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

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

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

Answer –

A = {(TTT), (TTH), (THT), (HTT), (HHH), (HHT), (HTH), (THH)}

P (A) = Number of favorable outcomes of A / Total number of outcomes

= 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 –

B = {(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)}

a) P (B) = 0/36

= 0

b) P (B) = {(1,1), (1,2), (1,3), (2,1), (2,2), (3,1)} / 36

= 6/36

= 1/6

c) P (B) = { (1,5), (2,4), (3,3), (4,2), (5,1), (6,6) } / 36

= 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?

Answer –

Total Number of balls = 7

(P) = 5/7

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 value = ∑ (Probability \* Value)

= 1(0.015) + 4(0.20) + 3(0.65) + 5(0.005) + 6(0.01) + 2(0.120)

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

= 3.09

Expected number of candies for a randomly selected child is 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**

Answer –

Mean = using df.mean()

Points 3.596563

Score 3.217250

Weigh 17.848750

Median = using df.median()

Points 3.695

Score 3.325

Weigh 17.710

Mode = Using from scipy import stats

stats.mode()

Points 3.07

Score 3.44

Weigh 17.02

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 –

Expected value = ∑ (Probability \* Value)

There are 9 patients = 108, 110, 123, 134, 135, 145, 167, 187, 199

Probability of choosing one patient is 1/9

E(x) - 108, 110, 123, 134, 135, 145, 167, 187, 199

P(x) - 1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9

Expected Value = (1/9)108 + (1/9)110 + (1/9)123 + (1/9)134 + (1/9)135 + (1/9)145 + (1/9)167 + (1/9)187 + (1/9)199

= (1/9) (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9) (1308)

= 145.33

Expected Value of weight of that one patient is 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

Answer-

Skewness  
Speed -0.117510

Distance 0.806895

**SP and Weight (WT)**

**Use Q9\_b.csv**

Answer –

Skewness  
SP 1.611450  
WT -0.614753

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



Answer –

Histogram:

Chick weight data is positively skewed because it has a tail towards the right.

More than 50% Chick weight is between 50 to 100.

Most of the Chick weight is between 50 to 100.

Boxplot:

There are outliers at the upper at the 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?

Answer –

Sample mean = 200, sample std deviation = 30, sample size= 2000

2000 > 30, so it is normal.

Standard error = 30/√2000

= 0.67

**For 94% interval:**

Computing alpha (α): α = 1 - (confidence level / 100) = 0.06

Critical probability (p\*): p\* = 1 - α/2 = 1 - 0.06/2 = 0.97

Degrees of freedom (df): df = n - 1 = 2000 - 1 = 1999

The critical value is the t score having 1999 degrees of freedom and a cumulative probability equal to 0.97.

Using stats.t.ppf(0.97,df=1999), we find that the critical value is 1.88.

Margin of error (ME): ME = critical value \* standard error = 1.88 x 0.67 = 1.25

Therefore, we can be 94% confident that the population mean falls within the interval 200 ± 1.25.

**For 98% interval:**

Computing alpha (α): α = 1 - (confidence level / 100) = 0.02

Critical probability (p\*): p\* = 1 - α/2 = 1 - 0.02/2 = 0.99

Degrees of freedom (df): df = n - 1 = 2000 - 1 = 1999

The critical value is the t score having 1999 degrees of freedom and a cumulative probability equal to 0.99.

Using stats.t.ppf(0.99,df=1999), we find that the critical value is 2.32.

Margin of error (ME): ME = critical value \* standard error = 2.32 x 0.67 = 1.55

Therefore, we can be 98% confident that the population mean falls within the interval 200 ± 1.55.

**For 96% interval:**

Computing alpha (α): α = 1 - (confidence level / 100) = 0.04

Critical probability (p\*): p\* = 1 - α/2 = 1 - 0.06/2 = 0.98

Degrees of freedom (df): df = n - 1 = 2000 - 1 = 1999

The critical value is the t score having 1999 degrees of freedom and a cumulative probability equal to 0.98.

Using stats.t.ppf(0.98,df=1999), we find that the critical value is 2.05.

Margin of error (ME): ME = critical value \* standard error = 2.05 x 0.67 = 1.37

Therefore, we can be 96% confident that the population mean falls within the interval 200 ± 1.37.

**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.
2. What can we say about the student marks?

Answer –

**Mean** = (34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56)/18

= 738/18

= 41

**Median** = (40+41)/2

= 81/2 = 40.5

**Variance** = (∑(34-41)^2 + (36-41)^2 + (36-41)^2 + (38-41)^2 + (38-41)^2 + (39-41)^2 + (39-41)^2 + (40-41)^2 + (40-41)^2 + (41-41)^2 + (41-41)^2 + (41-41)^2 + (41-41)^2 + (42-41)^2 + (42-41)^2 + (45-41)^2 + (49-41)^2 + (56-41)^2)/18-1

= (49 + 25 + 25 + 9 + 9 + 4 + 4 + 1 + 1 + 0 + 0 + 0 + 0 + 1 + 1 + 16 + 64 + 225)/17

= 434/17

= 25.52

**Standard deviation** = √Variance

= √25.52

= 5.05

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

Answer –

If the mean is equal to the median, hence the skewness is zero. If the distribution is symmetric, the mean equals the median, and the skewness of the distribution is zero.

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

Answer –

If the mean is greater than the median, the distribution is positively skewed.

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

Answer –

If the mean is less than the median, the distribution is negatively skewed.

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

Answer –

Positive values of kurtosis indicate that distribution is peaked and possesses thick tails. An extreme positive kurtosis indicates a distribution where more of the numbers are located in the tails of the distribution instead of around the mean.

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

Answer –

Negative excess values of kurtosis (<3) indicate that a distribution is flat and has thin tails. Platykurtic distributions have negative kurtosis values. A platykurtic distribution is flatter (less peaked) when compared with the normal distribution, with fewer values in its shorter (i.e. lighter and thinner) tails.

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



What can we say about the distribution of the data?

Answer – Data is distributed more on the right side, not normally distributed.

What is nature of skewness of the data?

Answer – Left side skewed.

What will be the IQR of the data (approximately)?   
Answer – Q3 - Q1

= 18-10

= 8 IQR

Q19) Comment on the below Boxplot visualizations?



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

Answer – The median of both the boxplots is same. Both are evenly distributed according to their size of the dataset.

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)

c. P (20<MPG<50)

Answer – Attached in email

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer – Yes, it follows normal distribution.

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

Dataset: wc-at.csv

Answer – Attached in email

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

Answer – Attached in mail

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

Answer – Attached in email

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 – Hypothesis Problem

We accept null hypothesis.