describe the difference between the predicted and observed values for the dependent variable. The values range from -0.5645 to 2.3512, with higher values indicating a larger difference between predicted and observed values.

The AIC (Akaike Information Criterion) is a measure of the model's goodness-of-fit, with lower values indicating a better fit. The AIC value for this model is 339.12.

In conclusion, the logistic regression model suggests that work status is a significant predictor of the likelihood of having a stroke, with employed individuals having a higher probability of having a stroke compared to unemployed or retired individuals. However, the effect size of work on stroke is smaller compared to age, as indicated by the smaller coefficient and smaller difference between null and residual deviances.

3. For blood glucose: non-diabetic and diabetic

```
model <- glm(stroke ~ blood_glucose, data = bs.samp,
             family = "binomial"(link = "logit"))
summary(model)
```

```
##
## Call:
## glm(formula = stroke ~ blood_glucose, family = binomial(link = "logit"),
##     data = bs.samp)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.8390  -0.3385  -0.3385  -0.3385   2.4034
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.8309     0.2058 -13.755  < 2e-16 ***
## blood_glucose   1.9678     0.3083   6.384 1.73e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 342.22  on 539  degrees of freedom
## Residual deviance: 303.65  on 538  degrees of freedom
## AIC: 307.65
##
## Number of Fisher Scoring iterations: 5
```

Interpretation:

The coefficients table shows that blood glucose is a significant predictor of the likelihood of having a stroke ($z = 6.384$, $p < 0.001$), with an estimated coefficient of 1.9678. The positive coefficient indicates that higher blood glucose levels are associated with a higher probability of having a stroke.

The deviance residuals describe the difference between the predicted and observed values for the dependent variable. The values range from -0.8390 to 2.4034, with higher values indicating a larger difference between predicted and observed values.

The AIC (Akaike Information Criterion) is a measure of the model's goodness-of-fit, with lower values indicating a better fit. The AIC value for this model is 307.65.

In conclusion, the logistic regression model suggests that blood glucose levels are a significant predictor of the likelihood of having a stroke, with higher levels of blood glucose being associated with a higher probability of having a stroke. The effect size of blood glucose on stroke is larger than that of work status or age, as indicated by the larger coefficient and the larger difference between null and residual deviance.

4. For blood glucose: low, normal, pre-diabetic, diabetic

```
model <- glm(stroke ~ blood_glucose2, data = bs.samp,
             family = "binomial"(link = "logit"))
summary(model)
```

```
##
## Call:
## glm(formula = stroke ~ blood_glucose2, family = binomial(link = "logit"),
##     data = bs.samp)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.8454  -0.3143  -0.3143  -0.3143   2.8529
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -4.0523     0.3677  -11.02  < 2e-16 ***
## blood_glucose2   1.0691     0.1620    6.60  4.1e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 342.22  on 539  degrees of freedom
## Residual deviance: 294.00  on 538  degrees of freedom
## AIC: 298
##
## Number of Fisher Scoring iterations: 6
```

Interpretation:

The model indicates that for a one unit increase in blood_glucose2, the odds of having a stroke increase by a factor of exp(1.0691) = 2.91, holding all other predictors constant. The intercept of -4.0523 represents the log odds of having a stroke when blood_glucose2 is equal to zero.

The p-value for the blood_glucose2 coefficient is less than 0.001, indicating that it is statistically significant at a 0.05 level of significance. The null deviance is the deviance of the model with only the intercept, and the residual deviance is the deviance of the model with the predictor variable. The difference between the null and residual deviances suggests that the blood_glucose2 variable improves the model fit.

The AIC value of 298 is lower than the AIC values for the other models, indicating that this model has the best balance between model fit and complexity among the four models. Overall, the model suggests that blood glucose level is a significant predictor of stroke risk.

Looking at the output of the for the logistic regression models with blood_glucose and blood_glucose2, we can see that the AIC for the blood_glucose model is 307.65, while the AIC for the blood_glucose2 model is 298. Therefore, the blood_glucose2 model is considered better because it has a lower AIC.

Therefore, including blood_glucose2 for combined and interaction effect.

