|  |  |
| --- | --- |
| 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) | 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?

Possible combinations HHH,HHT,HTH,THH,HTT,THT,TTH,TTT

PROBABILITY=No. of possible outcomes/Total no. of events

=3/8

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

Total no. of events=36

1. Equal to 1 : 0
2. Less than or equal to 4 :no.of possible outcomes=6

Probability=6/36

=1/6

1. Sum is divisible by 2and 3 :no. of possible outcomes=6

Probability= 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?

Total possibilities-21

Possible outcomes=10

Probability=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

Sol: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

= 3.09

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**

**Soluntion:**

Weigh : Mean = 17.848750000000003

Median = 17.71

Mode= 17.02, 18.90

Since mean and median are almost equal we can predict there are no outliers in dataset.

Standard Deviation= 1.786943

Variance= 3.19316

Range=8.4

Points: Mean=3.5965

Median=3.695

Mode=3.07,3.92

Standard Deviation= 1.786943

Variance= 0.28588

Range=2.17

Score : Mean = 3.217

Median = 3.325

Mode= 3.44

Standard Deviation= 0.9784574

Variance= 0.9784574

Range= 0.9784574

There are no outliers in any case as mean and median are almost equal.

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?

Sol: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 + 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**

**Sol: pls find the attached excel file**



**SP and Weight(WT)**

**Use Q9\_b.csv**

**Sol: pls find the attached excel file**



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



\*There are 200 chicks weighing between 50-100

\*Most of the mass is concentrated on left, hence it is positively skewed.

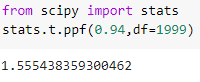


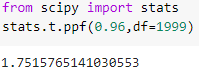
\*There are certain outlayers in the datapoints.

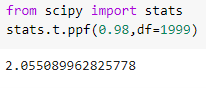
\*median is towards the left.

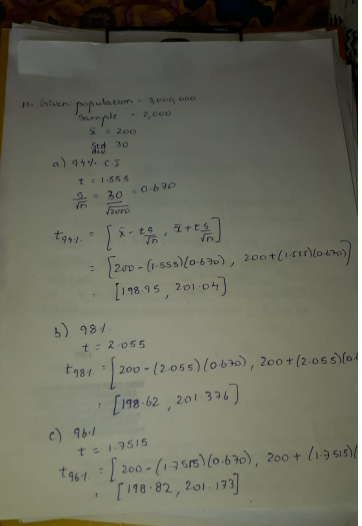
**Q11)**Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

Sol:









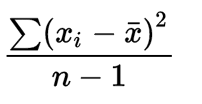
**Q12)**Below are the scores obtained by a student in tests

**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=sum of all observations/total no. of observations=41

Median=(40+41)/2=40.5

Variance== 25.5

Standard deviation=sqrt(variance)=5.0525

1. What can we say about the student marks?

Most of the students are scoring around 41.

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

Ans:zero skewness

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

Ans:Positively skewed

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

Ans:Negatively Skewed

Q16) What does positive kurtosis value indicates for adata ?

Ans: **the distribution has heavier tails than the normal distribution**.

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

**Ans:The distribution is flat and has thin tail**

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



What can we say about the distribution of the data?

* Middle 50% of the data is situated between 10 and 18
* Median is towars the right, we can say that the distribution is negatively skewed.
* Also there are no outliers in the data set.

What is nature of skewness of the data?

Negatively skewed

What will be the IQR of the data (approximately)?   
Soln: 8

Q19) Comment on the below Boxplot visualizations?



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

Ans:\*Median for both the data are the same.

\*Data points are widely distributed in 2nd case compared to 1st case.

\*Median is exactly at the middle, the distribution might be normal.

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

* 1. P(MPG>38)

33/82=.40

* 1. P(MPG<40)

61/82=.743

* 1. P (20<MPG<50)

68/82=.829

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

SOL:

|  |  |
| --- | --- |
| MEAN | 34.42208 |
| Median | 35.15273 |
| Mode | 29.62994 |

Since mean,median and mode are not equal we can say that mpg of cars do not follow normal distribution.

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

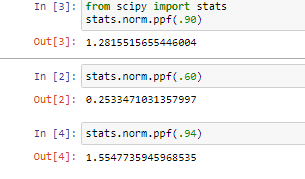
Dataset: wc-at.csv

Sol:

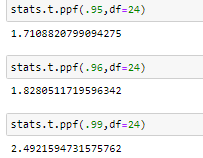
|  |  |  |
| --- | --- | --- |
|  | **waist** | **AT** |
| **Mean** | 91.90183 | 101.894 |
| **Median** | 90.8 | 96.54 |
| **Mode** | 94.5 | 121 |

Since mean ,median and mode are different it does not follow normal distribution.

Q 22) Calculate the Z scoresof 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 companyclaims 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:



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 = - 0.471

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.