**Inferential Statistics**

Instructions:

Please share your answers filled inline in the word document. Submit Python code and R code files wherever applicable.

Insights should be drawn from the plots about the data such as, is data normally distributed/not, outliers, measures like mean, median, mode, variance, std. deviation

Please ensure you update all the details:

**Name: ANANDAKRISHNAN K V**

**Batch Id: 190421**

**Topic: Basic Statistics**

**Problem Statements:**

Q1) Three Coins are tossed, find the probability that two heads and one tail are obtained?

Ans:-

Number of possible outcomes = 23 =8

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

Favourable events(2H & 1T) = {HHT, HTH, THH}

Number of favourable events = 3

**Probability of 2H & 1T condition** = (number of favourable outcomes)/ (number of possible events)

P(2H&1T) **= 3/8**

Q2) Two Dice are rolled, find the probability that sum is

1. Equal to 1
2. Less than or equal to 4
3. Sum is divisible by 2 and 3

Ans:-

Possible events = {(1,1), (1,2),(1,3),(1,4),(1,5),(1,6), (2,1), (2,2),(2,3),(2,4),(2,5),(2,6), (3,1), (3,2),(3,3),(3,4),(3,5),(3,6), (4,1), (4,2),(4,3),(4,4),(4,5),(4,6), (5,1), (5,2),(5,3),(5,4),(5,5),(5,6),(6,1), (6,2),(6,3),(6,4),(6,5),(6,6)}

Number of possible outcomes = 62 = 36

1. Number favourable events =0

**P(Sum=1) = 0**

1. Favourable events = {(1,1), (1,2),(1,3), (2,1), (2,2),(3,1)}

Number of faourable events = 6

**P(Sum<=4)** = 6/36 = **1//6**

1. **P(Sum is divisble by 2)** = {P(Sum=2) + P(Sum=4) + P(Sum=6) + P(Sum=8) + P(Sum=10) + P(Sum=12)}

= 1/36 + 3/36 + 5/36 + 5/36 + 3/36 + 1/36

= 18/36 = **½**

Possible events = {(1,1), (1,2),(1,3),(1,4),(1,5),(1,6), (2,1), (2,2),(2,3),(2,4),(2,5),(2,6), (3,1), (3,2),(3,3),(3,4),(3,5),(3,6), (4,1), (4,2),(4,3),(4,4),(4,5),(4,6), (5,1), (5,2),(5,3),(5,4),(5,5),(5,6),(6,1), (6,2),(6,3),(6,4),(6,5),(6,6)}

**P(Sum is divisble by 3) = {** P(Sum=3) + P(Sum=6) + P(Sum=9) + P(Sum=12)}

= 2/36 + 5/36 + 4/36 + 1/36

= 12/36 = **1/12**

**P(Sum is divisble by both 2 & 3) = {** P(Sum=6) + P(Sum=12)}

= 5/36 + 1/36 = 6/36 = **1/6**

Q3) A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?

Ans:-

Tota number of balls other than blue = Total Number of Red & Green balls,n = 5

Number of selections,r = 2

Total number of balls, N = 7 **combinations**

**nCr  = (n!)/(n-r)!\*(r!)**

number of chances getting None of the balls is blue = **nCr**

= 5C2 = 5!/[(5-2)!\*(2!)]

= 10

Total number of posible events = NCr = 7C2 = 21

Probability of getting none of them is blue, P(niether blue) = (2C2)/(7C2)

= **10/21**

Q4) Calculate the Expected number of candies for a randomly selected child:

Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)

i. Child A – probability of having 1 candy is 0.015

ii. Child B – probability of having 4 candies is 0.2

|  |  |  |
| --- | --- | --- |
| CHILD | Candies count | Probability |
| A | 1 | 0.015 |
| B | 4 | 0.20 |
| C | 3 | 0.65 |
| D | 5 | 0.005 |
| E | 6 | 0.01 |
| F | 2 | 0.12 |

Ans:-

Expected number of candies for a randomly selected child = sum of expected values of having candies for all childs = ∑ (X).P(X)

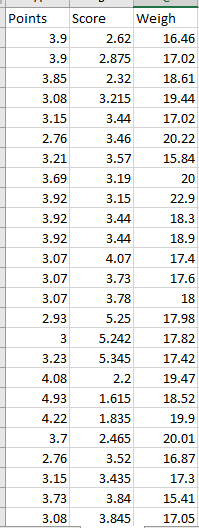
= 1\*.015 + 4\*0.20 + 3 \* .65 + 5 \*.005 + 6\* .01 + 2\*.12

= **3.09**

Q5) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points, Score, Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and comment about the values/ Draw some inferences.



Dataset: Refer to Hands-on Material In LMS - Data Types EDA assignment snap shot of dataset is given above.

