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

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

Answer - the probability of getting two heads and one tail when tossing three coins is 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

Ans: Number of possible outcomes for the above event is N (Event (Two dice rolled)) = 6^2 = 36

a.) P (sum is Equal to 1) = ‘0’ zero null nada none.

b.) P (Sum is less than or equal to 4) = N (Event (Sum is less than or equal to 4)) / N (Event (Two dice rolled)) = 6 / 36 = 1/6 = 0.166 = 16.66%

c.) P (Sum is divisible by 2 and 3) = N (Event (Sum is divisible by 2 and 3)) / N(Event (Two dice rolled)) = 6 / 36 = 1/6 = 0.16 = 16.66%

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?

Answer –

the probability that none of the balls drawn is blue is 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

Answer –

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

Ans: 0.015+0.8+1.95+0.025+0.06+0.24 = 3.09

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.

Answer-

Mean:

Mean of Points: 3.307

Mean of Score: 3.596

Mean of Weigh: 17.848

Median:

Median of Points: 3.695

Median of Score: 3.44

Median of Weigh: 17.71

Mode:

There is no clear mode for these datasets as no value repeats significantly.

Variance:

Variance of Points: 0.2859

Variance of Score: 0.957

Variance of Weigh: 3.193

Standard Deviation:

Standard Deviation of Points: 0.535

Standard Deviation of Score: 0.978

Standard Deviation of Weigh: 1.786

Range:

Range of Points: 2.17 (Max - Min)

Range of Score: 3.809 (Max - Min)

Range of Weigh: 7.46 (Max - Min)

**Use Q7.csv file**

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?

Answer

the expected weight of a randomly chosen patient is approximately 153.67 pounds.

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**ANSWER –**

**Use Q9\_a.csv Ans: q9a = pd.read\_csv("C:/Users/Moin Dalvi/Documents/EXcelR Study and Assignment Material/Data Science Assignments/Basic Statistics 1/Q9\_a.csv", index\_col = 'Index')**

**print('For Cars Speed', "Skewness value=", np.round(q9a.speed.skew(),2), 'and' , 'Kurtosis value=', np.round(q9a.dist.skew(),2)) For Cars Speed Skewness value= -0.12 and Kurtosis value= 0.81**

**print('Skewness value =', np.round(q9a.dist.skew(),2),'and', 'Kurtosis value =', np.round(q9a.dist.kurt(),2), 'for Cars Distance') Skewness value = 0.81 and Kurtosis value = 0.41 for Cars Distance**

**SP and Weight (WT) Use Q9\_b.csv Ans: q9b =pd.read\_csv("C:/Users/Moin Dalvi/Documents/EXcelR Study and Assignment Material/Data Science Assignments/Basic Statistics 1/Q9\_b.csv")**

**print('For SP Skewness =', np.round(q9b.SP.skew(),2), 'kurtosis =', np.round(q9b.WT.kurt(),2)) For SP Skewness = 1.61 kurtosis = 0.95**

**print('For WT Skewness =', np.round(q9b.SP.skew(),2), 'Kurtosis =', np.round(q9b.WT.kurt(),2)) For WT Skewness = 1.61 Kurtosis = 0.95**

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



**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?

The 94% confidence interval is approximately (198.74, 201.26) pounds.

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

The 96% confidence interval is approximately (198.62, 201.38) pounds.

**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?

Answer – mean = 40.78

Median= 41

The variance is approximately 49.80,

the standard deviation is approximately 7.07

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

Symmetrical

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

Positively skewed

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

Negatively skewed

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

Positive kurtosis value indicates leptokurtic distribution, indicating heavier tails and a more peaked central region compared to a normal distribution.

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

Negative kurtosis value indicates platykurtic distribution, indicating lighter tails and a flatter central region compared to a normal distribution.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

Q19) Comment on the below Boxplot visualizations?



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

ANS-

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)
  2. P(MPG<40)
  3. P (20<MPG<50)

ANSWER

Data \_set: Cars.csv Calculate the probability of MPG of Cars for the below cases. MPG <- Cars $ MPG a. P(MPG>38) Ans: Prob\_MPG\_greater\_than\_38 = np.round(1 - stats.norm.cdf(38, loc= q20.MPG.mean(), scale= q20.MPG.std()),3) print('P(MPG>38)=',Prob\_MPG\_greater\_than\_38)

P(MPG>38)= 0.348

b. P(MPG<40) Ans: prob\_MPG\_less\_than\_40 = np.round(stats.norm.cdf(40, loc = q20.MPG.mean(), scale = q20.MPG.std()),3) print('P(MPG<40)=',prob\_MPG\_less\_than\_40)

P(MPG<40)= 0.729

c. P (20<MPG<50) Ans: prob\_MPG\_greater\_than\_20 = np.round(1-stats.norm.cdf(20, loc = q20.MPG.mean(), scale = q20.MPG.std()),3) print('p(MPG>20)=',(prob\_MPG\_greater\_than\_20)) p(MPG>20)= 0.943

prob\_MPG\_less\_than\_50 = np.round(stats.norm.cdf(50, loc = q20.MPG.mean(), scale = q20.MPG.std()),3) print('P(MPG<50)=',(prob\_MPG\_less\_than\_50)) P(MPG<50)= 0.956

prob\_MPG\_greaterthan20\_and\_lessthan50= (prob\_MPG\_less\_than\_50) - (prob\_MPG\_greater\_than\_20) print('P(20<MPG<50)=',(prob\_MPG\_greaterthan20\_and\_lessthan50)) 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

ANSWER-

: a.) MPG of cars follows normal distribution

B): Adipose Tissue (AT) and Waist does not follow Normal Distribution

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

ANSWER-

**Z value for 90% confidence interval** –

print('Z score for 60% Conifidence Intervla =',np.round(stats.norm.ppf(.05),4)) Z score for 60% Conifidence Intervla = -1.6449

**z value for 94% confidence interval**

print('Z score for 60% Conifidence Intervla =',np.round(stats.norm.ppf(.03),4)) Z score for 60% Conifidence Intervla = -1.8808

**z value for 60% confidence interval**

print('Z score for 60% Conifidence Intervla =',np.round(stats.norm.ppf(.2),4)) Z score for 60% Conifidence Intervla = -0.8416

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

**T value for 95% confidence interval**

**print('T score for 95% Confidence Interval =',np.round(stats.t.ppf(0.025,df=24),4)) T score for 95% Confidence Interval = -2.0639**

**T value for 96% confidence interval**

**print('T score for 94% Confidence Inteval =',np.round(stats.t.ppf(0.03,df=24),4)) T score for 94% Confidence Inteval = -1.974**

**T value for 99% confidence interval**

**print('T score for 95% Confidence Interval =',np.round(stats.t.ppf(0.005,df=24),4)) T score for 95% 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(tscore,df)

df 🡪 degrees of freedom

**The probability that 18 randomly selected bulbs would have an average life of no more than 260 days, assuming the CEO's claim of 270 days, is approximately 0.0038, or about 0.38%.**