|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete/Nominal |
| Results of rolling a dice | Discrete/Nominal |
| 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 | Discrete/Nominal |
| Number of kids | Discrete/Nominal |
| Number of tickets in Indian railways | Continuous |
| Number of times married | Discrete/Nominal |
| Gender (Male or Female) | Discrete/Nominal |

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 | Ordinal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Interval |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Rati0 |
| Years of Education | Nominal |

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

**A:- The total possible outcomes 2^3=8**

**[HHH,HHT,HTT,THT,TTH,HTH,THH,TTT]**

**Number of favorable outcomes = 3**

**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

A:- **a) Sum is equal to 1 is not possible i.e. Probability is Zero(0)**

**b) <=4: can be possible only by (1,1) (1,2) (1,3) (2,1) (2,2) (3,1)**

**Favorable outcomes is 6**

**Number of possible outcomes are 36**

**Probability is 6/36 = 1/6**

1. **Sum is divisible by 2 and 3**

**Divisible by both 2 and 3 are 6 and 12**

**Favorable outcomes is (1,5) (2,4) (3,3) (4,2) (5,1) (6,6)**

**Favorable outcomes is 6**

**Number of possible outcomes are 36**

**Probability 6/36 = 1/6**

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?

**A:-** **5c2/7c2 = (5\*4/1\*2)/(7\*6/1\*2) = 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

**A:- Expected number of candies for a randomly selected child will be**

**E(x) = (1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120)**

**E(x) = 1.015+0.80+1.95+0.025+0.06+0.24**

**E(x) = 3.13**

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**

**A:- MEAN - Points 3.596563, Score 3.217250, Weigh 17.848750**

**MEDIAN - Points 3.695, Score 3.325, Weigh 17.710**

**MODE - Points Score Weigh**

**0 3.07 3.44 17.02**

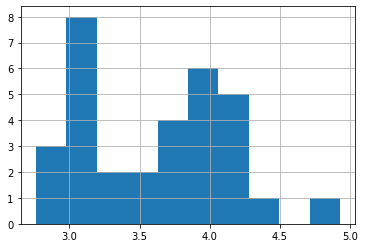
**1 3.92 NaN 18.90**

**VARIANCE - Points 0.285881, Score 0.957379, Weigh 3.193166**

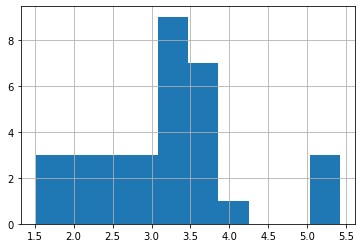
**STANDARD DEVIATION - Points 0.534679, Score 0.978457, Weigh 1.786943**

**RANGE – Points 2.17, Score 3.911, Weigh 8.399**

**Histogram of points-**

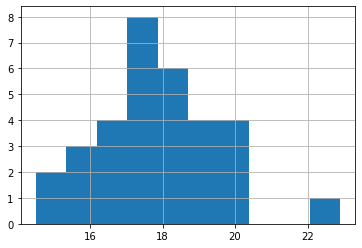


**Skewness factor of points - 0.29278021324083486**

**Histogram of score-**

**Skewness factor of score - 0.4659161067929858**

**Histogram of weigh –**



**Skewness factor of weigh - 0.4063466292404903**

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?

1. **Expected value for the random patient would be**

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

**E(X)=1/9\*(1298)**

**E(X)=144.22 pounds**

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

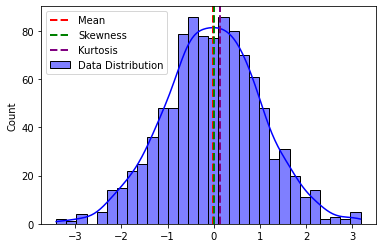
**Cars speed and distance**

**Use Q9\_a.csv**

**A) Skewness value: 0.18161654815794095**

**Kurtosis value: 0.09476390781488098**

**Inferences:**



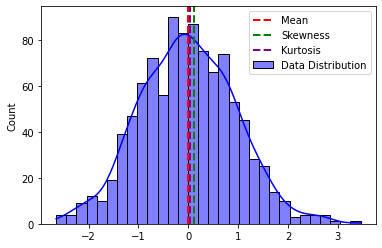
**SP and Weight(WT)**

**Use Q9\_b.csv**

**A) Skewness value: 0.10709790949037787**

**kurtosis\_value: 0.027597238324240436**

**Inferences**



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



**INFERENCES:-**

**Mean and median are far away from each other**

**It has widely distributed data**

**It is a multimodal histogram i.e it has no. of heads**

**It is a positively skewed histogram**

**And the data has outliers**



**Inferences :-**

**Line inside the box is closer to the bottom so it is called as the lower quartile, the data is distributed negatively skewed**

**And the boxplot has 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?

**A)** **94.0% Confidence Interval: (198.738325292158, 201.261674707842)**

**98.0% Confidence Interval: (198.43943840429978, 201.56056159570022)**

**96.0% 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?

**A)** **1)** **Mean: 41.0**

**Median: 40.5**

**Variance: 24.11111111111111**

**Standard Deviation: 4.910306620885412**

**2)Considering maximum marks as 60, the over all pass percentage of classroom is 68.33**

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

**A)**

* **Mean = Median: The distribution is symmetric**
* **Skewness = 0: The data points are symmetrically distributed around the mean. There is no skewness is the distribution**

**🡪So, if you observe that the mean and median are equal in a dataset, you can infer that the distribution is symmetric, and the skewness is zero.**

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

**A) When the mean>median than the graph is positively skewed (right side).**

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

**A) When the median>mean than the graph is negatively skewed (left side).**

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

**A) Positive kurtosis value indicates the heavier tail and sharper peak than the normal distribution.**

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

**A) Positive kurtosis value indicates the lighter tail and flatter peak than the normal distribution.**

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



What can we say about the distribution of the data?

**A) Data is not symmetrically distributed.**

What is nature of skewness of the data?

**A) Left whisker is longer than the right side, so NEGATIVELY SKEWED.**

What will be the IQR of the data (approximately)?   
**A) Variability is long**

Q19) Comment on the below Boxplot visualizations?



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

**A) In both the boxplots the data is distributed symmetrically but in the first case data is shortly distributed where as in second case data is widely distributed.**

**There is no outliers in both the boxplots.**

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)

**A)P(MPG > 38): 0.4074074074074074**

* 1. P(MPG<40)

**A)** **P(MPG < 40): 0.7530864197530864**

c. P (20<MPG<50)

**A)** **P(20 < MPG < 50): 0.8518518518518519**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**A)** **The data follows 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

**A)** **Z calculated value: 1.772**

**P-value : 0.078**

**H1 is rejected and Ho is accepted.**

**The data follows a normal distribution**

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

**A)** **90.0% Confidence Interval: Z-score: 1.6449**

**94.0% Confidence Interval: Z-score: 1.8808**

**60.0% Confidence Interval: Z-score: 0.8416**

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

**A)** **95.0% Confidence Interval: t-score: 2.0639**

**96.0% Confidence Interval: t-score: 2.1715**

**99.0% Confidence Interval: t-score: 2.7969**

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

**A)** **Probability that 18 randomly selected bulbs would have an average life of no more than 260 days: 0.32167253567098364**