|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete data |
| Results of rolling a dice | Discrete data |
| Weight of a person | Continuous data |
| Weight of Gold | Continuous data |
| Distance between two places | Continious data |
| Length of a leaf | Continuous data |
| Dog's weight | Continious data |
| Blue Color | 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) | 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 | 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 | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nomina |
| 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;

**By probability method**

**Total number of events** = 23=8

**Total number of interested events**

**=** ncr = 3c2 = 3!/(3-2)\*2! = 3

**Probability of interested events**

**P(X=2H)** = Total number of events / Total number of interested events

= 3/8 = 0.375%

= 3.75%

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;

When we roll two dices

**Total number of events**

**=** Number of possibilities per Experiments Number of Experiments = 62=36

we will get minimum sum of 2, Therefore

**Total Number of interested Events = 0**

**Probability of interested Events**

P**(**sum=1**)** = Total number of events/Total Numbers of interested events

= 0/36 = 0%

B) less than are equal to 4

When we roll two dies,

**Total Number of events**

**=** Number of possibilities per Experiments Number of Experiments = 62=36

**Total Number of interested events**

**=** number below that gives sum of 4 or less

= [(3,1), (2,2), (1,3)] = 3

**Probability of interested events**

P**(**sum<4**)** = Total number of events/Total Numbers of interested events

**=** 3/36 = 0.0833 = 8.33%

C) sum is divisible by 2 and 3

**Total Number of events**

**=** Number of possibilities per Experiments Number of Experiments = 62=36

Number below that divisible by 2&3 both = [6,12]

Number of combinations that gives sum 6

= [(5,1), (4,2), (3,3), (2,4), (1,5)] = 5

Number of combinations that gives sum 12

= (6,6) = 1

**Total Number of interested Events = 5+1 =6**

**Probability of interested Events**

= Total number of events/Total Numbers of interested events

**=** 6/36 = 0.166 = 1.66%

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 number of events**

**=** ncr = 7c2= 7\*6/2 = 21

**Total number of interested events**

**=** ncr = 5c2= 5\*4/2 = 10

**Probability of interested events**

P(none of the ball is blue)=total number of events/total number of interested events

= 10/21 = 0.476

= 47.6%

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;** Excepted random value = ∑p(Xi) \* Xi

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

= 3.09

Excepted number 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.

Ans;

**Use Q7.csv fill**

**Median:**

POINTS:3.6

SCORE:3.22

WEIGHT:17.85

**Median:**

POINTS:3.7

SCORE:3.33

WEIGHT:17.71

**Mode:**

POINTS:3.92

SCORE:3.44

WEIGHT:17.02

**VARIENCE:**

POINTS:0.29

SCORE:0.96

WEIGHT:3.19

**STANDARD DEVIATION:**

POINTS:0.53

SCORE:0.98

WEIGHT:1.79

**RANGE:**

POINTS:2.17

SCORRE:3.91

WEIGHT:8.4

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; Probability to pick a patient 1 = 1/9

Probability to pick a patient 2 = 1/9

Probability to pick a patient 3 = 1/9

Probability to pick a patient 4 = 1/9

Probability to pick a patient 5 = 1/9

Probability to pick a patient 6 = 1/9

Probability to pick a patient 7 = 1/9

Probability to pick a patient 8 = 1/9

Probability to pick a patient 9 = 1/9

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

= 145.3

**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**

**Ans;**

**INFERENCE:**

**SPEED:**

1. **Skewness =** -0.117

* Data is slightly negatively skewed or left skewed data (mass of data is on right side of median),
* Means data spared is more on left side of the median

1. **Kurtosis** = -0.508

* data has platykurtic distribution & has thin tails compared to normal dist.
* the distribution is flat as compared to normal distribution

**DISTANCE:**

1. **Skewness** = 0.806

* Data is slightly skewed positively or right skewed data (mass of data is on left side of median),
* Means data spared is more on right side of the median

1. **KURTOSIS** = 0.405

* data has Leptokurtic distribution & has thin tails compared to normal dist.,
* the distribution is peak as compared to normal distribution

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



Ans;

**Inferance:**

1. Positively skewed data (Right skewed data)
2. Spared of the data on right side of the distribution is more &mass of data is on left side median
3. Frequency of the data between 50-100 is more



**Inference:**

1. Positively skewed data or Right skewed data (Whisker is more on right side on medium)
2. Spared of the data on right side of the distribution is more & mass of data is on left side of the median
3. Positive outliers are there on right side of the distribution

**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**

Ans;

1) Find mean, median, variance, standard deviation.

