|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Numerical-discrete |
| Results of rolling a dice | Numerical-discrete |
| Weight of a person | Numerical-continues |
| Weight of Gold | Numerical-continues |
| Distance between two places | Numerical-continues |
| Length of a leaf | Numerical-continues |
| Dog's weight | Numerical-continues |
| Blue Color | Numerical- discrete |
| Number of kids | Numerical- discrete |
| Number of tickets in Indian railways | Numerical- continues |
| Number of times married | Numerical- discrete |
| Gender (Male or Female) | Numerical- continues |

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

3 coins are tossed

2 h

1 t

Answer

2^3 =8

H H H

H H T

H T H

H T T

T H H

T T T

T H T

T T H

P(2H and 1T) is 3/8

P(H)+P(H)+P(T)=1/8+1/8+1/8=3/8

Q4) Two Dice are rolled, find the probability that sum is

-2 dice possible outcome =6^2=36

1. Equal to 1

Only if there is 0 from one side and here there is no 0 so the probability will be 0/36 =0

1. Less than or equal to 4

Less than or equal means ,sum =4,+ sum=3, + sum =2, + sum=1

P(Sum2Dice)=4 only if {(2,2),(1,3),(3,1)} 3/36.

P(Sum2Dice)=3 only if {(1,2),(2,1)}= 2/36.

P(Sum2Dice)=2 only if {(1,1)}= 1/36.

P(Sum2Dice)=1 =0

3/36+2/36+1/36=**6/36**

1. Sum is divisible by 2 and 3

P(Sum2Dice)/2 and 3 P(Sum2Dice %2 and 3) only if {(3,3)}=**1/36**

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?

2---red 3---green 2---blue total is = 7 and balls with no blue is 5

if we take first ball and consider is not blue so the probability is 5/7one ball already taken and consider is not blue so the probability is 4/6

The probability that none of the balls drawn is blue 5/7\*4/6=20/42🡪**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

The probability of expected value is the sum of no of candies for each child multi by the probability of that event

1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120= **3.090**

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment on the values / draw inferences, for the given dataset

* For Points, Score,Weigh>

|  |  |  |  |
| --- | --- | --- | --- |
| Required | 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 | 0.276948 | 0.927461 | 3.09338 |
| SD | 0.526258 | 0.963048 | 1.758801 |
| Range | 2.17 | 3.911 | 8.4 |

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

Q8) Calculate the 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?

The probability that even will happen is P=1/9

So the expected value is ++ + + + + + + =145.33

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

**Cars speed and d Cars speed**

**Answer:**

**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
|  | **Cars speed** | **Cars speed** |
| **Skewness** | -0.117509861 | 0.806895 |
| **Kurtosis** | -0.50899442 | 0.405053 |

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | **Sp** | **Weight** |
| **Skewness** | -0.117509861 | 0.806895 |
| **Kurtosis** | -0.50899442 | 0.405053 |

