|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discreate |
| Results of rolling a dice | Discreate |
| Weight of a person | Countinous |
| Weight of Gold | Countinous |
| Distance between two places | Countinous |
| Length of a leaf | Countinous |
| Dog's weight | Countinous |
| Blue Color | Discreate |
| Number of kids | Discreate |
| Number of tickets in Indian railways | Discreate |
| Number of times married | Discreate |
| Gender (Male or Female) | Discreate |

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 | Discreate data - Nominal |
| High School Class Ranking | Discreate data - Nominal |
| Celsius Temperature | Continious Interval |
| Weight | Continious Ratio |
| Hair Color | Discreate Ratio |
| Socioeconomic Status | Countinious - interval |
| Fahrenheit Temperature | Countinious - ratio |
| Height | countinious - ratio |
| Type of living accommodation | Discreate - ordinal |
| Level of Agreement | Discreate - interval |
| IQ(Intelligence Scale) | Discreate - interval |
| Sales Figures | Discreate - interval |
| Blood Group | Discreate ratio |
| Time Of Day | Countinious interval |
| Time on a Clock with Hands | Continuous - interval |
| Number of Children | Discreate interval |
| Religious Preference | Discreate ratio |
| Barometer Pressure | Discrete interval |
| SAT Scores | Discrete ratio |
| Years of Education | Discrete nominal |

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

Ans:-the probability of getting two heads and one tail when three coins are tossed is p(x) = 3/8 = 0.375 .

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

1. Equal to 1

Ans:- 0.

1. Less than or equal to 4

Ans:- 1/6 = 0.167

1. Sum is divisible by 2 and 3

Ans:- 1/6 = 0.167

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

Let S be the sample space

Then, n(S) = Number of ways of drawing 2 balls out of 7

n(S)=7C2n(S)=(7×6)(2×1)n(S)=21

Let E = Event of 2 balls, none of which is blue

∴ n(E) = Number of ways of drawing 2 balls out of (2 + 3) balls

n(E)=5C2n(E)=(5×4)(2×1)n(E)=10

Answer = 10/27 = 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

Ans:- Expected value = x\*p(x) = 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. A. for points:**

**Mean (Average)3.5965625**

**Median3.695**

**Range 2.17**

**Mode 3.07, 3.92**

**Std 0.52625**

**Variance 0.2769**

**b. for score**

**Mean (Average)3.21725**

**Median3.325**

**Range3.911**

**Mode3.44**

**Standard Deviation0.96304770131079**

**Variance0.9274608750000031**

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:- 1st calculate the probability of patient weight

P(X)= Interested Event /Total no of event

P(X)= 1/9= 0.11

Calculate the Expected Value

E = X \* P(X) E = 145 \* 0.11=15.95

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans:-**

|  |  |  |
| --- | --- | --- |
|  | **speed** | **distance** |
| **Skewness** | **-0.1139** | **0.7824** |
| **kurtosis** | **2.4228** | **3.248** |

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Kurtosis for speed & distance are positive means there is higher peaks**

|  |  |  |
| --- | --- | --- |
|  | **speed** | **Weight** |
| **Skewness** | **1.58** | **-0.60** |
| **kurtosis** | **5.72** | **3.81** |

**Kurtosis for speed & weight are +ve means there is higher peaks.**

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



Histogram: - Chick weight data is right skewed or positively skewed 🡪 Yes.

More than 50% Chick Weight is between 50 to 150 🡪 Yes.

Most of the chick weight is between 50 to 100 🡪Yes.

The data is right skewed. - There areoutliers at upper side



**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) Confidence Interval = ´ X ±t(s/√n)

Where ´ X is Mean = 200,

s is standard deviation = 30,

N is sample size = 2000

CI94% :α = 1 - (confidence level / 100)

= 1- (94%/100) = 0.06

Critical probability (p\*) = 1 - α/2

= 1-0.03 = 0.97

Degrees of freedom =n-1=2000-1

=1999 t-score

=1.882 CI94%

= 200±1.2620

CI98% :α = 1 - (confidence level / 100) = 1-(98%/100)=0.02

Critical Probability (p\*) = 1 - α/2 = 0.99

Degrees of freedom =n-1=2000-1=1999

t-score = 2.328 CI98% = 200± 1.561

CI96%:α = 1 - (confidence level / 10)

= 1 – (96%/100) =0.04

Critical Probability (p\*) = 1-α/2 = 0.98

Degrees of freedom =n-1=2000-1=1999

t-score = 2.055

CI98%= 200±1.378.

**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.

Ans. Mean: 41, median: 40.5, Variance: 25.23, Std: 5.052.

1. What can we say about the student marks?

* Max no.of students got 35-40 range score.
* No one get 50-55 range score.

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

Ans. Normalized skewness.

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

Ans. Right skewed.

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

Ans. Left skewed.

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

Ans. Sharp peaks in the plot, less gap between tails to x-axis.

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

Ans. Broader peak under the curve and more gap between tails and visualization.

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



What can we say about the distribution of the data?

Ans. The data is distributed in De-assigned formate.

What is nature of skewness of the data?

Ans. Left side skewed

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

Ans. Q3-Q4 = 8

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

ANS. The box plot 1 designed with range =3,

2nd one range = 1.5.

Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

* 1. P(MPG>38)

Ans. 51.76%

* 1. P(MPG<40)

Ans. 0.527461

* 1. P (20<MPG<50)

Ans. 57%

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans. It’s not distribution, slightly left skewed.

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

Dataset: wc-at.csv

Ans. Wc-at $waist – Normal distribution

Wc-at$at – Right skewed

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

Ans. 90% = [321.63, -140.373]

94% = [369.63, -188.37]

60% = [135.7475, 45.512]

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

Ans:- t score of 95% confidence intervalis 2.06

t score of 95% confidence intervalis 2.17

t score of 95% confidence intervalis 2.8

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. X(bar ) = 260 ,

Mu = 270 ,

n = 18 ,

s = 90

t = x(bar) - mu / (s/(root n))

t = 260 - 270 / (90/4.24)

t = -10 / 21.23

t = -0.47

Pt(-0.47,17) = .322

Hence, there is 32.2% probability that average life of 18 selected bulbs are not more than 260 days.