|  |  |
| --- | --- |
| 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:

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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| 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?

Ans) possible outcomes:\_(HHH),(HHT),(HTT),(TTT),(THH),(TTH),(HTH),(THT)

probability that two heads and one tail are obtained =(HHT),(THH),(HTH) 3/8

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

* Equal to 1
* Less than or equal to 4
* Sum is divisible by 2 and 3

Ans) Total possible outcomes = 36

i) The probability that sum is equal to 1 = 0

ii) The probability that sum is Less than or equal to 4 = 6/36

iii) The probability that sum is Sum is divisible by 2 and 3 = 6/36

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?

Ans) Sample space = 7C2 = 21 , Event = 5C2 = 10

The probability that none of the balls drawn is blue = 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

Ans) 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

=3.090

=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**

Ans)

|  |
| --- |
| Unnamed: 0 Points Score Weigh  0 Mazda RX4 3.90 2.620 16.46  1 Mazda RX4 Wag 3.90 2.875 17.02  2 Datsun 710 3.85 2.320 18.61  3 Hornet 4 Drive 3.08 3.215 19.44  4 Hornet Sportabout 3.15 3.440 17.02  5 Valiant 2.76 3.460 20.22  6 Duster 360 3.21 3.570 15.84  7 Merc 240D 3.69 3.190 20.00  8 Merc 230 3.92 3.150 22.90  9 Merc 280 3.92 3.440 18.30  10 Merc 280C 3.92 3.440 18.90  11 Merc 450SE 3.07 4.070 17.40  12 Merc 450SL 3.07 3.730 17.60  13 Merc 450SLC 3.07 3.780 18.00  14 Cadillac Fleetwood 2.93 5.250 17.98  15 Lincoln Continental 3.00 5.424 17.82  16 Chrysler Imperial 3.23 5.345 17.42  17 Fiat 128 4.08 2.200 19.47  18 Honda Civic 4.93 1.615 18.52  19 Toyota Corolla 4.22 1.835 19.90  20 Toyota Corona 3.70 2.465 20.01  21 Dodge Challenger 2.76 3.520 16.87  22 AMC Javelin 3.15 3.435 17.30  23 Camaro Z28 3.73 3.840 15.41  24 Pontiac Firebird 3.08 3.845 17.05  25 Fiat X1-9 4.08 1.935 18.90  26 Porsche 914-2 4.43 2.140 16.70  27 Lotus Europa 3.77 1.513 16.90  28 Ford Pantera L 4.22 3.170 14.50  29 Ferrari Dino 3.62 2.770 15.50  30 Maserati Bora 3.54 3.570 14.60  31 Volvo 142E 4.11 2.780 18.60 |

**Mean**= Points e=3.596563

Score= 3.217250

Weigh=17.848750

**Median**=Points=3.695

Score= 3.325

Weigh=17.710

**Variance**=Points=0.285881

Score=0.957379

Weigh=3.193166

**Standard Deviation**=Points=0.534679

Score=0.978457

Weigh=1.786943

**Range**= \*points=4.93-3.59=1.3399999999999999

\*score=5.424-1.513=3.911000000000000 \*weigh=22.9-14.5=8.399999999999999

**Mode**=Points=0 3.07

1 3.92

Score=0 3.44

Weigh=0 17.02

1 18.90

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

* **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?**

Ans) Probability to pick a patient 1 = 1/9

Probability to pick a patient 2 = 1/9

Probability to pick a patient 3 = 1/9

Probability to pick a patient 4 = 1/9

Probability to pick a patient 5 = 1/9

Probability to pick a patient 6 = 1/9

Probability to pick a patient 7 = 1/9

Probability to pick a patient 8 = 1/9

Probability to pick a patient 9 = 1/9

Expected value = (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

= 145.3

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans)** Speed :- Skewness = -0.117510

Kurtosis = -0.508994

Distance:-Skewness= 0.806895

Kurtosis = 0.405053

**SP and Weight(WT)**

**Use Q9\_b.csv**

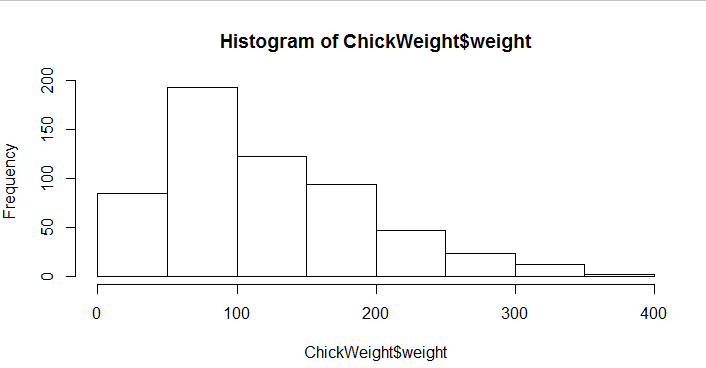
**Ans)** Sp :- Skewness = 1.611450

Kurtosis = 2.977329

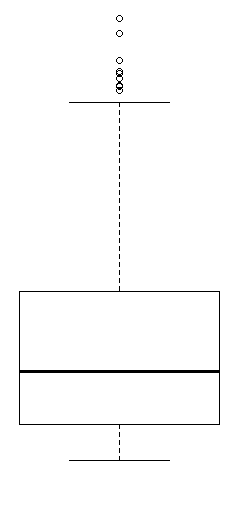
WT:-Skewness = -0.614753

Kurtosis = 0.950291

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



In the above histogram we can observe that most of the data lied between 0-200 & remaining lied in between 200-400. So, from this we can observe that, it is positively skewed