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

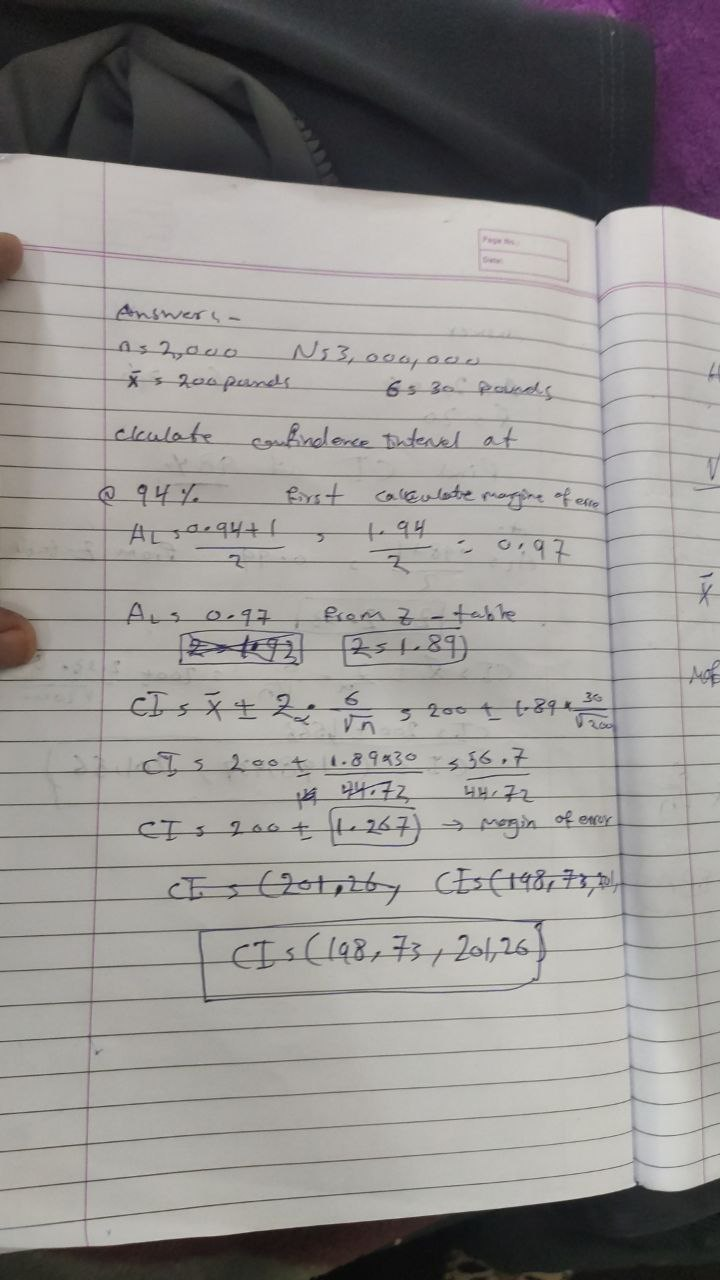


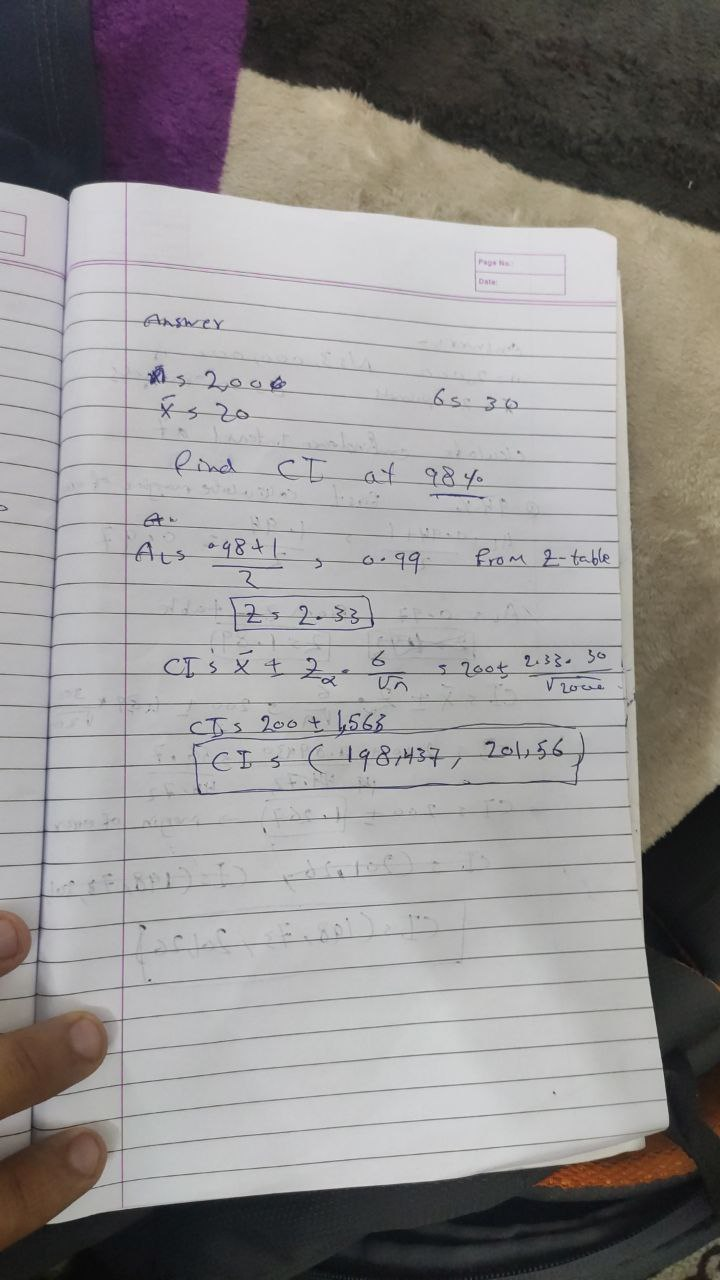
There are some spots in the histogram data where the check weight is on the x-axis and the frequency is on the y axis that indicates the weights are computed in frequency is dispersed, from 0 to 100 weights are occurring in between 50 and 100, and from 100 to 200. There are some weights on the x-axis the frequency the weights are disturbed x between 100 and 200, from 200 to 300 there are some weights in the frequency the data points fall between 0 and 100, from 300 to 400 there are some weights in the frequency the data points are fall between 0 and 100.

