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

**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 | Interval |
| Fahrenheit Temperature | Interval |
| 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 | Interval |

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

**Solution** : Sample Space (HHH,HHT,HTH,THH,HTT,THT,TTH,TTT)=8

Event A=( No of two heads and one tail)=( HHT,HTH,THH)=3

Probability=3/8=0.375

**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**

**Solution**: Sample Space S= {(1 1),(1 2),(1 3)…….(6 5),(6 6)}

n(S)=36

Event a= sum is equal to 1

n(a)=0

P(a)=0

**probability that sum is equal to 1 is 0 (zero**)

Event b= sum is less than or equal to 4

b={(1 1),(1 2) ,(1 3) ,(2 1), (2 2) ,(3 1) }

n(b)=6

P(b)=6/36=0.1667

**probability that sum is less than or equal to 4 is 6/36=0.1667**

Event c= sum is divisible by 2&3

C={(15) (24) (33) (42) (51) (66)}

n(c)=6

P(c)=6/36=0.1667

**probability that sum is divisible by 2&3 =6/36=0.1667**

**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?**

=

sample space =7c2 = 21. Let event E be none of the balls is blue = all balls are either red or green or both. n(E) = 5c2 =10 p(E) =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

1\*0.015+0.2\*4+0.65\*3+0.005\*5+0.01\*6+0.120\*2=3.09

**Q7) Calculate & 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**

**SOLUTION**

**#RESULT FOR POINTS**

Mean: 3.5965625000000006

Median: 3.6950000000000003

Mode: 0 3.07

1 3.92

dtype: float64

Varience 0.28588135080645166

Standard deviation 0.5346787360709716

**MEAN < MEDIAN LEFT SKEWED**

**#RESULT FOR SCORE**

Mean: 3.2172499999999995

Median: 3.325

Mode: 0 3.44

dtype: float64

Varience 0.9573789677419356

Standard deviation 0.9784574429896967

**MEAN < MEDIAN LEFT SKEWED**

**#RESULT FOR WEIGH**

Mean: 17.848750000000003

Median: 17.71

Mode: 0 17.02

1 18.90

dtype: float64

Varience 3.193166129032258

Standard deviation 1.7869432360968431

**MEAN > MEDIAN RIGHT SKEWED**

**Q8) Calculate Expected Value for the problem below**

1. **The (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?**

**SOLUTION : 145.33333333333331**

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

**Cars speed and distance**

**Use Q9\_a.csv**

Index 0.000000

speed -0.117510

dist 0.806895

dtype: float64

Index -1.200000

speed -0.508994

dist 0.405053

dtype: float64

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Solution**

skew : Unnamed: 0 0.000000

SP 1.611450

WT -0.614753

dtype: float64

kurt : Unnamed: 0 -1.200000

SP 2.977329

WT 0.950291

dtype: float64

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



The Highest frequency 200 relates to Chick Weight in the range of 50-100.

The frequency ranging 0 to 50(mostly around 0) relates to Chick Weight 350 to 400.

The frequency initially increases up to 200 from above 50 and after that there is a continuous fall. This resembles the distribution is Right Skewed



The Highest frequency 200 relates to Chick Weight in the range of 50-100.

The frequency ranging 0 to 50(mostly around 0) relates to Chick Weight 350 to 400.

The frequency initially increases up to 200 from above 50 and after that there is a continuous fall. This resembles the distribution is Right Skewed

**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?

N=2000, p=3000000, x(bar)=200 s=30

Alpha= 0.06,0.02,0.04

**Solution** Ref file attached

**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.

**Solution** Result

mean scores 41.0

dtype: float64

median scores 40.5

dtype: float64

varience scores 25.529412

dtype: float64

standard\_deviation scores 5.052664

dtype: float64

1. What can we say about the student marks?

Mean>median positively skewed

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

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.

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

The mean, mode and median can be used to figure out if you have a positively or negatively skewed distribution. ... If the mean is greater than the median, the distribution is positively skewed. If the mean is less than the median, the distribution is negatively 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?

A negative kurtosis means that your distribution is flatter than a normal curve with the same mean and standard deviation.

The easiest way to visualize this is to plot a histogram with a fitted normal curve

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



What can we say about the distribution of the data?

No outlier in given data mean<median & negatively skewed

What is nature of skewness of the data?

Left skewed, median is greater than mean.

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

Approx. -8

Q19) Comment on the below Boxplot visualizations?



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

Data almost symm. and mean equal to median no outliers in the data

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.

* 1. P(MPG>38)=
  2. P(MPG<40)=
  3. P (20<MPG<50)=

**Solution** refer .ipynb file

* 1. P(MPG>38)= 0.3475939251582705
  2. P(MPG<40)= 0.7293498762151616
  3. P (20<MPG<50)= 0.8988689169682046

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

Solution Refer attached .ipynb file

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

refer .ipynb file

1.2815515655446004

1.5547735945968535

0.2533471031357997

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

refer .ipynb file

2.0638985616280205

2.1715446760080677

2.796939504772804

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

Solution: refer .ipynb file