|  |  |
| --- | --- |
| 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 | ratio |
| Blood Group | nominal |
| Time Of Day | ordinal |
| Time on a Clock with Hands | interval |
| Number of Children | ratio |
| Religious Preference | nominal |
| Barometer Pressure | ordinal |
| SAT Scores | interval |
| Years of Education | ordinal |

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

Sol) 3 coins (head or tail) =2^3=8

probability of two heads+ probability of 1 tail

= 2/8+1/8 = 3/8.

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

1. Equal to 1

Sol) No. of combinations when 2 dice are rolled=6\*6=36

but minimum sum will be 1+1=2;

so sum will never be equal to 1 when 2 dies rolled;

so probability is 0.

1. Less than or equal to 4

Sol) 6\*6=36 combinations

(X<=4) => (X=1)+(X=2)+(X=3)=(X=4)

0+1+2+3 = 6

P= 6/36 = 1/6.

1. Sum is divisible by 2 and 3

Sol) n(S)= 36

possible outcomes are (1,5) (4,2) (3,3) (6,6) =4

so P= 4/36 = 1/9.

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?

Sol) Total no. of balls = 7

n(S)= 7C2 = 21

n(E)= n(none of balls drawn from blue) = 5C2 = 10

p(E)= n(E)/n(S)=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

Sol) Random Variable= Sum(xipi)

(1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.120)

Expected no. of candies for a randomly selected child = 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**

Sol) In Python file: Assignment1 Solutions.

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?

Sol)There are 9 patients

Probability of selecting each patients = 1/9

Expected value= summation of P(X).E(X)

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

=1/9(1308)

=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

Sol)In R file: Q9\_a solutions.

**SP and Weight(WT)**

**Use Q9\_b.csv**

Sol)In R file: Q9\_b solutions.

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

Histogram:

The above histogram shows that the data distribution is of ChickWeight$weight with respect to frequency.

The distribution of the data shows it was skewed to the right and is said to be positively skewed.

This kind of distribution has a large number of occurrences in the lower value cells (left side) and few in upper value cells (right side).

That means measures of central tendency are dispersed.

Mean, median and mode are not equal and all the collected data has values greater than zero.

Boxplot:

As we observed, median is closer to the lower/bottom quartile.

i.e Mean>Median.

It means the data constitutes higher frequency of high valued scores.

And boxplot contain outlier that means, the value is higher than 1.5\*IQR above the upper quartile else lower quartile. So it is positively skewed.



**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?

Sol) n=2000, N=3000000, x̅=200, s=30, df=1999

code: from scipy.stats import t

**for 94%:** t.interval(0.94,1999,loc=200,scale=30) #loc=mean, scale=std,

**output: (143.54415570565965, 256.45584429434035)**

**for 96%:** t.interval(0.96,1999,loc=200,scale=30)

**output: (138.34730111522666, 261.6526988847733)**

**for 98%:** t.interval(0.98,1999,loc=200,scale=30)

**output: (130.15355671679083, 269.84644328320917)**

**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. What can we say about the student marks?

Sol) In Python file: Assignment1 Solutions.

Q13) What is the nature of skewness when mean, median of data are equal? Distribution is symmetric with zero skewness.

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 ?

distribution is peaked and possess thick tails.

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

Distribution is flat and has thin tails with wider peaks.

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



What can we say about the distribution of the data?

Sol): In the given Boxplot, Mean is less than median as it is negatively skewed, median is closer to upper quartile range,

Here, lower extreme is 1 and upper extreme is 19, median=around 15,

lower quartile=10, upper quartile =18 and values above upper quartile or lower quartile i.e (Q1-1.5\*IQR or Q3-1.5\*IQR) are outliers.

And the IQR is our 50% of data i.e between 18 to 10.

What is nature of skewness of the data? Negatively skewed

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

Sol) In the following boxplots, the median is same.

They have equal proportions around the median, we can say that the distribution is symmetric or normally distributed.

And there are no outliers.

But the range of both is different. Boxplot1 has lower range than boxplot2.

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)

Sol) python file: Assignment1 Solutions.

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Sol) Since Skewness is -0.1746343, it is negatively skewed, so the given data doesn’t follow 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

Sol) Since Skewness for Adipose Tissue is 0.5767897, positively skewed. The given data doesn’t follow normal distribution.

Since Skewness for Waist Circumference is 0.132204, slightly positively skewed. The given data can follow normal distribution.

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

Sol) for 90%: A=(1+Confidence level)/2 = 1+0.90/2 = 0.95

check the row and column value adjacent to 0.95 in z-table

i.e 1.6(row)+0.05(column)= 1.65.

for 94%: A=(1+Confidence level)/2 = 1+0.94/2 = 0.97

check the row and column value adjacent to 0.97 in z-table

i.e 1.8(row)+0.09(column)= 1.89.

for 60%: A=(1+Confidence level)/2 = 1+0.60/2 = 0.8

check the row and column value adjacent to 0.8 in z-table

i.e 0.8(row)+0.05(column)= 0.85.

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

Sol) Given Sample size(n)=25; so degree of freedom(df)= n-1 = 24 (row)

For 95%: check the value in row(df=24) and column(95%) or α=0.05 from t table

= 2.064

For 99%: check the value in row(df=24) and column(99%) = 2.797

For 96%: check the value in row(df=24) and column(96%) = 2.074.

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

Sol) In python file: Assignment1 solutions.