|  |  |
| --- | --- |
| 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 |

**Basic Statistics\_Level 1**

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 | Discrete-Nominal |
| High School Class Ranking | Discrete-Nominal |
| Celsius Temperature | Continuous-Interval |
| Weight | Continuous-Ratio |
| Hair Color | Discrete-Ratio |
| Socioeconomic Status | Continuous-Interval |
| Fahrenheit Temperature | Continuous-Ratio |
| Height | Continuous-Ratio |
| Type of living accommodation | Discrete-Ordinal |
| Level of Agreement | Discrete-Interval |
| IQ(Intelligence Scale) | Discrete-Interval |
| Sales Figures | Discrete-Interval |
| Blood Group | Discrete-Ratio |
| Time Of Day | Continuous-Interval |
| Time on a Clock with Hands | Continuous-Interval |
| Number of Children | Discrete-Interval |
| Religious Preference | Discrete-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?

**Answer:** Sample Space={ HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}

n(S)=8

Number of getting two heads and one tail=3

The Probability of getting two heads and one tails in the toss of three coins simultaneously is 3/8 or 0.375.

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

**Answer:**

1. If two dices were rolled then total possible outcomes i.e S=36

Total Outcomes (Having sum=1) =0

As minimum sum is 2i.e(1,1)

P (sum is equal to 1) = 0/36= 0

1. Total outcomes of Sum is less than or equal to 4= {(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}=6

P (Sum is less than or equal to 4) =6/36=0.003

1. Total possible outcomes of Sum is divisible by 2 and 3

{(1, 5), (3 , 3) , (4 , 2) , (5 , 1) , (6 , 6)}=5

P(Sum is divisible by 2 and 3)= 5/36 =0.1388

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?

**Answer:**

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  
=7C2​=(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.  
=5C2​=(2×1)(5×4)​=10  
∴P(E)=n(S)n(E)​=21/10

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

**Answer:**

Expected number of candies for a randomly selected child

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

= 0.015 + 0.8  + 1.95 + 0.025 + 0.06 + 0.24= 3.090

=  3.09

**Expected number of candies for a randomly selected child  = 3.09**

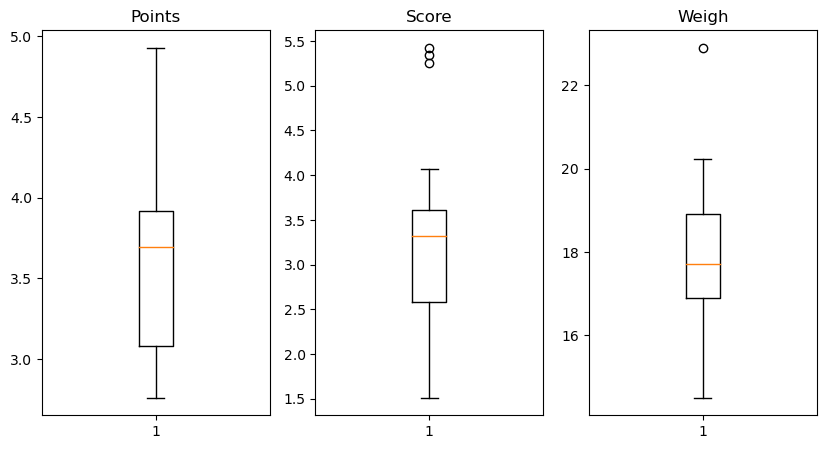
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Answer:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.596563 | 3.217250 | 17.848750 |
| Median | 3.695 | 3.325 | 17.710 |
| Mode | 3.92 | 3.44 | 17.02 |
| Variance | 0.285881 | 0.957379 | 3.193166 |
| Standard Deviation | 0.534679 | 0.978457 | 1.786943 |
| Range | 2.17 | 3.9110 | 8.3999 |



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?

**Answer:**

import numpy as np

Q8=np.array([108, 110, 123, 134, 135, 145, 167, 187, 199])

Q8.mean()

**Output: 145.33333333333334**

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

Cars speed and distance

**Ans:**

SP and Weight(WT)

Use Q9\_b.csv

Q10) Draw inferences about the following boxplot & histogram



**Answer:** The most of the data points are lies in the range 50-100 with frequency 200. And least range of weight is 400 somewhere 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**.**



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

**Answer:**

X+/-(Z1- α.σ/sqrt(n)

Degrees of freedom= 2000-1= 1999

Confidence interval= 94% (1- σ/2)

= 1-0.03) =0.97

for confidence interval for 94% is 1.882

Confidence interval for 98%= 2.33 Confidence interval for 96% = 2.05

**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?

**Answer: 1) Mean= 41, Median= 40, variance= 24.111, Standard deviation= 4.910**

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

Answer: Symmetrical

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

Answer: Right skewed

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

Answer: Left Skewed

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

Answer: The data is normally distributed and kurtosis value is 0.

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

**Answer:** The distribution of the data has lighter tails and a flatter peaks than the normal distribution.

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)?   
  
**Answer:** Let us assume that the boxplot is of student’s age.

From the above boxplot we can say 50% students are above 10yrs.

50 % students are bellow 10 yrs and around 40% student’s age is above 15yrs

IQR(Inter Quartile Range)=Q3-Q1=18-10=8 (Approx.)

Q19) Comment on the below Boxplot visualizations?



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

**Answer:**

By observing both the plots whisker’s level is high in boxplot 2, mean and median are equal hence distribution is symmetrical.

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)

**Answer:**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

**Answer:**

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

**Answer:**

Z scores of 90% confidence interval=1.644

Z scores of 94% confidence interval=1.880

Z scores of 60% confidence interval=0.841

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

**Answer:**

The t scores of 95% confidence interval=1.7108820799094275

The t scores of 96% confidence interval= 1.8280511719596342

The t scores of 99% confidence interval= 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

**Answer:**

**Given:**

µ=270, n=18, x bar=260, s=90

t score=(x bar-µ)/(s/sqrt(n))

=(260-270)/(90/sqrt (18))

=-10/21.23

=-0.47

Required probability=0.32