|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | ordinal |
| Weight of a person | ratio |
| Weight of Gold | ratio |
| Distance between two places | ratio |
| Length of a leaf | ratio |
| Dog's weight | ratio |
| Blue Color | nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | ordinal |
| Number of times married | discrete |
| 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 | ordinal |
| Celsius Temperature | Interval |
| Weight | ratio |
| Hair Color | nominal |
| Socioeconomic Status | ordinal |
| Fahrenheit Temperature | interval |
| Height | ratio |
| Type of living accommodation | ordinal |
| Level of Agreement | ordinal |
| IQ(Intelligence Scale) | ratio |
| Sales Figures | interval |
| Blood Group | ordinal |
| Time Of Day | ratio |
| Time on a Clock with Hands | ratio |
| Number of Children | ordinal |
| Religious Preference | nominal |
| Barometer Pressure | ratio |
| SAT Scores | ratio |
| Years of Education | interval |

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

HHT, HTH, THH

3 times

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

N(s) = 36

1. the sum is equal to 1 is zero because they starts with (1,1) .... likewise . other than in the dice we are not having zero.
2. possible outcomes are (1,3),(2,2),(3,1)

3/36 = 1/12

c)total posiblities 36

½ and 1/3

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 no of balls 7  
two balls drawn 2  
so 7C2 is 21  
No blue ball should drawn is 2  
except blue total no of ball is 5  
so 5C2is 10  
**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 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.



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 - The mean is:

Expected Value=(108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)/9 = 145.333

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

**Cars speed and distance**



**SP and Weight(WT)**



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



**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 ?

Find standard error.

The standard error (SE) of the mean is: SE = s / sqrt( n ) = 30 / sqrt2000 = 30/44.72 = 0.67

Find critical value.

Compute alpha (α): α = 1 - (confidence level (94%)/ 100) = 0.06

Find the critical probability (p\*): p\* = 1 - α/2 = 1 - 0.06/2 = 0.97

Find the degrees of freedom (df): df = n - 1 = 2000 - 1 = 1999

The critical value is the t score having 1999 degrees of freedom

cumulative probability = 0.975.

From the t Distribution Calculator, we find that the critical value is 1.88

Compute margin of error (ME): ME = critical value \* standard error

1.88 \* 0.67 = 1.25

200 +1.25

For 98%

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

Find the critical probability (p\*): p\* = 1 - α/2 = 1 - 0.02/2 = 0.99

Find the degrees of freedom (df): df = n - 1 = 2000 - 1 = 1999

The critical value is the t score having 1999 degrees of freedom

cumulative probability = 0.99.

From the t Distribution Calculator, we find that the critical value is 2.32

Compute margin of error (ME): ME = critical value \* standard error

2.32 \* 0.67 = 1.55

200 +1.55

For 96%

Compute alpha (α): α = 1 - (confidence level (96%)/ 100) = 0.04

Find the critical probability (p\*): p\* = 1 - α/2 = 1 - 0.04/2 = 0.98

Find the degrees of freedom (df): df = n - 1 = 2000 - 1 = 1999

The critical value is the t score having 1999 degrees of freedom

cumulative probability = 0.98.

From the t Distribution Calculator, we find that the critical value is 2.05

Compute margin of error (ME): ME = critical value \* standard error

2.05 \* 0.67 = 2.008

200 +2.008

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?

**ANS** – Mean – 41

Median – 40.5

Variance -25.52

Standard deviation – 5.052

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

ANS – **Usually Skewness can be positive, negative, or zero, If the mean, median are approximately equal to each other, the distribution can be assumed to be approximately symmetrical**

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

ANS - **If the mean is greater than the median, the distribution is positively skewed**

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

ANS -  **If the median is greater than the mean, the distribution is negatively skewed**

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

ANS - **A distribution with a positive kurtosis value indicates that the distribution has heavier tails and a sharper peak than the normal distribution**

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

ANS - **A distribution with a negative kurtosis value indicates that the distribution has lighter tails and a flatter peak 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?

**ANS** -Negative skewness

What will be the IQR of the data (approximately)?   
  
**ANS** – q3 – q1 = 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.

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) 31/38
  2. P(MPG<40)61/40

c. P (20<MPG<50)

Rcode:

MPG <-c(Cars$MPG)

MPG

sample(MPG)

a=subset(MPG,MPG>38)

b=subset(MPG,MPG<40)

c=subset(MPG,MPG>20 & MPG <50)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

We can interpret that the data of MPG of Cars follows the normal distribution by:

1.Conducting shapiro test (w=0,97797; p value =0,1764)

2.Evaluating kurtosis value which is -0,7054604

3.Finding of mean value (34,42208) which is not so far difference from median value (35, 15273)

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

Dataset: wc-at.csv

We can interpret that the data of Weight of WC\_AT follows the normal distribution by:

1.Conducting shapiro test (w=0,95586; p value =0,00117)

2.Evaluating kurtosis value which is -1,141846

3.Finding of mean value (91.902) which is not so far difference from median value (90.8),

We can interpret that the data of AT of WC\_AT follows the non-normal distribution by:

1.Conducting shapiro test (w=0,95234; p value =0,000654) which is significant lower than 0,05

2.Evaluating kurtosis value which is -0,3760059

3.Finding of mean value (101,894) which is quite far difference from median value (96.54)

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

Supposed the corresponding shape follows the normal distribution. In case of confidence interval 90% that covers 90% normal curve, the probability of observing value outside of this area is less than 0.1. Because the normal curve is symmetric then half of this area is in the left tail and other half is in the right tail of the curve. Therefore, the area of each tail of the curve is equal to (1-CI)/2 = 0.05. The Z scores is P(CI 90%) = 1,645; P (94%)= 1.56; P( 60% ) = 0.26.

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

The way to obtain t scores : (sample mean – population mean) / (std deviation – (sample size-1))t scores CI95% for n = 25 then the value lies in 2.063t scores CI96% for n = 25 then the value lies in 1.828t scores CI99% for n = 25 then the value lies in 2.797

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

mean = 270 days

sample size = 18

sample mean = 260

deviation sample = 90 days

