|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete type |
| Results of rolling a dice | Discrete type |
| 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 type |
| Number of kids | Discrete type |
| Number of tickets in Indian railways | Discrete type |
| Number of times married | Discrete type |
| Gender (Male or Female) | Discrete type |

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 | Ordinal |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ratio |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| 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?

Ans:-

H – for head, T –for tails

As we knw three coin tossed, P is,

(H,H,H),(H,H,T),(H,T,T),(T,T,T),(T,H,H),(T,T,H),(T,H,T),(H,T,H)

Total no.of outcome = 8

Probability that two heads and one tail are = 3

(H,H,T),(T,H,H),(H,T,H)

The no. of outcome is = 3

Thus probability of two head and one tail is = No. of outcome / Total no. of outcome

Is 3/8 = 0.375 = 37.5 %

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: - we have two dice, that why we have 6\*6=36 possible combination

Number.

The minimum possibilities of two dice throw is (1, 1) = sum 2

The maximum possibilities of two dice throw is (6, 6) = sum 12

1. If sum = 1,

Min possible sum is (1, 1) = 2

There for p(1) = (0)/(36) = 0 %

1. If sum is 4

It will achieved with possibilities no. of combination is

(2,2),(1,3),(3,1) only 3 combination of numbers

i.e p(4) = 4 / 36 = 1 / 40 is 0.25 is 25%

C) if sum is divided by 2 and 3

If 2 dice are rolled the the possibilities is 36

Outcome is (5,1),(1,5),(3,3),(4,2),(6,6)

There for outcome = 5

P = no.of outcome / total no.outcome

P = 5/36

Then the probability of sum is divided by 2 and 3 is 5 / 36

=0.13

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 :-

Here total no. of balls is 2+3+2 = 7

The sample space is (s)

Then n(s) = No. of ways of drawing 2 balls out of 7

=7C2 i.e 7 combination 2 = (7\*6)/(2\*1) = 21

E = event of drawing 2 balls no of which is blue

i.e n(E) = No. of ways of drawing 2 balls out of (2+3) balls

i.e 5C2 = (5\*4)/(2\*1) = 10

P(E) = n(E)/n(S) = 10/21 = 0.47 = 47.6 %

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 the no. 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

= 0.015+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 :-**

1. q\_seven.mean()
2. Points 3.596563
3. Score 3.217250
4. Weigh 17.848750
5. dtype: float64

2) q\_seven.median()

Points 3.695

Score 3.325

Weigh 17.710

dtype: float64

3) q\_seven.mode()

Points score weight

3.07 3.44 17.02

4) q\_seven.var()

Points 0.285881

Score 0.957379

Weigh 3.193166

dtype: float64

1. q\_seven.std()

Points 0.534679

Score 0.978457

Weigh 1.786943

dtype: float64

1. Range

Points 2.76

Score 1.51

Weigh 14.50

Plz ref. Math Assignment .csv

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 :-

There are 9 patient

So expected value = sum of probability \* value

i.e P(X) . E(X)

so probability of selecting each patient is 1 / 9

i.e 108, 110, 123, 134, 135, 145, 167, 187, 199

is 1/9, 1/9, 1/9, 1/9, 1/9 ,1/9, 1/9, 1/9, 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 i.e P(X) . E(X)

= (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:- Skewness for cars speed and dist. Is below**

1. speed.skew()

-0.11750986144663393

2) dist.skew()

0.8068949601674215

Kurtosis for cars speed and dist. Is below

1. q\_9A.kurtosis()

-speed -0.508994

Dist 0.405053

**SP and Weight(WT)**

**Use Q9\_b.csv**

q\_9B.skew()

SP 1.611450

WT -0.614753

dtype: float64

q\_9B.kurtosis( )

SP 2.977329

WT 0.950291

dtype: float64

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



**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 :- average weight of an adult male in Mexico. For 94%,98%,96% confidence interval is

stats.norm.interval(0.94,200,30/(2000\*\*0.5))

(198.738325292158, 201.261674707842)

stats.norm.interval(0.98,200,30/(2000\*\*0.5))

(198.43943840429978, 201.56056159570022)

stats.norm.interval(0.96,200,30/(2000\*\*0.5))

(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?

Ans:- 1) student = pd.Series([34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56])

student.mean() :- 41.0

student.median():- 40.5

student.var() :- 25.529411764705884

student.std() :- 5.05266382858645

1. There is a two Outlier in student data

(plzz ref.Math Assignment.csv)

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

Ans:- If mean and median of given data are equal then there should be no skewness i.e is zero and graph will follow normal distribution.

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

Ans:- skewness will tilt on right side

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

Ans:- skewness will tilt on left side

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

ANS :- The positive kurtosis value indicates that the peak value is thin and its tail is wide.

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

ANS :- The positive kurtosis value indicates that the peak value is wide and its tail is thin.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

Ans :- From above boxplot the maximum population lies between 10 to 15.2(approx..) that’s why this plot is not distributed normally.

From above plot median is slighting left hand side that’s why it is a –ve

Skewness

Here IQR is lies betn Q1 to Q3 is 10 to 18.2

Q19) Comment on the below Boxplot visualizations?



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

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)
  3. P (20<MPG<50)

Ans:-

stats.norm.cdf(38,loc=cars.MPG.mean(),scale=cars.MPG.std())

0.6524060748417295

stats.norm.cdf(40,loc=cars.MPG.mean(),scale=cars.MPG.std())

0.7293498762151616

stats.norm.cdf(20,loc=cars.MPG.mean(),scale=cars.MPG.std())

0.05712377632115936

stats.norm.cdf(50,loc=cars.MPG.mean(),scale=cars.MPG.std())

0.955992693289364

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:- Yes MPG of Cars 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:- Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set not following Normal Distribution.

(PLZ.. follow Math Assignment.ipynb attached file)

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

Ans:-

stats.norm.ppf(0.90)

1.281551565544600

stats.norm.ppf(0.94)

1.5547735945968535

stats.norm.ppf(0.60)

0.2533471031357997

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

Ans:- stats.t.ppf(0.95,df=20)

1.7247182429207857

stats.t.ppf(0.96,df=20)

1.8443309416068048

stats.t.ppf(0.99,df=20)

2.527977002740546

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:- here given is,

X is mean of sample bulb = 260

Mu is population mean = 270

S is standared deviation = 90

N is number of item in the sample = 18

Formula, T = (x-mu)/(sigma/root n)

So t = (260 - 270)/(90/root 18)

T = (-10)/(30/root 2)

T = (-1 \* root 2)/3

T = -0.471