Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| 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 |

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 | 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 | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ordinal |

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

1. The probability of two heads and one tail is 3/8 or 0.375.

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

Equal to 1

Less than or equal to 4

Sum is divisible by 2 and 3

aA) Zero.

bA) To get a number less than 4 on two dice is 6/36, i.e., 1/6.

cA) 1/6.

Q5) A bag 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?

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

7C2=21 contains 2 red, 3 green and

5C2=10

=10/21.

Q6) Calculate the Expected number of can1dies 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

A) Expected number of candies for a randomly selected child

= 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

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

|  |  |  |  |
| --- | --- | --- | --- |
| A) | Points | Score | Weigh |
| Mean | 3.596563 | 3.21725 | 17.84875 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.07 | 3.44 | 17.02 |
| Standard Deviation | 0.534679 | 0.978457 | 1.786943 |
| variance | 0.285881 | 0.957379 | 3.193166 |
| Range | 2.17 | 3.911 | 8.4 |

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?

A) 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 = ∑ (probability \* Value)

∑ P(x). E(x) = (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

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

Car’s speed and distance

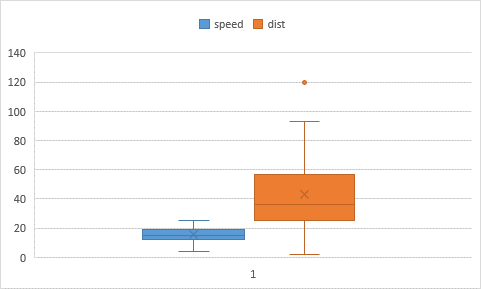
Use Q9\_a.csv

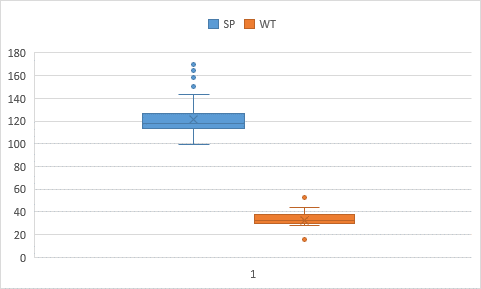
A) Skewness = -0.11751

Kurtosis = 0.405053

SP and Weight (WT)

Use Q9\_b.csv



A) Skewness = 1.61145 & Kurtosis = 0.950291

Q10) Draw inferences about the following boxplot & histogram

* The Histogram peak has right skew and tail on right. We have outliers on the higher side.



* 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?

A) 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).

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

i) Find mean, median, variance, standard deviation.

A) Mean - 41

Median – 40.5

Variance - 25.529411764705884

Standard Deviation - 5.05266382858645

ii) What can we say about the student marks?

1. 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?

1. Therefore, the distribution has approximately zero skew. In a distribution with zero skew, the mean and median are equal.

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

1. The mean of positively skewed data will be greater than the median.

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

1. The mean of negatively skewed data will be less than the median.

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

1. Positive values of kurtosis indicate that distribution is peaked and possesses thick tails.

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

1. 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.



i) What can we say about the distribution of the data?

A) The above Boxplot is not normally distributed the median is towards the higher value.

ii) What is nature of skewness of the data?

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

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

1. The Inter Quantile Range = Q3 Upper quartile – Q1 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.

1. There are no outliers. 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)

b) P(MPG<40)

c) P (20<MPG<50)

A) a) P(MPG>38) = 0.348

b) P(MPG<40) = 0.729

c) p(MPG>20) = 0.943

P(MPG<50) = 0.956

P(20<MPG<50) = 0.013

Q 21) Check whether the data follows normal distribution

i) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

A) MPG of cars follows normal distribution

A red line with blue dots

Description automatically generated

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

Dataset: wc-at.csv

1. Adipose Tissue (AT) and Waist does not follow Normal Distribution

A line with blue dots

Description automatically generatedA blue dotted line with a red line

Description automatically generated

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

A) z value for 90% confidence interval = -1.6449

z value for 94% confidence interval = -1.8808

z value for 60% confidence interval = -0.8416

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

A) t value for 95% confidence interval = -2.0639

t value for 94% confidence interval = -1.974

t value for 99% Confidence Interval = -2.7969

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 (t score, df)

df 🡪 degrees of freedom

1. t score = (x - pop mean) / (sample standard deviation / square root of sample size) = (260-270)/ (90/np.sqrt (18))

t score = -0.471

stats.t.cdf (t score, df = 17)

0.32 = 32%