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

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

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

P = (No.of desired outcomes)/(total no.of possible outcomes)

HHH, HHT, HTH, THH, TTH, THT, HTT, TTT

P =3/8

P=0.375

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)Probability that the sum is equal to 1:

There is no combination to get 1, It's impossible to get a sum of 1 since the minimum sum is 2 (if both dice show 1).

Probability = 0

(b) To get a sum less than or equal to 4, the combinations are : (1,1), (1,2), (2,1), (1,3), (3,1), (2,2).

There are 6 possible combinations.

Probability = Number of Favorable Outcomes / Total Possible Outcomes

= 6 / 36 = 1 / 6

(c) Probability that the sum is divisible by both 2 and 3:

For the sum to be divisible by 2 and 3, only number that satisfies this condition is 6.

The combinations resulting in a sum of 6 are: (1,5), (5,1), (2,4), (4,2), (3,3).

There are 5 possible combinations.

Probability = Number of Favorable Outcomes / Total Possible Outcomes

= 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?

Total no.of balls = 2+3+2 = 7

N(s) = no. of ways of taking 2 balls out of 7

Nc2 = 7C2 = 7\*6/2\*1 =42/2 = 21

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

N(E) = no of taking 2 balls out of 5

5C2 = 20/2 = 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 Value=∑(Candies count×Probability)

Expected candies=(1×0.015)+(4×0.20)+(3×0.65)+(5×0.005)+(6×0.01)+(2×0.120)

=0.015+0.80+1.95+0.025+0.06+0.24

=3.08

Therefore, the expected number of candies for a randomly selected child is 3.08.

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**

runfile('D:/exclr/Python/untitled0.py', wdir='D:/exclr/Python')

Mean of Points: 3.5965625

Mean of Score: 3.2172500000000004

Mean of Weigh: 17.848750000000003

Median of Points: 3.6950000000000003

Median of Score: 3.325

Median of Weigh: 17.71

Mode of Points: 0 3.07

1 3.92

Name: Points, dtype: float64

Mode of Score: 0 3.44

Name: Score, dtype: float64

Mode of Weigh: 0 17.02

1 18.90

Name: Weigh, dtype: float64

Variance of Points: 0.28588135080645166

Variance of Score: 0.9573789677419356

Variance of Weigh: 3.193166129032258

Standard Deviation of Points: 0.5346787360709716

Standard Deviation of Score: 0.9784574429896967

Standard Deviation of Weigh: 1.7869432360968431

Range of Points: 2.17

Range of Score: 3.9110000000000005

Range of Weigh: 8.399999999999999

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

Expected Value = ∑ (probability \* Value)

∑ P(x). E(x)

there are 9 patients

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

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Speed Skewness: -0.11395477012828319**

**Distance Skewness: 0.7824835173114966**

**Speed Kurtosis: -0.5771474239437371**

**Distance Kurtosis: 0.24801865717051808**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Skewness for SP: 1.5814536794423764**

**Skewness for WT: -0.6033099322115126**

**Kurtosis for SP: 2.7235214865269244**

**Kurtosis for WT: 0.8194658792266849**

**The SP data is positively skewed.**

**The WT data is negatively skewed.**

**The SP data is leptokurtic.**

**The WT data is leptokurtic.**

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



The distribution of chick weight is positively skewed. This means that there are more chicks with lower weights than there are chicks with higher weights. The tail of the distribution extends to the right, indicating that there are a few chicks with very high weights.

The mean and median chick weight is likely between 200 and 400 grams. The histogram is centered around this range, with the highest frequency of chicks in this weight range.

Most chicks have weights between 50 and 150 grams. This is the range of chick weights that is most represented in the histogram.

There are a few outliers at the upper end of the distribution. These are chicks with weights that are much higher than the average chick weight.

The majority of chicks (approximately 75%) have weights below 300 grams.

The 95th percentile chick weight is approximately 350 grams. This means that 95% of chicks have weights below 350 grams.

The 5th percentile chick weight is approximately 50 grams. This means that 95% of chicks have weights above 50 grams.



