|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | descrete |
| Results of rolling a dice | descrete |
| 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 | descrete |
| Number of kids | descrete |
| Number of tickets in Indian railways | descrete |
| Number of times married | descrete |
| Gender (Male or Female) | descrete |

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

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

Three coins are tossed(S) ={ HHT, HTT, THT, TTT, HHH, HTH, THH, TTH }

S1= {HHT, HTH, THH}

Probablity of getting two heads and one tail are p(S1)

= n(S1)/n(S)

= 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

Two Dice are rolled

Let sample space be (S) = { (1, 1) (1, 2) (1, 3) (1, 4) (1, 5) (1, 6)

(2, 1) (2, 2) (2, 3) (2, 4) (2, 5) (2, 6)

(3, 1) (3, 2) (3, 3) (3, 4) (3, 5) (3, 6)

(4, 1) (4, 2) (4, 3) (4, 4) (4, 5) (4, 6) (5, 1) (5, 2) (5, 3) (5, 4) (5, 5) (5, 6)

(6, 1) (6, 2) (6, 3) (6, 4) (6, 5) (6, 6) }

a) sum is equal to 1 p(a) = 0

p(a) = n(a)/n(S)

= 0/36 = 0

b) Sum is less than or equal to 4 p(b) = {(1,1) (1,2) (1,3) (2,1) (2,2) (3,1) }

p(b) = n(b)/n(S)

= 6/36 = 1/6 = 0.167

c) Sum is divisible by 2 and 3 p(c) = {(1, 5) (2, 4) (3, 3) (4, 2) (5, 1) (6, 6)}

= n(c)/ n(S)

= 6/36 = 1/6 = 0.167

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 numbers of balls = 7

Let S be the sample space.

Then, n(S) = Number of ways of drawing 2 balls out of 7 =7C2

= (7 x 6) / (2 x 1)

= 21

Let E = Event of drawing 2 balls, none of which is blue.

Therefore n(E) = Number of ways of drawing 2 balls out of (2 + 3) balls.

= 5 C2

= (5 x 4)/(2 x 1)

= 10

Therefore P(E) = 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

**=**

Answer=  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 :-**

Mean = 3.596563

Median = 3.695

Mode = 3.92

3.21725

MedianVariance = 0.285881

Standard Deviation = 0.534679

Range = 2.17

**Score :-**

Mean = = 3.325

Mode = 3.44

Variance = 0.957379

Standard Deviation = 0.978457

Range = 3.911

**Weight :-**

Mean = 17.84875

Median = 17.71

Mode = 17.02

Variance = 3.193166

Standard Deviation = 1.786943

Range = 8.4

**Inferences:**

1)The given data of points, scores, weights are shifted towards left and its tail on the right side.

2)it is a positively/right-skewed distribution.

3) Points dataset is a bimodal because it is having two modes and score, weight are unimodal because it having single mode.

Q8) Calculate Expected Value for the problem below

* 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?

=  ∑ ( probability  \* Value )

  = ∑ P(x).E(x)

Probability of selecting each patient = 1/9

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

= (1/9)  \* (  1308)

= 145.33

Weight of that patient = 145.33

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

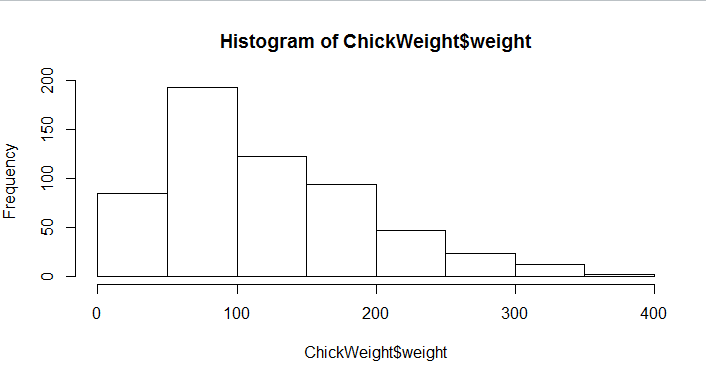
**Cars speed and distance**

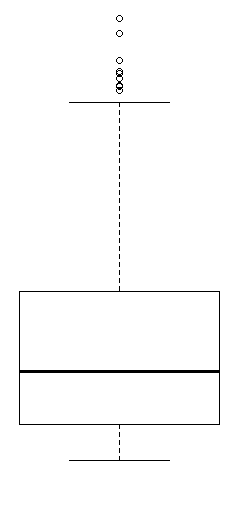
**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

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





**=**

from the above histogram, we can clearly say that most of given data are shifted towards left side and its tail on the right side, it is a positively or right-skewed distribution.

