|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continues |
| Weight of Gold | Continues |
| Distance between two places | Continues |
| Length of a leaf | Continues |
| Dog's weight | Continues |
| Blue Color | Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | 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 | Ratio |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ordinal |
| Years of Education | Ratio |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained? TTT, HHH, TTH, THT, HTT, HHT, HTH, THH

= 3/8 = 0.375

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

1. Equal to 1 = Zero
2. Less than or equal to 4 = (1,1) (1,2) (1,3) (2,1) (2,2) (3,1) = 6/36 = 1/6
3. Sum is divisible by 2 and 3 = (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) = 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?

10/21 = 0.476

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

= 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 = 3.09

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**

Median:

Points 3.695

Score 3.325

Weigh 17.710

Var:

Points 0.285881

Score 0.957379

Weigh 3.193166

Mode:

3.07 3.44 17.02

3.92 NaN 18.90

mean 3.596563 3.217250 17.848750

std 0.534679 0.978457 1.786943

Comments: Points - Left Skewed (-ve) then Weigh = Right Skewed = having some outliers (@ 22approx) and Scores = Left Skewed with outliers around 5points.

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?

145.333333

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Skewness:** speed -0.117510

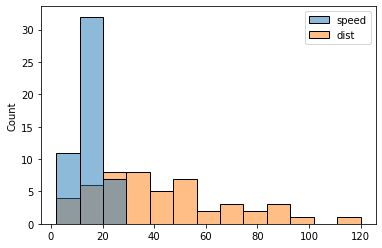
dist 0.806895

**Speed is negatively skewed and distance is positively skewed.**

**Kurtosis:** speed -0.508994

dist 0.405053

**Speed has negative kurtosis and distance has positive kurtosis.**

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Skewness:** SP 1.611450

WT -0.614753

**SP is positively skewed and WT is negatively skewed.**

**Kurtosis:** SP 2.977329

WT 0.950291

**SP has positive kurtosis and WT has positive kurtosis.**

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



In the above histograms graph the peak is right skewed on positive side. Then the Mean value is > Median. We have outliers are there around 300-400 range.



This plot is positively skewed with 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?

Con\_94 = 198.738325292158, 201.261674707842

Con\_96 = 198.62230334813333, 201.37769665186667

Con\_98 = 198.43943840429978, 201.56056159570022

**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?

Mean = 41, Median = 40.5, STD = 5.05 & var = 25.529412

Most of the students scored around 40, average score is 41. Having outliers around 50 and above.

Q13) What is the nature of skewness when mean, median of data are equal? - No Skewness

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

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

Q16) What does positive kurtosis value indicates for a data? - High and narrow peak at the mean

Q17) What does negative kurtosis value indicates for a data? - Wider peak at the mean

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



What can we say about the distribution of the data? Not evenly distributed, most of the data is on right side (Median > Mean)

What is nature of skewness of the data? Negative

What will be the IQR of the data (approximately)? Q3-Q1 = 8

Q19) Comment on the below Boxplot visualizations?



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

1. Both polt’s Data are Normally Distributed.
2. No Outliers.
3. Mean for both the plot is at same point. (same mean)
4. 1st plot is having +ve kurtosis (sharp peak, cover very small range)
5. 2nd plot is having -ve kurtosis (wider peak, covers wider range)

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) - 0.3475939251582705

P(MPG<40) - 0.7293498762151616

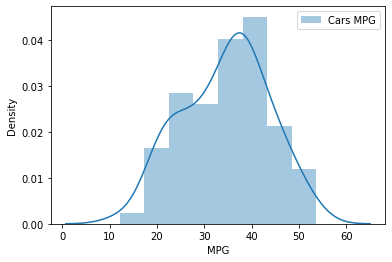
P (20<MPG<50) - 0.8988689169682046

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Yes, MPG in the dataset follows Normal Distribution.

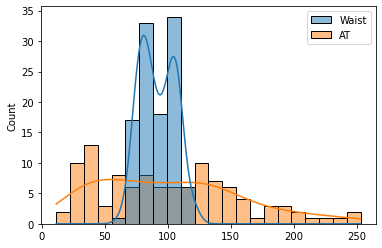


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

AT: Distribution is skewed positively and not normally distributed.

Waist: No Skewness in the distribution, but not normally distributed.



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

-> confidence interval 90 - 1.2815515655446004

confidence interval 94 - 1.5547735945968535

confidence interval 60 - 0.2533471031357997

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

stats.t.ppf(0.975,24) = 2.0638985616280205

In [58]:

stats.t.ppf((1-.96)/2,24) = -2.1715446760080677

stats.t.ppf((1-.99)/2,24) = -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)

tscore = (260-270)/(90/18\*\*.5) = -0.4714045207910317

stats.t.cdf(tscore, 17) = 0.32167253567098364

32%

df 🡪 degrees of freedom = 17