|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | DISCRETE |
| Results of rolling a dice | DISCRETE |
| Weight of a person | CONTINUOS |
| Weight of Gold | CONTINUOS |
| Distance between two places | CONTINUOS |
| Length of a leaf | CONTINUOS |
| Dog's weight | CONTINUOS |
| 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 | 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 | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

These combinations are HHH, HHT, HTH, THH, TTH, THT, HTT, TTT.(8 no)

The number of combinations which have two heads and one tail are: HHT, HTH, THH (3 no)

P(H,H,T)= 3/8=o.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

Ans:

1. **P= 0/36 = 0**
2. **Events={(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}**

**Therefore P= 6/36=1/6**

**So probability for sum is Less than or equal to 4 is =1/6**

1. **Events= {(1,1),(1,2),(1,3),(1,5),(2,1),(2,2),(2,4),(2,6),(3,1),(3,3),(3,5),(3,6),**

**(4,2),(4,4),(4,5),(4,6),(5,1),(5,3),(5,4),(5,5),(6,2),(6,3),(6,4),(6,6)**

**}**

**P= 24/36=2/3**

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 number of events= nCr=7C2=7!2!\*5!=21**
* **Interested events=5C2=5!2!\*3!=10**
* **Probability that none of  the balls drawn is blue =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

Ans :

**E(x)= X1P1+ X2P2+…+ XnPn**

**=1\*0.015+4\*0.2+3\*0.65+5\*0.005+6\*0.001+2\*0.120=3.09**

**Therefore, the Expected number of candies for a randomly selected child is 3.09**

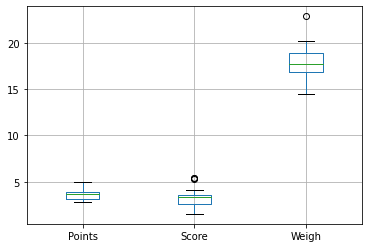
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* 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**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weigh** |
| **Mean** | 3.596563 | 3.217250 | 17.848750 |
| **Median** | 3.695 | 3.325 | 17.710 |
| **Mode** | 3.07, 3.92 | 3.44 | 17.02, 18.90 |
| **Std deviation** | 0.534679 | 0.978457 | 1.786943 |
| **variance** | 0.285881 | 0.957379 | 3.193166 |
| **Range** | 2.71 | 3.910 | 8.399 |

****

**Observations:**

**1)Score and Weight have outliers**

**2) Weigh have highest median value**

**3)Point is positively skewed**

**4) Score & weigh are negatively skewed**

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  = ∑ P(x).E(x)**

**there are 9 patients**

**Probability of selecting each patient = 1/9**

**Ex  108, 110, 123, 134, 135, 145, 167, 187, 199**

**P(x)  1/9  1/9   1/9  1/9   1/9   1/9   1/9   1/9  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**

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

**= 145.33**

**Expected Value of the Weight of that patient = 145.33**

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**

* **Scale of skewness for Speed & Distance lies between -0.5 to 0.5. Hence the data is low skewed i.e approx. symmetric**
* **Distance is moderately skewed**

**Kurtosis**

**speed -0.508994**

**dist 0.405053**

* **Spped has light tail distribution**
* **Distance has heavy tail distribution**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Skewness**

**SP 1.611450**

**WT -0.614753**

* **SP is highly skewed**
* **WT is moderately skewed**

**Kurtosis:**

**SP 2.977329**

**WT 0.950291**

**SP & WT both are heavily tail distribution**

Q10) Draw inferences about the following boxplot & histogram



**1) Frequency of having 50 to 100 Chick Weight is high**

**2) This is rightly skewed graph hence most weights fall under left side of graph i.e frequency of having weight ranging from 0 to 200 is high as compared to from 200 to 400**



**1) This box plot has outliers**

**2) Minimum value is nearest to median as compare to maximum value**

**3) It has positively skewed distribution**

**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:**

X-bar=200, s=30 ,n=2000

#Avg weight of Adult in Mexico with 94% CI

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

(198.738325292158, 201.261674707842)

#Avg weight of Adult in Mexico with 98% CI

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

(198.738325292158, 201.261674707842)

#Avg weight of Adult in Mexico with 96% CI

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

(198.62230334813333, 201.37769665186667)