So from boxplot, we can conclude that their are some outliers exists on right side of 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?

**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**

* Find mean, median, variance, standard deviation.
* What can we say about the student marks?

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

= 738 / 18

= 41

**Median =** (40+41)/2

= 81/2

= 40.5

**Variance (S^2) =** (49+25+25+9+9+4+4+1+1+0+0+0+0+1+1+16+64+225)/18

= 434/18

= 24.111

**Standard Deviation =** 4.91

**Inference:** The given data are shifted towards left and its tail on the right side, it is a positively/right-skewed distribution.

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

**=** If the nature of skewness when mean, median of data are equal The distribution is both symmetric and unimode, skewness is zero and mean = median = mode

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

= The distribution is positively skewed.

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

**=** The distribution is negatively skewed.

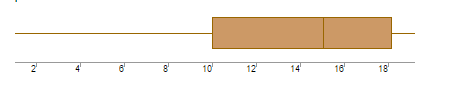
Q16) What does positive kurtosis value indicates for a data ?

**=** Positive value of kurtosis indicates that the distribution is peaked and possesses thick tails

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

= Negative value of kurtosis indicates that the distribution is flat and has thin tails.

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



What can we say about the distribution of the data?

**= :** The peak of the given boxplot or the most of data points are situated towards right side and the tail is at left side of boxplot.

What is nature of skewness of the data?

= Negative Skewness

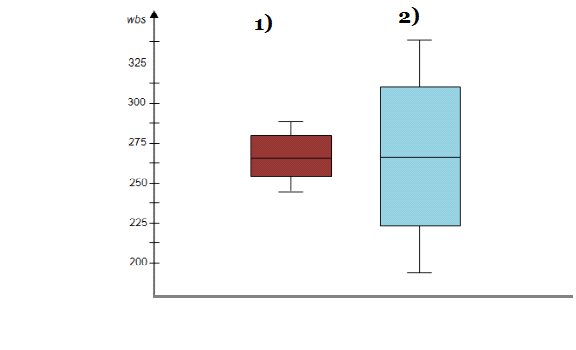
What will be the IQR of the data (approximately)?

**:** IQR = upper quartile – lower quartile

= 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.

= For boxplot 1 and boxplot 2, value of median is same.

Value of IQR(inter quartile range) for boxplot 1 is less than value of IQR for 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

* P(MPG>38)
* P(MPG<40)

c. P (20<MPG<50)

Q 21) Check whether the data follows normal distribution

* Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

=

MPG of Cars does not follows Normal Distribution rule, data distributed randomly.

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

Dataset: wc-at.csv

=

Data does not follows bell curve, data distributed randomly. Mean ≠ Median ≠ Mode

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

=

**:** for 90% , z\_score = (1 - 0.90)

=0.1

Z0.1 = 1.28 (Using Z-Table)

For 94 % , z\_score = (1 – 0.94)

= 0.06

Z0.06 = 1.55 (Using Z-Table)

For 60% , z\_score = (1 - 0.60)

= 0.4

Z0.4 = 0.25 (Using Z-Table)

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

=

: t\_score of 95% = 1 – 0.95 / 2

= 0.05 / 2

= 0.025

T0.025 for sample size 25 = 2.060

: t\_score of 96% = 1 – 0.96 / 2

= 0.04 / 2

= 0.02

T0.02 for 25 = 2.485

: t\_score of 99% = 1 – 0.99 / 2

= 0.01 / 2

= 0.005

T0.005 for 25 = 2.787

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

= Formula: t = x - µ / ϭ / √n