|  |  |
| --- | --- |
| 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 | DISCRETE DATA-NOMINAL DATA |
| High School Class Ranking | DISCRETE DATA-ORDINAL DATA |
| Celsius Temperature | CONTINOUS DATA-INTERVAL DATA |
| Weight | CONTINOUS DATA-RATIO |
| Hair Color | DISCRETE DATA-NOMINAL DATA |
| Socioeconomic Status | CONTINOUS DATA-ORDINAL |
| Fahrenheit Temperature | CONTINOUS DATA-INTERVAL DATA |
| Height | CONTINOUS DATA-RATIO |
| Type of living accommodation | DISCRETE DATA-NOMINAL DATA |
| Level of Agreement | DISCRETE DATA-ORDINAL |
| IQ(Intelligence Scale) | CONTINUOUS DATA-INTERVAL DATA |
| Sales Figures | DISCRETE DATA-RATIO |
| Blood Group | DISCRETE DATA-NOMINAL DATA |
| Time Of Day | CONTINOUS DATA-ORDINAL |
| Time on a Clock with Hands | CONTINOUS DATA-INTERVAL DATA |
| Number of Children | DISCRETE DATA- RATIO |
| Religious Preference | DISCRETE DATA-NOMINAL DATA |
| Barometer Pressure | CONTINOUS DATA-INTERVAL |
| SAT Scores | DISCRETE DATA-INTERVAL DATA |
| Years of Education | CONTINOUS DATA-RATIO |

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

The probability when 3 coins are tossed are{HHH,HHT,HTH,THH, TTT,TTH,THT,HTT}

So the probability of getting 2 heads and 1 tail is 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
4. The probability that sum is equal to 1 is ‘0’
5. The probability that sum is lessthan or equal to 4 is

{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),

(2,1),(2,2), (2,3),(2,4), (2,5),(2,6),

(3,1),(3,2), (3,3),(3,4), (3,5),(3,6),

(4,1),(4,2), (4,3),(4,4), (4,5),(4,6),

(5,1),(5,2), (5,3),(5,4), (5,5),(5,6),

(6,1),(6,2), (6,3),(6,4), (6,5),(6,6)}

The probability that sum is less than or equal to 4 is 6/36=0.16

1. The probability that sum is divisible by 2 and 3

{(1,5), (5,1),(2,4),(4,2),(3,3),(6,6)}

i.e, 6/36=0.16

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 no of balls=7

A=when two balls are drawn randomly=7c2=(7\*6)/(2\*1)=21

B=when two balls are drawn and none of the ball is blue=5c2=(5\*4)/(2\*1)=10

probability that none of the balls drawn is blue =B/A=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

|  |  |  |  |
| --- | --- | --- | --- |
| CHILD | Candies count(X) | ProbabilityP(X) | XP(X) |
| A | 1 | 0.015 | 0.015 |
| B | 4 | 0.2 | 0.8 |
| C | 3 | 0.65 | 1.95 |
| D | 5 | 0.005 | 0.025 |
| E | 6 | 0.01 | 0.06 |
| F | 2 | 0.12 | 0.24 |
|  |  |  | 3.09 |

\*\*SO THE EXPECTED MEAN 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**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| MEAN: | 3.596563 | 3.21725 | 17.84875 |
| MEDIAN: | 3.695 | 3.325 | 17.71 |
| MODE: | 3.92 | 3.44 | 17.02 |
| STD: | 0.534679 | 0.978457 | 1.786943 |
| VARIANCE: | 0.285881 | 0.957379 | 3.193166 |
| MAX | 4.93 | 5.424 | 22.9 |
| MIN | 2.76 | 1.513 | 14.5 |
| RANGE | 2.17 | 3.911 | 8.4 |

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?

The expected value of weight of the patient would be 145.33(in pounds)

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

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
|  | speed | dist |
| skew | -0.11751 | 0.806895 |
| kurtosis | -0.50899 | 0.405053 |

Here ,

The skewness of speed is 0.11751 as it lies between -0.5 and +0.5 the skewness is symmetrical .

The skewness of dist is 0.806895 as it lies between +0.5 and +1 the skewness is moderately positive skewed

The kurtosis of speed is 0.50899 as it is lessthan 3 it is platykurtic.

The kurtosis of dist is 0.405053 as it is lessthan 3 it is platykurtic

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| skewness | 1.61145 | -0.61475 |
| kutosis | 2.977329 | 0.950291 |

Here ,

The skewness of SP is 1.61145 as is greater than 1 so the skewness is highly positive skewed .

The skewness of WT is -0.61475 as it is less than -1 so the skewness is highly negative skewed.

The kurtosis of SP is 2.977329 as it is approximately 3 it is mesokurtic.

The kurtosis of WT is 0.9502 as it is lessthan 3 it is platykurtic

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



From the above histogram we can say it is the right skewed data and the chickweights between 50 to 100 are of higher frequency 200



From the above boxplot we can say thatit is a right skewed data and the data is having outliers .

