HW5_Madhu

Madhu Jagdale

3/3/2021

```
library(tigerstats)
```

```
## Loading required package: abd
```

```
## Loading required package: nlme
```

```
## Loading required package: lattice
```

```
## Loading required package: grid
```

```
## Loading required package: mosaic
```

```
## Registered S3 method overwritten by 'mosaic':
##   method                           from
##   fortify.SpatialPolygonsDataFrame ggplot2
```

```
##
## The 'mosaic' package masks several functions from core packages in order
to add
## additional features.  The original behavior of these functions should not
be affected by this.
```

```
##
## Attaching package: 'mosaic'
```

```
## The following objects are masked from 'package:dplyr':
##
##     count, do, tally
```

```
## The following object is masked from 'package:Matrix':
##
##     mean
```

```
## The following object is masked from 'package:ggplot2':
##
##     stat
```

```
## The following objects are masked from 'package:stats':
##
##     binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,
##     quantile, sd, t.test, var
```

```
## The following objects are masked from 'package:base':
##
##     max, mean, min, prod, range, sample, sum

## Welcome to tigerstats!
## To learn more about this package, consult its website:
##  http://homerhanumat.github.io/tigerstats
```

Part ONE: Clarifying the concepts

1. Explore the t-value and the t distribution

a) What does t0.975=2.144787 mean? Assume that T is a random variable. that follows a t-distribution with n−1=14 degrees of freedom.

Answer: *T takes values less than or equal to 2.144787 exactly for 97.5%. There is a sample size of 15. We can not use CLT because Signa value is not mentioned.*

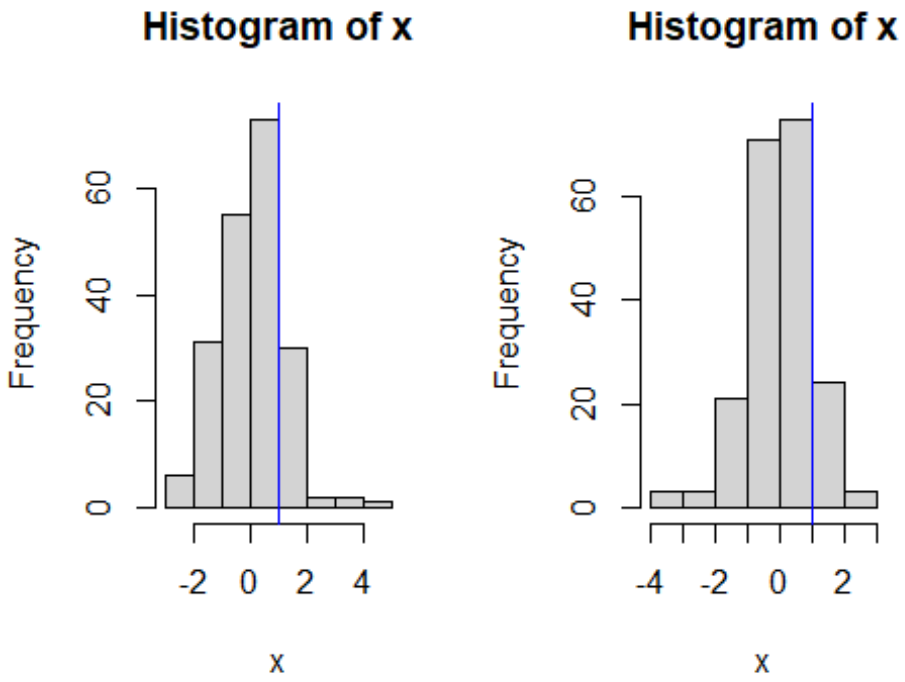
b) Checking the t-value of -2.145 using tigerstats or the tdist function:

Answer:

If we changes degree of freedom to n-1=140, Histogram will look like following:

```
par(mfrow = c(1, 2))
x =  x = rt(200,df=14)
hist(x)
abline(v=1,col="blue")


x =  x = rt(200,df=140)
hist(x)
abline(v=1,col="blue")
```



The curve for degree of freedom of n-1=14 and n-1=140:

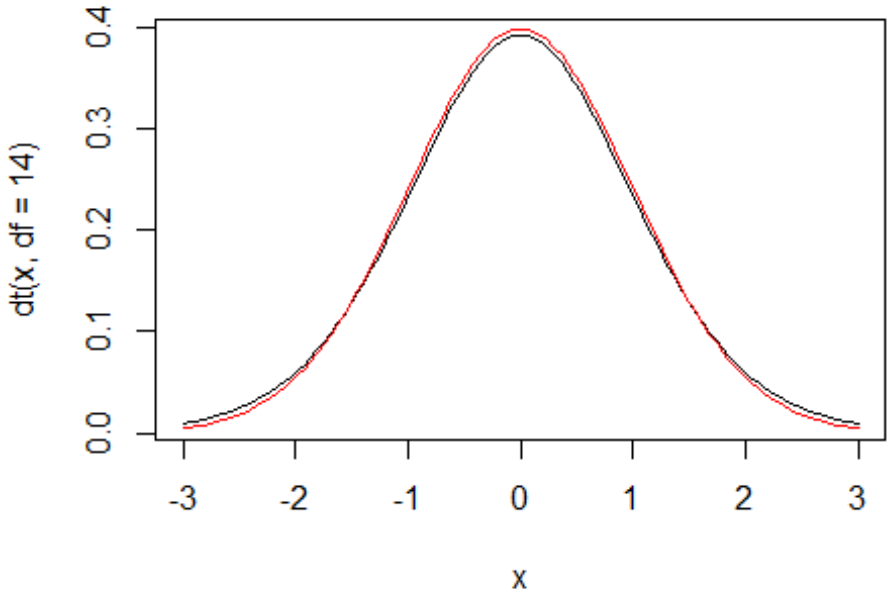
```
qt(0.975, df=14)

## [1] 2.144787

qt(0.975, df=140)

## [1] 1.977054

curve(dt(x, df=14), from=-3, to=3, col="black")
curve(dt(x, df=140), from=-3, to=3, col="red", add=TRUE)
```


dt(x, df = 14)
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3
x

c) What happens if you change the degrees of freedom to, say, n−1=140? Does the probability corresponding to the t-value go up or down?

