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

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained? = 0.375

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 2 Head | 1 Tail | 2 Head & 1 Tail |
| H | H | H | 3 | - |  |
| H | H | T | 2 | 1 | 1 |
| H | T | H | 2 | 1 | 1 |
| H | T | T | 1 | 2 |  |
| T | H | H | 2 | 1 | 1 |
| T | H | T | 1 | 2 |  |
| T | T | H | 1 | 2 |  |
| T | T | T | - | 3 |  |

|  |  |  |
| --- | --- | --- |
| Total outcomes | 2^3 | 8 |
| 2 Heads & 1 Tail Count |  | 3 |
| P (x) |  | 0.375 |

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

1. Equal to 1 = 0
2. Less than or equal to 4 = 6/36
3. Sum is divisible by 2 and 3 = 6/36

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 | 2 | Sum | Equal to 1 | Less than or equal to 4 | Sum is divisible by 2 and 3 |
| 1 | 1 | 2 | - | 1 | - |
| 1 | 2 | 3 | - | 1 | - |
| 1 | 3 | 4 | - | 1 | - |
| 1 | 4 | 5 | - | - | - |
| **1** | **5** | 6 | - | - | 1 |
| 1 | 6 | 7 | - | - | - |
| 2 | 1 | 3 | - | 1 | - |
| 2 | 2 | 4 | - | 1 | - |
| 2 | 3 | 5 | - | - | - |
| **2** | **4** | 6 | - | - | 1 |
| 2 | 5 | 7 | - | - | - |
| 2 | 6 | 8 | - | - | - |
| 3 | 1 | 4 | - | 1 | - |
| 3 | 2 | 5 | - | - | - |
| **3** | **3** | 6 | - | - | 1 |
| 3 | 4 | 7 | - | - | - |
| 3 | 5 | 8 | - | - | - |
| 3 | 6 | 9 | - | - | - |
| 4 | 1 | 5 | - | - | - |
| **4** | **2** | 6 | - | - | 1 |
| 4 | 3 | 7 | - | - | - |
| 4 | 4 | 8 | - | - | - |
| 4 | 5 | 9 | - | - | - |
| 4 | 6 | 10 | - | - | - |
| **5** | **1** | 6 | - | - | 1 |
| 5 | 2 | 7 | - | - | - |
| 5 | 3 | 8 | - | - | - |
| 5 | 4 | 9 | - | - | - |
| 5 | 5 | 10 | - | - | - |
| 5 | 6 | 11 | - | - | - |
| 6 | 1 | 7 | - | - | - |
| 6 | 2 | 8 | - | - | - |
| 6 | 3 | 9 | - | - | - |
| 6 | 4 | 10 | - | - | - |
| 6 | 5 | 11 | - | - | - |
| **6** | **6** | 12 | - | - | 1 |
|  |  |  | **-** | **6** | **6** |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Count | P |
| Total outcomes |  | 36 | - |
| Sum Equal to 1 |  | 0 | - |
| Sum Less than or equal to 4 |  | 6 | 6/36 |
| Sum is divisible by 2 and 3 |  | 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?= 0.47619

nCr = n! / r! \* (n – r)!

Total no of balls =2+3+2=7

n(S) = no of ways od drawing 2 balls out of 7 balls=7C2=(7\*6)/(2\*1)=21

## n(E)= Number of ways of drawing 2 balls out of (2 + 3) balls

=5C2=(5\*4)/(2\*1)=10

P(E)=10/21=0.47619

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

=3.09

|  |  |  |  |
| --- | --- | --- | --- |
| CHILD | Candies count | Probability |  |
| A | 1 | 0.0 | 0.015 |
| B | 4 | 0.2 | 0.8 |
| C | 3 | 0.7 | 1.95 |
| D | 5 | 0.0 | 0.025 |
| E | 6 | 0.0 | 0.06 |
| F | 2 | 0.1 | 0.24 |
| Total | **21** | **1.0** | **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.

