|  |  |
| --- | --- |
| 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 | Discrete |
| 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 | Ordinal |
| Celsius Temperature | Ratio |
| Weight | Interval |
| Hair Color | Nominal |
| Socioeconomic Status | Nominal |
| Fahrenheit Temperature | Ratio |
| Height | Interval |
| Type of living accommodation | Ordinal |
| 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?

Sol;- The number of possibility by tossing 3 coins

N(S)={HHH,HHT,HTH,THH,THT,TTT,TTH,HTT}

N(S)=8

The number of 2 heads and 1 tail are

N(P)={HHT,HTH,THH}

N(P)=3

N(E)=N(P)/N(S)

N(E)=3/8=0.4

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

Sol;- When 2 dice are rolled the possible cases =6x6 =36

1. Total favorable cases(having sum =1)=0

Hence, minimum sum is 2 for outcome(1,1)

So, probability is 0

1. When roll 2 dice, possibility of getting number 4 is(1,3),(2,2),(3,1)

so the outcomes =3=N(P)

N(S)=36

N(E)=N(P)/N(S)=3/36

N(E)=1/12

So, probability of getting a sum of 4 is 1/12

1. N(S)=36

Favorable outcomes = (1,5),(3,3),(4,2),(5,1),(2,4)

N(P)=5

N(E)=N(P)/N(S)=5/36

So, probability of sum divisible by 2 & 3 is 5/36

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?

Sol;- A bag contains (2+3+2)=7 balls

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

=7C2

=(7x6)/(2x1)

=21

Let n(E)=number of ways of drawing 2 balls, none of which is blue

--🡪 n(E)=Number of ways of drawing 2 balls out from (2+3)balls

=5C2

=(5x4)/(2x1)

=10

---🡪P(E)=n(E)/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 = 0.20

Expected number of candies for randomly selected students is

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

=0.015+0.80+1.65+0.025+0.06+0.240

=3.090

S0, expected number of candies for a randomly selected child is = 3.090

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 | Scores | Weight |
| Mean | 3.596 | 3.217 | 17.848 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.891 | 3.54 | 17.43 |
| Variance | 0.285 | 0.957 | 3.19 |
| SD | 0.534 | 0.978 | 1.786 |
| Range | 2.76,4.83 | 1.513,5.424 | 14.5,22.9 |

**Sol;-**

**Chart, box and whisker chart

Description automatically generated**

**Points Sports Weigh**

3.90 2.620 16.46

3.90 2.875 17.02

3.85 2.320 18.61

3.08 3.215 19.44

3.15 3.440 17.02

2.76 3.460 20.22

3.21 3.570 15.84

3.69 3.190 20.00

3.92 3.150 22.90

3.92 3.440 18.30

3.92 3.440 18.90

3.07 4.070 17.40

3.07 3.730 17.60

3.07 3.780 18.00

2.93 5.250 17.98

3.00 5.424 17.82

3.23 5.345 17.42

4.08 2.200 19.47

4.93 1.615 18.52

4.22 1.835 19.90

3.70 2.465 20.01

2.76 3.520 16.87

3.15 3.435 17.30

3.73 3.840 15.41

3.08 3.845 17.05

4.08 1.935 18.90

4.43 2.140 16.70

3.77 1.513 16.90

4.22 3.170 14.50

3.62 2.770 15.50

3.54 3.570 14.60

4.11 2.780 18.60

Q8) [Calculate](file:///C:\Users\91805\Downloads) 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?