Answer: *If we the degrees of freedom to, say, n−1=140, the probability corresponding to the t-value goes down.*

By using abline() function we can add one or more horizontal, vertical and regression lines to plot.

2. Review the approach to location problems for one and two populations

a) For inference on population mean, which of the following could we potentially use?

Answer: *It is depending on what data we have about pupulation we can use The t-distribution (with the T statistic) or The normal distribution (with the Z statistic).*

b) For inference on population mean when population variance is known, which of the following should we use? (In this part, suppose X_1, … , X_1000 is a random sample (of size 1000) from some unknown distribution.)

Answer: *The normal distribution (with the Z statistic). In CLT x values does not matter, Sample mean is always centered and normal to the pupulation mean.*

c) For inference on population mean when population variance is unknown, which of the following should we use? (In this part, suppose X_1, … , X_1000 is a random sample (of size 1000) from some unknown distribution.)

Answer: *The t-distribution (with the T statistic), but ONLY if X comes from a normal distribution. When population variance is unknown We uses s instead of sigma.*

d) A maker of a certain brand of low-fat cereal bars claims that the average saturated fat content is 0.5 gram. In a random sample of 8 cereal bars of this brand, the saturated fat content was 0.6, 0.7, 0.7, 0.3, 0.4, 0.5, 0.4, and 0.2. Would you agree with the claim? Assume a normal distribution.

Answer: *Mean=0.5 ,*

Sample mean = sum of all the data/total number of data

Sample mean = (0.6 + 0.7 + 0.7 + 0.3 + 0.4 + 0.5 + 0.4 + 0.2)/8 = 3.8/8 = 0.475

*Sample Standard deviation : sqrt((1/n-1)*square(SUM(x-$\bar{x}$))) = sqrt(0.033571428571429) = 0.18322507626258*

Population Standard deviation is unknown. Lets use the t-test:

t = (0.475-0.5)/(0.1833/sqrt(n)) = -0.386

df = n-1 = 7 2(P(t<-0.386)) = 0.7081 = p value (high value)

The claim of average saturated fat content 0.5 is not correct.

e) If you are interested in a difference of means between two populations, what should you keep in mind? [Select all that apply.] (Suppose X_1, … , X_1000 is a random sample (of size 1000) from one population, and Y_1, …, Y_500 a random sample (of size 500) from another population.)

Answer: *If we are interested in a difference of means between two populations, we should keep in mind the followings:*

The difference of means and $\bar{x}$-$\bar{y}$ are normal after re-scaling.

Sample size not much matter, Even If sample size is large enough, we can use Z.

Part TWO: Working with small data sets

3. Checking out some small data sets that come with R

Two data samples are independent if they come from unrelated populations and the samples does not affect each other. Here, we assume that the data populations follow the normal distribution. In the data frame column mpg (which stands for "miles per galon") of the data set mtcars, there are gas mileage data of various 1974 U.S. automobiles. Let's take a look:

```
mtcars$mpg
```

```
##  [1] 21.0 21.0 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 17.8 16.4 17.3 15.2
10.4
## [16] 10.4 14.7 32.4 30.4 33.9 21.5 15.5 15.2 13.3 19.2 27.3 26.0 30.4 15.8
19.7
## [31] 15.0 21.4
```

Meanwhile, another data column in mtcars, named am, indicates the transmission type of the automobile model (0 = automatic, 1 = manual):

```
mtcars$am
```

```
##  [1] 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1
```

```
L = mtcars$am == 0
mpg.auto <- mtcars[L,]$mpg
mpg.auto
```

```
##  [1] 21.4 18.7 18.1 14.3 24.4 22.8 19.2 17.8 16.4 17.3 15.2 10.4 10.4 14.7
21.5
## [16] 15.5 15.2 13.3 19.2
```

```
mpg.manual <- mtcars[!L,]$mpg
mpg.manual
```

```
##  [1] 21.0 21.0 22.8 32.4 30.4 33.9 27.3 26.0 30.4 15.8 19.7 15.0 21.4
```

```
mean.diff <- mean(mpg.manual)-mean(mpg.auto)
mean.diff
```

```
## [1] 7.244939
```

