|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuous |
| Weight of Gold | Discrete |
| Distance between two places | Discrete |
| Length of a leaf | Discrete |
| Dog's weight | Continuous |
| 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 | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ordinal |
| 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 | Ratio |
| Years of Education | Interval |

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

Total outcomes for 3 coins = 2^3 = 8

HHT,HTH,THH

P(2H,1T) = 3/8 = 0.375

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

Total outcomes for 2 dice = 6^2 = 36

1. Equal to 1

P(Sum=1) = **0**

1. Less than or equal to 4

P(Sum<= 4) = 6/36 = 1/6 = **0.167**

1. Sum is divisible by 2 and 3

Sum(6) = 5

Sum(12) = 1

P(Sum divisible by 2 and 3) = 6/36 = **0.167**

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?

Total no of balls = 7

Total outcome for pulling out 2 balls = 7C2 = 7\*6/2\*1 = 21

Number of ways of pulling out 2 non blue balls = 5C2 = 5\*4/2\*1 = 10

P(2 NB) = 10/21 = 0.47

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

= 1 \* 0.015 + 4\*0.20 + 3 \*0.65 + 5\*0.005 + 6 \*0.01 + 2 \* 0.12

= 3.10

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**

Mean(Points,Scores,Weigh) =(3.60, 3.22, 17.85)

Median(Points,Scores,Weigh) = (3.70, 3.325, 17.71)

Mode(Points,Scores,Weigh) = (3.07 & 3.92, 3.44, 17.02 & 18.90)

Variance(Points,Scores,Weigh) = (0.29, 0.96, 3.19)

Std Dev (Points,Scores,Weigh) = (0.534, 0.978, 1.787)

Range (Points,Scores,Weigh) = (2.17, 3.911, 8.399)

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  =  (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)

= 1308/9 = 145.33

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

1. Cars speed and distance

**Use Q9\_a.csv**

**Skewness of speed:** -0.118

**Kurtosis of speed:**  -0.508

This means distribution of speed is negatively skewed and the

tail is larger at left side of the distribution. There are

small outliers in the distribution.

**Skewness of distance:**  0.807

**Kurtosis of distance:**  0.405

This means distribution of speed is positively skewed and the

tail is larger at right side of the distribution. There are

large amount of outliers in the distribution.

1. **SP and Weight(WT)**

**Use Q9\_b.csv**

**Skewness of SP:** 1.661

**Kurtosis of SP:**  2.977

This means distribution of SP is highlyly skewed and the

tail is larger at right side of the distribution. The

distribution is too peaked

**Skewness of WT:**  -0.615

**Kurtosis of WT:**  0.950

This means distribution of WT is moderately skewed and the

tail is larger at left side of the distribution. The

distribution is too peaked

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



ANSWER:

1. Median of chick weight lies between 50 to 100
2. Normal distribution is positively skewed
3. There are around 10 chicks that are outliers
4. Minimum weight is around 25 grams
5. Maximum weight is around 385 grams
6. 50% weight ranges from 50 to 200 grams
7. Q3-Q2>Q2-Q1

**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:

N=2000

P=3000000

Mean of n= 200

SD = 30

1. stats.norm.interval(0.94,200,30/(2000\*0.5))

Population weight at 94% interval: 199.95 to 200.056

1. stats.norm.interval(0.96,200,30/(2000\*0.5))

Population weight at 96% interval: 199.94 to 200.061

1. stats.norm.interval(0.98,200,30/(2000\*0.5))

Population weight at 98% interval: 199.93 to 200.069

**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 = 40.5

Variance = 25.5294

SD = 5.052

1. What can we say about the student marks?

Most of the students scored 41 marks

Students who scored 49 and 56 are outliers

ND is evenly distributed

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

There is no skewness in the distribution

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

Positively Skewed

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

Negatively Skewed

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

The peak is too high in distribution and there is huge amount of outliers

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

The distribution is flat and there is small amount of outliers

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



1. What can we say about the distribution of the data?

Median of the data is around 15.25

There is no outlier in this data

1. What is nature of skewness of the data?

The data is positively skewed

1. What will be the IQR of the data (approximately)?   
   Inter Quartile Range will be 10 to 18.1

Q19) Comment on the below Boxplot visualizations?



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

ANSWER:

1. Median is same for both the box plots
2. There are no outliers in both the cases
3. The IQR is small for red box when compared to blue box
4. Both the distribution are symmetric in nature
5. Red box: IQR = 253 to 277, first quartile is at 250 and third quartile is in 275
6. Blue box: IQR = 225 to 308, first quartile is at 225 and third quartile is in 310

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)

Ans: 33/81=0.4074

* 1. P(MPG<40)

Ans: 61/81=0.753

* 1. P (20<MPG<50)

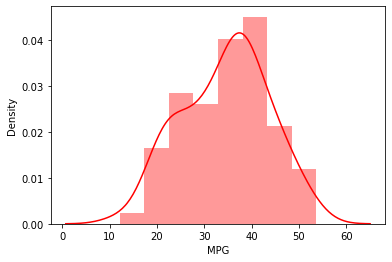
Ans: 69/81=0.851

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

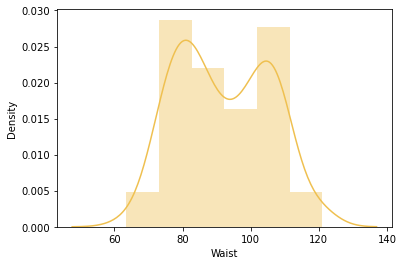
**Yes the MPG is normally distributed**

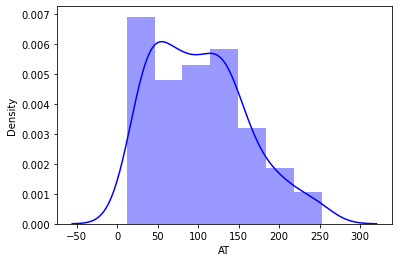


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Due to outliers AT and Waist is not normally distributed





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

1. stats.norm.ppf(0.90) = 1.281
2. stats.norm.ppf(0.94) = 1.5547
3. stats.norm.ppf(0.60) = 0.2533

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

1. stats.t.ppf(0.975,df=24)= 2.0638
2. stats.t.ppf(0.98,df=24) = 2.1715
3. stats.t.ppf(0.995,df=24)= 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

n=18

x=260

SD of sample = 90

Pop mean = 270

t = 260-270/(90/sqrt18)

t= -10/(90/3\*sqrt2)

t = -10/(30/sqrt2)

t = -10 \* sqrt2/30

t= -10 \* 1.41/30

t= -0.4714

df = 17

p(18) = stats.t.sf(abs(-0.4714), df=17) = 0.32167

Therefore 32% of randomly selected 18 bulbs will have an average life of 260 days