Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| **Activity** | **Data Type** |
| Number of beatings from Wife | **DISCRETE DATA TYPE** |
| Results of rolling a dice | **DISCRETE DATA TYPE** |
| Weight of a person | **CONTINOUS DATA TYPE AND RATIO** |
| Weight of Gold | **CONTINOUS DATA TYPE AND RATIO** |
| Distance between two places | **CONTINOUS DATA TYPE AND INTERVAL SCALE** |
| Length of a leaf | **CONTINOUS DATA TYPE** |
| Dog's weight | **CONTINOUS DATA TYPE AND RATIO** |
| Blue Color | **DISCRETE DATA TYPE & NOMINAL** |
| Number of kids | **DISCRETE & ORDINAL** |
| Number of tickets in Indian railways | **INTERVAL SCALE** |
| Number of times married | **DISCRETE DATA TYPE** |
| Gender (Male or Female) | **DISCRETE DATA TYPE AND NOMINAL** |

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 | INTERVAL |
| Hair Color | NOMINAL |
| Socioeconomic Status | NOMINAL |
| Fahrenheit Temperature | RATIO |
| Height | INTERVAL |
| Type of living accommodation | NOMINAL |
| Level of Agreement | NOMINAL |
| IQ(Intelligence Scale) | INTERVAL |
| Sales Figures | INTERVAL |
| Blood Group | NOMINAL |
| Time Of Day | ORDINAL |
| Time on a Clock with Hands | NOMINAL |
| Number of Children | NOMINAL |
| Religious Preference | ORDINAL |
| Barometer Pressure | RATIO |
| SAT Scores | INTERVAL |
| Years of Education | INTERVAL |

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

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  | trails | 3 |  | **head** | **tails** |  |
|  |  |  |  | 2 | 1 |  |
|  |  |  |  |  |  |  |
| solution |  |  |  |  |  |  |
|  | possible outcomes | |  |  |  |  |
|  |  | H,H,H |  |  |  |  |
|  |  | H,H,T |  |  |  |  |
|  |  | H,T,H |  |  |  |  |
|  |  | H,T,T |  |  |  |  |
|  |  | T,H,H |  |  |  |  |
|  |  | T,H,T |  |  |  |  |
|  |  | T,T,H |  |  |  |  |
|  |  | T,T,T |  |  |  |  |
| total num of outcomes = 8 | | |  |  |  |  |
|  |  |  |  |  |  |  |
|  | num of outcomes with 2 head and 1 tail = 3 | | | | |  |
|  |  | (H,H,T),(H,T,H),(T,H,H) | | |  |  |
|  | I.e. num of favorable outcomes =3 | | | |  |  |
|  |  |  |  |  |  |  |
|  | Probability of getting 2 heads and 1 tail | | | | 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 2and 3