Sol:- Expected Value  =  ∑ ( probability  \* Value )

 ∑ P(x).E(x)

there are 9 patients

Probability of selecting each patient = 1/9

E(x) =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

Expected Value of the Weight of that patient = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
| Index | speed | dist |
| 1 | 4 | 2 |
| 2 | 4 | 10 |
| 3 | 7 | 4 |
| 4 | 7 | 22 |
| 5 | 8 | 16 |
| 6 | 9 | 10 |
| 7 | 10 | 18 |
| 8 | 10 | 26 |
| 9 | 10 | 34 |
| 10 | 11 | 17 |
| 11 | 11 | 28 |
| 12 | 12 | 14 |
| 13 | 12 | 20 |
| 14 | 12 | 24 |
| 15 | 12 | 28 |
| 16 | 13 | 26 |
| 17 | 13 | 34 |
| 18 | 13 | 34 |
| 19 | 13 | 46 |
| 20 | 14 | 26 |
| 21 | 14 | 36 |
| 22 | 14 | 60 |
| 23 | 14 | 80 |
| 24 | 15 | 20 |
| 25 | 15 | 26 |
| 26 | 15 | 54 |
| 27 | 16 | 32 |
| 28 | 16 | 40 |
| 29 | 17 | 32 |
| 30 | 17 | 40 |
| 31 | 17 | 50 |
| 32 | 18 | 42 |
| 33 | 18 | 56 |
| 34 | 18 | 76 |
| 35 | 18 | 84 |
| 36 | 19 | 36 |
| 37 | 19 | 46 |
| 38 | 19 | 68 |
| 39 | 20 | 32 |
| 40 | 20 | 48 |
| 41 | 20 | 52 |
| 42 | 20 | 56 |
| 43 | 20 | 64 |
| 44 | 22 | 66 |
| 45 | 23 | 54 |
| 46 | 24 | 70 |
| 47 | 24 | 92 |
| 48 | 24 | 93 |
| 49 | 24 | 120 |
| 50 | 25 | 85 |

Car speed Distance

Skewness -0.1139548 0.7824835

Kurtosis 2.422853 3.24801

**SP and Weight(WT)**

**Use Q9\_b.csv**

Sol:-

|  |  |  |
| --- | --- | --- |
| Sl.no | SP | WT |
| 1 | 104.1854 | 28.76206 |
| 2 | 105.4613 | 30.46683 |
| 3 | 105.4613 | 30.1936 |
| 4 | 113.4613 | 30.63211 |
| 5 | 104.4613 | 29.88915 |
| 6 | 113.1854 | 29.59177 |
| 7 | 105.4613 | 30.30848 |
| 8 | 102.5985 | 15.84776 |
| 9 | 102.5985 | 16.35948 |
| 10 | 115.6452 | 30.92015 |
| 11 | 111.1854 | 29.36334 |
| 12 | 117.5985 | 15.75353 |
| 13 | 122.1051 | 32.81359 |
| 14 | 111.1854 | 29.37844 |
| 15 | 108.1854 | 29.34728 |
| 16 | 111.1854 | 29.60453 |
| 17 | 114.3693 | 29.53578 |
| 18 | 117.5985 | 16.19412 |
| 19 | 114.3693 | 29.92939 |
| 20 | 118.4729 | 33.51697 |
| 21 | 119.1051 | 32.32465 |
| 22 | 110.8408 | 34.90821 |
| 23 | 120.289 | 32.67583 |
| 24 | 113.8291 | 31.83712 |
| 25 | 119.1854 | 28.78173 |
| 26 | 114.5985 | 16.04317 |
| 27 | 120.7605 | 38.06282 |
| 28 | 119.1051 | 32.83507 |
| 29 | 99.56491 | 34.48321 |
| 30 | 121.8408 | 35.54936 |
| 31 | 113.4846 | 37.04235 |
| 32 | 112.289 | 33.23436 |
| 33 | 119.9211 | 31.38004 |
| 34 | 121.3926 | 37.57329 |
| 35 | 111.289 | 32.70164 |
| 36 | 115.0131 | 31.91122 |
| 37 | 114.0934 | 28.754 |
| 38 | 116.9094 | 27.87992 |
| 39 | 116.9094 | 28.6305 |
| 40 | 128.4613 | 30.11543 |
| 41 | 116.3926 | 37.39252 |
| 42 | 115.7488 | 35.02718 |
| 43 | 117.4613 | 30.52743 |
| 44 | 114.0934 | 28.34398 |
| 45 | 114.381 | 33.07863 |
| 46 | 117.1051 | 32.62192 |
| 47 | 118.2087 | 36.49862 |
| 48 | 116.4729 | 33.91006 |
| 49 | 127.9094 | 28.0706 |
| 50 | 118.289 | 33.45847 |
| 51 | 118.289 | 33.21395 |
| 52 | 118.289 | 33.43671 |
| 53 | 120.4043 | 40.39816 |
| 54 | 143.3926 | 37.62069 |
| 55 | 135.3926 | 37.25439 |
| 56 | 126.4043 | 40.58907 |
| 57 | 110.4613 | 30.14754 |
| 58 | 118.289 | 32.73452 |
| 59 | 112.6452 | 30.61528 |
| 60 | 115.5766 | 37.66287 |
| 61 | 130.2087 | 36.88815 |
| 62 | 117.6685 | 37.86041 |
| 63 | 126.0481 | 43.39099 |
| 64 | 125.3123 | 40.72283 |
| 65 | 128.1284 | 40.15948 |
| 66 | 126.5985 | 15.71286 |
| 67 | 132.4846 | 37.97996 |
| 68 | 133.6802 | 41.57397 |
| 69 | 133.3123 | 40.47204 |
| 70 | 158.3007 | 37.14173 |
| 71 | 164.5985 | 15.82306 |
| 72 | 133.416 | 44.01314 |
| 73 | 133.1401 | 43.35312 |
| 74 | 124.7152 | 52.99775 |
| 75 | 121.8642 | 42.6187 |
| 76 | 132.8642 | 42.77822 |
| 77 | 169.5985 | 16.13295 |
| 78 | 150.5766 | 37.92311 |
| 79 | 151.5985 | 15.76963 |
| 80 | 167.9445 | 39.4231 |
| 81 | 139.8408 | 34.94861 |

