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

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 | Discrete data-nominal |
| High School Class Ranking | Discrete data-nominal |
| Celsius Temperature | Continuous data-interval |
| Weight | Continuous data-nominal |
| Hair Color | Discrete data-ratio |
| Socioeconomic Status | Discrete data-ratio |
| Fahrenheit Temperature | Continuous data-ratio |
| Height | Continuous data-ratio |
| Type of living accommodation | Discrete data-ordinal |
| Level of Agreement | Discrete data-interval |
| IQ(Intelligence Scale) | Discrete data-interval |
| Sales Figures | Discrete data-interval |
| Blood Group | Discrete data-ratio |
| Time Of Day | Continuous data-interval |
| Time on a Clock with Hands | Continuous data-interval |
| Number of Children | Discrete data-interval |
| Religious Preference | Discrete data-ratio |
| Barometer Pressure | Continuous data-interval |
| SAT Scores | Continuous data- ratio |
| Years of Education | Discrete data-no |

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

Sol= given that, 3 coins are tossed i.e.2^3=8, P(HHT)+P(THH)+P(HTH)=1/8+1/8+1/8=3/8

Q4) 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

Sol= a)-there no outcomes that probability of sum is equal to 1 i.e. 0/36. Probability=0

b) (1,3) (2,2) (3,1)=3 outcomes, i.e. 3/36=1/12

c) (1,5) (2,4) (3,3) (4,2) (5,1) (6,6)= 6 outcomes i.e. 6/36=1/6

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?

Sol= P(2R, 3G, 2B), total no. of balls=(2+3+2)=7

S be the sample space and n(s)= number of ways of drawing 2 balls out of 7= nCr =7C 2 =7!/2!(5-2)!= 7\*6\*5!/2!(5!)=21

n(S)=21

n(E)=5C2 =5!/2!(3!)=10

P(E)=n(S)/n(E)=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

Sol= probability of expected no. of candies for randomly selected child=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.

Sol = Mean= 3.59 3.22 17.85

Median=3.69 3.33 17.71

Sd= 0.53 0.98 1.79

Variance= 0.29 0.96 3.19

**Use Q7.csv file**

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=Expected value of the weight of that patient is 145.3

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Speed of car skewness = -vely skewed = -0.117510**

**Distance skewness = =+vely skewed=0.806895**

**Speed of car kurtosis= -ve kurtosis= -0.508994**

**Distance kurtosis= +ve kurtosis= 0.405053**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Skewness(SP)=+vely skewed=1.611450**

**Skewness(WT)=-vely skewed= -0.614753**

**Kurtosis(SP)=+ve kurtosis=2.977329**

**Kurtosis(WT)= +ve kurtosis=0.950291**

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



Ans= the most of the data points are concentrated in the range 50 to 100 with freq 200. And least range of weight is 400.

Skewness = here we notice it is right skewed i.e mean is greater than median



Ans= here median is less than mean which means that it is right skewed and we have outliers in upper side of the boxplot.

**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?

Sol= std dev=30

n=2000, sample mean=3000000, std error= 30/sqrt(2000)= 0.67

alpha= 1-(confidence level/100)=0.06, degree of freedom =DF=2000-1=1999. Critical probability= 1-(alpha/2)=0.97

by using t distribution , critical value=1.881 , error= critical value\*std error=1.26,

FOR 94% CONFIDENCE INTERVAL =(200+1.26, 200-1.26)= (201.26, 198.74)

NOW FOR 98% confidence interval,

Alpha= 1-(confidence level/100)=0.02, Critical probability= 1-(0.02/2)=0.99,

DF=1999, critical value at df= 2.33, error= critical value\*std error= 1.56

Confidence interval of 94%= (200+1.56,200-1.56)=(201.56, 198.44)

Also FIND 96% CONFIDENCE INTERVAL,

STD error= 0.67, alpha= 1-(confidence level/100)=1-0.96=0.04, Critical probability= 1-(alpha/2)=0.98.

by using t distribution , critical value=2.055 , error= critical value\*std error=1.37,

FOR 96% CONFIDENCE INTERVAL =(200+1.37, 200-1.37)= (201.37, 198.62)

**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.
2. What can we say about the student marks?

Sol= 1) mean= 41.0

Median=40.5

Std deviation= 5.052664

Var= 25.529412

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

Sol= Symmetric

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

Sol= Right Skewed

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

Sol= Left skewed

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

Sol=Implies that data has sharp peak and wide tails.

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

Sol=Implies that has data has wider peak and thin tails.

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



What can we say about the distribution of the data?

Sol= it is normally distributed

What is nature of skewness of the data?

Sol= left skewed. Median is greater than mean

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

Interquartile range(IQR)=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.

Sol= Observing both the graph the whiskers level is high in boxplot 2. Mean and median are equal hence distribution is symmetrical.

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

Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

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

c. P (20<MPG<50)

sol=1) P(MPG>38)=0.3475

2) P(MPG<40)= 0.7293

3) P (20<MPG<50)=0.8988

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Sol= yes, the data set follows normal distribution

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

Dataset: wc-at.csv

Sol= yes, the data set follows normal distribution

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

Sol= Z score of 90% CI=1.6

Z score of 94% CI=1.8

Z score of 60% CI=0.8

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

Sol= T score of 95% CI= 2.06

T score of 96% CI= 2.1

T score of 99% CI= 2.7

Q 24**)** A Government company claims 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

sol= s=90, population mean=260 , n=18, sample mean=x= 270

t=( 270-260)/90/sqrt(18)= 0.4714, DF= n-1= 18-1=17 ,

P value= 0.3221