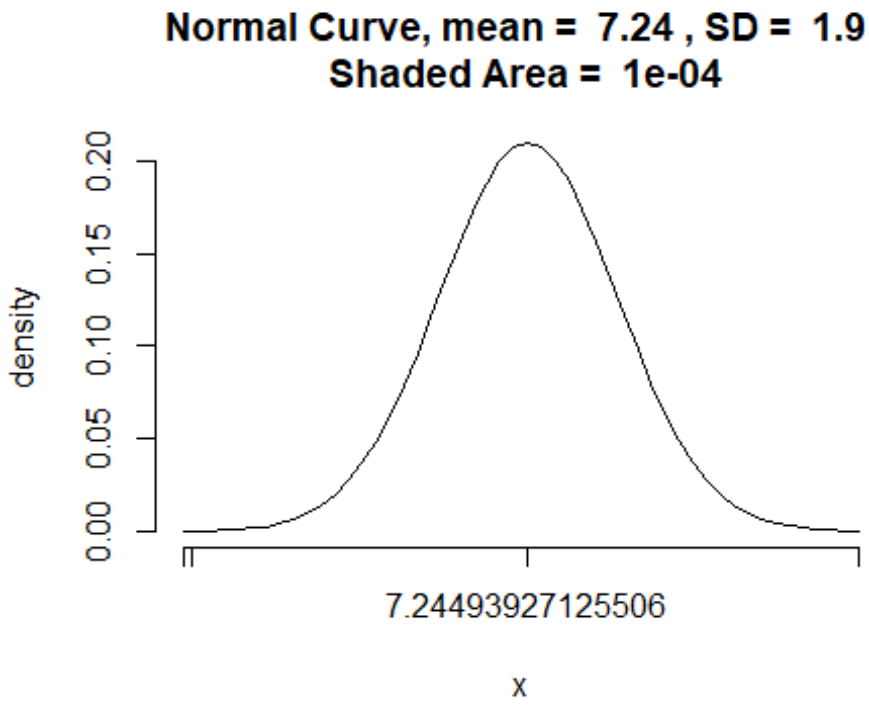
```
fac <- ((16/length(mpg.auto))+(36/length(mpg.manual)))
fac
```

```
## [1] 3.611336
```

```
std <- sqrt(fac)
std
```

```
## [1] 1.900352
```

```
pnormGC(0.05,mean=mean.diff,sd=std,region="below",graph=T)
```




```
## [1] 7.65121e-05
```

By looking at the curve we can say that the difference between the mean gas mileage of manual and automatic transmissions seems to be statistically significant.

4. Review the approach to scale problems for one and two populations

For inference on population variance, which of the following distributions will be useful?

Answer: *χ2*

For comparing variances between two populations, which of the following distributions will be useful?

Answer: *F*

Why do we use a quantile plot?

Answer: *As a heuristic test whether sample might be coming from a normal population. We only need to do this when we have a small sample or don't know population variances.*

5. Functions

```
dataset <- read.csv("https://campuspro-uploads.s3-us-west-
2.amazonaws.com/a9d789c2-6b5e-4020-a941-69984947f1ee/d2c0b7ab-df96-4891-b40f-
392d348c30dc/bank_marketing_training")
bank_train <- dataset
head(bank_train)

##   age         job  marital   education default housing loan   contact
month
## 1  56   housemaid  married    basic.4y      no      no   no telephone
may
## 2  57    services  married high.school unknown      no   no telephone
may
## 3  41 blue-collar  married     unknown unknown      no   no telephone
may
## 4  25    services   single high.school      no     yes   no telephone
may
## 5  29 blue-collar   single high.school      no      no  yes telephone
may
## 6  57   housemaid divorced    basic.4y      no     yes   no telephone
may
##   day_of_week duration campaign days_since_previous previous
previous_outcome
## 1         mon      261        1                999        0
nonexistent
## 2         mon      149        1                999        0
nonexistent
## 3         mon      217        1                999        0
nonexistent
## 4         mon      222        1                999        0
nonexistent
```

```
## 5         mon      137        1                 999        0
nonexistent
## 6         mon      293        1                 999        0
nonexistent
##   emp.var.rate cons.price.idx cons.conf.idx euribor3m nr.employed response
## 1          1.1         93.994         -36.4     4.857        5191       no
## 2          1.1         93.994         -36.4     4.857        5191       no
## 3          1.1         93.994         -36.4     4.857        5191       no
## 4          1.1         93.994         -36.4     4.857        5191       no
## 5          1.1         93.994         -36.4     4.857        5191       no
## 6          1.1         93.994         -36.4     4.857        5191       no
```

Lets calculate sample mean and sample variance of age:

```
mean(bank_train$age)

## [1] 39.98776

var(bank_train$age)

## [1] 108.1659
```

a) Tasks:

1) Write a function in R to a small sample from your data set (say, 100 entries):

