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

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

= HHT,HTH,THH

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1:

If two dices were rolled, then total possible cases =36

Total Favorable cases (Having sum =1) = 0

As minimum sum is 2 for outcome (1,1).

Hence, probability is 0.

**No (0)**

1. Less than or equal to 4

= {1+1, 1+2, 3+1, 2+1, 2+2, 3+1}

**1/6**

1. **Sum is divisible by 2 and 3**

**= Favorable outcomes = (1 , 5) , (2,4), (3 , 3) , (4 , 2) , (5 , 1) , (6 , 6)**

**6/36**

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?

Total number of outcome 7\*6/2 = 21

Number of ways of drawing 2 balls out of (2 + 3) balls 5\*4/2= 10

**Ans 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

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

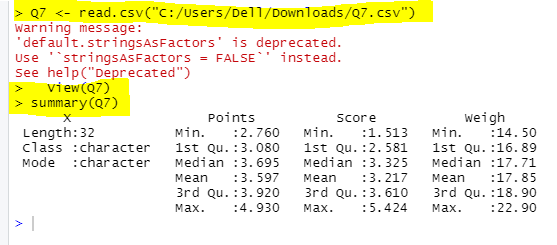
= 3.09

**Ans= Expected no. of candies for 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**

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?

there are 9 patientsProbability of selecting each patient = 1/9Ex 108, 110, 123, 134, 135, 145, 167, 187, 199P(x) 1/9 1/9 1/9 1/9 1/9 1/9 1/9 1/9 1/9Expected 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 the Weight of that patient = 145.33**

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

**Cars speed and distance**

**= Speed: Skewness = -0.1139548, Seed: 2.422853**

**=Distance : Skewness =0.7824835, Distance: 3.248019**

**Inferences:**

1. In given data of car speed and distance the skewness value of car speed is negative so, the data are negatively skewed or skewed left, meaning that the left tail is long, and In kurtosis case car speed is negative so, simply means that more data values are located near the mean and less data values are located on the tails. negative kurtosis is the uniform distribution, which has no peak at all and is a completely flat distribution. In Distance data skewness value is positive so, the mean and median is greater than mode. In kurtosis case of distance data is positive so, simply means that fewer data values are located near the mean and more data values are located on the tails.

**Use Q9\_a.csv**

**SP and Weight(WT)**

**SP: Skewness =1.581454, 5.723521**

**Weight: Skewness =-0.6033099, 3.819466**

**Use Q9\_b.csv**

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





Ans. In above figure histogram of ChickWeight$weight is shown. In this figure x axis shown chickweight$weight and in y-axis shown frequency. In above histogram figurer and boxplot figure is positively skewed on right side. That means Mean and median of the data is greater than mode.

**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?

* The **94%** confidence interval is **(198.73, 201.27).**
* The **96%** confidence interval is **(198.61, 201.39).**
* The **98%** confidence interval is **(198.43, 201.57).**

We are given the standard deviation for the sample, which is why the **t-distribution** is used to solve this question.

The **information given** is:

* Sample **mean**of .
* Sample **standard deviation** of .
* Sample **size**of .

The **interval** is:



* In which **t** is the critical value for the two-tailed confidence interval.

Considering a **94%** confidence level, using a calculator, with 200 - 1 = **199 df**, the critical value is **t = 1.8916**, hence:





The **94%** confidence interval is **(198.73, 201.27).**

Considering a **96%** confidence level, using a calculator, with 200 - 1 = **199 df**, the critical value is **t = 2.0673**, hence:





The **96%** confidence interval is **(198.61, 201.39).**

Considering a **98%** confidence level, using a calculator, with 200 - 1 = **199 df**, the critical value is **t = 2.3452**, hence:





The **98%** confidence interval is **(198.43, 201.57).**

**Ans. Using the t-distribution it is found that : Sample Mean = 200 Sample standard deviation = 30 Sample size = 2000 1. The 94% confidence interval is (198.73 , 201.39) (t= 1.8916) 2. The 96% confidence interval is (198.61 , 201.39) (t= 2.0623) 3. The 98% confidence interval is (198.43 , 201.57) (t= 2.3452)**

**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

Standard Deviation : 4.9103

Variance : 24.111

1. What can we say about the student marks?

**Ans: Mass of students marks between 38-42. Skewness(1.52) is positive because mass of marks in left side of plot.**

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

**Ans: If the mean is equal to the median as well as the mode, hence the skewness is zero. If the distribution is symmetric, the mean equals the median, and the skewness of the distribution is zero.(0**)

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

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

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

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

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

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?

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?

The distribution of a data set is the shape of the graph when all possible values are plotted on a frequency graph (showing how often they occur). Usually, we are not able to collect all the data for our variable of interest. Therefore, we take a sample. This sample is used to make conclusions about the whole data set

**What is nature of skewness of the data?**

**What will be the IQR of the data (approximately)?**

Ans. In above boxplot visualization

1. A negatively skewed distribution is a type of distribution in which more values are concentrated on the right side (tail) of the distribution graph while the left tail of the distribution graph is longer.

2. The nature of skewness of the data is the Mean of negatively skewed data will be less than the Median.

3. The IQR describes the middle 50% of values when ordered from lowest to highest. In above example of data the IOR = (-10-18)

Q19) Comment on the below Boxplot visualizations?



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

Ans. In above figure of boxplot. Boxplot 1 is positively skewed that’s means Mean and Median is greater than Mode. And In boxplot 2 is shows normal distribution that means the skewness for a normal distribution is zero, and any symmetric data should have a skewness near zero

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)

Ans. From Given Data of cars \

MPG <-CarsMPG \

1. P(MPG>38)-

1- pnorm(38,34.422,9.13144) = 0.3475908

b. P(MPG<40)

p-norm(40,34.422,9.13144)= 0.7293527

c. P (20<MPG<50)-

p-norm (50,34.422,9.13144) - (1-pnorm(20,34.422,9.13144))= 0.01311818

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans. In above data MPG of Cars is highly negative skewed.

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. Confidence interval Z scores

Z score of 60% = 0.8416212

Z score of 90% = 1.644854

Z score of 94% = 1.880794

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

Ans. Confidence Interval t score

95% 2.063899

96% 2.171545

99% 2.79694

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. 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

Solution=0.32%