Homework_2

Madhu Jagdale

2/11/2021

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.5     v dplyr   1.0.3
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Q1. Types of random variables

Instructions

Review the definition of discrete and continuous random variables.

Tasks

Part A. Classify the following random variables as discrete or continuous:

X: the number of automobile accidents per year in Virginia.

Answer: Discrete

Y : the length of time to play 18 holes of golf.

Answer: Continuous

M: the amount of milk produced yearly by a particular cow.

Answer: Continuous

N: the number of eggs laid each month by a hen.

Answer: Discrete

P: the number of building permits issued each month in a certain city

Answer: Discrete

Q: the weight of grain produced per acre.

Answer: Continuous

Part B.

A coin is flipped until 3 heads in succession occur. List only those elements of the sample space that require 6 or less tosses. Is this a discrete sample space? Explain.

Answer:

List={HHH,THHH,TTHHH,TTTHHH,,HHTHHH,THTHHH,HTHHH,HTTHHH}

Yes, this is a discrete sample space as we can count the total number of outcomes in this set.

Q2. Sample space and experiment outcomes

Tasks

An overseas shipment of 5 foreign automobiles contains 2 that have slight paint blemishes. If an agency receives 3 of these automobiles at random, list the elements of the sample space S, using the letters B and N for blemished and nonblemished, respectively; then to each sample point assign a value x of the random variable X representing the number of automobiles with paint blemishes purchased by the agency.

Answer:

Let X be the number of automobiles with paint blemishes purchased by the agency.

The Automobile set be A={B1,B2,N1,N2,N3}

X={0,1,2}

The sample space = S

When X=0, then S={N1N2N3} When X=1, then S={B1N1N2,B2N1N2,B1N1N3,B2N1N3,B1N2N3,B2N2N3 When X=2, then S={B1B2N1,B1B2N2,B1B2N3}

Q3. Sample space and experiment outcomes, continued

Tasks

Let W be a random variable giving the number of heads minus the number of tails in three tosses of a coin. List the elements of the sample space S for the three tosses of the coin and to each sample point assign a value w of W.

Answer:

Let S be the sample space of tossing a coin three times. Let W be a random variable giving the number of heads minus the number of tails in three tosses of a coin.

S={HHH,HHT,HTT,HTH,TTT,TTH,THH,THT}

Now assigning random variable value to each sample point in the sample space : HHH-> W=3 HHT-> W=1 HTT-> W=-1 HTH-> W=1 TTT-> W=-3 TTH-> W=-1 THH-> W=1 THT-> W=-1

Q4

A certain polymer is used for evacuation systems for aircraft. It is important that the polymer be resistant to the aging process. Twenty specimens of the polymer were used in an experiment. Ten were assigned randomly to be exposed to an accelerated batch aging process that involved exposure to high temperatures for 10 days. Measurements of tensile strength of the specimens were made, and the following data were recorded on tensile strength in psi: No aging: 227 222 218 216 218 217 225 229 228 221 Aging: 219 214 218 203 215 211 209 204 201 205

(a) Do a dot plot of the data.

```
#### no_aging<-c(227,222,218,216,218,217,225,229,228,221)
```

```
#### aging<-c(219,214,218,203,215,211,209,204,201,205)
```

```
#### par(mfrow = c(2,3))
```

```
#### x=c(1:10)
```