* Mean = 738/18 = 41
* Median = 81/2 = 40.5
* Variance = 1/18-1\*434 = 25.52941
* Standard deviation = √25.52941 = 5.052664

2) What can we say about the student marks?

**Mean = 41**

* Most of student’s marks are nearer to 41

**Median = 40.5 ~mean**

* There is no too high (like 98,76) & too low marks (like0,2) (outliers)present

**Standard deviation = 5.05**

As mean is approximately equal to median follows, normal distribution,

* 1σ = (41-5=36,41+5=47)
* 68% of students are scored between 36 to 47
* 2σ = (41-10=31,41+10=51)
* 95% of students are scored between 41 to 51
* 3SD = (41-15 = 26,41+16 = 57)
* All most all (99.7%) students are scored between 26 to 57

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

Ans;

**when mean =median,** we can say data is normally distributed. which has zero skewness.

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

Ans; **:** **when mean>median** we can say that positively skewed data (Right skewed data)

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

Ans;

**when mean <median** we can say that Negatively skewed data (left skewed data)

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

**Ans;**  Positive kurtosis indicates that:

* Distribution is peak of bell curve is more as compared to normal distribution.
* Spread there are more values around mean

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

**Ans:** Negative kurtosis indicates that

* Distribution is peak of bell curve is less as compared to normal distribution.
* Spread of the data is more far values from the mean.

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



What can we say about the distribution of the data?

* Most of the data lies between 10 to 18
* Q1= 10
* Q2=15 = Median
* Q3=18

What is nature of skewness of the data?

**Negatively skewed data:** there are negative outliers present in the data

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.

Ans;

**BOXPLOT 1:**

* Data ranges between 240 to 280
* Mean = median = mode = Qurtile2(Q2) = 260
* It is Normally distributed
* Quartile 1 = 255
* Quartile 3 = 280
* IQR (INTER QUARTIL RANGE) IS LESS = 280-255 = 25

**BOXPLOT 2:**

* Data ranges between 190 to 130
* Mean = median = mode = Quartile2 (Q2) = 260
* It is normally distributed
* Quartile1 = 220
* Quartile2 = 310
* IQR (INTER QUARTILE RANGE) is more = 310-220 = 90

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)

Ans;

* 1. P(MPG>38)  
     **ANS:**

P(MPG>38) = 34.83%

* 1. P(MPG<40)

**ANS:**

P(MPG<40) = 72.91%

c. P (20<MPG<50)

**ANS:**

P (20<MPG<50) = 89.93%

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans;

Since,

Mean != Median

Skewness =0.177

Kurtosis = 0.6116

In Box plot Q2 is not at center ,whisker is more negative side, Median (Q2) is nearer to Q3 and in bell curve skewed towards negative

numbers

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

Dataset: wc-at.csv

Ans;

Since,

Mean = Median = 91.9018

Skewness = 0.134~0

Kurtosis = -1.01

In Box plot Q2 is approximately at center

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;

Confidence interval of Z score (Z table)

90% = (1.64)

94% = (1.88)

60% = (0.84)

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

Claim: an average light bulb lasts 270days

=μ=2720

Number of sample bulbs =n=18

Average days of sample =X=260days

Standard deviation of sample =S=90 days

To find probability that 18 randomly selected bulbs would have an average life of no more than 260days, we need to calculate t statistics for given data,

t = (x - μ)/s/√n

t = (260-270)/90/√18

t = 0.4714

P t=0.47, df=17

18 randomly selected bulbs would haver an average life of no more than 260 days = 0.3216 = 32.16%