|  |  |
| --- | --- |
| 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:ss

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| Data | Data Type |
| Gender | Nominal |
| High School Class Ranking | Ratio |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ratio |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Ordinal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

P=1/4

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 2and 3

a)P=0

b)P=3/36

c)P=3/36

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?

P=4/9

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

Expected number of candies for a randomly selected child ∑Xp(X)=1\*0.015+4\*0.20+3\*0.65+5\*0.65+5\*0.005+6\*0.01+2\*0120

∑Xp(X)=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

🡪Solution in Assignment pract

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

🡪Solution in Assignment pract

As mean,median,mode values are nearly equal hence dataet follows normal distribution.

**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?

Expected value:145.3333

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

**Cars speed and distance**

**Use Q9\_a.csv**

For speed-i)kurtosis=-0.5089,skewness=-0.1175

The speed distribution is negatively skewed or right skewed,mean & median of the data is on the left side

For Dist-i)kurtosis=0.4050,skewness=0.8068

The distance distribution is positively skewed as mean,median ofdata is on the right side

**SP and Weight(WT)**

**Use Q9\_b.csv**

For SP-i)kurtosis=2.9773,skewness=1.6114

The SP distribution of datais positively skewed or right skewed mean & median of the data is on the right side.

For WT-i)kurtosis=0.9502,skewness=-0.6147

The WT distribution of data is Negatively skewed or left skewed,mean & median of the data is on the left side.

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



From above histogram of ChickWeight$Weight vs Frequency we can conclude that ,Initially Frequency increases as the Chickweight increases But then it decreases gradually with increase in Chickweight & histogram is positively skewed or right skewed.



Above boxplot is positively skewed or right skewed & There are some outliers at the head of the boxplot.

**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

#Sample Mean=200

#standard deviation of sample=30

#Avg weight of adult male in mexico with 94% confidence interval

stats.norm.interval(0.94,200,30/2000\*\*0.5)=**(198.7383,201.2616)**

# Avg weight of adult male in mexico with 98% confidence interval

stats.norm.interval(0.98,200,30/2000\*\*0.5)=**(198.4394,201.5605)**

# Avg weight of adult male in mexico with 96% confidence interval

stats.norm.interval(0.98,200,30/2000\*\*0.5)=**(198.6223,201.3776)**

**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=41,median=40,Variance= 25.52

1. What can we say about the student marks?

🡪Data is positively skewed & there are 2 outliers.

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

🡪symmetrical

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

🡪Right Skewed

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

🡪Left Skewed

Q16) What does positive kurtosis value indicates for adata ?

🡪Data is normally distributed &kurtosis value is 0

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

🡪The distribution has lighter tail & flatter peaks than normal distribution.

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



What can we say about the distribution of the data?

From the data it is clear that as the mean,median &mode are not equal hence data do not follow normal distribution.

What is nature of skewness of the data?

Left skewed since, median>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.

  
From boxplot 1 &2 we can say that since mean & median are equal hence overall distribution is symmetric.

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)

stats.norm.cdf(38,cars.MPG.mean(),cars.MPG.std())=0.6524

* 1. P(MPG<40)

stats.norm.cdf(40,cars.MPG.mean(),cars.MPG.std())=0.7293

* 1. P (20<MPG<50)

stats.norm.cdf(50,cars.MPG.mean(),cars.MPG.std())- stats.norm.cdf(20,cars.MPG.mean(),cars.MPG.std())

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Mean**=cars.MPG.mean()=34.422

**Median**= cars.MPG.median()=35.152

**Mode=** cars.MPG.mode()=29.629

🡪From the boxplot of ‘MPG’ is not symmetric & also as the values of mean,median are not equal hence,we can conclude that following data set 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

Since the values of mean,median,mode are not equal hence,we can conclude that both AT & WAIST columns in the following dataset do not follow normal distribution.

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

**🡪1)Z-score for 90% confidence interval**

stats.norm.ppf(0.95) #df=n-1=25-1=24🡪1.6448

**2)Z-score for 94% confidence interval**

stats.norm.ppf(0.97) #df=n-1=25-1=24🡪1.8807

3**)Z-score for 60% confidence interval**

stats.norm.ppf(0.80) #df=n-1=25-1=24🡪0.8416

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

1)**t score for 95% confidence interval:**

🡪 stats.norm.ppf(0.975,24)=2.0638, df=n-1=25-1=24

2)**t score for 96% confidence interval:**

🡪 stats.t.ppf(0.98,24)=2.1715,df=n-1=25-1=24

3)**t score for 99% confidence interval:**

🡪 stats.t.ppf(0.995,24)=2.7969,df=n-1=25-1=24

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

🡪Given:Population Mean=270

Sample Mean=260

Sample Size=18

Sample standard deviation=90

#Assume Null hypothesis as Ho, avg life of bulb>=260 days

#Assume Alternate hypothesis as Ha, avg life of bulb<260 days

#Find t-scores at x=260

t=(s\_mean-P\_mean)/(s\_SD/sqrt(n))

t = (260-270)/(90/18\*\*0.5) = -0.4714045

P\_value = stats.t.cdf(-0.4714045207910317, df=17)

= 0.32167