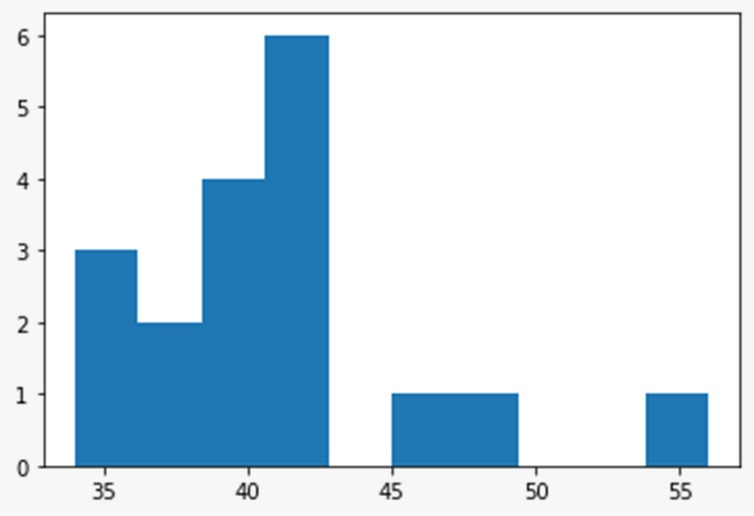
**Q12)** Below are the scores obtained by a student in tests

**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.

* Mean 🡪41
* Median🡪40.5
* Variance🡪25.52
* SD🡪5.05

1. What can we say about the student marks?



* High no of students scored between 35-45 marks
* The students who scored between 45-55 marks is very less

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

\*Skewness is 0

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

\*Data is negatively skewed

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

\*Data is positively skewed

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

\*Thinner peak wide tails

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

\*Wider Peak thinner tails

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



What can we say about the distribution of the data?

* Data is distributed towards right side

What is nature of skewness of the data?

* Negative skweness

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

* IQR of the data is between 10 to 18

Q19) Comment on the below Boxplot visualizations?



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

1. Median of both plots are same
2. The IQR for **BOX PLOT 1** is less Than **BOX PLOT 2**

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)
  2. P(MPG<40)

c. P (20<MPG<50)

# P(MPG>38)

1-stats.norm.cdf(38,cars.MPG.mean(),cars.MPG.std())

0.3475939251582705

# P(MPG<40)

stats.norm.cdf(40,cars.MPG.mean(),cars.MPG.std())

0.7293498762151616

# P (20<MPG<50)

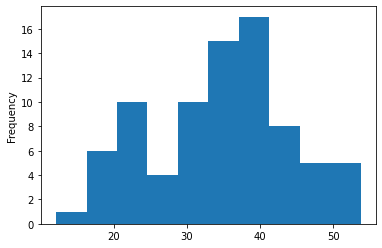
stats.norm.cdf(50,34.422076,9.131445)-stats.norm.cdf(20,34.422076,9.131445)

0.8988689076273199

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



Histogram Of MPG

a. from this graph we can say that this graph does not follow bell shaped curve

b. skewness = -0.17 , skeweness is not equal to 0

c. mean🡪34.42

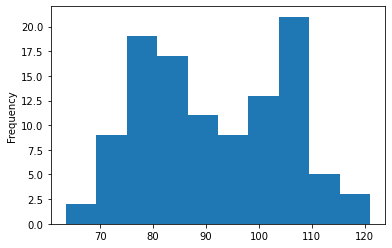
Median🡪35.15

Mode🡪29.62 , mean, median,mode does not have same value.

d. Prob of any random variable is 0.95 which is not equal to 0

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

Dataset: wc-at.csv



Histogram od Waist

Skweness🡪0.13 which is not equal to zero

mean🡪91.901835

Median🡪90.80

Mode🡪94.5 Mean,Median<mode does not have similar values

Hence waist does not follow normal distribution.



Histogram Of AT

Skewness🡪 0.58 which is not equal to zero

Mean🡪 101.89

Median🡪 96.54

Mode🡪121 ,123

mean, median, mode these does not have similar values.

Hence we can say that AT does not follow normal distribution.

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

i) Z scores of 90% confidence interval-- 1.65

ii) Z scores of 94% confidence interval – 1.88

iii) Z scores of 60% confidence interval-- 0.85

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

i) T scores for 95% CI – 2.06

ii) T scores for 96% CI – 2.17

iii) T scores for 99% CI – 2.79

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

Ho = Avg life of Bulb >= 260 days

Ha = Avg life of Bulb < 260 days

from scipy import stats

from scipy.stats import norm

t=(260-270)/(90/18\*\*0.5)

t

-0.4714045207910317

p\_value=1-stats.t.cdf(abs(-0.4714),df=17)

p\_value

0.32167411684460556

Probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 32.17%Assuming significance value α = 0.05

Thus, as p-value > α ; Accept Ho i.e. The CEO claims are false and the avg life of bulb > 260 days