|  |  |
| --- | --- |
| 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 | Nominal |
| 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 | Ratio |
| Number of Children | Nomianl |
| Religious Preference | Nominal |
| 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?

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

N(S)=8

A(two heads and one tail)={HHT,HTH,THH}

N(A)=3

P(A)=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

Ans.

S={(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)}

N(S)=36

A(sum is equal to one)=0

B(sum is less than or equal 4)={1,1), (1,2),(1,3),(2,1),(2,2), (3,1)}

N(B)=6

P(B)=6/36

=1/6

C(sum is divisible by 2 and 3)={(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)}

N(C)=6

P(N)=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?

Ans.

Total number of balls =(2+3+2)=7

N(S)=Numbers of ways of drawing 2 balls out of 7

=7C2

=(2\*1)/(7\*6)

= 21

A= Drawing 2 balls none of them is blue

N(A)=5C2

= (2\*1)/(5\*4)

= 10

P(A)=N(A)/N(S)

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

Ans.

Expected number of candies for a randomly selected child

=1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

= 0.0015+0.8+1.95+0.025+0.06+0.24

= 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.

**Use Q7.csv file**

**Ans.**

**Points(Mean=3.596563,Median=3.659,Mode=numeric,variance=0.957379,**

**SD=0.5346787)**

**Score(Mean=3.217225,Median=3.325,Mode=numeric,variance=0.957379,**

**SD=0.9784574)**

**Weigh(Mean=17.84875,Median=17.71,Mode=numeric,variance=3.193166,**

**SD=1.786943)**

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?

ANS.

Expected value= **∑ (probability\*value)**

**∑ P(x).E(x)**

**Probability of selecting each patient=1/9**

**Ex. 108,110,123,134,135,145,167,187,199**

P(x) 1/9

Expected value = (1/9)(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)(108+110+123+134+135+145+167+187+199)

=(1/9)(1308)

=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**ANS.**

**Df1[’speed’].skew()**

**-0.1175098**

**Df1[’dist’].skew()**

**0.8068949**

**Df1[’speed’].kurtosis()**

**-0.5089944**

**Df1[’dist’].kurtosis()**

**0.4050525**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**ANS.**

**Df2[’SP’].skew()**

**1.611450**

**Df2[’WT’].skew()**

**-0.614753**

**Df2[’SP’].kurtosis()**

**2.977328**

**Df2[’WT’].kurtosis()**

**0.950291**

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



Here the maximum data lies between 0 to 200 interval and the maximum frequency goes upto 200 and here outliers can be seen towards right which means that the histrogram is positively skewed.



Here we can see that median is less then then mean so it concludes the the data is positively skewed.

And also some outliers are present at 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?

ANS. Using t distribution

98% confidence interval is 198.43,201.57

x-t (s)^n = 200-1.8916(30)^2000

= 198.43

x+t(s)^n= 200 +1.816(30)^2000

=201.57

96% confidence interval is 198.61,201.39

x-t (s)^n = 200-2.3452(30)^2000

= 198.61

x+t(s)^n= 200 +2.3452(30)^2000

=201.39

94% confidence interval is 198.73,201.27

x-t (s)^n = 200-2.0673(30)^2000

= 198.73

x+t(s)^n= 200 +2.0673(30)^2000

=201.27

**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?

ANS.

Mean=x.mean()

= 41.2131

Median=x.median()

= 40.5

Variance=x.var()

= 28.2990

Standard deviation=x.std

= 5.3194

Here maximum students have obtain the marks around 40-41 and some of them have obtain marks which vary by 5 from each other.

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

ANS. Symetrical

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

ANS. Right skewed (positive skewness)

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

ANS. Left skewed (negative skewness)

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

ANS. Data has more in tails than the normal distribution.

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

ANS. Data has less in tails than the normal distribution.

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



What can we say about the distribution of the data?

ANS. Here we can see that maximum data (50%) falls between 10 to 18 and minimum value is 2 and maximum value is 18 its shows that maximum points are arounds it,here the median shows that measure 50% of data is divided into 60-40%.z

What is nature of skewness of the data?

ANS. Left skewed (median> mean)

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

ANS. 8

Q19) Comment on the below Boxplot visualizations?



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

ANS. Both the boxplot have same mean and median due to which it has normal

distribution. Second boxplot has high whiskers which indicates that

Minimum and maximum data points are far away from the median than the

First boxplot.

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)

ANS. (MPG>38)

1-stats.norm.cdf(38,cars.MPG.mean(),cars.MPG.std())

0.3475

(MPG<40)

Stats.norm.cdf(40,cars.MPG.mean(),cars.MPG.std())

0.7293

(MPG<50)

Stats.norm.cdf(0.50,cars.MPG.mean(),cars.MPG.std())-stats.norm.cdf(0.20,cars.

MPG.mean(),cars.MPG.std())

1.2430

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

ANS. Using this data set and for MPG of cars mean is 34.4 and medain is 35.1 which approx.the same and distribution of data forms the bell shape curve which indicates its a 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

ANS. Using this data set and for Adipose Tissue (AT) mean is 101.89 and median is 96.54 which is approx. the same and distribution of data forms bell shaped curve which indicates its a normal distribution.

Using this data set and for Waist Circumference (waist) mean is 91.90 and median is 90.8 which is approx. the same and distribution of data forms bell shaped curve which indicates its a normal distribution.

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

ANS. Z score for

90% confidence interval

Stats.norm.ppf(0.95)

= 1.644

94% confidence interval

Stats.norm.ppf(0.97)

= 1.880

60% confidence interval

Stats.norm.ppf(0.8)

= 0.841

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

ANS. T score for

95% confidence interval

Stats.t.ppf(0.975,24)

=2.063

96% confidence interval

Stats.t.ppf(0.98,24)

=2.171

99% confidence interval

Stats.t.ppf(0.99,24)

=2.796

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

ANS.Assume Ho=avg life of bulb >=260 days

X=260

T = (s\_mean-p\_mean)/(s\_sd/sqrt(n))

T= (260-270)/(90/18\*\*05)

T score = - 0.4714

P value = 0.3216

P= 0.005

So reject Ho,therefore an avg life of bulb <=260 days .