|  |  |
| --- | --- |
| 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 | 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 | Ratio |
| Years of Education | Interval |

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

When 3 coins are tossed,

The outcomes are given by sample space, S = {HHH, TTT, HTT, THT, TTH,

THH, HTH, HHT}

Number of elements in sample space, n(S) = 8

The probability of getting two heads and one tail= {HHT, HTH, THH}

So, the Probability of getting two head and one tail = 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

a) Equal to 1 is = zero

b) Less than or equal to 4 = (1,1),(1,2),(1,3),(2,1),(2,2),(3,1) /36

= 6/36 = 1/6

c) Sum divisible by 2 and 3 = 25/36

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?

Total number of balls = (2 + 3 + 2) = 7

Let S be the sample space.

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

= (7 x 6) / (2x 1) = 21

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

∴n(E) = Number of ways of drawing 2 balls out of (2red + 3green) balls = 5C2​

= (5×4) ​/ (2×1) = 10

∴P(E) = n(S) / n(E)​ = 21/10​

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 No of candies for a randomly selected child = 1/6

Expected Value = (1/6) (1 + 4 + 3 + 5 + 6 + 2)

= (1/6) (21)

= 21/6

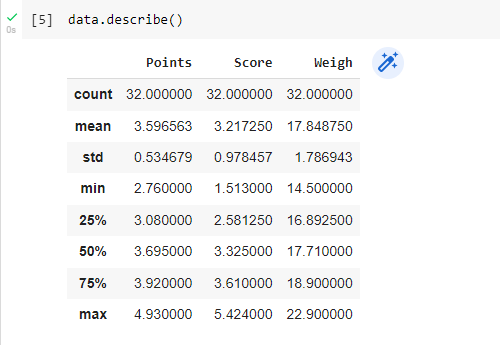
Expected No of candies for a randomly selected child = 3.5

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

* For Points,Score,Weigh>

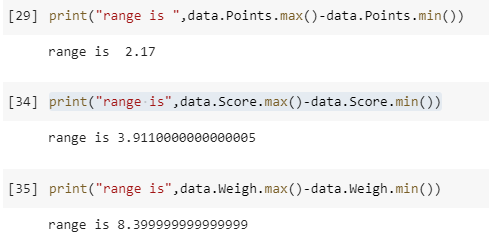
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

****

**Text

Description automatically generated with medium confidence**

****

**Inference:**

**- Points has the Least Variance, Standard Deviation, and Range**

**- Points &amp; Score data - Negative skewness, Weigh has Positive Skewness**

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?

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)

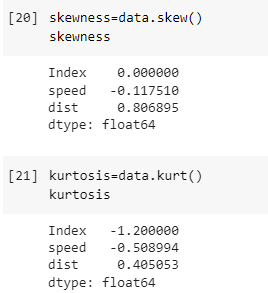
= 145.33

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**

****

**Inferences:**

**Skewness:**

**speed = - 0.117510 – Fairly Symmetrical**

**distance = 0.806895 - Moderately Skewed**

**Kurtosis:**

**speed = - 0.508994 - Platykurtic distribution**

**distance = 0.405053 - leptokurtic distribution**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**A screenshot of a computer

Description automatically generated with low confidence**

**Inferences:**

**Skewness:**

**SP = 1.611450 - Highly Skewed**

**WT = -0.614753 - Moderately Skewed**

**Kurtosis:**

**SP = 2.977329 - leptokurtic distribution**

**WT = 0.950291 - leptokurtic distribution**

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



**- Data is Positively Skewed**

**- Data is present on the Left side and has a long tail on the right side**

**- Mode of the data is 100**

**- Data Follows Unimodal and approximately normal distribution**



**- The above boxplot has a long tail on the upper quartile**

**- Most of the data points are in first quartile**

**- It has 7 outliers**

**- Median of the data lies close to the 1st quartile**

**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?

Population mean = 3000000

Sample mean = 2000

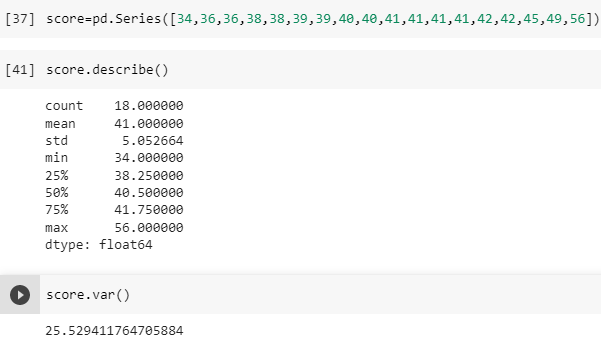
Sample = 200

Standard Deviation = 30

**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.



1. What can we say about the student marks?

**- The average score obtained by the students is 41 and the probability value of students scoring above 41 is about 57%**

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

**-Skewness is a measure of asymmetry and it may be zero or +ve or –ve. If the mean, median and mode of data are equal then there is no skewness.**

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

**-If the mean > median than the distribution is positively skewed.**

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

**-If the median > mean than the distribution is negatively skewed.**

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

**-Positive values of kurtosis indicate that a distribution is peaked and possess thick tails.**

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

**-A distribution with a negative kurtosis value indicates that the distribution has lighter tails than the normal distribution**

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



What can we say about the distribution of the data?