SP WT

Skewness 1.581454 -0.6033099

Kurtosis 5.723521 3.81946

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



Ans: The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.



Ans: The boxplot has outliers on the maximum side.

**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?

Sol:-

The **information given** is:

* Sample **mean**of .
* Sample **standard deviation** of .
* Sample **size**of .

The **interval** is:



* In which **t** is the critical value for the two-tailed confidence interval.

Considering a **94%** confidence level, using a calculator, with 200 - 1 = **199 df**, the critical value is **t = 1.8916**, hence:





The **94%** confidence interval is **(198.73, 201.27).**

Considering a **96%** confidence level, using a calculator, with 200 - 1 = **199 df**, the critical value is **t = 2.0673**, hence:





The **96%** confidence interval is **(198.61, 201.39).**

Considering a **98%** confidence level, using a calculator, with 200 - 1 = **199 df**, the critical value is **t = 2.3452**, hence:





The **98%** confidence interval is **(198.43, 201.57).**

Using the **t-distribution**, it is found that:

* The **94%** confidence interval is **(198.73, 201.27).**
* The **96%** confidence interval is **(198.61, 201.39).**
* The **98%** confidence interval is **(198.43, 201.57).**

**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.

Sol:- Mean=41

Median=40.5

Standard deviation=5.05

Variance=25.52

1. What can we say about the student marks?

Ans:- we don’t have outliers and the data is slightly skewed towards right because mean is greater than median.

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

Ans:- If the distribution is symmetric, then the mean is equal to the median and the distribution has zero skew

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

Ans:-Skewness and tail is towards Right

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

Ans:-Skewness and tail is towards Left

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

Ans:-Positive kurtosis value indicates that thinner peak and wider tails.

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

Ans:- Negative kurtosis value indicates that wider peak and thinner tails.

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



What can we say about the distribution of the data?

Ans:-The above Boxplot is not normally distributed the median is towards the higher value

What is nature of skewness of the data?