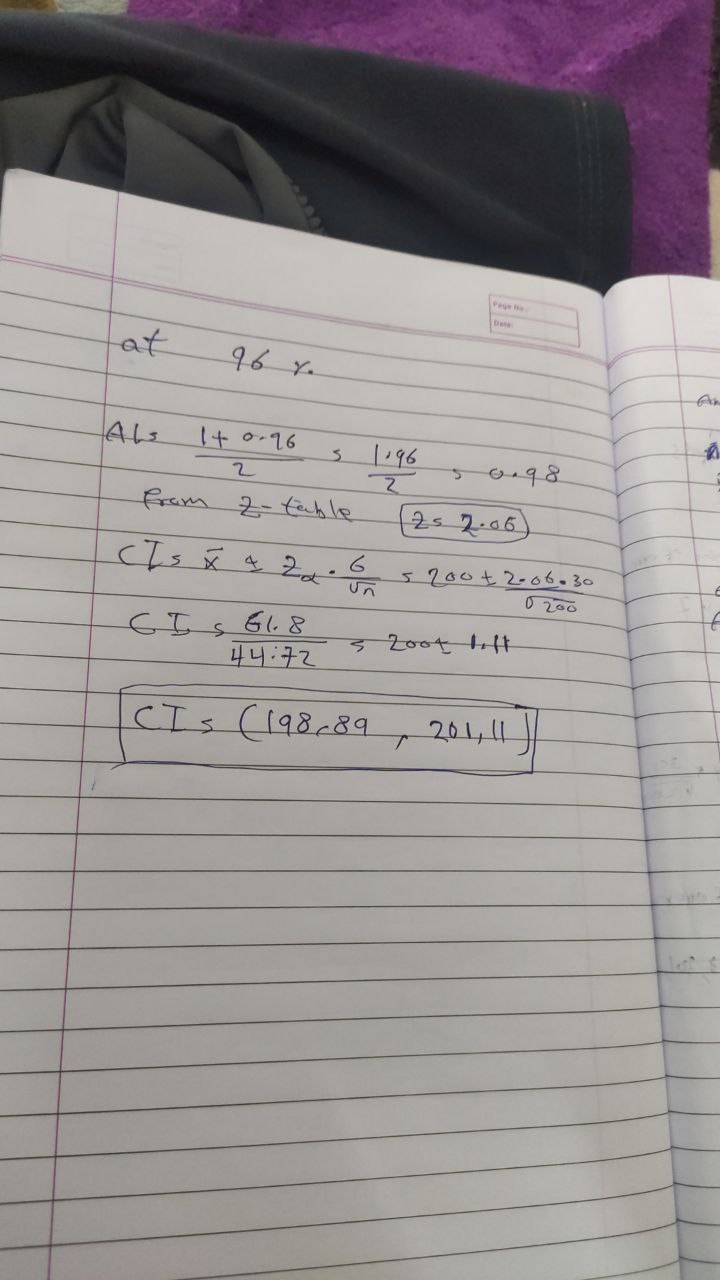


There are some probable outliers in the boxplot diagram, where the minimum point of the box plot is extremely close to quantile 1, the median of the box blot is far from quantile 3, and the maximum is far from quantile 3.

