|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| 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 |
| 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 | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Interval |

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

No of possibility=2 pow(3)=8

No.of events={HHT,HTH,THH}

Probability=3/8=0.375

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

1. Equal to 1=0/36=0
2. Less than or equal to 4  
   total outcomes={11,12,21,22,13,31}  
   probability=6/36=1/6=0.16666667
3. Sum is divisible by 2 and 3  
   Outcome={15,24,33,51,42,66}  
   Probabilty=6/36=0.166666667

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?

7C2=42/2=21

5C2=20/2=10  
Probability=10/21=0.476190

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- 1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0120  
 =0.015+0.80+1.95+0.025+0.06+0.240  
 =3.090

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

* For Points,Score,Weigh>

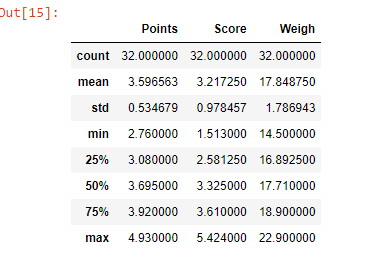
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

Mean Median Mode Variance Std Range

Points=3.596563 3.695 3.07 0.285881 0.534679 2.17

Score=3.217250 3.325 3.44 0.957379 0.978457 1.964

Weigh=17.848750 17.71 17.02 3.193166 1.786943 2.68

**Using Q7.csv file **

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?

Ans:-There are 9 patients and each data are different

So Probability of each person weight is 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(1308)

= 145.333333

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

**Cars speed and distance**

**Using Q9\_a.csv** > skewness(data9$speed)

[1] -0.1139548

Negative skewed=left Skew and here mean is less than median.

> skewness(data9$dist)

[1] 0.7824835  
positive skew means Right skew and here mean is greater than median.

> kurtosis(data9$speed)

[1] 2.422853

Hence it is positive, distribution is sharp

> kurtosis(data9$dist)

[1] 3.248019

It is also sharp distribution also 0.248019 excess kurtosis.

**SP and Weight(WT)**

**Use Q9\_b.csv**

skewness(data10$SP)

[1] 1.581454

Hence Right skew,mean>median

> kurtosis(data10$SP)

[1] 5.723521

Sharp distribution also with excess kurtosis.

skewness(data10$WT)

[1] -0.6033099

Negative skew means left skew median>mean

> kurtosis(data10$WT)

[1] 3.819466

Positive kurtosis.

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



It is Right skew here mean>median.here most of datapoints in left side.

IQR(Inter Quartile Range)=Q3-Q1

It has 4Parts

Lower whisker=25%

Upper whisker=25%

Q2-Q1=25%

Q3-Q2=25%

Lower limt=Q1-1.5\*IQR

Upper limit=Q3+1.5\*IQR



**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:- for 94% CI

(198.738325292158, 201.261674707842)

for 98% CI

(198.43943840429978, 201.56056159570022)

for 96% CI

(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.

Ans:- mean= 41.0

median= 40.5

variance= 25.529411764705884

standard deviation= 5.05266382858645

1. What can we say about the student marks?

Mean>median So skewness is also +ve.  
kurtosis also greater than 3 so excess kurtosis.

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

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

Ans- Right skew

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

Ans-Left Skew

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

Ans-sharp distribution of data

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

Ans- Flatter distribution of data.

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



What can we say about the distribution of the data?

Ans- mean<median

What is nature of skewness of the data?

Ans- Left skew

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

Ans- IQR=Q3-Q1=18-10=8

Q19) Comment on the below Boxplot visualizations?



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

IQR(1)=275-250=25

Lower Extreme=250-25\*1.5=212.5  
 upper Extreme=275+25\*1.5=312.5

IQR(2)=300-225=75  
 Upper Extreme=300+75\*1.5=412.5  
 Lower Extreme=225-75\*1.5

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)

stats.norm.cdf(38,mean,std)

* + 1. 0.6524060748417295
  1. P(MPG<40)
     1. 0.7293498762151616
  2. P (20<MPG<50)
     + 1. stats.norm.cdf(50,car.MPG.mean(),car.MPG.std())-stats.norm.cdf(20,car.MPG.mean(),car.MPG.std())
       2. 0.8988689169682046

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

No, it does not follow in Normal distribution because mean is not equal to median.(mean= 34.422075728024666

And median= 35.15272697)if mean=median then it can be in 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

With the help of distplot on AT-Density and Waist-Density it looks not in Normal distribution. Mean and Median are also not equal.

For AT mean, Median= (101.89403669724771, 96.54)  
 For Waist mean, Median= (91.90183486238533, 90.8)

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

**Z score for 90% 94% 60%**

1.6448536269514722 1.8807936081512509 0.8416212335729143

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

t-Score 95% 96% 99%

stats.t.ppf(cl,df) 1.7108820799094275 2.1715446 2.796939

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 Statistical value=260-270/90/18\*\*0.5= -0.47140452079103173

P value=pt(-0.471,17)= 0.321814