Answer:

```
small.sample <- sample(1:nrow(bank_train), 100)
small.sample

##   [1] 12321 19436 11395 18399 17885  5595  4381  6391  7346 26512 12725
25081
##  [13]  6876 17767  5615 13649 10039 14833 24747 12600  3627  5879  4655
10189
##  [25]  5858 14100 17863 18301 23101  1253  9975 11493  3767  1039 12676
22287
##  [37]  2912  2402 10255 25872 25250 15182 22809 25798 24170  3993  5315
11001
##  [49] 10760 19495  4560   425 14813 13612 21833 18589  2725 22429  7194
23568
##  [61] 11379 10521  9837  9598 24592 20777 21351  9625  1933 17926 21188
8618
##  [73]  7133 25469  1676  9238 26759 17252  3790 10571 15762 22606  6833
17429
##  [85] 24452 11312 21401  8022 10381 26000  5077  3513  9746 23315 10473
4427
##  [97] 14568 14521  7912 12224

sample <- bank_train[small.sample, ]
sample
```

```
##       age           job  marital           education default housing
loan
## 12321  43        admin.  married   university.degree      no      no
no
## 19436  38        admin.  unknown   university.degree      no      no
no
## 11395  43    technician divorced         high.school      no     yes
yes
## 18399  38  entrepreneur  married professional.course      no     yes
yes
## 17885  31        admin.   single         high.school      no     yes
yes
## 5595   42        admin.  married   university.degree      no      no
no
## 4381   45        admin.  married            basic.9y      no     yes
no
## 6391   38    technician  married         high.school      no      no
no
## 7346   38    technician   single professional.course      no      no
no
## 26512  34      services divorced         high.school      no     yes
no
## 12725  29    management  married   university.degree      no      no
yes
## 25081  54     housemaid  married professional.course      no      no
no
## 6876   34   blue-collar  married            basic.6y unknown      no
no
## 17767  51    technician   single            basic.9y      no     yes
no
## 5615   34  entrepreneur  married            basic.4y      no     yes
no
## 13649  33    technician  married         high.school      no      no
no
## 10039  32      services  married         high.school      no     yes
no
## 14833  31    management  married   university.degree      no      no
no
## 24747  31    management  married   university.degree      no     yes
yes
## 12600  34        admin.  married   university.degree      no      no
no
## 3627   50  entrepreneur divorced   university.degree      no     yes
no
## 5879   42   blue-collar  married            basic.4y unknown      no
no
## 4655   47        admin.   single             unknown unknown      no
no
## 10189  26   blue-collar   single            basic.9y unknown     yes
no
```

```
## 5858   50   blue-collar  married           basic.4y      no     yes
no
## 14100  43        admin.  married   university.degree      no      no
yes
## 17863  35        admin. divorced   university.degree      no     yes
no
## 18301  29        admin.  married   university.degree      no     yes
no
## 23101  33    technician   single   university.degree      no      no
no
## 1253   27      services   single        high.school      no      no
no
## 9975   33        admin.  married   university.degree      no      no
no
## 11493  56      services divorced        high.school unknown     yes
no
## 3767   38       unknown  married           basic.6y      no     yes
no
## 1039   38   blue-collar  married           basic.4y unknown     yes
yes
## 12676  59    technician  married professional.course      no      no
no
## 22287  29   blue-collar  married        high.school      no     yes
no
## 2912   38        admin.  married   university.degree      no     yes
no
## 2402   40   blue-collar  married           basic.6y unknown      no
no
## 10255  35        admin.   single   university.degree      no     yes
no
## 25872  52    technician  married   university.degree      no      no
no
## 25250  37  entrepreneur divorced        high.school      no     yes
yes
## 15182  37      services  married        high.school      no      no
no
## 22809  52 self-employed divorced   university.degree      no     yes
no
## 25798  59       retired divorced        high.school      no     yes
no
## 24170  41        admin.  married   university.degree      no     yes
no
## 3993   35  entrepreneur divorced   university.degree      no      no
no
## 5315   34   blue-collar  married           basic.6y unknown      no
no
## 11001  52       retired   single           basic.9y      no      no
no
## 10760  48      services  married           basic.9y      no     yes
no
```

```
## 19495  27   blue-collar  married            basic.9y unknown     yes
no
## 4560   56        admin. divorced             unknown unknown     yes
no
## 425    38   blue-collar  married            basic.9y      no     yes
no
## 14813  29        admin.   single   university.degree      no     yes
no
## 13612  33    technician  married professional.course unknown     yes
no
## 21833  45   blue-collar  married            basic.6y      no     yes
no
## 18589  30   blue-collar  married            basic.4y      no     yes
no
## 2725   46   blue-collar divorced            basic.4y      no      no
no
## 22429  24    technician  married professional.course      no     yes
yes
## 7194   28   blue-collar  married            basic.9y      no     yes
no
## 23568  25        admin.   single   university.degree      no      no
no
## 11379  44   blue-collar   single         high.school unknown      no
no
## 10521  38     housemaid  married            basic.4y unknown      no
no
## 9837   45        admin.   single   university.degree      no     yes
no
## 9598   58        admin.  married   university.degree      no     yes
no
## 24592  31        admin.   single         high.school      no     yes
no
## 20777  36    technician  married professional.course      no     yes
no
## 21351  30   blue-collar  married            basic.9y      no     yes
yes
## 9625   25   blue-collar   single         high.school      no     yes
no
## 1933   31    technician  married   university.degree      no      no
no
## 17926  41 self-employed  married professional.course unknown     yes
no
## 21188  42   blue-collar  married professional.course      no     yes
no
## 8618   48   blue-collar  married            basic.4y      no      no
no
## 7133   40   blue-collar divorced            basic.9y unknown unknown
unknown
## 25469  54    technician divorced   university.degree      no unknown
unknown
```

```
## 1676   34    management  married   university.degree      no     yes
no
## 9238   31         admin.   single            basic.9y      no      no
no
## 26759  37     housemaid  married   university.degree      no      no
no
## 17252  30         admin.   single   university.degree      no     yes
no
## 3790   36         admin.  married         high.school      no      no
no
## 10571  44         admin.  married   university.degree      no     yes
no
## 15762  31   blue-collar  married            basic.9y      no     yes
no
## 22606  24       student   single            basic.9y      no     yes
no
## 6833   45    management  married             unknown unknown      no
yes
## 17429  50      services  married             unknown      no     yes
no
## 24452  66    technician  married professional.course      no     yes
no
## 11312  35    unemployed   single            basic.9y      no     yes
yes
## 21401  34         admin.  unknown   university.degree      no      no
no
## 8022   46      services  married         high.school      no      no
no
## 10381  37         admin.   single         high.school      no     yes
no
## 26000  37    unemployed  married         high.school      no     yes
no
## 5077   52  entrepreneur  married            basic.9y      no     yes
yes
## 3513   53    technician divorced professional.course      no      no
no
## 9746   30   blue-collar  married            basic.9y      no     yes
no
## 23315  46         admin.  married         high.school      no     yes
yes
## 10473  31         admin.   single   university.degree      no     yes
no
## 4427   31    unemployed   single professional.course      no      no
no
## 14568  56         admin.  married         high.school unknown      no
no
## 14521  60 self-employed  married   university.degree      no      no
no
## 7912   34         admin.  married   university.degree      no      no
no
```

```
## 12224  48        admin.   single   university.degree      no     yes
yes
