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

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

ANS=probability of getting 2 heads and 1 tail = 3/8

(HHT),(HTH),(THH)

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

1. Equal to 1

ANS=There is no such probability that equals to 1.therefore 0

1. Less than or equal to 4

ANS=(1,1)(1,2)(1,3)(2,1)(2,2)(3,1)=>6/36=1/6

1. Sum is divisible by 2 and 3

ANS=(1,5)(2,4)(3,3)(4,2)(5,1)(6,6)=>6/36=1/6

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?

ANS=Total no of Chances to draw 2 Balls at Random From 7 Colored Balls, **N(S)=7C2=21**

Let **E** be an event to draw 2 balls other than blue.

No of Chances to draw two balls other than blue,

**N(E)=2C2+3C2+2C1.3C1=1+3+6=10**

The Probability that none of the balls drawn is blue is,

**P(E)=N(E)/N(S)=10/21**

**P(E)=0.476**

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=0.015x1+0.20x4+0.65x3+0.005x5+0.01x6+0.120x2

=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**

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?

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

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



ANS= In this histogram,Mostly the data is on left side of the graph ,so the this is right skewed data.The value of skewness will be positive.The concentration of frequency lies more at scale 50-100.



ANS= Boxplot is used to find the OUTLIERS and DATAPOINTS.It is going to give information about 5 points, They are Mean ,Upper Quartile(Q3),Lower Quartile(Q1),Upper extreme,Lower extreme and also Outliers.The outliers/Tails are more after upper extreme.This boxplot also gives info that it is right skewed data,And the IQR range is very less.

**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?

**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 = Sum of obtained scores/Total no. of scores

(**34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56)/18**

**==>738/18**

**==>Mean/average = 41**

Median=Middle most observation is median.if we find 2 values in odd obervations,Then average of both 2 scores is median

**==>median=40.5**

**==>Variance= 25.4117**

**==>Standard deviation=5.0410**

1. What can we say about the student marks?

=The average marks that students getting among all the tests is 40.5,The deviation of marks is 5.In every test student is varying with 25 marks from mean(41 marks).

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

ANS= perfectly symmetric and we get bell curve, skewness will be “0”,we can apply all statistics concepts.

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

ANS= The nature of skewness will be positive,right skewed data and left biased data.

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

ANS= The nature of skewness will be negative,left skewed data and right biased data.

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

ANS=High peakedness and fat tails

\*positive kurtosis value ranges from 1-infinity

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

ANS=very low peakedness and thin tails

\*negative kurtosis value ranges from -2 to -values

==>For Normal distribution of data kurtosis value should be “3”.

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



What can we say about the distribution of the data?

==>The data is completely lying on right side,so right biased with a mean of 15.5(approx) and the major data is in between 10,18 scale.

What is nature of skewness of the data?

==>The nature of skewness is left skewed and right biased data,the skewness value will be in negative.

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

==>Interquartile Range =Quartile3-Quartile1

===>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.

ANS= >>Mean of both 1 and 2 boxplot is same

>> IQR is more in 2nd boxplot compared to that of 1stbox plot

>>Q1 and Q3 of 1ST box plot is 250 and 280(apprx)

>>Q1 and Q3 of 2ST box plot is 225 and 350(apprx)

\*boxplot 2 has more data than compared to that of boxplot 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)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

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

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

ANS=# T critical value for 95% confidence level with sample size=25==>2.0638985616280205

# T critical value for 96% confidence level with sample size=25==>2.1715446760080677

# T critical value for 99% confidence level with sample size=25==>2.796939504772804

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= Assuming null hypothesis as h0 = avg life of bulb = 260 days

Alternative hypothesis as h1 = avg life of bulb ≠ 260 days

The probability that 18 randomly selected bulbs would have an average life of no

more than 260 days is 0.32167271782057305

---------------------------------------------------------------------------------------------------------------------

NOTE: Few answers in SMD ASSIGNMENT-1(BASICS STATISTICS-1).ipynb attached