**Ans:-**

**Rcode:-**

####inferential statistics\_Q5\_Score

library(readxl)

library(readr)

### import data set Assignment\_module02(1)

plot1 <- as.data.frame(Assignment\_module02\_1\_)

### select the column name as required

### analysis on score column

p <- plot1$Score

hist(p)

a <- list("mean=", "median=","mode=")

b <- list("variance=", "std. deviation=", "minimum & maximum =", "range=")

c <- list("skewness=", "excess kurtosis=", "outliers=", "number of outliers=")

#custom mode function

mode <- function(m)

{

m1 <- unique(m)

m2 <- match(m,m1)

m3 <- tabulate(m2)

m4 <- max(m3)

m5 <- m1[m4==m3]

}

md <- mode(p)

m1 <- unique(p)

#first moment business decision

a[[1]][2] <- mean(p)

a[[2]][2] <- median(p)

ml <- length(md)+1

for(i in 2:ml)

{

a[[3]][i] <- md[i-1]

}

##no mode condition

nm <- length(m1)==length(md)

if(nm==TRUE)

{

a[3] <- ("mode:- all unique values are repeating equal number of times" )

}

#second moment business decision

b[[1]][2] <- var(p)

b[[2]][2] <- sd(p)

b[[3]][2:3] <- range(p,na.rm=FALSE)

maximum <- as.numeric(b[[3]][3])

minimum <- as.numeric(b[[3]][2])

b[[4]][2] <- maximum-minimum

#install e1071 package

library(e1071)

#third & fourth moment business decision

bx <- boxplot(p)

##skewness###

c1 <- skewness(p)

print(mean(p))

print(median(p))

s <- c(1:3)

s[1]= mean(p)>median(p)

s[2]= mean(p)<median(p)

s[3]= mean(p)==median(p)

sk <- c("+ve skewness: because mean > median", "-ve skewness: because mean<median ", "normal or zero: mean = median ")

for(i in 1:3)

{

if(s[i]==TRUE)

{

actual\_skew <- sk[i]

}

}

##kurtosis

c2 <- kurtosis(p)

k <- c(1:3)

k[1]= c2>0

k[2]= c2<0

k[3]= c2==0

ku <- c("leptokurtic", "platykurtic", "mesokurtic")

for(i in 1:3)

{

if(k[i]==TRUE)

{

kurt <- ku[i]

}

}

##outliers

c3 <- bx$out

c4 <- length(bx$out)+1

for(i in 2:c4)

{

c[[3]][i] <- bx$out[i-1]

}

c[[1]][2]=c1

c[[2]][2]=c2

c[[4]][2]=c4-1

##normality checking

qq <- qqnorm(p)

qq1 <- qqline(p)

###conclusion

#first moment business decision

print("first moment business decisions are")

print(a)

##second moment business decision

print("second moment business decisions are")

print(b)

##third & fourth moment business decisions & conclusions

print("third & fourth moment business decisions are")

print(c)

print("actual skewness is")

print(actual\_skew)

print("type of kurtosis is")

print(kurt)

**Analysis:-**

**a)Points :-**

**FIRST MOMENT BUSINESS DECISIONS:-**

"mean=" "3.5965625"

"median=" "3.695"

"mode=" "3.92" "3.07"

**SECOND MOMENT BUSINESS DECISIONS:-**

"variance=" "0.285881350806452"

"std. deviation=" "0.534678736070971"

"minimum & maximum =" "2.76" & "4.93"

"range=" "2.17"

**b) Score:-**

**FIRST MOMENT BUSINESS DECISIONS:-**

"mean=" "3.2115625"

"median=" "3.325"

"mode=" "3.44"

**SECOND MOMENT BUSINESS DECISIONS:-**

"variance=" "0.932502576612903"

"std. deviation=" "0.965661729910067"

"minimum & maximum =" "1.513" & "5.345"

"range=" "3.832"

**3)Weigh:-**

**FIRST MOMENT BUSINESS DECISIONS:-**

"mean=" "3.5965625"

"median=" "3.695"

"mode=" "3.92" "3.07"

**SECOND MOMENT BUSINESS DECISIONS:-**

"variance=" "3.19316612903226"

"std. deviation=" "1.78694323609684"

"minimum & maximum =" "14.5" "22.9"

"range=" "8.4"

Q6) Calculate Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

Ans:-

Total number of patients = 9

So, Probabillity of each patients to be selected , P(x)= 1/9

the Expected Value of the Weight of that patient E(x)= ∑ X.P(x)

= 108\*1/9 + 110\*1/9 + 123\*1/9 + 134\*1/9 + 135\*1/9 + 145\*1/9 + 167\*1/9 + 187\*1/9 + 199\*1/9