##         contact month day_of_week duration campaign days_since_previous
## 12321  cellular   aug         mon      116        1                  999
## 19436  cellular   apr         mon      517        1                  999
## 11395  cellular   jul         mon      200        1                  999
## 18399  cellular   apr         mon      278        2                  999
## 17885  cellular   nov         fri       50        1                  999
## 5595  telephone   jun         tue      604       14                  999
## 4381  telephone   may         wed      173        1                  999
## 6391  telephone   jun         mon      479        4                  999
## 7346  telephone   jun         thu      574        4                  999
## 26512 telephone   sep         mon       11        1                  999
## 12725  cellular   aug         wed      181        2                  999
## 25081 telephone   oct         mon     1745        3                  999
## 6876  telephone   jun         mon       75        7                  999
## 17767 telephone   nov         fri      218        3                  999
## 5615  telephone   jun         wed      250        1                  999
## 13649  cellular   aug         thu      152        4                  999
## 10039  cellular   jul         fri       91        2                  999
## 14833  cellular   aug         fri       58        3                  999
## 24747  cellular   aug         mon      271        2                  999
## 12600  cellular   aug         wed      224        1                  999
## 3627  telephone   may         mon      131        2                  999
## 5879  telephone   jun         thu      164        2                  999
## 4655  telephone   may         thu      127        1                  999
## 10189  cellular   jul         mon       82        4                  999
## 5858  telephone   jun         thu      152        1                  999
## 14100  cellular   aug         tue      209        1                  999
## 17863  cellular   nov         fri       28        3                  999
## 18301  cellular   apr         wed       73        1                  999
## 23101  cellular   may         fri      204        3                  999
## 1253  telephone   may         fri       86        4                  999
## 9975   cellular   jul         thu     1018        1                  999
## 11493  cellular   jul         mon      761        8                  999
## 3767  telephone   may         mon      240        1                  999
## 1039  telephone   may         fri      383        3                  999
## 12676  cellular   aug         wed      378        4                  999
## 22287  cellular   may         wed      177        5                  999
## 2912  telephone   may         tue      137        2                  999
## 2402  telephone   may         fri       33        1                  999
## 10255  cellular   jul         mon      156        1                  999
## 25872  cellular   may         thu      211        1                    3
## 25250  cellular   nov         wed      258        2                  999
## 15182  cellular   aug         tue      377        7                  999
## 22809  cellular   may         fri      158        1                  999
## 25798  cellular   apr         thu      247        2                  999
## 24170  cellular   jul         tue      826        1                  999
## 3993  telephone   may         tue      124        1                  999
## 5315  telephone   jun         mon      479        3                  999
```

```
## 11001 telephone   jul         thu      239        6                  999
## 10760  cellular   jul         wed      178        1                  999
## 19495  cellular   apr         mon      294        2                  999
## 4560  telephone   may         thu      134        1                  999
## 425   telephone   may         tue      276        2                  999
## 14813  cellular   aug         fri      602        7                  999
## 13612  cellular   aug         wed       31        1                  999
## 21833 telephone   may         tue      291        1                  999
## 18589  cellular   apr         thu      179        1                  999
## 2725  telephone   may         mon      235        7                  999
## 22429  cellular   may         thu      428        3                  999
## 7194  telephone   jun         wed      200        1                  999
## 23568  cellular   may         tue       80        1                  999
## 11379  cellular   jul         mon      699        2                  999
## 10521  cellular   jul         tue      136        2                  999
## 9837   cellular   jul         thu      544        1                  999
## 9598   cellular   jul         tue      275        4                  999
## 24592  cellular   aug         tue      243        2                  999
## 20777  cellular   may         thu      258        2                  999
## 21351  cellular   may         mon      231        1                  999
## 9625  telephone   jul         tue     1142        4                  999
## 1933  telephone   may         wed      326        3                  999
## 17926  cellular   nov         fri     1571        1                  999
## 21188  cellular   may         fri      262        2                  999
## 8618   cellular   jul         wed      128        1                  999
## 7133  telephone   jun         wed      104        1                  999
## 25469  cellular   dec         mon      164        1                  999
## 1676  telephone   may         tue      230        3                  999
## 9238   cellular   jul         mon      221        3                  999
## 26759  cellular   oct         thu      549        2                   11
## 17252  cellular   nov         thu       34        1                  999
## 3790  telephone   may         mon       76        2                  999
## 10571  cellular   jul         tue      263        2                  999
## 15762 telephone   nov         fri      115        1                  999
## 22606 telephone   may         thu       25        8                  999
## 6833  telephone   jun         mon       73        1                  999
## 17429  cellular   nov         thu      153        3                  999
## 24452 telephone   aug         wed      150        2                  999
## 11312  cellular   jul         fri       50        3                  999
## 21401  cellular   may         mon      257        1                  999
## 8022  telephone   jul         wed       83        3                  999
## 10381  cellular   jul         mon      291        1                  999
## 26000 telephone   jun         tue       29        1                  999
## 5077  telephone   may         fri      390        2                  999
## 3513  telephone   may         fri      327        2                  999
## 9746   cellular   jul         wed      547        2                  999
## 23315  cellular   may         mon      196        3                  999
## 10473  cellular   jul         tue       72        1                  999
## 4427  telephone   may         wed       83        1                  999
## 14568  cellular   aug         thu       87        2                  999
```

```
## 14521  cellular   aug         thu      115        1                 999
## 7912  telephone   jun         fri      198        2                 999
## 12224 telephone   jul         thu       55        6                 999
##       previous previous_outcome emp.var.rate cons.price.idx cons.conf.idx
## 12321        0      nonexistent          1.4         93.444         -36.1
## 19436        0      nonexistent         -1.8         93.075         -47.1
## 11395        0      nonexistent          1.4         93.918         -42.7
## 18399        1          failure         -1.8         93.075         -47.1
## 17885        0      nonexistent         -0.1         93.200         -42.0
## 5595         0      nonexistent          1.4         94.465         -41.8
## 4381         0      nonexistent          1.1         93.994         -36.4
## 6391         0      nonexistent          1.4         94.465         -41.8
## 7346         0      nonexistent          1.4         94.465         -41.8
## 26512        1          failure         -1.1         94.199         -37.5
## 12725        0      nonexistent          1.4         93.444         -36.1
## 25081        1          failure         -3.4         92.431         -26.9
## 6876         0      nonexistent          1.4         94.465         -41.8
## 17767        0      nonexistent         -0.1         93.200         -42.0
## 5615         0      nonexistent          1.4         94.465         -41.8
## 13649        0      nonexistent          1.4         93.444         -36.1
## 10039        0      nonexistent          1.4         93.918         -42.7
## 14833        0      nonexistent          1.4         93.444         -36.1
## 24747        0      nonexistent         -2.9         92.201         -31.4
## 12600        0      nonexistent          1.4         93.444         -36.1
## 3627         0      nonexistent          1.1         93.994         -36.4
## 5879         0      nonexistent          1.4         94.465         -41.8
## 4655         0      nonexistent          1.1         93.994         -36.4
## 10189        0      nonexistent          1.4         93.918         -42.7
## 5858         0      nonexistent          1.4         94.465         -41.8
## 14100        0      nonexistent          1.4         93.444         -36.1
## 17863        1          failure         -0.1         93.200         -42.0
## 18301        0      nonexistent         -1.8         93.075         -47.1
## 23101        1          failure         -1.8         92.893         -46.2
## 1253         0      nonexistent          1.1         93.994         -36.4
## 9975         0      nonexistent          1.4         93.918         -42.7
## 11493        0      nonexistent          1.4         93.918         -42.7
## 3767         0      nonexistent          1.1         93.994         -36.4
## 1039         0      nonexistent          1.1         93.994         -36.4
## 12676        0      nonexistent          1.4         93.444         -36.1
## 22287        0      nonexistent         -1.8         92.893         -46.2
## 2912         0      nonexistent          1.1         93.994         -36.4
## 2402         0      nonexistent          1.1         93.994         -36.4
## 10255        0      nonexistent          1.4         93.918         -42.7
## 25872        4          success         -1.8         93.876         -40.0
## 25250        1          failure         -3.4         92.649         -30.1
## 15182        0      nonexistent          1.4         93.444         -36.1
## 22809        0      nonexistent         -1.8         92.893         -46.2
## 25798        1          failure         -1.8         93.749         -34.6
## 24170        1          failure         -2.9         92.469         -33.6
## 3993         0      nonexistent          1.1         93.994         -36.4
```