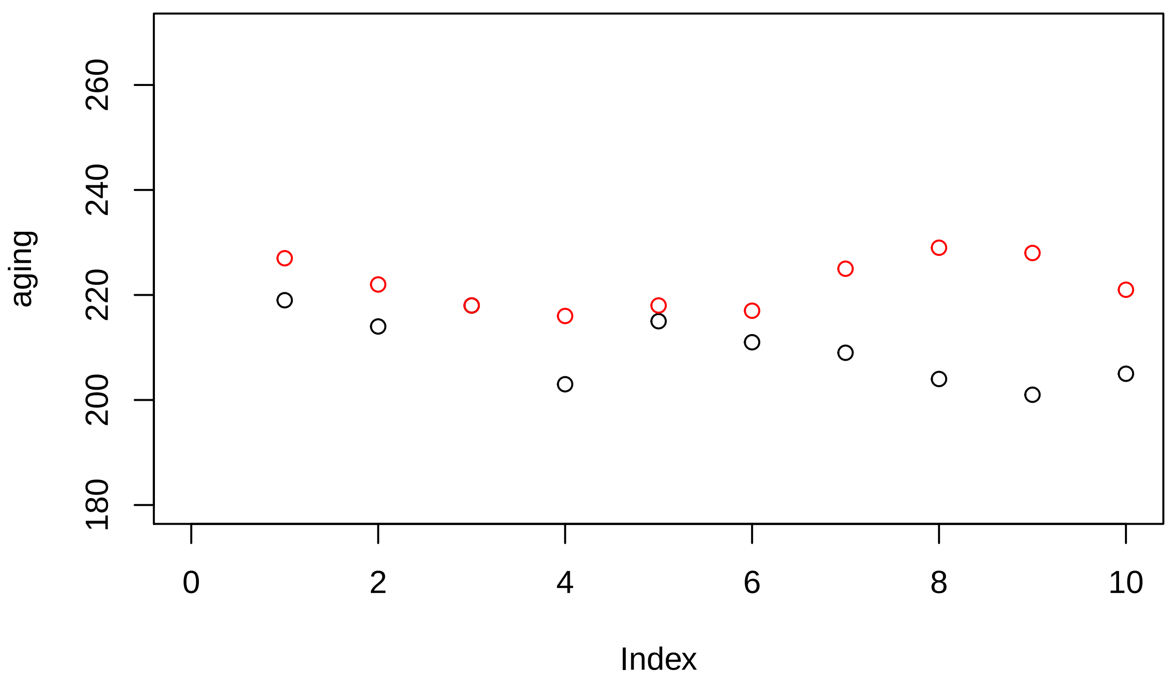
```r
#### plot(x, no_aging, col = "black", pch = 20, ylab="no_aging",xlab="specimens")
```

```r
#### plot(x, aging, col = "red", pch = 20, ylab="aging",xlab="specimens")
```

```r
no_aging<-c(227,222,218,216,218,217,225,229,228,221)
```

```r
aging<-c(219,214,218,203,215,211,209,204,201,205)
```

```r
plot(aging,col="black",xlim = c(0,10),ylim = c(180,270))
points(no_aging,col="red")
```



Answer:

(b) From your plot, does it appear as if the aging process has had an effect on the tensile strength of this polymer? Explain.

Answer:

From the dot plot, we can see that the ageing process has affected the tensile strength of the polymer because the least strength of the untreated polymer is more than one of the best strengths of the affected polymers. Moreover, the average strength of the untreated sample is 222.1 whereas the average strength of the treated polymer is 209.9 . The best polymer which is untreated has a strength of 229 units, whereas the best polymer of the treated group is only 219 units. Thus, we can conclude that the ageing process has had an effect on the tensile strength.

(c) Calculate the sample mean tensile strength of the two samples.

```
mean(no_aging)
```

Answer:

```
## [1] 222.1
```

```
mean(aging)
```

```
## [1] 209.9
```

(d) Calculate the median for both. Discuss the similarity or lack of similarity between the mean and median of each group

```
median(no_aging)
```

Answer:

```
## [1] 221.5
```

```
median(aging)
```

```
## [1] 210
```

The mean value in no_aging is higher than median value because data set has a symmetrical distribution, the mean and the median are close in aging because the middle value in the aging data set, when we arrange the dataset in ascending order we get the values as 209, 211.

