|  |  |
| --- | --- |
| 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 | ordinal |
| Level of Agreement | ordinal |
| IQ(Intelligence Scale) | ordinal |
| Sales Figures | ordinal |
| Blood Group | nominal |
| Time Of Day | interval |
| Time on a Clock with Hands | interval |
| Number of Children | ordinal |
| Religious Preference | ordinal |
| Barometer Pressure | ratio |
| SAT Scores | ordinal |
| Years of Education | interval |

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

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

FOR every single event the probability is 1/8

Event where two heads and tail are obtained are described as follows E={HHT,HTH,THH}

P(E)=P{HHT}+P{HTH}+P{THH}

1/8+1/8+1/8=3/8=0.375

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

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)}

Here, the probability of any outcome is 1/36

A :- events representing sum = 1 are not seen here so,

P(A) = 0

Event B :-sum of two dices is less than or equal to 4

B = {(1,1),(1,2),(2,1),(1,3),(3,1)(2,2)}

P(B) =p(1,1) + p(1,2) + p(1,3) + p(2,2)+p(2,1)+p(3,1)

= (1/36) + (1/36) + (1/36) + (1/36)+(1/36)+1(36)

=6/36

=1/6

Event C = sum is divisable by 2 and 3 = 6

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

P(C)=p{(1,5)}+p{(5,1,)}+p{(2,4)}+p{(4,2)}+p{(3,3)}+p{(6,6}

P(c) = 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 5 :

Total no of balls are 7 out of which the probability of selecting 1 ball is 1/7.

There are two ted balls so probability of selection of red balls p(r) is 2/7.

There are three balls green so probability of selection of green balls p(g) is 3/7.

And the same way for probability of two blue balls to b selected are p(b) is 2/7.

Total probability is always 1 so probability for none of the balls to b drawn blue is 1-(2/7)=(7-2)/7=5/7=0.714

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 |

Ans 6

Expected number of candies for a randomly selected child is

[1(0.015)+4(0.20)+3(0.65)+5(0.005)+6(0.01)+2(0.120)]/6

(0.015+0.8+1.95+0.025+0.06+0.24)/6 = 3.09/6=0.051

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 (a python file is attached here )**

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: - Total number of patients = 9

So, the probability of selecting each patient is = 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

Q**9) 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**

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



Answer 10

(Histogram )This is positively skewed data means a right tailed distribution and data is heavily spread on leftside ..here mean>median>mode.

* More than 50% of Chick Weight is between 50 to 150.
* Most of the chick weight is between 50 to 100.
* Outlier is at upper side so there are outliers mainly on the maximum side …



Boxplot:--The data representing this boxplot contains lots of outliers. Its again right tailed and positively skewed distribution. Here mean>median>mode

**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?

94% of confidence interval

(198.738325292158, 201.261674707842)

98% of confidence interval

(198.43943840429978, 201.56056159570022)

96% of confidence interval

(198.62230334813333, 201.37769665186667)

**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?

Mean=(34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56)/18

Median=(40+41)/2 = 40.5

Variance = 25.52

Std deviation = 5.05

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

Ans 13

There is no skewness or skewness=0. in data means data is normally distributed where mean and median are equal.

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

Ans.14

When mean >median data is skewed positively (positive skewness).it’s a right tailed distribution .

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

Ans.15

When median > mean the nature of skewness is negative and the distribution is left tailed

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

Ans.16.

Kurtosis refers to degree os flatness or pickedness of a frequency of distribution.

The positive kurtosis value indicates that pickedness of data is high and its said to b good .It refers as large kurtosis.Its known as leptokurtic curve.

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

Negative kurtosis refers as small kurtosis which indicates that pickedness of data is very low and frequency distribution curve is mostly flat type known as platykurtic curve.

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



What can we say about the distribution of the data?

* Here data is heavily spread on right hand side ie. Its asymmetric distribution.

What is nature of skewness of the data?

* Skewness is left tailed as well as negative.Here mean<median<mode.

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

* IQR=(q3-q1)=(18-10)=8(approximately)

Q19) Comment on the below Boxplot visualizations?



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

1. There are no outliers present.

2. Both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: The data 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

Ans. The data follows normal distribution .

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

The Zscore of 90% confidence interval – 1.6448536269514722

The Zscore of 94% confidence interval - 1.8807936081512509

The Zscore of 60% confidence interval - 0.8416212335729143

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

**t score for 95% confidence interval**

print('T score for 95% Confidence Interval =',np.round(stats.t.ppf(0.025,df=24),4)) T score for 95% Confidence Interval = -2.0639

**t value for 94% confidence interval**

print('T score for 94% Confidence Inteval =',np.round(stats.t.ppf(0.03,df=24),4)) T score for 94% Confidence Inteval = -1.974

**t value for 99% Confidence Interval**

print('T score for 95% Confidence Interval =',np.round(stats.t.ppf(0.005,df=24),4)) T score for 95% Confidence Interval = -2.7969

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

df ◊ degrees of freedom

Ans: import numpy as np Import scipy as stats

t\_score = (x - pop mean) / (sample standard daviation / square root of sample size)

(260-270)/90/np.sqrt(18))

t\_score = -0.471

stats.t.cdf(t\_score, df = 17) 0.32 = 32%

population mean = 270

sample size - = 18

sample mean = 260

stardard daviation = 90 days ….