**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?

For 94% confidence interval: (198.738325292158, 201.261674707842)

For 98% confidence interval: (198.43943840429978, 201.56056159570022)

For 96% confidence interval: (198.62230334813333, 201.37769665186667)

**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?

1)

|  |  |
| --- | --- |
| MEAN: | 41 |
| MEDIAN: | 40.5 |
| MODE: | 41 |
| VARIANCE: | 25.52941176 |
| STD: | 5.052663829 |

2)From the above student marks first we caluculate mean ,median and mode.As mean=mode=median(approx).The data is symmetric distribution data.

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

* As mean=median=mode the data will be symmetrical distribution data/zero skewness.

Mean=median=mode

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

* As mean> median the nature of skewness is positive skewed.

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

* As median>mean the nature of skewness is negatively skewed

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

* The kurtosis is >3 it is called positive kurtosis.
* It is also called as leptokurtic.
* The positive kurtosis will have high peak curve.

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

* The kurtosis is <3 it is called negative kurtosis.
* It is also called as platykurtic.
* The negative kurtosis will have more flat curve.

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



What can we say about the distribution of the data?

* From the above distribution of data the min value =1
* Q1=10
* MEDIAN=15
* Q3=18
* Max value=19

What is nature of skewness of the data?

* The nature of skewness is left skewed.

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

* The IQR (INTER QUARTILE RANGE ) is 8(approx)i.e,max-min(18-10)  
    
    
    
  Q19) Comment on the below Boxplot visualizations?



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

* From the box plot 1 we get the

min val=243.74,Q1=252,median=263,Q3=278.125,max val=287.125.

* From boxplot 2 we get the

min val=203.125,Q1=225,median=263,Q3=312,max val=325.

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

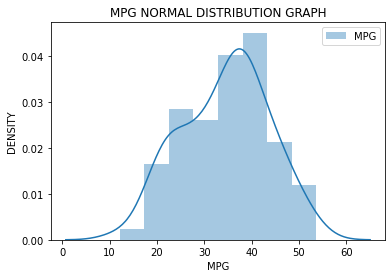
* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

|  |  |
| --- | --- |
| P(MPG>38) | 0.347594 |
| P(MPG<40) | 0.72935 |
| P(20<MPG<50) | 0.898869 |
|  |  |

Q 21) Check whether the data follows normal distribution

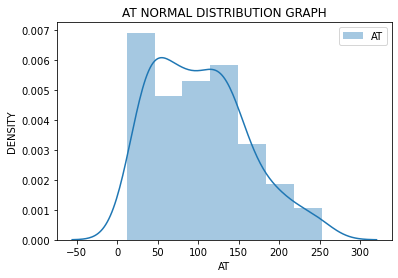
1. Check whether the MPG of Cars follows Normal Distribution

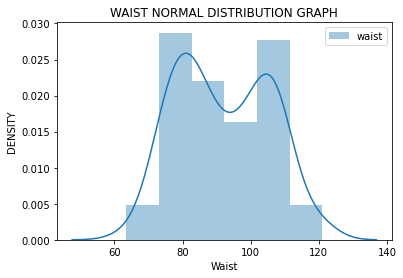
Dataset: Cars.csv



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

#FOR 90% CONFIDENCE INTERVAL

x=1-alpha=1-0.90=0.10

x/2=0.05

value to look in ztable=1-0.05=0.95

Z scores of 90% confidence interval is: 1.6448536269514722

#FOR 94% CONFIDENCE INTERVAL

x=1-alpha=1-0.94=0.06

x/2=0.03

value to look in ztable=1-0.03=0.97

Z scores of 94% confidence interval is: 1.8807936081512509

#FOR 60% CONFIDENCE INTERVAL

x=1-alpha=1-0.60=0.40

x/2=0.20

value to look in ztable=1-0.20=0.80

Z scores of 60% confidence interval is: 0.8416212335729143

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

degrees of freedom= n-1=25-1=24

#FOR 95% CONFIDENCE INTERVAL

x=1-alpha=1-0.95=0.05

x/2=0.025

value to look in ztable=1-0.025=0.975

t scores of 95% confidence interval is: 2.0638985616280205

#FOR 96% CONFIDENCE INTERVAL

x=1-alpha=1-0.96=0.04

x/2=0.02

value to look in ztable=1-0.02=0.98

t scores of 96% confidence interval is: 2.1715446760080677

#FOR 99% CONFIDENCE INTERVAL

x=1-alpha=1-0.99=0.01

x/2=0.005

value to look in ztable=1-0.005=0.995

t scores of 99% confidence interval is: 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

# H0: AVERAGE LIFE OF BULB<= 260DAYS

# Ha: AVERAGE LIFE OF BULB>260DAYS

x=260,mu=270,std=90,n=18,df=n-1=17

alpha=0.05,pval=0.321

As pval>alpha,we fail to reject the null hypothesis and consider the alternate hypothesis