|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete data |
| Results of rolling a dice | Discreate data |
| Weight of a person | Continuous data |
| Weight of Gold | Continuous data |
| Distance between two places | Continuous data |
| Length of a leaf | Continuous data |
| Dog's weight | Continuous data |
| Blue Color | [categorical ]Discrete data |
| Number of kids | Discrete data |
| Number of tickets in Indian railways | Discrete data |
| Number of times married | Discrete data |
| Gender (Male or Female) | [Nominal]Discrete data |

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

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

No of outcomes:- HHH, HHT, HTH, THH, HTT, THT, TTH, TTT.

Two heads and one tail= [HHT, HTH, THH]=3

:- 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
4. Equal to 1= 0
5. Less than or equal to 4= 6
6. Sum is divisible by 2 or 3 = 24/36 =2/3

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?

:- Total no of balls:-7

S= Two balls are drawn at random

N(s) = 7c2=7\*6/2\*1=21

N(s)=21

A= none of the ball drawn is blue

N(a)=5c2=5\*4/2\*1=20/2=10

P(A)=n(a)/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 =0.8

Answer:

Mathematical expectation / mean of randome variable X

E(X) = ∑xp(x)

=(1×0.015)+(4×0.2)+(3×0.65)+(5×0.005)+(6×0.01)

+(2×0.120)

=3.09

The variance of random variable X is:

Var(x) = E(x^2) – [E(x)]^2

E(x^2) = (1×0.015)+(16×0.2)+(9×0.65)+(25×0.005)+(36×0.01)

+(4×0.12)

= 10.03

E(x^2) – [E(x)]^2 = 10.03 – (3.09)^2

= 0.4819

The Standard deviation of random variable X is

S.D(X) =

=0.69419

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?

:- Expected value =108+110+123+134+135+145+167+187+199=1308

Total no of patients N=9

Expected value =1308/9

Expected value =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**



* Data is not following Normal distribution.

From the above histogram, we can conclude that it’s a right skewed distribution. In which the maximum is <=200 and minimum is <=10.



A :- [Box plot] From the above box plot we can conclude that its right skewed, and it has some of the outliers .

* Data is not normal distributed.

**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%, and 96% confidence interval.

Confidence Interval = X̄ ±Z \* (σ/√n)

Mean=X̄=200 pounds

SD=σ=30pounds

N=sampel size=2000

Z score=z=?

For 94% confidence interval Z score=1.88

96% confidence interval Z score =2.05

98% confidence interval Z score=2.33

a)94% confidence interval =200±1.88\*(30/√2000)

=200±1.88\*(30/44.72)

=200±1.26

Confidence interval=198.74, 201.26.

b) 96% confidence interval=200 ± 2.05 \* (30/√2,000)

=200 ± 2.05 \* (30/44.72)

=200 ± 1.37

Confidence interval = 198.63, 201.37.

c)98% confidence interval=200 ± 2.33 \* (30/√2,000)

= 200 ± 2.33 \* (30/44.72)

= 200 ± 1.55

Confidence interval =198.45, 201.55.

**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?

a:- Mean= 34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56

Mean=738/18=41

Median=(40+41)/2=40.5

Varience= Σ(xi - X̄)² / (n - 1)

(xi - X̄)²=434 X̄=41

(n-1)=17

Varience=434/17=25.52

Standard deviation= √[Σ(xi - X̄)² / (n - 1)]or Root of varience

SD= √25.52=5.025

b:- overall by seeing the marks obtained the majority of them are near or in between 40-42, which comes near to the mean.

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

:- When mean and median are equal,it implies that symmetric data[both side equal distribution of data ], there is no skewness

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

:- Right skewed

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

:- left skewed

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

:-Positive kurtosis=Lepto kurtic, outliers can more occur in the dataset

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

:- Negative kurtosis=platy kurtic,outlier are less in the dataset

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



What can we say about the distribution of the data?

A :- considering the graph of Boxplot we can that Q1 > Q3 , because its unequal distribution of data. Where q1=10,q2=15.3,q3=18.2.

\*Left skewed data

What is nature of skewness of the data?

A :- the nature of the skewness is Let skewed[negative skewness]

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

A :- IQR = q3 – q1 , q1=10,q2=15.3,q3=18.2

IQR = 18.2-10

IQR = 8.2.

Q19) Comment on the below Boxplot visualizations?



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

Boxplot 1= range =max-min=287.5-237.5=50 IQR=Q3-Q1=287.5-237.5=50

Boxplot 2= range=max-min=337.5-187.5=150 IQR=Q3-Q1=312.5-225=87.5

* Both boxplots follows normal distribution and not have outliers.

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)

Answers:=

A = P(MPG > 38)=0.4074

B= P(MPG < 40)=0.7531

C = P(20 < MPG < 50)=0.8519

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

From the above data we can conclude that its approximately Normal distributed with the values kurtosis = -0.6116786559430913 and skewness=

-0.17794674747025727

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

Dataset: wc-at.csv

A:-the waist data is not normally distributed, kurt = -1.1026666011768886, skew = 0.1340560824786468.

B:= the AT data follows Normal distribution. kurt = -0.28557567504584425,

Skew = 0.584869324127853.

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

Answer:-

90% confidence interval = 1.645

94% confidence interval = 1.881

60% confidence interval = 0.8416

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

Answer:- degree of freedom = 25-1 = 24

1. 95% Confidence Interval = 2.064
2. 96% Confidence Interval = 2.178
3. 99% Confidence Interval = 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

Answer:= *t*=​*x*ˉ−*μ*​/(s/sqrt(n))  
*x*ˉ = sampel mean = 260days

*μ* = population mean = 270days

s = std = 90 days

n = sampel size = 18

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

t-Score: -0.4714045207910317

Probability: 0.3217