In Score column : mean < median means left skew data

In Weigh column : mean > median means right skew data

**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? =145.33

N=9

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 108 | 110 | 123 | 134 | 135 | 145 | 167 | 187 | 199 |  |
| P | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 |  |
| Expected value | 12.00 | 12.22 | 13.67 | 14.89 | 15.00 | 16.11 | 18.56 | 20.78 | 22.11 | **145.33** |

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

**Cars speed and distance**

**Use Q9\_a.csv**

Speed column Mean > median, - Right skew data

Dist column Mean > median, - Right skew data

Speed column having –ve Kurtosis data, wider peak and thinner tails in histogram graph

Dist column having +ve Kurtosis data, narrow peak and over tails in histogram graph

**SP and Weight(WT)**

**Use Q9\_b.csv**

SP column Mean > median, - Right skew data

WT column Mean < median, - Left skew data

SP column having +ve Kurtosis data, narrow peak and over tails in histogram graph

WT column having +ve Kurtosis data, narrow peak and over tails in histogram graph

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



Ans : Right skew data



Ans : Mean < median, Left skew data

**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?

Sample mean x bar =200

Sample **standard deviation s = 30**

**Sample size n = 2000**

**Population size = 3000000**

**Interval = xbar +/-t \* s/sqrt n**

**Df=200-1=199**

**t=1.8916 for 94%** confidence level

xbar-t\*s/sqrtn=200-1.8916\*30/sqrt2000=198.73

xbar+t\*s/sqrtn=200+1.8916\*30/sqrt2000=201.27

(198.73, 201.27)

**t=** **2.0673 for 96%** confidence level

(198.61, 201.57)

**t= 2.3452 for 98% confidence level**

**(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?

Mean : 41.0

Median : 40.5

Standard Deviation : 4.9103

Variance : 24.111

Student average score is 41.0.

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

Ans : Normal distribution

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

Ans : Right skew data

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

Ans : Left skew data

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

Ans : Narrow peak and over tails

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

Wider peak and thinner tails

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



What can we say about the distribution of the data?

Right skew data. Mean > Median

What is nature of skewness of the data?

Right skew data. Mean > Median

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

IQR=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.

Ans : In both box plots Mean = Median, follows the normal distribution curve

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) = 33/81
  2. P(MPG<40) = 61/81

c. P (20<MPG<50) = 69/81

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Skewness = -0.17795 is acceptable

Kurtosis = -0.61168 is acceptable

Hence follow the normal distribution

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

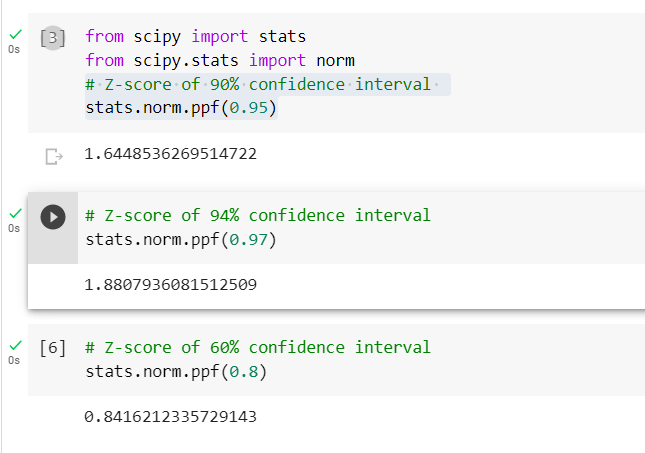
Dataset: wc-at.csv

Skewness is acceptable

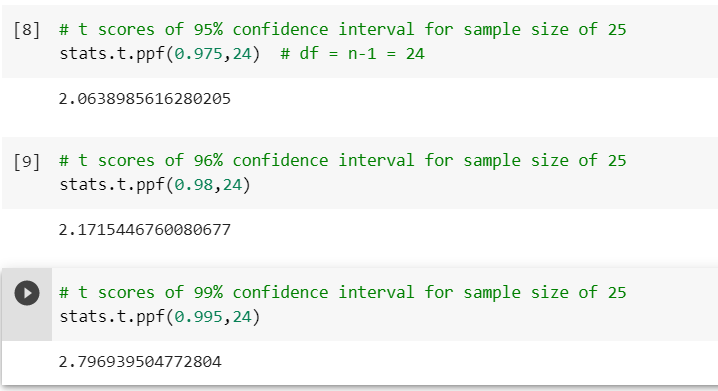
Kurtosis is acceptable

Hence follow the normal distribution

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



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



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

t - statistics for the data is given as follows

https://tex.z-dn.net/?f=t%3D%5Cdfrac%7Bx-%5Cmu%7D%7B%5Cfrac%7Bs%7D%7B%5Csqrt%20n%7D%7D

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/sqrt18)

t=-0.471

For probability calculations, the number of degrees of freedom is n - 1,

need the t-distribution with 17 degrees of freedom.