Q5

The previous problem showed tensile strength data for two samples, one in which specimens were exposed to an aging process and one in which there was no aging of the specimens. (a) Calculate the sample variance as well as standard deviation in tensile strength for both samples.

```
summary(no_aging)
```

Answer:

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   216.0   218.0   221.5   222.1   226.5   229.0
```

```
summary(aging)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   201.0   204.2   210.0   209.9   214.8   219.0
```

```
var(no_aging)
```

```
## [1] 23.65556
```

```
var(aging)
```

```
## [1] 42.1
```

```
sd(no_aging)
```

```
## [1] 4.863698
```

```
sd(aging)
```

```
## [1] 6.488451
```

(b) Does there appear to be any evidence that aging affects the variability in tensile strength?

Answer: Variability is the amount of spread data around the mean. The variance before aging is 23.65556 and the variance after aging is 42.1. As per the values we can see that after aging variability is more.

Part Two: Working With Data

Tasks

For the following exercises, work with the bank_marketing_training data set. Use either Python or R to solve each problem.

Type a comment stating that you are working on a random data set we downloaded.

Locate the "Run" button and note whether there is a keyboard shortcut.

Execute the comment from the previous exercise. What is the output? Explain your answer.

Import the following packages:

For Python, import the pandas and numpy packages. Rename the pandas package "pd" and rename the numpy package "np". For R, import the ggplot2 package. Make sure you both install and open the package. Import the bank_marketing_training data set and name it bank_train.

```
dataset <- read_csv("https://campuspro-uploads.s3-us-west-2.amazonaws.com/a9d789c2-6b5e-4020-a941-69984
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_character(),
##   age = col_double(),
##   duration = col_double(),
##   campaign = col_double(),
##   days_since_previous = col_double(),
##   previous = col_double(),
##   emp.var.rate = col_double(),
##   cons.price.idx = col_double(),
##   cons.conf.idx = col_double(),
##   euribor3m = col_double(),
##   nr.employed = col_double()
## )
## i Use `spec()` for the full column specifications.
```

```
bank_train <- dataset
```

```
t1<-table(bank_train$response,bank_train$previous_outcome)
t1
```

Create a table of the response and previous_outcome variables from the bank_train data set. Do not save the output from the code.

```
##
##       failure nonexistent success
##   no     2390       21176     320
##   yes     385        2034     569
```

```
bank_train[c(1,2,3,4,5,6,7,8,9), ]
```

Save the first nine records of the bank_train data set as their own data frame.

```
## # A tibble: 9 x 21
##     age job   marital education default housing loan  contact month day_of_week
##   <dbl> <chr> <chr>   <chr>     <chr>   <chr>   <chr> <chr>   <chr> <chr>
## 1    56 hous~ married basic.4y  no      no      no    teleph~ may   mon
## 2    57 serv~ married high.sch~ unknown no      no    teleph~ may   mon
## 3    41 blue~ married unknown   unknown no      no    teleph~ may   mon
```

```
## 4    25 serv~ single  high.sch~ no      yes     no    teleph~ may   mon
## 5    29 blue~ single  high.sch~ no      no      yes   teleph~ may   mon
## 6    57 hous~ divorc~ basic.4y  no      yes     no    teleph~ may   mon
## 7    35 blue~ married basic.6y  no      yes     no    teleph~ may   mon
## 8    39 mana~ single  basic.9y  unknown no      no    teleph~ may   mon
## 9    30 unem~ married high.sch~ no      no      no    teleph~ may   mon
## # ... with 11 more variables: duration <dbl>, campaign <dbl>,
## #   days_since_previous <dbl>, previous <dbl>, previous_outcome <chr>,
## #   emp.var.rate <dbl>, cons.price.idx <dbl>, cons.conf.idx <dbl>,
## #   euribor3m <dbl>, nr.employed <dbl>, response <chr>
```

```r
am_data <- bank_train[, c("age","marital")]
am_data
```

Save the age and marital records of the bank_train data set as their own data frame.

```
## # A tibble: 26,874 x 2
##      age marital
##    <dbl> <chr>
##  1    56 married
##  2    57 married
##  3    41 married
##  4    25 single
##  5    29 single
##  6    57 divorced
##  7    35 married
##  8    39 single
##  9    30 married
## 10    55 single
## # ... with 26,864 more rows
```

```r
am <- bank_train[, c("age","marital")][1:3,]
am
```

Save the first three records of the age and marital variables as their own data frame.

```
## # A tibble: 3 x 2
##     age marital
##   <dbl> <chr>
## 1    56 married
## 2    57 married
## 3    41 married
```