|  |  |
| --- | --- |
| 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 | Continuous |
| Blue Color | Discrete |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Continuous |
| 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 | Interval |
| 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) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| 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?

Solution: {H,H,H},{H,H,T}, {H,T,H}, {H,T,T}{T,T,T},{T,T,H},{T,H,H},{T.H,T}.

Hence the probability of two heads and one tail is 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

Solution: Probability of two dice when rolled

{(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)}=36

1. Equal to 1

Probability is 0/36 = 0.

1. Less than or equal to 4

{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}

Probability of less than or equal to 4 is 6/36 = 1/6.

1. Sum is divisible by 2 and 3.

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

Probability of sum is divisible by 2 and 3 is 6/36 = 1/6.

.

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: Red=2, green=3,blue=2

Two balls can be selected from 7 balls 7C2 = 21ways

Find no. of ways no blue ball is selected, both ball is selected from Red and Green set of 5, two balls can be selected 5C2 = 10 ways

Required probability = 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

Solution: Expected no. of candies = sum (X \* P(x))

= (1 candy by Child A) \* P(1 candy by Child A) + (4 candies by Child B) \* P(4 candies by Child B) + (3 candies by Child C) \* P(3 candies by Child C)

+ (5 candies by Child D) \* P(5 candies by Child D) + (6 candies by Child E) \* P(6 candies by Child E) + (2 candies by Child F) \* P(2 candies by Child F)

= 1 \* 0.015 + 4 \* 0.20 + 3 \* 0.65 + 5 \* 0.005 + 6 \* 0.01 + 2 \* 0.12

Expected no. of candies for randomly selected child is 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**

Solution: Below is the Mean, Median, Mode, variance, Standard Deviation and Range for given data set.

|  |  |  |  |
| --- | --- | --- | --- |
|  | 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 |
| SD | 0.534679 | 0.978457 | 1.786943 |
| Range | 2.17 | 3.911 | 8.4 |

The average value for Points, Score, Weight is 3.597, 3.217, 17.84 respectively. The standard deviation for all variable is low, this implies average reliable measure of central tendency.

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: Expected Value = ∑ (probability \* Value)

 ∑ P(x). E (x)

there are 9 patients

Probability of selecting each patient = 1/9

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

P(x) 1/9 1/9   1/9 1/9   1/9   1/9   1/9   1/9 1/9

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

Random patient weight = 145.33 pounds

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

**Cars speed and distance**

**Use Q9\_a.csv**

Solution:

|  |  |  |
| --- | --- | --- |
|  | Speed | dist |
| Skewness | -0.114 | 0.7825 |
| Kurtosis | -0.5771 | 0.248 |

**SP and Weight(WT)**

**Use Q9\_b.csv**

Solution:

|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| Skewness | 1.5815 | -0.6033 |
| Kurtosis | 2.72352 | 0.8195 |

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



Solution: Both Histogram and Boxplot is right skewed or positively skewed distribution as per above diagram.

**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: CI = Confidence Interval = x̄ ± z (1-α) (σ/√n)

1. For 94% confidence interval

Lower = 198.74

Upper = 201.26

1. For 98% confidence interval

Lower = 198.44

Upper = 201.56

1. For 96% confidence interval

Lower = 198.62

Upper = 201.38

**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.

Solution:

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.5 |
| Variance | 25.53 |
| SD | 5.05 |

1. What can we say about the student marks?

Solution: Student marks are right skewed or positively skewed where mean > median.

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

Solution: When mean and median are equal, the skewness is zero

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

Solution: The distribution is right skewed or positively skewed.

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

Solution: The distribution is left skewed or negatively skewed

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

Solution: When kurtosis is positive, the data has heavy tails.

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

Solution: If the kurtosis is negative (less than 0), then the distribution has a flat top.

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



What can we say about the distribution of the data?

Solution: Not normally distributed

What is nature of skewness of the data?

Solution: Negative skewness

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

Solution: IQR = 10-18 = 8

Q19) Comment on the below Boxplot visualizations?



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

Solution: Data in boxplot 1 has a lower variability than data in boxplot 2 , Both boxplots exhibit a normal distribution since all sections of the boxplots are even.

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)

Solution: From given data set Cars.csv ,the fearture MPG mean=34.422, SD=9.13144, using R

1. P(MPG>38) where X=38

1-pnorm(38, 34.422, 9.1344) = 0.3475.

1. P(MPG<40) X=40

pnorm(40, 34.422, 9.1344) = 0.72935.

1. P (20<MPG<50)

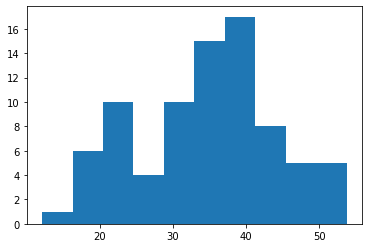
pnorm(20, 34.422, 9.1344) – (1- pnorm(50, 34.422, 9.1344)) = 0.01311818

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

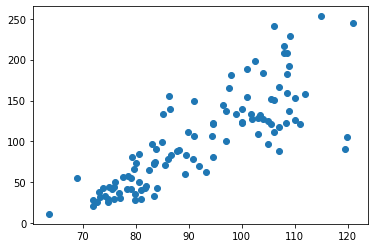
Solution:



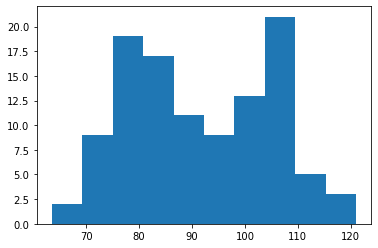
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Solution: Waist Vs AT



Waist Circumference(Waist) normal distributed



Adipose Tissue (AT) not normally distributed



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

Solution:

for 90% z score=1.64

for 94% z score=1.88

for 60% z score=0.84

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

Solution:

95% Confidence Interval

So, Significance Level﻿(α)﻿ = 1 - 0.95 = 0.05

﻿t score=tα/2​, df​﻿ ﻿=t20.05​,24​=t0.025,24​﻿ ﻿=2.064﻿ ﻿﻿ (From t table)

96% Confidence Interval

So, Significance Level﻿(α)﻿ = 1 - 0.96 = 0.04

﻿t score=tα/2​, df​﻿ ﻿=t20.04​,24​=t0.02,24​﻿ ﻿=2.492﻿ ﻿﻿

99% Confidence Interval

So, Significance Level﻿(α)﻿ = 1 - 0.99 = 0.01

﻿t score=tα/2​, df​﻿ ﻿=t20.01​,24​=t0.005,24​﻿ ﻿=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

Solution:Refer QA24.R script enclosed

The probability that **t < - 0.471 with 17 degrees of freedom**, there is a 67% chance that the average bulb life for 18 randomly selected bulbs would be less than or equal to 260days.