Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| 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 | **Nominal** |
| Number of kids | **Discrete** |
| Number of tickets in Indian railways | **Discrete** |
| Number of times married | **Discrete** |
| Gender (Male or Female) | **Nominal** |

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 | **Ordinal** |
| 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 | **Nominal** |
| Time Of Day | **Ratio** |
| Time on a Clock with Hands | **Interval** |
| Number of Children | **Ratio** |
| Religious Preference | **Ordinal** |
| Barometer Pressure | **Interval** |
| SAT Scores | **Interval** |
| Years of Education | **Ratio** |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

**Probability = 3/8**

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

**Answer:**

1. **Equal to 1 = 0**
2. **Less than or equal to 4 = 5/36**
3. **Sum is divisible by 2 and 3 = 29/36**

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?

**Probability = 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.

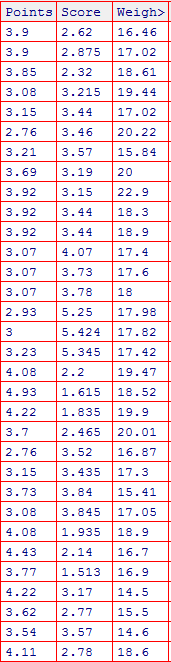
Child B – probability of having 4 candies = 0.20

**Answer = Expected No. -> (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.

****

**Answer=**

**#measures of central tendencies**

**# For points**

**> mean(ex1$points)**

**[1] 3.596563**

**> median(ex1$points)**

**[1] 3.695**

**>getmode(ex1$points)**

**[1] 3.92**

**> var(ex1$points)**

**[1] 0.2858814**

**> sd(ex1$points)**

**[1] 0.5346787**

**> range(ex1$points)**

**[1] 2.76 4.93**

**>#For Score**

**>mean(ex1$score)**

**[1] 3.21725**

**>median(ex1$score)**

**[1] 3.325**

**>getmode(ex1$score)**

**[1] 3.44**

**>var(ex1$score)**

**[1] 0.957379**

**>sd(ex1$score)**

**[1] 0.9784574**

**>range(ex1$score)**

**[1] 1.513 5.424**

**>#For Weight**

**>mean(ex1$weight)**

**[1] 17.84875**

**>median(ex1$weight)**

**[1] 17.71**

**>getmode(ex1$weight)**

**[1] 17.02**

**>var(ex1$weight)**

**[1] 3.193166**

**>sd(ex1$weight)**

**[1] 1.786943**

**>range(ex1$weight)**

**[1] 14.5 22.9**

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?

**Answer= Expected Wight -> Sum (Weight)/No. of Patients**

**= 1308/9**

**= 145.33**

Q9) Calculate Skewness, Kurtosis & draw inferences on the following data

Cars speed and distance

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

Ans)

#using e1071 package

>skewness(ex3\_csv$SP)

[1] -0.3898407

>skewness(ex3\_csv$WT)

[1] -1.230919

>kurtosis(ex3\_csv$SP)

[1] -1.034207

>kurtosis(ex3\_csv$WT)

[1] 0.5979244

|  |
| --- |
| #using moments package  skewness(ex3\_csv$SP)  [1] -0.4076944  >skewness(ex3\_csv$WT)  [1] -1.287292  >kurtosis(ex3\_csv$SP)  [1] 2.086738  > kurtosis(ex3\_csv$WT)  [1] 3.819284 |

**Q10) Draw inferences about the following boxplot & histogram**



**Answer: The above boxplot suggests that the distribution has lots of outliers towards upper extreme**

**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

Answer: n=2000

= 200

s= 30

Confidence Interval Estimate= Z => 200 Z

94% Confidence: qnorm(0.97)

[1] 1.880794=Z

200 1.88\* =**198.74 – 201.26**

98% Confidence: >qnorm(0.99)

[1] 2.326348=Z

200 2.33\* =**198.44-201.56**

96% Confidence: >qnorm(0.98)

[1] 2.053749

200 2.05\* = **198.62-201.38**

**Q12)** Below are the scores obtained by a student in tests

34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56

1. Find mean, median, variance, standard deviation.

**Answer : Mean = 41**

**Median = 20.5**

**SD = 5.052664**

**Variance = 25.52941**

1. What can we say about the student marks?

**Answer:  Mean > Median, This implies that the distribution is slightly skewed towards right. No outliers are present.**

Q13) what is the nature of skewness when mean, median of data are equal?

**Answer = When mean and median if data are equal, at that time the distribution is symmetric and skewness is Zero.**

Q14) What is the nature of skewness when mean > median ?

**Answer = When Mean is greater than median, at that time the skewness is positive. Means the Data distribution has large right tail than left.**

Q15)What is the nature of skewness when median > mean?

**Answer = When Median is greater than mean, at that time the skewness is negative. Means the Data distribution has large left tail than right.**

Q16) What does positive kurtosis value indicates for a data?

**Answer = Positive Kurtosis Value indicates that the data has less variation, data distribution is peaked and has thick tails.**

Q17) What does negative kurtosis value indicates for a data?

**Answer = Negative Kurtosis Value indicates that the data has more variation, distribution is flat and has thin tails.**

Q18) Answer the below questions using the below boxplot visualization.



What can we say about the distribution of the data?

**Answer= It’s not a normal distribution of data.**

What is nature of skewness of the data?

**Answer= It is left skewed.**

What will be the IQR of the data (approximately)?

**Answer: IQR= (Upper Quartile – Lower Quartile) = 18-10 = 8.**

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

**Answer= 1. There are no outliers in both the boxplots.**

**2. The median of both the boxplot is approximately same i.e. 270.**

**3. The skewness is neither in positive nor negative direction.**

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

**Answer:** >mean(Cars$MPG)

[1] 34.42208

P(MPG>38):

>sd(Cars$MPG)

[1] 9.131445

>pnorm(38,34.42,9.13)

[1] 0.652513

P(MPG>38)=1-P(MPG<38)(PS: Z-table gives you only less than probabilities)

>1 - 0.65

[1] 0.35

P(MPG<40):

pnorm(40,34.42,9.13)

[1] 0.7294571

P (20<MPG<50):

>pnorm(50,34.42,9.13)-pnorm(20,34.42,9.13)

[1] 0.8989178

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Answer: Follows Normal distribution as indicated by qq-plot.**



1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

**Answer: waist follows Normal Distribution from the below QQ-plot**

>qqnorm(wc\_at$Waist)

>qqline(wc\_at$Waist)



Adipose Tissue follows normal distribution

qqnorm(wc\_at$AT)

qqline(wc\_at$AT) 

Q 22) Calculate the Z scoresof 90% confidence interval,94% confidence interval, 60% confidence interval

Answer:

90% 🡪>qnorm(0.95)

[1] 1.644854

94% 🡪>qnorm(0.97)

[1] 1.880794

60% 🡪>qnorm(0.8)

[1] 0.8416212

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

Answer:

95% 🡪>qt(0.975,24)

[1] 2.063899

96% 🡪>qt(0.98,24)

[1] 2.171545

99% 🡪qt(0.995,24)

[1] 2.79694

Q 24**)**A Government companyclaims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode🡪pt(tscore,df)

df 🡪 degrees of freedom

**Answer:**

**µ=270, =260, SD=90, n=18, df=n-1=18-1= 17**

**tscore= = = -10/21.23= -0.47**

>pt(-0.47,17)

[1] 0.3221639

**Required probability = 0.32=32%**