|  |  |
| --- | --- |
| 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 | Continuous |
| 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 | Ratio |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Interval |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Interval |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Ordinal |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ordinal |

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

Ans: P(X)={HHH,HHT,HTH,THH,THT,TTH,HTT,TTT}

N(X)=8

Probability for two heads and one tail:

P(for two heads and one tail)= {HHT,HTH,THH}

N(for two heads and one tail)= 3/8

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

Probability for two dice are rolled:

(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)

Total outcome=6×6=36

1. Equal to 1

Probability of getting sum equal to less than

0/36.probability is 0.

b) Less than or equal to 4

(1,3), (2,2), (3,1) (1,1),(1,2),(2,1)=6 outcomes,6/36 i.e. 1/6

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?

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

=

7C 2

=

(2×1)/(7×6)

=21

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

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

5 C 2

​= (2×1)/(5×4)

​=10

∴P(E)=

n(S)/n(E)

= 10/21

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

Ans: Expected value of candies for children:

=1\*0.015+ 4\*0.20+ 3\*0.65 +5 \*0.005 +6\*0.01 +2\*0.12

=0.0015 + 0.8+1.95+0.025+0.06+0.24

=3.090

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**- mean(Q7$Points)

[1] 3.596563

> mean(Q7$Score)

[1] 3.21725

> mean(Q7$Weigh)

[1] 17.84875

> median(Q7$Points)

[1] 3.695

> median(Q7$Score)

[1] 3.325

> median(Q7$Weigh)

[1] 17.71

Mode(Q7$Points)

[1] 3.07 3.92

attr(,"freq")

[1] 3

> Mode(Q7$Score)

[1] 3.44

attr(,"freq")

[1] 3

> Mode(Q7$Weigh)

[1] 17.02 18.90

attr(,"freq")

[1] 2

> var(Q7$Points)

[1] 0.2858814

> var(Q7$Score)

[1] 0.957379

> var(Q7$Weigh)

[1] 3.193166

> sd(Q7$Points)

[1] 0.5346787

> sd(Q7$Score)

[1] 0.9784574

> sd(Q7$Weigh)

[1] 1.786943

> range(Q7$Points)

[1] 2.76 4.93

> range(Q7$Score)

[1] 1.513 5.424

> range(Q7$Weigh)

[1] 14.5 22.9

**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?

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

**Cars speed and distance**

**ANS-** kurtosis(Q9\_a$speed)

[1] 2.422853

> kurtosis(Q9\_a$dist)

[1] 3.248019

> skewness(Q9\_a$speed)

[1] -0.1139548

> skewness(Q9\_a$dist)

[1] 0.7824835

**Use Q9\_a.csv**

**SP and Weight(WT)**

**ANS-** kurtosis(Q9\_b$SP)

[1] 5.723521

> kurtosis(Q9\_b$WT)

[1] 3.819466

> skewness(Q9\_b$SP)

[1] 1.581454

> skewness(Q9\_b$WT)

[1] -0.6033099

**Use Q9\_b.csv**

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



Ans: The most of the data points are Concerted in the range 50-100 with frequency

200. And least range of weight is 400 Somewhere around 0-10. Around 0-10.

So the expected value the above distribution is 75.

Skewness -we can notice a long tail towards right so it is heavily right skewed.



Ans: Median is less than mean right skewed and we have outlier on the upper side of

the box plot and there is less data points between Q1 and bottom point.

**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: At 94% confidence interval = (143.57619175546247, 256.42380824453755)

At 98% confidence interval= (130.2095637787748, 269.7904362212252)

At 96% confidence interval = (138.38753268104531, 261.61246731895466)

**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.

Mean=41.0

Median=40.5

Variance=25.529412

Standard deviation=5.052664

1. What can we say about the student marks?

Ans: Most of the students score 41 marks.

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

Ans: Skewness is zero

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: Thinner peak and wider tails.

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

Ans: Wider peak and thinner tails.

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



What can we say about the distribution of the data?

Ans: After distribution we can state that the plot is negatively skewed and mean is less than the median.

What is nature of skewness of the data? = Negatively Skewed

What will be the IQR of the data (approximately)? = Between 10 to 18

Q19) Comment on the below Boxplot visualizations?



Ans: By observing both the plots whisker’s level is high in boxplot 2.Mean

and median are equal hence distribution is symmetrical.

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): 0.652405801935003
2. P(MPG<40): 0.7293499868728933
3. P(20<MPG<50): 0.8988706389453319

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: Normal Distribution of dataset is Positive and its wider.

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

Dataset: wc-at.csv

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

Ans=Z score at 90% confidence interval is 1.6448536269514722

Z score at 94% confidence interval is 1.8807936081512509

Z score at 60% confidence interval is 0.8416212335729143

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

Ans= T score at 95% confidence interval is (-2.0638985616280205, 2.0638985616280205)

T score at 96% confidence interval is (-2.1715446760080677, 2.1715446760080677)

T score at 99% confidence interval is (-2.796939504772804, 2.796939504772804)

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: Assume H0= An average life of bulb&gt;=260 days

X=260

Using python code

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

t-0.4714045207910317

P value=0.32167411684460556

P0.005 so reject h0

So, an average life of bulb&lt;260 days