|  |  |
| --- | --- |
| 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 | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | categorical |

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 | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Interval |

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

HHH,HHT,HTH,THH,HTT,THT,TTH,TTT=8

Probability of two head and one tail=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

ANSWER

N(S)=36

a)probability of sum getting equal to 1=0

b) number of favourable outcomes=6

probability of sum getting Less than or equal to 4=6/36=0.166

c)no. of favourable outcomes=6

probability of sum Sum is divisible by 2 and 3=6/36 =0.166

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?

ANSWER:

Two balls drawn random=7c2=7\*6/1\*2 =21

None of the ball is blue = 5c2=5\*4/2\*1 =10

Probability =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

ANSWER:

Probability=(1\*0.015+4\*.20+3\*.65+5\*.005+6\*.01+2\*.120)

=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**

**Answer**

**Mean:** Points-3.596563 Score-3.217250 Weigh-17.848750

Explanation: data.mean()

**Median:** Points-3.695 Score-3.325 Weigh-17.710

Explanation: data.median()

**Standard deviation:** Points-0.534679 Score-0.978457 Weigh-1.786943

Explanation: data.std()

**Variance:** Points-0.285881 Score-0.957379 Weigh-3.193166

Explanation: data.var()

Range: Points-2.17 Score-3.911 Weigh-8.399

Range=highest value -lowest value

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?

ANSWER: (108+110+123+134+135+145+167+187+199)/9=145.33333333333334

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

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |
| --- | --- |
| Answer: |  |

Skewness: speed -0.117510 it is left skew

dist 0.806895 it is positive skew or right

Kurtosis: speed -0.508994 it is left skew

dist 0.405053 it is positive skew or right

**SP and Weight(WT)**

**Answer:**

skewness: SP 1.611450 it is right skew WT -0.614753 left skew

kurtosis: SP 2.977329 it is rightskew WT 0.950291 it is also right dkew

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



Inferences: most of the data lies between 50-100 with frequency of 200

Minimum value is at 400 it is around 0-10

The expected value is above 50

It as outliers



Inferences:It as ouliers of upper box plot.

Mean is greater than medain.

Left extreme is less than right extreme.

**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?

94%- (143.57619175546247, 256.42380824453755)

stats.norm.interval(0.94,200,30)

**98%-** (130.2095637787748, 269.7904362212252)

stats.norm.interval(0.98,200,30)

**96%-** (138.38753268104531, 261.61246731895466)

stats.norm.interval(0.96,200,30)

**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

Data.mean(),

 Median= 40 data.median(),

variance= 24.111, data.var()

Standard deviation= 4.910 data.std()

1. What can we say about the student marks?

Average mark is 40

Mean is slightly equal to median

It may have outliers

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

It is normally distributed(normal distribution).It is symmetrical

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

Skewness is towards right side(+ve skewers)

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

Skewness is towards left side(-ve skewers)

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

It distributed is peaked and posses thick tails.

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

It has lighter tails compared to normal distribution.

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



What can we say about the distribution of the data?

It has tail.It has negative skewer. Mean<median

What is nature of skewness of the data? Negative skewer(it has tail on left side)

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

IQR=upper quartile-lower quartile

=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.

Both the box plot does not have any outliers.

Both are normally distributed due to mean and median is equal.

1st one from around 235-290 and 2nd one from 200-330

Both have same mean and median.

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: [1-stats.norm.cdf(38,34.42,9.13)]

0.34748702501304063

* 1. P(MPG<40)

Ans: stats.norm.cdf(40,34.42,9.13)

0.729457127955707

* 1. P (20<MPG<50)

Ans: stats.norm.cdf(50,34.42,9.13)-stats.norm.cdf(20,34.42,9.13)

0.8989177824549222

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Median>mean so it is not normally distributed

Mean=34 median=35

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

Dataset: wc-at.csv

In waist mean >median so it is not normally distributed

Mean=90 median=91

In AT mean> median so it is also not normally distributed

Mean=101 median=96

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

z-score of 90%confidence level-(-1.6448536269514729,1.644853626951472)

z-score of 94%confidence level(-1.8807936081512509,1.8807936081512509) z-score of 60%confidence level(-0.8416212335729142,0.8416212335729143)

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

T scores

96%- 0.8266861501961364

95%- 0.8242092124949681

99%- .8339745443912898

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

Ans: (1-stats.t.cdf(0.05,17))

0.48035258147818183

0.48/90/np.sqrt(18)

0.0012570787221094178