**-Here’s a box and whisker plot of the distribution from above**

**that does not show outliers.**

**-Here most of the data points lie in the right side.**

What is nature of skewness of the data?

**-It is negatively skewed because all the data points lie on the right side**

**and tail is present in the left side.**

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

**IQR = Q3 – Q1**

**IQR = 8 (approx.)**

Q19) Comment on the below Boxplot visualizations?



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

**- Both the box plots have the same median value lie at same point (262.5)**

**- Both Plots have equal distribution of data above and below Median which**

**means plots have Normal Distribution, to be specific data is symmetric**

**about mean value.**

**- Kurtosis is negative in 1st case as it has thinner tails and positive in the 2nd**

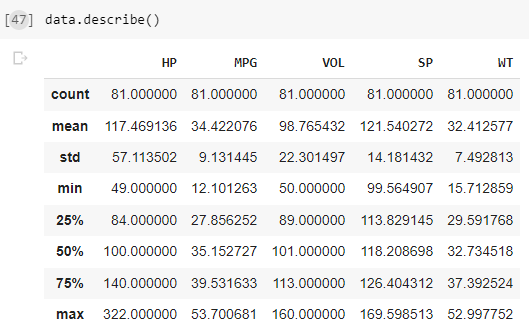
**box plot.**

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

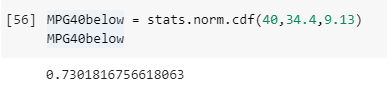


* 1. P(MPG>38) - 0.346

Graphical user interface, text, application

Description automatically generated

* 1. P(MPG<40) - 0.730



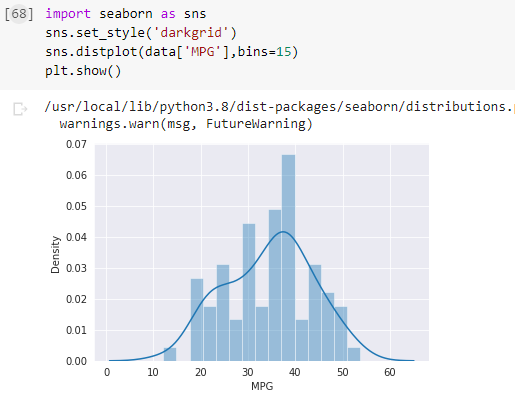
* 1. P (20<MPG<50) - 0.898

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

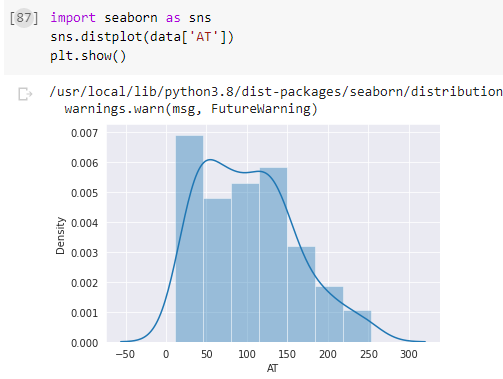
Dataset: Cars.csv

MPG of cars follows a Normal Distribution



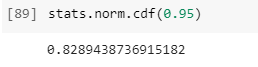
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv



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

1+0.90/2 = 0.95



1+0.94/2 = 0.97

Text

Description automatically generated

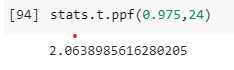
1+0.60/2 = 0.8

Graphical user interface, application, website

Description automatically generated

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

1+0.95/2 = 0.975 sample size = n-1 sample = 24



1+0.96/2 = 0.98

Graphical user interface, text

Description automatically generated

1+0.99/2 = 0.995

A picture containing text

Description automatically generated

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

Ans: The t statistic value for the above problem is

population mean=270

sample mean=260

sample std=90

sample=18

t= (sample mean-population mean)/ (sample std/ np.sqrt (sample))

t value is = -0.4714

8

stats.t.cdf (- 0.471, df=17)

0.3218140331685075

The probability that on an avg bulb lasts less than 260 days is 0.3218140331