|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | DISCRETE |
| Results of rolling a dice | DISCRETE |
| Weight of a person | CONTINOUS |
| Weight of Gold | CONTINOUS |
| Distance between two places | CONTINOUS |
| Length of a leaf | CONTINOUS |
| Dog's weight | CONTINOUS |
| 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 | NOMINAL |
| Level of Agreement | NOMINAL |
| IQ(Intelligence Scale) | INTERVAL |
| Sales Figures | RATIO |
| Blood Group | NOMINAL |
| Time Of Day | ORDINAL |
| Time on a Clock with Hands | INTERVAL |
| Number of Children | RATIO |
| 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?

{HHT+HTH+THH+HHH+TTT+TTH+HHT+THT}

=3/8

=0.375

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

1. Equal to 1

=ZERO,Because none of the outcomes will be one.

1. Less than or equal to 4

=(1,3)(2,2)(3,1) = 3/36 i.e 1/12 and Ans is 0.083

1. Sum is divisible by 2 and 3

P(sum is divisible by 2 and 3) = number of desired outcomes/

total number of possible outcomes

=11/36

=0.3056

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?

=p(2R,3G,2B)

=p(5/7,4/6) =20/42

=10/21 i.e 0.47

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

= 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.

Ans:- The Q7.csv file has 4 rows and 32 coloumns,

1st column is objective and remaning are float data types.

**The mean is**

Points 3.596563

Score 3.217250

Weigh 17.848750

**The median is**

Points 3.695

Score 3.325

Weigh 17.710

**The mode is**

Points 3.07

Score 3.44

Weigh 17.02

**The variance is**

Points 0.285881

Score 0.957379

Weigh 3.193166

**The std is**

Points 0.534679

Score 0.978457

Weigh 1.786943

**The range is**

Points 2.170

Score 3.911

Weigh 8.400

**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?

Ans :- given the weights of 9 patients at a clinic

X = 108,110,123,134,135,145,167,187,199

The probability of selecting a patient is P(X) = 1/9

Expected value of weight of choosing one patient at random is

E(X) = ∑ X P(X)

EX = 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)

=12+12.22+13.66+14.88+15+16.11+18.55+20.77+22.11

=145.3

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Skewness**

speed -0.117510

dist 0.806895

**kurtosis**

speed -0.508994

dist 0.405053

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Skewness**

SP 1.611450

WT -0.614753

**Kurtosis**

SP 2.977329

WT 0.950291

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



Ans :- The data is positively or right skewed.

More than 50% chick weight is between 50 -150.

Most of chick weight is between 50-100 with frequency 200.

There are outliers at upperside.

**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?

**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?

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

**Ans :- mean=median=0 data is symmetric.**

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

**Ans :- positively skewed**

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

**Ans :- negatively skewed.**

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

**Ans:- distribution is peaked and possess thick tails**.

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

**Ans:- distribution has lighter tails.**

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



What can we say about the distribution of the data?

**Ans:- The above box plot is not normally distributed median is towards higher value.**

What is nature of skewness of the data?

**Ans:- The data is skewed towards left the whisker range of min value is greater than max.**

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

**Ans:- IQR=Q3-Q1**

**=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.

Ans:-

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Ans: There are no outliers in the first Boxplot.**  **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 nor 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)

ANS **0.6524060748417286**

* 1. P(MPG<40)

ANS **0.7293498762151609**

c. P (20<MPG<50)

**ANS 1.2430968797327491e-05**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

ANS ;- It does not follow a normal distribution

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

Dataset: wc-at.csv

ANS :- It follow a normal distribution

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

ANS:- 90% 1.6448536269514722

94% 1.8807936081512509

60% 0.8416212335729143

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

ANS :- 95% 2.063898561680205

96% 201715446760080677

99% 2.4921594731575762

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 :- from scipy import stats

From scipy.stats imports norm

T = (260-270)/(90/18\*\*0.5)

T

-0.4714045207910317

P=1-stats.t.cdf(abs(-0.4714),df =17)

Print(“p value”,p)

P value 0.32167411684460556

If p <0.05:

Print(‘HO is rejected, H1 is accepted’)

Else:

Print(“H1 is rejected,H0 is accepted”)

H1 is rejected, H0 is accepted.