In the above Boxplot we can observe that the box is short, where the mean lies and the median too. But the median line seems to shifted towards bottom of the box . So, the from this we can see mean is greater than median and there are multiple outliers present beyond the max value

**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?

Ans) 94% CI=(197.74162011566807, 202.25837988433193)

98% ci=(199.24783863179837, 200.75216136820163)

96% ci=(198.49520384079835, 201.50479615920165)

**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**

* Find mean, median, variance, standard deviation.
* What can we say about the student marks?

Ans) Mean = 41. 0

Median = 40.5

Variance= 25.529411764705884

Standard deviation = 5.052664

\* There are 2 Outliers in Student's marks: 49 and 56

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

Ans) if disribution is same, then the mean is equal to the median. Which has zero skewness.

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

Ans)If the mean is greater than the median, the distribution is positively skewed

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

Ans) if the mean is lesser than the median, the distribution is negetively skewed

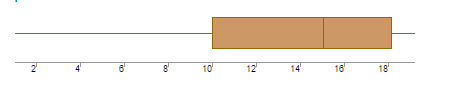
**Q16) What does positive kurtosis value indicates for a data ?**

Ans) Positive kurtosis value indicates that thinner peak and wider tails.

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

Ans) Negative values of kurtosis indicate that a distribution is flat and has thin tails

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



**What can we say about the distribution of the data?**

**Ans)** Here, data lies in scale between 2 to 20, where the median of the lies in between 14 to 16. And we can see clearly, there are no outliers in our data. And if we observe more, the upper Quartile is in 10, Lower Quartile is in 18.

**What is nature of skewness of the data?**

**Ans)** Here we can see median of the data lies towards lower Quartile. So, the data is positively skewed data

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

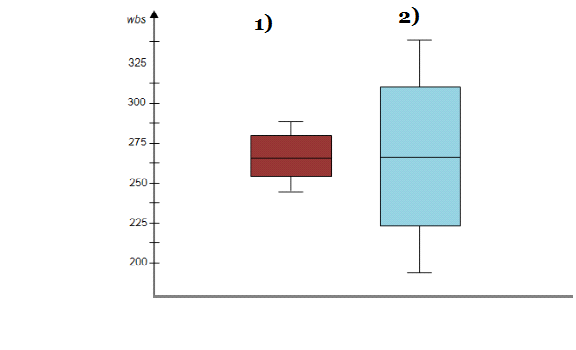
**Ans)**IQR=Q3-Q1

Q1=(12+13)/2 => Q1=12.5

Q3=(16)

IQR=16-12.5

IQR=3.5  
  
 v  
**Q19) Comment on the below Boxplot visualizations?**

****

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

**Ans)** BOXPLOT1 : -

* No outliers
* Median line is lies towards lower Quartile
* So, it is positive skewed

BOXPLOT 2 :-

* No outliers
* Median line lies in middle
* This data follows symmetry
* Based on the boxplot, it may became as normalized 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.**

**MPG <- Cars$MPG**

* **P(MPG>38)**
* **P(MPG<40)**
* **P (20<MPG<50)**

**Ans)**

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**

* **Check whether the MPG of Cars follows Normal Distribution**

**Dataset: Cars.csv**

* **Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from w c-at data set follows Normal Distribution**

**Dataset: wc-at.csv**

Ans)

* A) Here we can observe in boxplot that there are no outliers, at the same time, we can see that the median is slightly towards upper quartile, so we can say that MPG of cars doesn’t follow Normal distribution
* B) Here we can observe in boxplot that there are no outliers in both AT and Waist. at the same time, we can see that the median is approximatly presented in the middle in both the columns. But by observing histograms and curves, we can say that they are not in a bell shape curve.

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

Ans) Z score of 90% ci = 1.6448536269514722

Z score of 94% ci = 1.5547735945968535

Z score of 60% ci = 0.2533471031357997

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

Ans) t score of 95%=25.644853626951473

T score of 96%=25.75068607125217

T score of 99%=26.32634787404084

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

Ans) t=(x-μ)/s/√n

=(260-270)/90/√18

t=0.4714

p t=0.47, df=17

18 bulbs selected randomly, average life of no more than 260 days = 0.3216=32.16%