**= 145.33**

Q7) Look at the data given below. Plot the data, find the outliers and find out

**Hint:** [Use a plot which shows the data distribution, skewness along with the outliers; also use R/Python code to evaluate measures of centrality and spread]

|  |  |
| --- | --- |
| **Name of company** | **Measure X** |
| Allied Signal | 24.23% |
| Bankers Trust | 25.53% |
| General Mills | 25.41% |
| ITT Industries | 24.14% |
| J.P.Morgan & Co. | 29.62% |
| Lehman Brothers | 28.25% |
| Marriott | 25.81% |
| MCI | 24.39% |
| Merrill Lynch | 40.26% |
| Microsoft | 32.95% |
| Morgan Stanley | 91.36% |
| Sun Microsystems | 25.99% |
| Travelers | 39.42% |
| US Airways | 26.71% |
| Warner-Lambert | 35.00% |

**Ans:-**

Here each percentage of “measure X” belongs to different company. Actually it’s a bivariate data. We gonna do analysis by considering only the “measure X” data.

((((((((((((((((( Data set is not available ))))))))))))))))))))

Q8) AT&T was running commercials in 1990 aimed at luring back customers who had switched to one of the other long-distance phone service providers. One such commercial shows a businessman trying to reach Phoenix and mistakenly getting Fiji, where a half-naked native on a beach responds incomprehensibly in Polynesian. When asked about this advertisement, AT&T admitted that the portrayed incident did not actually take place but added that this was an enactment of something that “could happen.” Suppose that one in 200 long-distance telephone calls is misdirected.

What is the probability that at least one in five attempted telephone calls reaches the wrong number? (Assume independence of attempts.)

**Hint:** [Using Probability formula evaluate the probability of one call being wrong out of five attempted calls]

Ans:-

Probability of misdirected calls, P(misdirected calls) = 1/200

 Probability of call not any Misdirecting , P(not any misdirecting calls)= 1 - 1/200 = 199/200

Number of Calls = n

Probability of at least one in five attempted telephone calls reaches the wrong number

= 1  -  none of the call reaches the wrong number

= 1  - P(0) P(x) = ⁿCₓpˣqⁿ⁻ˣ

= 1   -  ⁵C₀(1/200)⁰(199/200)⁵⁻⁰ p = 1/200

= 1  -  (199/200)⁵= **0.02475** q = 199/200

n = 5

Q9) Returns on a certain business venture, to the nearest $1,000, are known to follow the following probability distribution

|  |  |
| --- | --- |
| X | P(x) |
| -2,000 | 0.1 |
| -1,000 | 0.1 |
| 0 | 0.2 |
| 1000 | 0.2 |
| 2000 | 0.3 |
| 3000 | 0.1 |

1. What is the most likely monetary outcome of the business venture?

**Hint:** [The outcome is most likely the expected returns of the venture]

Ans:-

Most likely monetary outcome of the business venture = Expected value

E(x) = ∑X.P(x)

= -2000\*.1 + -1000\* .1 + 0\*.2 + 1000\*.2 + 2000\*.3 + 3000\*.1

=  **$** **800**

1. Is the venture likely to be successful? Explain.

**Hint:** [Probability of % of venture being a successful one]

Ans:-

Since it has a positive expected value of earning or return, indicate the venture is successful in whole. Means the venture is not in loss.

1. What is the long-term average earning of business ventures of this kind? Explain.

**Hint:** [Here, the expected returns to the venture is considered as the

the required average]

Ans:-

Long-term average earning of business ventures = Expected value of earning of business ventures

= =  **$** **800**

((( Expected value never gonna change along with the passing time. Its would be costant or same vale for every time moment ))))

1. What is the good measure of the risk involved in a venture of this kind? Compute this measure.

**Hint:** [Risk here stems from the possible variability in the expected returns, therefore, name the risk measure for this venture]

**Ans:-**

Some part there happening losses and zero return either. As because its probability comparitively less as that of positive return there have not happening any losses in the business. Here +ve & -ve returns (profit & loss) are very close in range both by its weightage & probability. So any changes either in weightage or in probabilty of the return amount may leads to lossage. So take remedial actions either decrease weightage or probability of lossage that’s happening in the business.

**Hints:**

For each assignment, the solution should be submitted in the below format

1. Research and Perform all possible steps for obtaining solution.

2. For Statistics calculations, explanation of the solutions should be documented detail along with codes. Use the same word document to fill in your explanation

Must follow these guidelines:

2.1. Be thorough with the concepts of Probability, Central Limit Theorem and Perform the

calculation stepwise

2.2. For True/False Questions, or short answer type questions explanation is must.

2.3. R & Python code for Univariate Analysis (histogram, box plot, bar plots etc.) the data

distribution to be attached

3. All the codes (executable programs) should execute without errors

4. Code modularization should be followed

5. Each line of code should have comments explaining the logic and why you are using that function