|  |  |
| --- | --- |
| 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 | Continuous |
| 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 | Ordinal |
| 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?

**ANSWER:**

**Sample space= HHH, HHT,HTH,HTT,THH,THT,TTH,TTT**

**Probability=3/8**

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

**ANSWER:**

**(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)**

**For a) 0**

**b)6/36=1/6**

**c)12/36**

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?

**ANSWER**

**Probability that none of the balls drawn is blue=(1-Probability that all of the balls drawn is blue)=(1-2/7)=5/7**

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

**ANSWER:**

**EXPECTED NUMBER=1\*.015+4\*.20+3\*.65+5\*.005+6\*.01+2\*.120**

**=.015+.8+1.95+.025+.06+.240**

**=3.225**

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**

**ANSWER:**

**MEAN**

**Points 3.596563**

**Score 3.217250**

**Weigh 17.848750**

**MEDIAN**

**Points 3.695**

**Score 3.325**

**Weigh 17.710**

**MODE:**

**points 3.07, 3.92**

**score 3.44**

**weigh 17.02 ,18.90**

**HERE POINTS AND WEIGH ARE BIMODAL**

**HERE MEAN, MODE AND MEDIAN OF EACH CASES ARE APPROX. THE SAME. THUS IT IS A NORMAL DISTRIBUTION.**

**VARIANCE**

**Points 0.285881**

**Score 0.957379**

**Weigh 3.193166**

**STANDARD DEVIATION**

**Points 0.534679**

**Score 0.978457**

**Weigh 1.786943**

**RANGE (MAX VALUE-MIN VALUE)**

**Points 2.17(4.93-2.76)**

**Score 3.911(5.424-1.513)**

**Weigh 8.4(22.9-14.5)**

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?

**ANSWER**

**Probability of one patient at random=1/9**

**Expected Value of the Weight of that patient=1/9\*108+1/9\*110+1/9\* 123+1/9\*134+1/9\*135+1/9\*145+1/9\*167+1/9\*187+1/9\*199**

**=1/9\*(108+ 110+123+ 134+135+145+ 167+187+199)**

**=145.3333**

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

**Car’s speed and distance**

**Use Q9\_a.csv**

**ANSWER**

**SKEWNESS:**

**speed -0.117510****(left skewed)-concentrations more on the right-hand side**

**distance- 0.806895(right skewed****)-concentrations are more on the left-hand side**

**KURTOSIS**

**speed -0.508994** **-since kurtosis is negative, we can conclude that distribution gives a wider curve**

**distance- 0.405053-****since kurtosis is positive, we can conclude that distribution gives a sharp peak in the curve**

**SP and Weight (WT)**

**Use Q9\_b.csv**

**SKEWNESS:**

**SP -1.611450(right skewed)-concentrations are more on the left-hand side**

**WT -0.614753(left skewed)-concentrations more on the right-hand side**

**KURTOSIS**

**SP 2.977329**

**WT 0.950291**

**since kurtosis is positive, we can conclude that distribution gives a sharp peak in the curve in both cases**

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



**ANSWER:Here it is right skewed beacuse distributions are mostly concenrated at the left.So the nature of skewness is POSITIVE.**



**ANSWER: It has got 5 points, basically upper extreme, lower extreme, q1, q2 and q3. (q3-q1) is the inter quartile region where the median i.e. q2 exists. In a symmetric distribution both the mean and the median remain exactly at the middle of inter quartile region if not it is asymmetric i.e here exists a skewness. Since here the median is not at the center of IQR there exists a skewness and also there are some distributions above the extreme so there exist outliers also.**

**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?

**ANSWER: Here we are not given the value of standard deviation of the population, so we need to find out the value of t.**