```
## 5315         0      nonexistent          1.4         94.465         -41.8
## 11001        0      nonexistent          1.4         93.918         -42.7
## 10760        0      nonexistent          1.4         93.918         -42.7
## 19495        0      nonexistent         -1.8         93.075         -47.1
## 4560         0      nonexistent          1.1         93.994         -36.4
## 425          0      nonexistent          1.1         93.994         -36.4
## 14813        0      nonexistent          1.4         93.444         -36.1
## 13612        0      nonexistent          1.4         93.444         -36.1
## 21833        0      nonexistent         -1.8         92.893         -46.2
## 18589        0      nonexistent         -1.8         93.075         -47.1
## 2725         0      nonexistent          1.1         93.994         -36.4
## 22429        0      nonexistent         -1.8         92.893         -46.2
## 7194         0      nonexistent          1.4         94.465         -41.8
## 23568        1          failure         -1.8         92.893         -46.2
## 11379        0      nonexistent          1.4         93.918         -42.7
## 10521        0      nonexistent          1.4         93.918         -42.7
## 9837         0      nonexistent          1.4         93.918         -42.7
## 9598         0      nonexistent          1.4         93.918         -42.7
## 24592        0      nonexistent         -2.9         92.201         -31.4
## 20777        0      nonexistent         -1.8         92.893         -46.2
## 21351        0      nonexistent         -1.8         92.893         -46.2
## 9625         0      nonexistent          1.4         93.918         -42.7
## 1933         0      nonexistent          1.1         93.994         -36.4
## 17926        0      nonexistent         -0.1         93.200         -42.0
## 21188        0      nonexistent         -1.8         92.893         -46.2
## 8618         0      nonexistent          1.4         93.918         -42.7
## 7133         0      nonexistent          1.4         94.465         -41.8
## 25469        0      nonexistent         -3.0         92.713         -33.0
## 1676         0      nonexistent          1.1         93.994         -36.4
## 9238         0      nonexistent          1.4         93.918         -42.7
## 26759        4          success         -1.1         94.601         -49.5
## 17252        1          failure         -0.1         93.200         -42.0
## 3790         0      nonexistent          1.1         93.994         -36.4
## 10571        0      nonexistent          1.4         93.918         -42.7
## 15762        0      nonexistent         -0.1         93.200         -42.0
## 22606        0      nonexistent         -1.8         92.893         -46.2
## 6833         0      nonexistent          1.4         94.465         -41.8
## 17429        0      nonexistent         -0.1         93.200         -42.0
## 24452        0      nonexistent         -2.9         92.201         -31.4
## 11312        0      nonexistent          1.4         93.918         -42.7
## 21401        0      nonexistent         -1.8         92.893         -46.2
## 8022         0      nonexistent          1.4         93.918         -42.7
## 10381        0      nonexistent          1.4         93.918         -42.7
## 26000        0      nonexistent         -1.7         94.055         -39.8
## 5077         0      nonexistent          1.1         93.994         -36.4
## 3513         0      nonexistent          1.1         93.994         -36.4
## 9746         0      nonexistent          1.4         93.918         -42.7
## 23315        0      nonexistent         -1.8         92.893         -46.2
## 10473        0      nonexistent          1.4         93.918         -42.7
## 4427         0      nonexistent          1.1         93.994         -36.4
```

```
## 14568        0      nonexistent          1.4         93.444         -36.1
## 14521        0      nonexistent          1.4         93.444         -36.1
## 7912         0      nonexistent          1.4         94.465         -41.8
## 12224        0      nonexistent          1.4         93.918         -42.7
##       euribor3m nr.employed response
## 12321     4.970        5228       no
## 19436     1.405        5099       no
## 11395     4.962        5228       no
## 18399     1.466        5099       no
## 17885     4.021        5195       no
## 5595      4.864        5228       no
## 4381      4.857        5191       no
## 6391      4.961        5228       no
## 7346      4.961        5228       no
## 26512     0.882        4963       no
## 12725     4.967        5228       no
## 25081     0.739        5017       no
## 6876      4.960        5228       no
## 17767     4.021        5195       no
## 5615      4.864        5228       no
## 13649     4.964        5228       no
## 10039     4.957        5228       no
## 14833     4.964        5228       no
## 24747     0.821        5076      yes
## 12600     4.967        5228       no
## 3627      4.857        5191       no
## 5879      4.866        5228       no
## 4655      4.860        5191       no
## 10189     4.960        5228       no
## 5858      4.866        5228       no
## 14100     4.963        5228       no
## 17863     4.021        5195       no
## 18301     1.498        5099       no
## 23101     1.250        5099       no
## 1253      4.855        5191       no
## 9975      4.958        5228       no
## 11493     4.962        5228       no
## 3767      4.857        5191       no
## 1039      4.855        5191       no
## 12676     4.967        5228       no
## 22287     1.281        5099       no
## 2912      4.856        5191       no
## 2402      4.859        5191       no
## 10255     4.960        5228       no
## 25872     0.677        5008      yes
## 25250     0.719        5017       no
## 15182     4.965        5228       no
## 22809     1.250        5099       no
## 25798     0.644        5008       no
## 24170     1.044        5076       no
```

```
## 3993     4.857        5191       no
## 5315     4.865        5228       no
## 11001    4.962        5228       no
## 10760    4.963        5228       no
## 19495    1.405        5099       no
## 4560     4.860        5191       no
## 425      4.857        5191       no
## 14813    4.964        5228       no
## 13612    4.965        5228       no
## 21833    1.291        5099       no
## 18589    1.435        5099      yes
## 2725     4.858        5191       no
## 22429    1.266        5099       no
## 7194     4.962        5228       no
## 23568    1.266        5099       no
## 11379    4.962        5228      yes
## 10521    4.961        5228       no
## 9837     4.958        5228       no
## 9598     4.961        5228       no
## 24592    0.859        5076       no
## 20777    1.327        5099       no
## 21351    1.299        5099       no
## 9625     4.961        5228       no
## 1933     4.859        5191       no
## 17926    4.021        5195      yes
## 21188    1.313        5099       no
## 8618     4.962        5228       no
## 7133     4.962        5228       no
## 25469    0.717        5023       no
## 1676     4.856        5191       no
## 9238     4.962        5228       no
## 26759    1.025        4963      yes
## 17252    4.076        5195       no
## 3790     4.857        5191       no
## 10571    4.961        5228       no
## 15762    4.474        5195      yes
## 22606    1.266        5099       no
## 6833     4.960        5228       no
## 17429    4.076        5195       no
## 24452    0.879        5076       no
## 11312    4.962        5228       no
## 21401    1.299        5099       no
## 8022     4.956        5228       no
## 10381    4.960        5228       no
## 26000    0.713        4991       no
## 5077     4.864        5191       no
## 3513     4.857        5191       no
## 9746     4.957        5228       no
## 23315    1.244        5099       no
## 10473    4.961        5228       no
```

```
## 4427      4.857         5191        no
## 14568     4.963         5228        no
## 14521     4.963         5228        no
## 7912      4.947         5228        no
## 12224     4.968         5228        no
```

