|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuous |
| Weight of Gold | Continuous |
| Distance between two places | Continuous |
| Length of a leaf | Continuous |
| Dog's weight | Continuous |
| Blue Color | Discrete |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Discrete |

Q1) Identify the Data type for the Following:

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| Data | Data Type |
| Gender | Nominal |
| High School Class Ranking | Ordinal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Ordinal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ordinal |
| Years of Education | Ratio |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained? -🡪 3/8

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1🡪0/36
2. Less than or equal to 4 🡪(1,1)(1,2)(1,3)(2,1)(2,2)(3,1) ==6/36
3. Sum is divisible by 2 and 3 🡪 (1 , 5) , (3 , 3) , (4 , 2) , (5 , 1) , (6 , 6) == 5/36

Q5) 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? 🡪 10/21

Q6) 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)

|  |  |  |
| --- | --- | --- |
| 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 | 0.120 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

Expected Value => ex\*px = 1\*0.015 +4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120=3.09

Q7) 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 also Comment about the values/ Draw some inferences.



**For Points :**

> median(cars$Points)

[1] 3.695

> sd(cars$Points)

[1] 0.5346787

> mean(cars$Points)

[1] 3.596563

> range(cars$Points)

[1] 2.76 4.93

> var(cars$Points)

[1] 0.2858814

>

> # MODE

> getmode <- function(x){

+ uniq <- unique(x)

+ uniq[which.max(tabulate(match(x,uniq)))]

+ }

>

> result <- getmode(cars$Points)

> result

[1] 3.92

**For Score:**

median(cars$Score)

[1] 3.325

> sd(cars$Score)

[1] 0.9784574

> mean(cars$Score)

[1] 3.21725

> range(cars$Score)

[1] 1.513 5.424

> var(cars$Score)

[1] 0.957379

>

> # MODE

> getmode <- function(x){

+ uniq <- unique(x)

+ uniq[which.max(tabulate(match(x,uniq)))]

+ }

>

> result <- getmode(cars$Score)

> result

[1] 3.44

**For Weigh:**

median(cars$Weigh)

[1] 17.71

> sd(cars$Weigh)

[1] 1.786943

> mean(cars$Weigh)

[1] 17.84875

> range(cars$Weigh)

[1] 14.5 22.9

> var(cars$Weigh)

[1] 3.193166

>

> # MODE

> getmode <- function(x){

+ uniq <- unique(x)

+ uniq[which.max(tabulate(match(x,uniq)))]

+ }

>

> result <- getmode(cars$Weigh)

> result

[1] 17.02

Q8) 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?

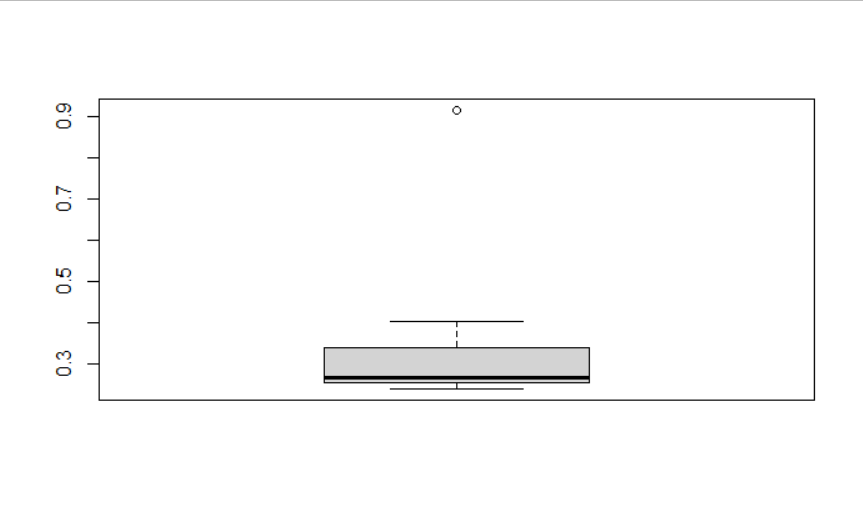
Sol: > x <- c(108, 110, 123, 134, 135, 145, 167, 187, 199)

> mean(x)

[1] 145.3333

Q9) Look at the data given below. Plot the data, find the outliers and find out

|  |  |
| --- | --- |
| **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% |



comp <- read.csv(file.choose())

View(comp)

attach(comp)

data2 <- as.numeric(sub("%", "",comp$Measure.X,fixed=TRUE))/100

boxplot(data2)

\*\*\*\*\*\*

> sd(data2,na.rm = TRUE)

[1] 0.169454

> mean(data2,na.rm = TRUE)

[1] 0.3327133

> var(data2,na.rm = TRUE)

[1] 0.02871466

Q10) 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.)

probability of call misdirecting  p = 1/200

     Probability of call not Misdirecting = 1 - 1/200 = 199/200

Number of Calls = 5

P(x) = ⁿCₓpˣqⁿ⁻ˣ

n = 5

p = 1/200

q = 199/200

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

= 1  -  none of the call reaches the wrong number

= 1  - P(0)

= 1   -  ⁵C₀(1/200)⁰(199/200)⁵⁻⁰

= 1  -  (199/200)⁵

= 0.02475

Q11) Returns on a certain business venture, to the nearest $1,000, are known to follow the following probability distribution

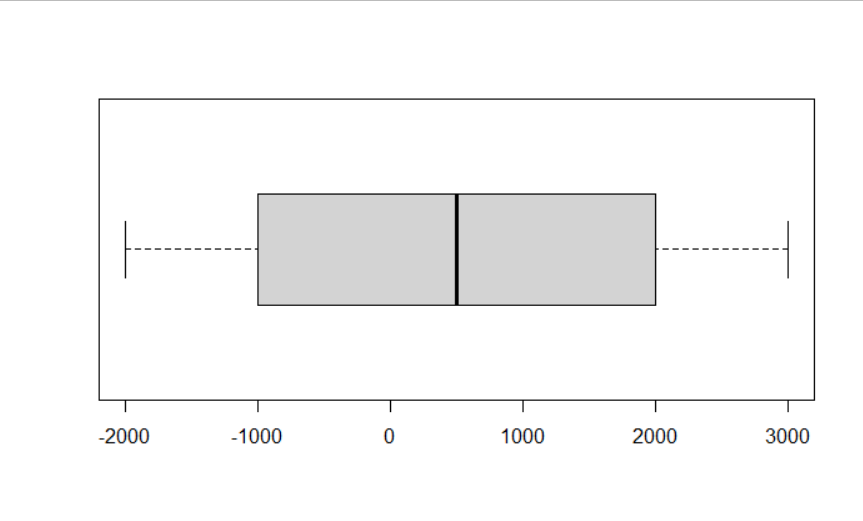
|  |  |
| --- | --- |
| x | P(x) |
| -2,000 | 0.1 |
| -1,000 | 0.1 |
| 0 | 0.2 |
| 1000 | 0.2 |
| 2000 | 0.3 |
| 3000 | 0.1 |

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

🡪x=2000 as px is 0.3

1. Is the venture likely to be successful? Explain

🡪It is likely to be successful as median is more in the middle and means its normal distribution And x=2000 where p(x) is 0.3



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

🡪Expected value = (0.1) (−2,000) + (0.1) (−1,000) + (0.2) (0) + (0.2) (1,000) + (0.3) (1,000) + (0,1) (3,000) = 800

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

🡪Good measure in terms of skewness is more right skewed ( positive) and risk involved is standard deviation.

sd(x,na.rm =TRUE)

[1] 1870.829