|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continous |
| Weight of Gold | Continous |
| Distance between two places | Continous |
| Length of a leaf | Continous |
| Dog's weight | Continous |
| 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 | Nominal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Ratio |
| Socioeconomic Status | Interval |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Interval |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Ratio |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Interval |
| Religious Preference | Ratio |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Nominal |

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

Ans: 3/8

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

1. Equal to 1

Ans :0

1. Less than or equal to 4

Ans: 1/6

1. Sum is divisible by 2 and 3

Ans: 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?

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

**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

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.

**Use Q7.csv file**

**Ans :**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weight** |
| **Mean** | 3.596563 | 3.21725 | 17.84875 |
| **Median** | 3.695 | 3.325 | 17.71 |
| **Mode** | numeric | Numeric | numeric |
| **Variance** | 0.2858814 | 0.957379 | 3.193166 |
| **Standard deviation** | 0.5346787 | 0.9784574 | 1.786943 |

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:**

Expected Value  =  ∑ ( probability  \* Value )

 ∑ P(x).E(x)

there are 9 patients

Probability of selecting each patient = 1/9

Ex  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

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans :** Skewness for speed= -0.1139548

skewness value is negative so it is left skewed. Since magnitude is slightly greater than 0 it is slightly left skewed.

And for distance= 0.7824835

right skewed (Positive) slight magnitude to right

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



**Ans:**

The most of the data points are concentrated 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?

**Ans:**

Using the t-distribution, it is found that:

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 x̄= 200.

Sample standard deviation of s = 30.

Sample size of n = 2000.

The interval is:

x̄+- t(s)/(√(n))

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:

x̄- t(s)/(√(n)) = 200 - 1.8916(30)/(√(2000)) = 198.73

x̄+ t(s)/(√(n)) = 200 + 1.8916(30)/(√(2000)) = 201.27

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:

x̄- t(s)/(√(n)) = 200 - 2.0673(30)/(√(2000)) = 198.61

x̄+ t(s)/(√(n)) = 200 + 2.0673(30)/(√(2000)) = 201.39

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:

x̄- t(s)/(√(n)) = 200 - 2.3452(30)/(√(2000)) = 198.43

x̄+ t(s)/(√(n)) = 200 + 2.3452(30)/(√(2000)) = 201.57

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

**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

variance= 24.111

Standard deviation= 4.910

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

**Ans:**

When the mean and median of a dataset are equal, the skewness of the data is symmetrical or approximately symmetrical. In a perfectly symmetrical distribution, the mean and median coincide, and the skewness is zero.

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

**Ans:**

When the mean is greater than the median, the distribution is said to be right-skewed or positively skewed. This indicates that there are some relatively high values pulling the mean to the right. The tail on the right side of the distribution is longer compared to the left side.

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

Ans:

When the median is greater than the mean, the distribution is said to be left-skewed or negatively skewed. This indicates that there are some relatively low values pulling the median to the left. The tail on the left side of the distribution is longer compared to the right side.

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

**Ans:**

A positive kurtosis value indicates that the distribution of the data has heavier tails and a sharper peak than the normal distribution. That means there are more extreme values in the tails of the distribution.

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

**Ans:**

A negative kurtosis value indicates that the distribution of the data 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?

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

1.From the above box plot lets assume it is about age’s of the students in a school.

50% of the people are above 10 yrs old and remaining are less .And students who’s age is above 15 are approx 40%.

2. It is left skewed as median is greater than mean.

3. IQR = -8

Q19) Comment on the below Boxplot visualizations?



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

**Ans:**

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)

c. P (20<MPG<50)

Ans:

By using filter command and installing the dplyr package into the ‘R’.

a) There are 33 observations in MPG which are greater than 38.

b) There are 61 observations in MPG which are lesser than 40.

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

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

**Ans:**

Z score of 90% confidence interval is 1.65

Z score of 94% confidence interval is 1.55

Z score of 60% confidence interval is 0.85

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

**Ans:**

For 95%= 1.96

For 96%= 2.5

For 99% = 2.47

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:**

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.

t - statistics for the data is given as follows:

t=x-μ/8/√ n

x = mean of the sample of bulbs = 260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

t=260-270/90√ 18

t= -10/90/3√ 2

t=-10/30/√ 2

t=-1\*√ 2/3

Therefore,

t = - 0.47

For probability calculations, the number of degrees of freedom is n - 1, so here you need the t-distribution with 17 degrees of freedom.

The probability that t < - 0.471 with 17 degrees of freedom assuming the population mean is true, the t-value is less than the t-value obtained With 17 degrees of freedom and a t score of - 0.471, the probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of the bulbs is 300 days.

POOJASHREE – ASSIGNMENT -1