Lets check whether the person in the data set is over 40 years old

```
AgeGrouping <- function(x)
{
if (x>=40)
{
  age_group = "Yes"
}
  else {age_group = "No"}
age_group
}

sapply(sample$age, FUN=AgeGrouping)
```

```
##   [1] "Yes" "No"  "Yes" "No"  "No"  "Yes" "Yes" "No"  "No"  "No"  "No"
"Yes"
##  [13] "No"  "Yes" "No"  "No"  "No"  "No"  "No"  "No"  "Yes" "Yes" "Yes"
"No"
##  [25] "Yes" "Yes" "No"  "No"  "No"  "No"  "No"  "Yes" "No"  "No"  "Yes"
"No"
##  [37] "No"  "Yes" "No"  "Yes" "No"  "No"  "Yes" "Yes" "Yes" "No"  "No"
"Yes"
##  [49] "Yes" "No"  "Yes" "No"  "No"  "No"  "Yes" "No"  "Yes" "No"  "No"
"No"
##  [61] "Yes" "No"  "Yes" "Yes" "No"  "No"  "No"  "No"  "No"  "Yes" "Yes"
"Yes"
##  [73] "Yes" "Yes" "No"  "No"  "No"  "No"  "No"  "Yes" "No"  "No"  "Yes"
"Yes"
##  [85] "Yes" "No"  "No"  "Yes" "No"  "No"  "Yes" "Yes" "No"  "Yes" "No"
"No"
##  [97] "Yes" "Yes" "No"  "Yes"
```