**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**

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

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

Median **34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

= =40.5

Mode = 41

Variance = 31.35294

Standard deviation = 1.3580

1. What can we say about the student marks?

We can sat they are normally distributed cause mean = median =mode

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

If the distribution is symmetric, then the mean is equal to the median, and the distribution has zero skewness. If the distribution is both symmetric and unimodal, then the mean = median = mode.

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

When the mean is greater than medina we get a positively skewed distribution we can call it right-skewed also

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

When the mean is lesser than medina we get a negatively skewed distribution we can call it left-skewed also

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

Positive kurtosis values suggest that the distribution is peaked with thick tails. An extreme positive kurtosis suggests that more numbers are distributed towards the tails of the distribution rather than around the mean.

Q17) What does negative kurtosis value indicates for data?

Negative kurtosis is interpreted as follows: If your distribution has a negative kurtosis, it is flatter than a normal curve with the same mean and standard deviation.

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



What can we say about the distribution of the data?

As per the given above boxplot

Q1 is 10,

median is 15

Q3 is 18

What is nature of skewness of the data?

Negatively skewed

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

As we can see from the above plot the box number 1 max and min value are closely to the Q1 and Q3 from the box1 while the Q2 are equal to the Q2 from the box 2, and there is no outliers noticed.

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)

there are 33 possible outcomes out of 82 entries which are greater than 38

* 1. P(MPG<40)

there are 61 possible outcomes, out of 82 entries which are less than 40

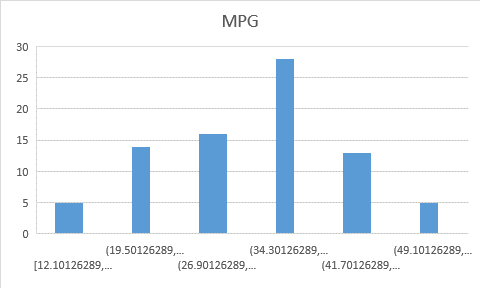
* 1. P (20<MPG<50)

there are 76 possible outcomes which lies between 20 and 50

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



Mean = 34.42208

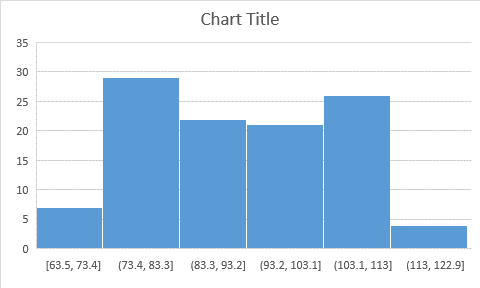
Median = 35.1523

Mode = 29.629

Hence median >mean>mode

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

Dataset: wc-at.csv



Mean = 91.90183

Median =90.08

Mode = 94.5

Hence mode>mean>median

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

Formula 1+confidence interval

------------------------------

2

At the confidence interval for 90 % is = 1+0.9

----------- = 0.95

2

At the confidence interval for 94% is = 1+0.94

----------- = 0.97

2

At the confidence interval for 94% is = 1+0.6

----------- = 0.8

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

Sample size = 25

Degree of freedom = n-1 = 24

At the 95% of confidence interval = 2.06390

At the 99% of confidence interval = 3.7454

At the 96% of confidence interval = 2.492

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

https://tex.z-dn.net/?f=t%3D%5Cdfrac%7Bx-%5Cmu%7D%7B%5Cfrac%7Bs%7D%7B%5Csqrt%20n%7D%7D

While t is t- distribution

X sample mean =260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

= - 0.471

t = - 0.471

Because the number of degrees of freedom in probability calculations is n - 1, the t-distribution with 17 degrees of freedom is required.

The probability that t - 0.471 with 17 degrees of freedom is less than the t-value derived if the population mean is true. With 17 degrees of freedom and a t score of -0.471, the probability of the bulbs lasting less than 260 days on average is 0.3218, assuming a 300-day mean life.