2. For combined effect

```
model <- glm(stroke ~ age_groups + work + blood_glucose2,
             data = bs.samp, family = "binomial"(link = "logit"))
summary(model)
```

```
##
## Call:
## glm(formula = stroke ~ age_groups + work + blood_glucose2, family = binomial(link = "logit"),
##     data = bs.samp)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0159  -0.4346  -0.2834  -0.1718   2.5794
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -7.4546     1.2533  -5.948 2.72e-09 ***
## age_groups       0.9183     0.3056   3.005  0.00266 **
## work            -0.1423     0.2841  -0.501  0.61638
## blood_glucose2   0.8709     0.1684   5.171 2.33e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 342.22  on 539  degrees of freedom
## Residual deviance: 282.00  on 536  degrees of freedom
## AIC: 290
##
## Number of Fisher Scoring iterations: 6
```

Interpretation:

The coefficients show that age and blood glucose levels are significant predictors of stroke occurrence, while work is not as significant as the prior two. Specifically, for every one-unit increase in age groups, the odds of stroke increase by a factor of 2.5. For every one-unit increase in blood glucose levels, the odds of stroke increase by a factor of 2.4.

Final Conclusion: The current model has the least AIC score i.e., 290 compared to previous models. Therefore, the model with combined effect of work, age groups, and blood glucose levels is the best model so far.

3. For interaction effect

```
model <- glm(stroke ~ age_groups * work * blood_glucose2,
             data = bs.samp, family = "binomial"(link = "logit"))
summary(model)
```

```
##
## Call:
## glm(formula = stroke ~ age_groups * work * blood_glucose2, family = binomial(link = "logit"),
##     data = bs.samp)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.54928  -0.51072  -0.21194  -0.08197   2.75946
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    -12.7519    13.8471  -0.921    0.357
## age_groups                       1.7202     2.9546   0.582    0.560
## work                            -1.7135     6.2873  -0.273    0.785
## blood_glucose2                  -1.2923     6.4558  -0.200    0.841
## age_groups:work                  0.4971     1.3266   0.375    0.708
## age_groups:blood_glucose2        0.6941     1.3881   0.500    0.617
## work:blood_glucose2              3.2156     2.9722   1.082    0.279
## age_groups:work:blood_glucose2  -0.7973     0.6352  -1.255    0.209
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 342.22  on 539  degrees of freedom
## Residual deviance: 264.34  on 532  degrees of freedom
## AIC: 280.34
##
## Number of Fisher Scoring iterations: 10
```

Interpretation:

Intercept: -12.7519, indicating the log odds of stroke for the reference category when all other predictors are at zero.

Age groups: The coefficient for age groups is 1.7202, which is not statistically significant (p = 0.560), suggesting that the effect of age groups on stroke occurrence is not significant after accounting for the other predictors in the model.

Work: The coefficient for work is -1.7135, which is also not statistically significant (p = 0.785), suggesting that work may not be a significant predictor of stroke occurrence in this model.

Blood glucose levels: The coefficient for blood glucose levels is -1.2923, which is not statistically significant (p = 0.841), indicating that blood glucose levels may not be a significant predictor of stroke occurrence in this model.

Age groups:work: The interaction term between age groups and work is 0.4971, which is not statistically significant (p = 0.708), suggesting that the effect of the interaction between age groups and work on stroke occurrence is not significant after accounting for the other predictors in the model.

Age groups:blood_glucose2: The interaction term between age groups and blood glucose levels is 0.6941, which is not statistically significant (p = 0.617), suggesting that the effect of the interaction between age groups and blood glucose levels on stroke occurrence is not significant after accounting for the other predictors in the model.

Work:blood_glucose2: The interaction term between work and blood glucose levels is 3.2156, which is not statistically significant (p = 0.279), indicating that the effect of the interaction between work and blood glucose levels on stroke occurrence may not be significant in this model.

Age groups:work:blood_glucose2: The interaction term between age groups, work, and blood glucose levels is -0.7973, which is not statistically significant (p = 0.209), suggesting that the effect of the interaction between age groups, work, and blood glucose levels on stroke occurrence is not significant after accounting for the other predictors in the model.

Overall, the model does not provide strong evidence that any of the included variables or interaction terms are significant predictors of stroke occurrence after accounting for the other variables in the model. However, it is important to note that this model may not be the best-fitting or most appropriate model for this data set, and further analysis may be needed to determine the most appropriate predictors of stroke occurrence.

