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

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

Answer 3 -

Possible Events = 8

HHH,

HHT, HTH,THH

HTT,THT,TTH

TTT

Favorable Event - 3

**Probability = 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

Answer 4 –

Possible outcome : 11,12,13,14,15,16

21,22,23,24,25,26

31,32,33,34,35,36

… 41,42,43,44,45,46,

51,52,53,54,55,56

61,62,63,64,65,66

1. Favorable event = 0 , P(Sum=1) = 0
2. P(Sum <=4) = 6/36 = 1/6
3. P(Sum divisible by 2&3) = P(Sum divisible by 6) = P(Sum = 6 or 12) = 1/36

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?

**Answer 5 :**

**Total Number of balls = 2+3+2 = 7**

**Possible Outcome = Number ways of drawing 2 balls = 7C2 = 7!/2!(7-2)! = 7\*6/2 = 21**

**Favorable Outcome (Drawing ball from red(2) or green(3) balls) = 5C2 = 5!/2!3! = 5\*4/2\*1 = 10**

**P(None drawn are blue) = 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

Answer 6

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

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**

**Points Score Weigh**

**count 32.000000 32.000000 32.000000**

**mean 3.596563 3.217250 17.848750**

**std 0.534679 0.978457 1.786943**

**min 2.760000 1.513000 14.500000**

**25% 3.080000 2.581250 16.892500**

**50% 3.695000 3.325000 17.710000**

**75% 3.920000 3.610000 18.900000**

**max 4.930000 5.424000 22.900000**

Mean of Points: 3.5965625

Mean of Score: 3.21725

Mean of Weigh: 17.84875

Median of Points: 3.6950000000000003

Median of Score: 3.325

Median of Weigh: 17.71

Mode of Points: 3.92

Mode of Score: 3.44

Mode of Weigh: 17.02

Range of Points: 2.76 - 4.93

Range of Score: 1.513 - 5.424

Range of Weigh: 14.5 - 22.9

Std Deviation of Points: 0.5347

Std Deviation of Score: 0.9785

Std Deviation of Weigh: 1.7869

Variance of Points: 0.2859

Variance of Score: 0.9574

Variance of Weigh: 3.1932

**For Points Median > Mean 🡪 Left Skewed**

**For Score Median > Mean 🡪 Left Skewed**

**For Weigh Median < Mean 🡪 Right Skewed**

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 8:

Expected value = ∑Porbability\*value

Probability = 1/9

Expected value = 145.33 =Mean

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Answer 9:**

**Skewness**

1. speed -0.117510 – in range of -0.5 , 0.5, so it data are nearly symmetrical
2. dist 0.806895 – in range of 0.5 – 1, so data is slightly Positively skewed

**Kurtosis:**

1. Speed -0.508994 – short tailed
2. dist 0.405053 – heavily tailed – data has outliers

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Skewness**

SP 1.611450, 🡪 extremely positively Skewed

WT -0.614753, 🡪 Slightly Negatively Skewed

**Kurtosis**

SP 2.977329 🡪 Long Tailed

WT 0.950291 🡪 Long tailed

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



Answer 10:

1. Data is positively skewed.
2. Mode ~75
3. Data has outliers

**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 11:

Confidence Interval is

94% 🡪 (143.57619175546247, 256.42380824453755)

96% 🡪 (138.38753268104531, 261.61246731895466)

98% 🡪 (130.2095637787748, 269.7904362212252)

**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:

Mean: 41.0

Median: 40.5

Variance: 24.11111111111111

Std deviation: 4.910306620885412

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

Answer 13 – Symmetrical

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

Answer 14 – Positively Skewed or Right skewed

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

Answer 15 – Negatively Skewed or Left Skewed

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

Ans 16 – Positive Kurtosis signifies data is Long tailed i.e. data as outliers

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

Ans 17 – Negative Kurtosis means – short tailed data

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



Answer 18 –

What can we say about the distribution of the data?

1. 25% data has less tha ~10’
2. 50% data has less than ~15’
3. 75% data has less than ~18’

What is nature of skewness of the data?

Right skewed data

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

Answer 19 :

1. Boxplot 1 has same median as Boxplot 2 = 262.5
2. Boxplot 1 & Boxplot 2 both are symmetrical
3. Boxplot1 has lower kurtosis than Boxplot 2
4. Boxplot 1 has lower range than 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)

0.3475939251582705

* 1. P(MPG<40)

0.7293498762151616

* 1. P (20<MPG<50)

0.8988689169682046

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer:

Mean = 34.422075728024666

Median = 35.15272697

Mode = 29.629936

There is not much difference in Mean and Median, we can say data follows normal distribution though not completely

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

Dataset: wc-at.csv

Answer : - Data follows Normal Distribution

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

Answer:

z Score of 90%: 1.2815515655446004

z Score of 94%: 1.5547735945968535

z Score of 60%: 0.2533471031357997

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

t Score of 95%: -1.7108820799094282

t Score of 96%: -1.8280511719596344

t Score of 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

Answer 24:

sample\_mean = 260

population\_mean = 270

Sample\_std\_deviation = 90

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

Probability (life < 260) = stats.t.cdf((Tscore),17)

0.32167253567098364