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| solution | |  | |  | |  | | |  | | |  | | |
|  | | the total outcome is 6p6 i.e. = 6\*6 = 36 | | | | | | | | | | | | |
|  | | min possible sum (1,1) = 2 | | | | | | | | | | | | |
|  | | max possible sum (6,6) = 12 | | | | | | | | | | | | |
| 1,1 | | 2,1 | | 3,1 | | 4,1 | 5,1 | | | 6,1 | | |
| 1,2 | | 2,2 | | 3,2 | | 4,2 | 5,2 | | | 6,2 | | |
| 1,3 | | 2,3 | | 3,3 | | 4,3 | 5,3 | | | 6,3 | | |
| 1,4 | | 2,4 | | 3,4 | | 4,4 | 5,4 | | | 6,4 | | |
| 1,5 | | 2,5 | | 3,5 | | 4,5 | 5,5 | | | 6,5 | | |
| 1,6 | | 2,6 | | 3,6 | | 4,6 | 5,6 | | | 6,6 | | |
|  | |  | | total=36 | |  |  | | |  | | |
| a) Equal to 1 | | |  | |  | | |  | | |  | | |  | |  |  |  |
|  | the total outcome is 6p6 i.e. = 6\*6 = 36 | | | | | | | | | | | | |  | |  |  |  |
|  | The min possible sum (1,1) = 2 | | | | | | |  | | |  | | |  | |  |  |  |
|  | equal to 1 = 0 | | | |  | | |  | | |  | | |  | |  |  |  |
|  |  | |  | |  | | |  | | |  | | |  | |  |  |  |
|  | P(1) = possible outcome / total outcome | | | | | | | | | | 0/36= | | | 0 | |  |  |  |
|  |  | |  | |  | | |  | | |  | | |  | |  |  |  |
| b) less than or equal to 4 | | | | |  | | |  | | |  | | |  | |  |  |  |
|  | the total outcome is 6p6 i.e. = 6\*6 = 36 | | | | | | | | | | | | |  | |  |  |  |
|  | the min possible sum outcome = | | | | | | | (1,1)(1,2)(1,3),(2,1)(2,2)(3,1) | | | | | | | |  |  |  |
|  |  | |  | | = | | | 6 | | |  | | |  | |  |  |  |
|  |  | |  | |  | | |  | | |  | | |  | |  |  |  |
|  | P(<=4) = possible outcome/ total outcome = | | | | | | | | | | 6/36= | | | 0.16666667 | |  |  |  |
|  |  | |  | |  | | |  | | |  | | |  | |  |  |  |
|  |  | |  | |  | | |  | | |  | | |  | |  |  |  |
| c) sum is divisible by 2 & 3 | | | | |  | | |  | | |  | | |  | |  |  |  |
|  | the total outcome is 6p6 i.e. = 6\*6 = 36 | | | | | | | | | | | | |  | |  |  |  |
|  | the min possible sum outcome divisible by 2and 3 = | | | | | | | | | | | | | (1,5)(2,4)(3,3)(4,2)(5,1)(6,6) | | | |  |
|  |  | |  | |  | | |  | | | = | | | 6 | |  |  |  |
|  |  | |  | |  | | |  | | |  | | |  | |  |  |  |
|  |  | |  | |  | | |  | | |  | | |  | |  |  |  |
|  | P(sum is divisible by 2&3) = | | | | | | | the min possible sum outcome divisible by 2and 3 / total outcome | | | | | | | | | | |
|  |  | |  | |  | | | 6/36= | | | 0.16666667 | | |  | |  |  |  |

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?

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| solution |  |  |  |  |  |  |  |  |
|  | let S be the total possible outcome | | | |  |  |  |  |
|  | total num of balls [2+3+2] = | | | 7 |  |  |  |  |
|  | n[S]=7C2 = | [7\*6]/[2\*1] | |  |  |  |  |  |
|  | n[S]=7C2 = | **21** |  |  |  |  |  |  |
|  |  | total outcome =21 | |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | n[E] = none of the balls drawn is blue[2+3]= 5 | | | |  |  |  |  |
|  | n[E]=5c2= | [5\*4]/92\*1] | |  |  |  |  |  |
|  | n[E]=5c2= | **10** |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | total possible outcome =10 | | |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | therefore P(non of the balls drawn are blue) = possible outcome/ total outcome | | | | | | | |
|  |  | P[blue]= | 10/21= | **0.47619** |  |  |  |  |

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

Solution

|  |
| --- |
| Expected number 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 |
| =  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**   |  |  |  |  | | --- | --- | --- | --- | |  | Points | Score | Weigh | | Mazda RX4 | 3.9 | 2.62 | 16.46 | | Mazda RX4 Wag | 3.9 | 2.875 | 17.02 | | Datsun 710 | 3.85 | 2.32 | 18.61 | | Hornet 4 Drive | 3.08 | 3.215 | 19.44 | | HornetSportabout | 3.15 | 3.44 | 17.02 | | Valiant | 2.76 | 3.46 | 20.22 | | Duster 360 | 3.21 | 3.57 | 15.84 | | Merc 240D | 3.69 | 3.19 | 20 | | Merc 230 | 3.92 | 3.15 | 22.9 | | Merc 280 | 3.92 | 3.44 | 18.3 | | Merc 280C | 3.92 | 3.44 | 18.9 | | Merc 450SE | 3.07 | 4.07 | 17.4 | | Merc 450SL | 3.07 | 3.73 | 17.6 | | Merc 450SLC | 3.07 | 3.78 | 18 | | CadillacFleetwood | 2.93 | 5.25 | 17.98 | | LincolnContinental | 3 | 5.424 | 17.82 | | Chrysler Imperial | 3.23 | 5.345 | 17.42 | | Fiat 128 | 4.08 | 2.2 | 19.47 | | Honda Civic | 4.93 | 1.615 | 18.52 | | Toyota Corolla | 4.22 | 1.835 | 19.9 | | Toyota Corona | 3.7 | 2.465 | 20.01 | | Dodge Challenger | 2.76 | 3.52 | 16.87 | | AMC Javelin | 3.15 | 3.435 | 17.3 | | Camaro Z28 | 3.73 | 3.84 | 15.41 | | Pontiac Firebird | 3.08 | 3.845 | 17.05 | | Fiat X1-9 | 4.08 | 1.935 | 18.9 | | Porsche 914-2 | 4.43 | 2.14 | 16.7 | | Lotus Europa | 3.77 | 1.513 | 16.9 | | Ford Pantera L | 4.22 | 3.17 | 14.5 | | Ferrari Dino | 3.62 | 2.77 | 15.5 | | Maserati Bora | 3.54 | 3.57 | 14.6 | | Volvo 142E | 4.11 | 2.78 | 18.6 |  |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | | points | score | weight | | Mean | | 3.596563 | 3.21725 | 17.84875 | | Median | | 3.695 | 3.325 | 17.71 | | Mode | | 3.92 | 3.44 | 17.02 | | Variance | | 0.285881 | 0.957379 | 3.193166 | | Standard Deviation | | 0.534679 | 0.978457 | 1.786943 | | Range | MAX | 4.93 | 5.424 | 22.9 | | MIN | 2.76 | 1.513 | 14.5 | |