Studying interaction effect of hypertension and blood glucose on work type's association with stroke:

```
model <- glm(stroke ~ work * hypertension * blood_glucose2,
             data = bs.samp, family = "binomial"(link = "logit"))
summary(model)
```

```
##
## Call:
## glm(formula = stroke ~ work * hypertension * blood_glucose2,
##     family = binomial(link = "logit"), data = bs.samp)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1710  -0.4416  -0.2911  -0.1681   2.9214
##
## Coefficients:
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                        -7.7361     1.4482  -5.342 9.19e-08 ***
## work                                1.7415     0.5915   2.944  0.00324 **
## hypertension                        2.4800     4.1502   0.598  0.55013
## blood_glucose2                      2.8734     0.7164   4.011 6.05e-05 ***
## work:hypertension                  -0.8358     1.8845  -0.443  0.65741
## work:blood_glucose2                -0.8802     0.3070  -2.867  0.00415 **
## hypertension:blood_glucose2        -1.4210     1.7832  -0.797  0.42553
## work:hypertension:blood_glucose2    0.6376     0.8187   0.779  0.43611
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 342.22  on 539  degrees of freedom
## Residual deviance: 280.25  on 532  degrees of freedom
## AIC: 296.25
##
## Number of Fisher Scoring iterations: 7
```

Interpretation:

The intercept coefficient is -7.7361, indicating that the log odds of stroke when all the independent variables are at 0 is -7.7361.

The coefficient for work is 1.7415, indicating that the log odds of stroke increase by 1.7415 when work increases by 1 unit, holding all other variables constant.

The coefficient for hypertension is 2.4800, indicating that the log odds of stroke increase by 2.4800 when hypertension is present, holding all other variables constant.

The coefficient for blood glucose is 2.8734, indicating that the log odds of stroke increase by 2.8734 when blood glucose increases by 1 unit, holding all other variables constant.

The coefficient for work:hypertension is -0.8358, indicating that the effect of work on stroke is moderated by hypertension such that the log odds of stroke decrease by 0.8358 when work and hypertension increase by 1 unit, holding blood glucose constant.

The coefficient for work:blood_glucose2 is -0.8802, indicating that the effect of work on stroke is moderated by blood glucose such that the log odds of stroke decrease by 0.8802 when work and blood glucose increase by 1 unit, holding hypertension constant.

The coefficient for hypertension:blood_glucose2 is -1.4210, indicating that the effect of hypertension on stroke is moderated by blood glucose such that the log odds of stroke decrease by 1.4210 when hypertension and blood glucose increase by 1 unit, holding work constant.

The coefficient for work:hypertension:blood_glucose2 is 0.6376, indicating that the effect of work on stroke is moderated by both hypertension and blood glucose such that the log odds of stroke increase by 0.6376 when work, hypertension, and blood glucose increase by 1 unit.

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.2.3

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 4.2.3

## Loading required package: lattice

##
## Attaching package: 'caret'

## The following object is masked from 'package:sampling':
##
##     cluster
```

```r
# Split the data into training and testing sets (70% training, 30% testing)
set.seed(123)
trainIndex <- createDataPartition(bs.samp$stroke, p = 0.7, list = FALSE)
train_set <- bs.samp[trainIndex, ]
test_set <- bs.samp[-trainIndex, ]


# Model : For all variables
model4 <- glm(stroke ~ age_groups + work + blood_glucose2, data = bs.samp, family = "binomial"(link = "l
pred_probs4 <- predict(model4, newdata = test_set, type = "response")
pred_labels4 <- ifelse(pred_probs4== "Yes", 1, 0)
confusionMatrix4 <- confusionMatrix(factor(pred_labels4), factor(test_set$stroke))
```

```
## Warning in confusionMatrix.default(factor(pred_labels4),
## factor(test_set$stroke)): Levels are not in the same order for reference and
## data. Refactoring data to match.
```

```r
print(confusionMatrix4)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 147  15
##          1   0   0
##
##                Accuracy : 0.9074
##                  95% CI : (0.8519, 0.9472)
##     No Information Rate : 0.9074
##     P-Value [Acc > NIR] : 0.5681702
##
##                   Kappa : 0
##
##  Mcnemar's Test P-Value : 0.0003006
##
##             Sensitivity : 1.0000
##             Specificity : 0.0000
##          Pos Pred Value : 0.9074
##          Neg Pred Value :    NaN
##              Prevalence : 0.9074
##          Detection Rate : 0.9074
##    Detection Prevalence : 1.0000
##       Balanced Accuracy : 0.5000
##
##        'Positive' Class : 0
##
```