**Here n=2000, s=30, sample mean(x)=200,std deviation=30,df=1999**

**FOR 94% CONFIDENCE INTERVAL**

**t=1.88186147647801**

**Confidence Interval****= [x(t\*std/sqrt of n)]**

**=****2001.88186147647801\*30/44.72136)**

**=200(1.8818614764780113\*** **0.6708204) =2001.2596**

**So, our CONFIDENCE INTERVAL is [200+1.2596,200-1.2596]**

**For 98% CONFIDENCE INTERVAL**

**t= 2.328214776106972**

**Confidence Interval= [xt\*std/sqrt of n]**

**= [2002.328214776106972\*30/44.72136)]**

**= [2002.328214776106972\* 0.6708204)] =2001.55976**

**So,our CONFIDENCE INTERVAL is [200+1.55976,200-1.55976]**

**For 96% CONFIDENCE INTERVAL**

**t= 2.055089962825778**

**Confidence Interval= [xt\*std/sqrt of n)]**

**=2002.055089962825778\*30/44.72136)**

**=2002.055089962825778\* 0.6708204)=****1.3735**

**So, our CONFIDENCE INTERVAL is [200+1.3735,200-1.3735]**

**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.

**ANSWER:**

**MEAN:41**

**MEDIAN:40.5**

**VARIANCE: 25.52941**

**STANDARD DEVIATION: 5.052664**

1. What can we say about the student marks?

**ANSWER:**

**The average marks scored by the students is 41.**

**The max no of students scored 41 marks which is the mode of this distribution**.

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

**ANSWER: Symmetric distribution where skewness is zero.**

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

**ANSWER: RIGHT SKEWED i.e., skewness is positive.**

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

**ANSWER: LEFT SKEWED i.e., skewness is negative.**

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

**ANSWER: It means the curve gives a SHARP PEAK i.e.all the distributions are at the center and can give outliers also sometimes.**

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

**ANSWER: It means the curve is flat i.e.all the distributions are almost equally distributed.**

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



What can we say about the distribution of the data?

**ANSWER: Here the MEAN<MEDIAN**

What is nature of skewness of the data?

**ANSWER: Here the skewness is NEGATIVE since it is left skewed.**

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

**ANSWER: Here IQR is approx. 8**

Q19) Comment on the below Boxplot visualizations?



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

**ANSWER**: **Approx. values**

**In fig1)**

**Upper extreme=287.5, q3=275, q2=262.5, q1=250, Lower extreme=237.5**

**Thus IQR=(q3-q1) =25**

**Infig2)**

**Upper extreme=337.5, q3=312.5, q2=262.5, q1=225, lower extreme=200**

**Thus IQR=(q3-q1) =87.5**

**Hence we can conclude that the IQR ( 2)>1) ).Since here mean=median in both cases, they are symmetric distribution and skewness=0.Ans the outliers present in both cases are 0.**

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.

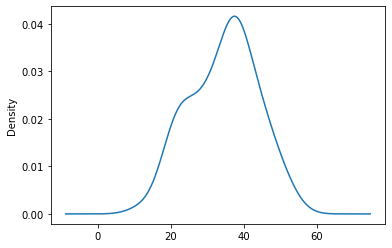
MPG <- Cars$MPG

* 1. P(MPG>38) =**66/81**
  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

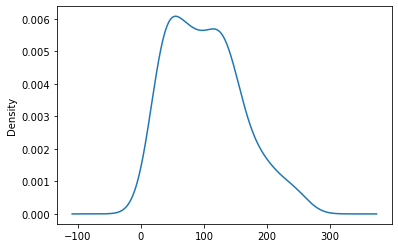
Dataset: Cars.csv

**ANSWER:**  **Here the distribution plotted is not symmetrical hence this is not a normal distribution**.

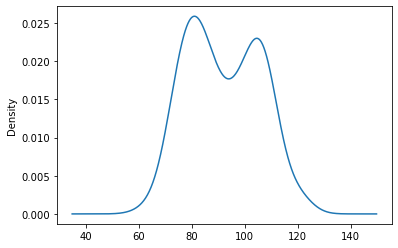
1. Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

**ANSWER: Waist Circumference (Waist)**



**Here the distribution plotted is not symmetrical hence this is not a normal distribution**

- **Here the distribution plotted is not symmetrical hence this is not a normal distribution**.

Q 22) Calculate the Z scores

of 90% confidence interval,94% confidence interval, 60% confidence interval

**ANSWER:**

**Z scores of 90% confidence interval= 1.644853626951472**

**Z scores of 94% confidence interval= 1.88079360**

**Z scores of 60% confidence interval= 0.8416212335729143**

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

**ANSWER:**

**t scores of 95% confidence interval 2.0638985616280205**

**t scores of 96% confidence interval 2.171544676008067**

**t scores of 99% confidence interval 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

**Here,the hypothesis testing is done where α=.05(standard)(H0,H1)**

**mean of population()=270,sample mean)=260,standard deviation of sample(sd)=90,n=18.**

**H0=Government claim is true**

**H1=Government claim is false**

**since std of population is not given, we go for,**

**t score=(270-260)/90/sqrt(18) = 10/90/4.242640= 0.47140444(Assuming the population mean is true)**

**probability that 18 randomly selected bulbs would have an average life of**

**no more than 260 days=1-0.6783274360719023=.3217**

**Since .3217>.05, we go for H0 and no change or action can be taken, hence the probability remains 32.17%**