Inferences: a) For Points dataset: 1) The data is concentrated aroound Median 2) There are no outliars 3) The distribution is Right skewed b) For Score dataset: 1) The data is concentrated around Median 2) There are 3 Outliars: 5.250, 5.424, 5.345 3) The distribution is Left skewed c) For Weigh dataset: 1) The data is concentrated around Median 2) There is 1 Outliar: 22.90 3) The distribution is Left skewed

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?

Solution

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

No of variables = 9

Probability of selecting each patient = 1/9 = P(x)

Expected Value  =  ∑ ( probability  \* Value )

 ∑ P(x)\*E(x)

Expected 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

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

= (1/9)  (  1308)

= 145.33

**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 ?

Solution

Sample mean = 200 Sample SD = 30 n = 2000

CI = sample mean +- (z\*(stdev/squrt(n)

Z(94%) = (0.94+1)/2

Z(94%) = 0.97

Z(94%) by table = 1.89

CI(94%) =200+(1.89(30/sqrt(2000) =201.26

CI(94%) =200-(1.89(30/sqrt(2000) =198.73

Z(98%) = (0.98+1)/2

Z(98%) = 0.99

Z(98%) by table = 2.33

CI(98%) =200+(2.33(30/sqrt(2000) =201.56

CI(98%) =200-(2.33(30/sqrt(2000) =198.43

Z(96%) = (0.96+1)/2

Z(96%) = 0.98

Z(96%) by table = 2.06

CI(96%) =200+(2.06(30/sqrt(2000) =201.38

CI(96%) =200-(2.06(30/sqrt(2000) =198.61

**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?

Solution

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

**Mean** =41.0

Median=(40+41)/2=40.5

Variance= 434/(18-1)

=25.52

Stddev =squrt(variance)

Stdev=(squrt(25.52)

=5.052

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

Data is normal and normal distribution

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

Distribution will be on RIGHT side soo –ve skewness

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

Distribution will be on LEFT side soo +ve skewness.

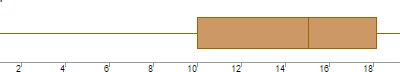
Q16) What does positive kurtosis value indicates for a data ?

Long tails

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

Short tails

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



What can we say about the distribution of the data?

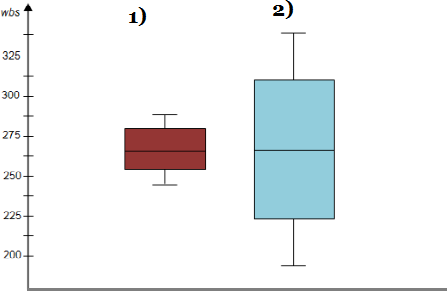
Not normally distributed

What is nature of skewness of the data?

Negative skewness

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

10-18

Q19) Comment on the below Boxplot visualizations? 

An Inference from the distribution of data of 1 & 2.

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.

Ci(0.95) 1.96

Ci(0.96) 2.05

CI(0.99) 2.34

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

Solution

x = mean of the sample of bulbs =  260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18