The median chick weight is approximately 250 grams. This is the value that divides the data set in half, with half of the chicks having weights below 250 grams and half of the chicks having weights above 250 grams.

The interquartile range (IQR) is approximately 100 grams. This means that the middle 50% of chicks have weights between 150 grams and 250 grams.

The whiskers extend from the lower quartile (Q1) to the upper quartile (Q3), plus 1.5 times the IQR. This means that the majority of chicks have weights within the range of 50 grams to 350 grams.

There are a few outliers. These are chicks with weights that are more than 1.5 times the IQR below Q1 or above Q3.

**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?

Confidence Interval=Sample Mean±(Critical Value X Sample Size/Sample Standard Deviation)

For a 94% confidence interval, the critical value is

Critical value: 1.96

Standard error of the mean: 30 / sqrt 2000 = 0.098

Margin of error: 1.96 \* 0.098 = 0.193

Confidence interval: (200 - 0.193, 200 + 0.193)

= (198.73, 201.27)

and for a 98% confidence interval,

Critical value: 2.326

Standard error of the mean: 30 / sqrt2000 = 0.098

Margin of error: 2.326 \* 0.098 = 0.229 pounds

Confidence interval: (200 - 0.229, 200 + 0.229)

= (198.43, 201.57)

for 96% confidence intervals:

Critical value: 2.054

Standard error of the mean: 30 / sqrt 2000 = 0.098

Margin of error: 2.054 \* 0.098 = 0.202 pounds

Confidence interval: (200 - 0.202, 200 + 0.202)

= (198.61, 201.39)

**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.
2. What can we say about the student marks?

Mean = (Sum of all values) / (Number of values) =

34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56 =738

738/18

=41

Median = Middle value after sorting in ascending order =(40+41)/2 = 40.5

Variance = ∑(X = mean)^2 / N

=(34-41)^2+(36-41)^2+......(56-41)^2 / 18

=25.529

2)Overall, the student's marks are relatively good, with a mean and median that are close to each other and a low variance. This tells that the student is performing consistently at a high level.

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

The distribution is approximately symmetrical. This means that the data is distributed evenly on both sides of the central tendency. The skewness of the distribution is approximately zero.

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

When the mean is greater than the median, the distribution is positively skewed. This means that the data is pulled out to the right, more than the left.

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

When the median is greater than the mean, the distribution is negatively skewed. This means that the data is pulled out to the left, than right.

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

A positive kurtosis value indicates that the distribution is more peaked than a normal distribution.This means more extreme values in the distribution, both on the high and low sides.

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

A negative kurtosis value indicates that the distribution is flatter than a normal distribution, .This means that there are fewer extreme values in the distribution, both on the high and low sides.

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



What can we say about the distribution of the data?

The distribution of the data is symmetrical, meaning that it is evenly distributed on both sides of the central tendency.

What is nature of skewness of the data?

The skewness of the data is zero,

What will be the IQR of the data (approximately)?   
  
The IQR of the data is approximately 8.

Q19) Comment on the below Boxplot visualizations?



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

Characteristic Boxplot 1 Boxplot 2

Median 275 225

Lower quartile 250 200

Upper quartile 300 250

Whisker range 200-325 175-275

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)
  3. P (20<MPG<50)

1. a. P(MPG > 38): 0.4074074
2. b. P(MPG < 40): 0.7530864
3. c. P(20 < MPG < 50): 0.8518519

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Shapiro-Wilk Test for Normality:

p-value: 0.1763985

The MPG data appears to be normally distributed

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

Confidence interval = sample mean +/- Z score \* standard deviation

= 68 inches +/- 1.645 \* 2 inches

Here are the Z scores for the given confidence intervals:

Confidence interval | Z scores

90% (-1.645, 1.645)

94% (-1.555, 1.555)

60% (-0.253, 0.253)

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

For a 95% confidence interval (two-tailed), the critical t value for a sample size of 25 and

df = 25-1

=24

For a 96% confidence interval (two-tailed), the critical t value for a sample size of 25 and

df=24 is approximately 2.171.

=2.171

For a 99% confidence interval (two-tailed), the critical t value for a sample size of 25 and

df=24 is approximately 2.797.

=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

0.3216725