|  |  |
| --- | --- |
| 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 | Discrete |
| Blue Color | Continuous |
| 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 | Nominal |
| Hair Color | Nominal |
| Socioeconomic Status | Nominal |
| Fahrenheit Temperature | Ordinal |
| Height | Nominal |
| Type of living accommodation | Nominal |
| Level of Agreement | Ratio |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Ordinal |
| Number of Children | Ordinal |
| Religious Preference | nominal |
| Barometer Pressure | Interval |
| SAT Scores | Nominal |
| Years of Education | Ordinal |

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

Ans: Total no. of possible outcomes = 8

p(a)= p(getting 2 head and 1 tail)

p(a) = 3/8

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

Ans : a

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 balls = 7 = 7c2 = 7!/2!5! = 7\*6/2 = 21

total favorble cases = 5c2 = 5!/2!3! = 5\*4/2 = 10

p = 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

Ans : child A - probability of having 1candy = 0.015

child B = probability of having 4 candies = 0.20

expected no. of candies for a randomly selected child = 1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.12

0.015+0.8+1.95+0.025+0.06+0.24

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.

**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?

Ans : #FIND EXPECTED VALUE

#solution:

expected value = (probability value)

there are 9 pateint

probablity of selecting each patient = 1/9

ex 108,110,123,134,135,145,167,187,199

p(x) 1/9

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

= 1/9(1308)

expected value of the weight of that patient = 145.33

**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 histogran in the x axis it represent the SOCRES.Also, it shows the decreasing trend.100 is very huge histogram and in front of that it represent the frequency

In boxplot it shows that there are outliers on the top of the boxplot.

**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 : random sample (n) = 2000

standard deviation = 30

x = 3,000,00

z-score for 94% confidence interval = 1.55

interval estimate is = 300000\_+1.55(30/sqrt(2000))

z-score for 96% confidence interval = 1.75

interval estimate is = 300000+-1.75(30/sqrt(2000))

z-score for 98% confidence interval = 2.32

interval estimate is = 300000+-2.32(30/sqrt(2000))

**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?

Ans : (PART1)

MEAN = All the items/no. of items

mean = 738/18

mean = 41

MEDIAN = mid position value

sort the data = 18/2 = 9

calculate average 2 values = (40+41)/2 = 81/2

VARIANCE

mean = 41

sqrt(x-u) = 434

var = result/18 = 434/18 = 24.11

STANDARD DEVIATION = sqrt(variance) = sqrt(24.11) = 581.29

ANSWER(PART2)

IT IS GREAT AS THEY ACHIEVE GOOD MARKS.

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

Ans : if the mean ,median is equal then the nature of skewness is 0

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

Ans: The nature of the skewness is positive when the mean is > median

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

Ans: the nature of the skewness is negative when the median > mean

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

Ans: positive values of kurtosis indicates that distributuon is peaked and posseses thick tails.

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

Ans: negative valuees of kurtosis indicates that the distribution has lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

What will be the IQR of the data (approximately)?   
  
  
Ans: (A): The data is distributed at the upper bound and lying from middle To upper quartile and it covers 2nd and 3rd quartile,respectively.

(B):

(C): IQR of the data is Q3-Q1 = 18-2 = 16

Q19) Comment on the below Boxplot visualizations?



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

Ans : boxplot 1 there is no outliers present

boxplot 2 the middle half of the dataset falls within the inter quartile range.

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

Ans : #z score of confidence interval

#90% ci

stats.norm.interval(0.90)

#94% CI

stats.norm.interval(0.94)

#60% CI

stats.norm.interval(0.60)

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

Ans: # t score for 95% and sample size = 25

stats.t.ppf(0.975,25)

t score for 96% and sample size = 25

stats.t.ppf(0.98,25)

#t score for 99% and sample size = 25

stats.t.ppf(0.995,25)

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-statistics for the data is given as follow:

t = x-population mean/s/sqrt(n)

x = mean of the sample of bulbs = 260

population mean = 270

s = standard deviation of the sample = 90

n = numbers of items = 18

t = 260-270/90/sqrt(18)

t = -10/90/3sqrt(2)

t = -10/30/sqrt(2)

t = -1\*sqrt(2)/3

t = 0.471

The number of degrees of freedom in n-1,so here the t-distribution with 17 degrees of freedom.

assuming that the ,mean life of the bulbs is 300 days.