Lets check whether the person in the data set falls into which age group.

```
AgeGrouping <- function(x)
{
if (x>=56)
{
  age_group = "Old_age"
}
else if (x>=36)
{
  age_group = "Middle_age"
}
  else if (x>=18)
{
```

```
  age_group = "Young_age"
}
else {age_group = "Teenager"}

age_group
}

sapply(sample$age, FUN=AgeGrouping)
```

```
##   [1] "Middle_age" "Middle_age" "Middle_age" "Middle_age" "Young_age"
##   [6] "Middle_age" "Middle_age" "Middle_age" "Middle_age" "Young_age"
##  [11] "Young_age"  "Middle_age" "Young_age"  "Middle_age" "Young_age"
##  [16] "Young_age"  "Young_age"  "Young_age"  "Young_age"  "Young_age"
##  [21] "Middle_age" "Middle_age" "Middle_age" "Young_age"  "Middle_age"
##  [26] "Middle_age" "Young_age"  "Young_age"  "Young_age"  "Young_age"
##  [31] "Young_age"  "Old_age"    "Middle_age" "Middle_age" "Old_age"
##  [36] "Young_age"  "Middle_age" "Middle_age" "Young_age"  "Middle_age"
##  [41] "Middle_age" "Middle_age" "Middle_age" "Old_age"    "Middle_age"
##  [46] "Young_age"  "Young_age"  "Middle_age" "Middle_age" "Young_age"
##  [51] "Old_age"    "Middle_age" "Young_age"  "Young_age"  "Middle_age"
##  [56] "Young_age"  "Middle_age" "Young_age"  "Young_age"  "Young_age"
##  [61] "Middle_age" "Middle_age" "Middle_age" "Old_age"    "Young_age"
##  [66] "Middle_age" "Young_age"  "Young_age"  "Young_age"  "Middle_age"
##  [71] "Middle_age" "Middle_age" "Middle_age" "Middle_age" "Young_age"
##  [76] "Young_age"  "Middle_age" "Young_age"  "Middle_age" "Middle_age"
##  [81] "Young_age"  "Young_age"  "Middle_age" "Middle_age" "Old_age"
##  [86] "Young_age"  "Young_age"  "Middle_age" "Middle_age" "Middle_age"
##  [91] "Middle_age" "Middle_age" "Young_age"  "Middle_age" "Young_age"
##  [96] "Young_age"  "Old_age"    "Old_age"    "Young_age"  "Middle_age"
```

2) Write a function in Python to a small sample from your data set (say, 100 entries):

Answer:

```
import pandas as pd
import numpy as np
dataset = pd.read_csv("https://campuspro-uploads.s3-us-west-
2.amazonaws.com/a9d789c2-6b5e-4020-a941-69984947f1ee/d2c0b7ab-df96-4891-b40f-
392d348c30dc/bank_marketing_training")
sample1 = dataset.head(100)
print(sample1)
```

```
##     age          job   marital  ... euribor3m nr.employed response
## 0    56    housemaid   married  ...     4.857        5191       no
## 1    57     services   married  ...     4.857        5191       no
## 2    41  blue-collar   married  ...     4.857        5191       no
## 3    25     services    single  ...     4.857        5191       no
## 4    29  blue-collar    single  ...     4.857        5191       no
## ..  ...          ...       ...  ...       ...         ...      ...
## 95   37  blue-collar    single  ...     4.857        5191       no
## 96   40  blue-collar   married  ...     4.857        5191       no
```

```
## 97   42  blue-collar   married  ...     4.857        5191       no
## 98   39     services  divorced  ...     4.857        5191       no
## 99   38     services   married  ...     4.857        5191       no
##
## [100 rows x 21 columns]
```