|  |  |
| --- | --- |
| 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) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Interval |

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

P=HHH, HHT, HTH, HTT, THT, TTH, THH, TTT

P(2H&1T)=3/8

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

1. Equal to 1 :

Soln: P(X)=0

1. Less than or equal to 4: P(X<=4)

Soln: Sum of observations / Possibility

P(2)=1/36

P(3)=2/36

P(4)=3/36

P(X<=4)=6/36 or 1/6

1. Sum is divisible by 2 and 3

P(2&3) = P(6)\*P(12)

=5/36\*1/36

=5/36

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?

Soln: R=2, G=3, B=2.

P(B=0)=2C0/7C2

=1/21

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

= (1\*0.015+4\*0.2+3\*0.65+5\*0.005+6\*0.01+2\*0.12)

=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.



Result

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.60 | 3.22 | 17.85 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 17.02 |
| Variance | 0.28 | 0.96 | 3.19 |
| Std Dev | 0.53 | 0.98 | 1.79 |
| Range | 2.17 | 3.911 | 8.4 |

Conclusion: The set of data points in weigh column has very wide range. Hence variance is high as compared to the points and score.

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?

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

**Cars speed and distance**



|  |  |  |
| --- | --- | --- |
| Skewness | -0.89542 | 1.290763 |
| Kurtosis | 0.249561 | 2.464546 |

**Inference: Skewness of Speed is negative which means majority data is concentrated on the right side of the distribution curve. Whereas distance skewness is >1 hence data is high positively skewed towards left.**

**Kurtosis value for speed indicates the distribution is normal since the value is close to zero. But for distance the kurtosis value is 2.4 which shows data points are very close to each other.**

**SP and Weight(WT)**



|  |  |  |
| --- | --- | --- |
| Skewness | -0.43626 | -1.34755 |
| Kurtosis | -0.85582 | 1.152953 |

**Inference: Skewness of SP is negative but close to zero, which means data is less skewed and follows normal distribution on the right side of the distribution curve. Whereas WT skewness is >-1 hence data is high negatively skewed towards right.**

**Kurtosis value for SP indicates the distribution is normal since the value is close to zero. But for distance the kurtosis value is 1.15 which shows data points are close to each other.**

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



More weights are concentrated towards left. Hence this is positively skewed data. Weight more than 300 can be categorized as outliers.



A large number of data set is concentrated on right side of the data. The number of outliers also concentrated on maximum end of the chart. Spread of whiskers is more on the Q3 quartile.

**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 ?

X-Bar = 200

N=2000

Sigma= 30

|  |  |  |  |
| --- | --- | --- | --- |
| C.I | 94% | 98% | 96% |
| t(1-alpha, n-1) | 1.881 | 2.328 | 2.055 |
| Min | 198.73 | 198.43 | 198.62 |
| Max | 201.26 | 201.56 | 201.37 |

**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.

|  |  |
| --- | --- |
| 41 | Mean |
| 40.5 | Median |
| 25.53 | Variance |
| 5.05 | S.D |

1. What can we say about the student marks?

Ans: The marks obtained by students have the same mean and median. We can infer the scores follow normal distribution.

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

Ans: The data set follows normal distribution.

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

The distribution is positively skewed.

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

The distribution is negatively skewed.

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

The distribution has heavier tails and sharper peak than the normal distribution.

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

The distribution has lighter tails and a flattened peak 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: Lot of data is concentrated on the right side of the distribution curve.

What is nature of skewness of the data?

Ans: The data is negatively skewed.

What will be the IQR of the data (approximately)?   
=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: Spread is more for Boxplot2 as compared to Boxplot1. Hence 2 has more variance than 1.

Values cannot be compared since the nature of box plot1 is horizontal and boxplot 2 is vertical.

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) = 0.348
  2. P(MPG<40) = 0.729

c. P (20<MPG<50)=0.90

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Yes. The data set follows normal distribution. With some outliers.

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

Dataset: wc-at.csv

Ans. No

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

90% = qnorm(0.95)=1.644

94% = qnorm(0.97)=1.880

60% = qnorm(0.80)=0.841

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

95%=qt(0.975,24) = 2.063

96% = qt(0.98,24) = 2.171

99% = qt(0.995,24) = 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

n=18

x=260

s=90

Mu=270

T score =2.109 (Industry standard 95%)

Pt(2.109,18)=0.975

(x-mu)/s/sqrt(n)

-10/21.21

T=-0.47

rcode 🡪 pt(tscore,df)

pt(-0.47,18)

0.321