Ans:- The data is a skewed towards left. The whisker range of minimum value is greater than maximum

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

Ans:-IQR=Q3-Q1

IQR=18-10

IQR=8

Q19) Comment on the below Boxplot visualizations?



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

Ans:- First there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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)

Sol:- P(MPG>38)=0.348

* 1. P(MPG<40)

Sol:- P(MPG<40)=0.729

* 1. P (20<MPG<50)

Sol:- p(MPG>20)= 0.943

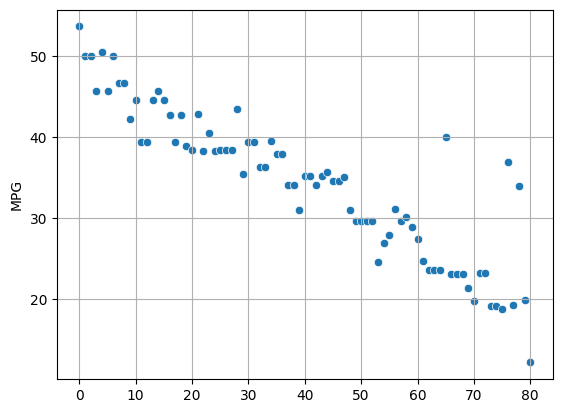
P(MPG<50)= 0.956

P(20<MPG<50)= 0.013000000000000012

Q 21) Check whether the data follows normal distribution

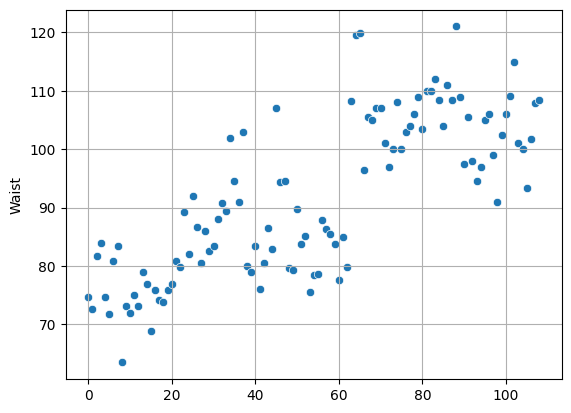
1. Check whether the MPG of Cars follows Normal Distribution

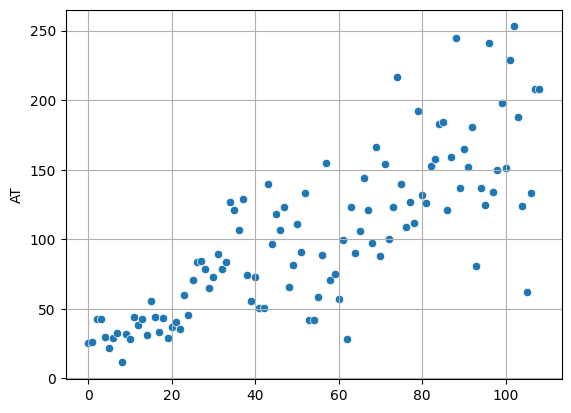
Dataset: Cars.csv



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

Dataset: wc-at.csv





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

**Sol:-**For 90% confidence interval:

We have the significance level at 5 % ( as it is a two tailed test)

that is:

α = 5 % = 0.05

z at α = 0.05 from the z table will be:

z = 1.645.

For 94 % confidence interval, we get:

We have the significance level at 3 % ( as it is a two tailed test)

that is:

α = 3 % = 0.03

z at α = 0.03 from the z table will be:

z = 1.555.

For 60 % confidence interval, we get:

We have the significance level at 20 % ( as it is a two tailed test)

that is:

α =20 % = 0.2

z at α = 0.2 from the z table will be:

z = 0.253

**Therefore, we get that the z score at 90 % confidence interval is 1.645, at 94 % confidence interval is 1.555 and at 60 % confidence interval is 0.253.**

Q23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

Sol:-

Confidence interval T scores

95% 2.063899

96% 2.171545

99% 2.79694

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

Sol:- 0.3216