|  |  |
| --- | --- |
| 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ordinal |

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

S={{HTT},{HTH},{HHH},{THH},{TTH},{TTT},{THT},{HHT}}

P=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

S={(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)}

1. P(sum=0) = 0
2. P(sum<=4) = P(sum=1)+P(sum=2)+P(sum=3)+P(sum=4)

=0 + 1/36 + 2/36 + 3/36

=1/6 = 0.1667

1. P(Sum is divisible by 2 and 3) = P(sum is divisible by 6) =6/36=1/6=0.1667

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?

P(ball is not blue)=5C2/7C2

=10/21

=0.4761

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

Child C – probability of having 4 candies = 0.65

Child D – probability of having 4 candies = 0.005

Child E – probability of having 4 candies = 0.01

Child F – probability of having 4 candies = 0.120

=0.015 + 0.80 + 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,Weight

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

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weight |
| Mean | 3.596 | 3.217 | 17.848 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.07 | 3.44 | 17.02 |
| Variance | 0.277 | 0.927 | 3.0934 |
| STD | 0.526 | 0.963 | 1.759 |
| Range | 2.76 -4.93 | 1.513 -5.424 | 14.5 -22.9 |

**Use 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?

Expected value of weight of random patient = 145.33

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

**Cars speed and distance**

**A]Speed:**

**skewness: 0.8068 Kurtosis:** **-0.508**

**B]Distance:**

**Skewness:** **-0.1175 Kurtosis : 0.405**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**For SP**

**Skewness:1.611 Kurtosis:2.977**

**For Weight**

**Skewness:** **0.615 Kurtosis: 0.95**

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



Histogram showing positively skewed data (right skewed data)

Also Mean > Median.

We have outliers on the right side.

ss



Boxplot has outliers on higher side

**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?

N=2000

Mu=200

Sigma=30

from scipy import stats

#mu=200, N=2000, std dev= 30 find CI for 94% 96% 98%

print("94% ",stats.norm.interval(0.94,200,30))

print("96% ",stats.norm.interval(0.96,200,30))

print("98% ",stats.norm.interval(0.98,200,30))

94% (143.57619175546247, 256.42380824453755)

96% (138.38753268104531, 261.61246731895466)

98% (130.2095637787748, 269.7904362212252)

**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. import numpy as np
3. data=list([34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56])
4. print("Mean",np.mean(data))
5. np.sort(data)
6. print("Median",np.median(data))
7. print("std deviation",round(np.std(data),2))
8. print("variance",round(np.var(data),2))

Mean 41.0

Median 40.5

std deviation 4.91

variance 24.11

2)What can we say about the student marks?

Mean > Median hence data is right skewed.

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

If mean , median of data are equal then skewness is zero

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

If mean>median then it is right skewed data or positively skewed data

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

If mean<median then it is left skewed data or negatively skewed data

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

Positive kurtosis indicates that data is located around mean and curve is leptokurtic

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

Negative kurtosis indicates that data is located away from mean and curve is platykurtik

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



What can we say about the distribution of the data?

Its not normally distributed data as median is not at center.

What is nature of skewness of the data?

Data is negatively skewed as whisker range is maximum at left side as compared to right.

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



Both box plots have the same median.

Both are normally distributed

And whisker range is same at minimum and maximum in both plots.

No outliers.

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)
  3. P (20<MPG<50)

1. import numpy as np
2. from scipy import stats
3. print("P(MPG>38)",round((1-stats.norm.cdf(38,np.mean(cars['MPG']),np.std(cars['MPG']))),2))
4. print("P(MPG<40)",round(stats.norm.cdf(40,np.mean(cars['MPG']),np.std(cars['MPG'])),2))
5. print("P (20<MPG<50)",round(stats.norm.cdf(50,np.mean(cars['MPG']),np.std(cars['MPG']))-stats.norm.cdf(20,np.mean(cars['MPG']),np.std(cars['MPG'])),2))
6. P(MPG>38) 0.35
7. P(MPG<40) 0.73

P (20<MPG<50) 0.9

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

As mean and median of data are almost equal

And distribution plot gives bell shaped curve , its 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

Waist Circumference(Waist) and AT does not follows normal dist.

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

import numpy as np

from scipy import stats

print('Z score for 90% Conifidence Intervl =',np.round(stats.norm.ppf(.05),3))

print('Z score for 94% Conifidence Intervl =',np.round(stats.norm.ppf(.03),3))

print('Z score for 60% Conifidence Intervl =',np.round(stats.norm.ppf(.20),3))

Z score for 90% Conifidence Intervl = -1.645

Z score for 94% Conifidence Intervl = -1.881

Z score for 60% Conifidence Intervl = -0.842

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

print('T score for 95% Conifidence Intervl =',np.round(stats.t.ppf(.025,df=24),3))

print('T score for 96% Conifidence Intervl =',np.round(stats.t.ppf(.02,df=24),3))

print('T score for 99% Conifidence Intervl =',np.round(stats.t.ppf(.005,df=24),3))

Z score for 95% Conifidence Intervl = -2.064

Z score for 96% Conifidence Intervl = -2.172

Z score for 99% Conifidence Intervl = -2.797

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

H0:avg life of bulb>=260

H1:avg life of bulb<260

T\_score=(260-270)/(90/sqrt(17))

T\_score= -0.4714045207910317

Probability that H1 is true =

mu=270

n=18

avg=260

std=90

t\_score=(260-270)/(90/np.sqrt(18))

print(t\_score)

#H0:mu>260

#H1:mu<=260

print("probability that H1 is true ",stats.t.cdf(t\_score,df=17))

-0.4714045207910317

probability 0.32167253567098364