|  |  |
| --- | --- |
| 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 | 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 | Interval |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| 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?

Ans: possible outcomes{HHH,HHT,HTT,TTT,THT,THH,TTH,HTH}

we want 2heads and one tail so; {HHT,THH,HTH}-favourable outcomes are 3

so, favorable outcomes/total outcomes: i.e 3/8

therefore 3/8 is the probability. i.e=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

Ans: when 2 dices are rolled possible outcomes are 36

1. Equal to 1; There is no such outcome, as minimum sum of 2 dices is 2(1,1)

Therefor

/.

e probability of getting the sum =1 is **‘0’.**

1. Less than or equal to 4; there are only 6 outcomes whose sum is equal to 4 or less than 4:{1,1; 1,2; 1,3; 2,1; 2,2; 3,1}

So by formula, it is 6/36 i.e **1/6.**

1. Sum is divisible by 2 and 3; there are only 6 outcomes whose sum is divisible by both 2 and 3 i.e {1,5; 2,4; 3,3; 4,2; 5,1; 6,6}

So by formula it is 6/36 i.e **1/6.` `**

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?

Ans: Total there are 7 balls(2+3+2) in the bag: n!/r!(n-r)! 7C2 = 7\*6/2= **21.**

But we don’t want blue balls so 7-2=5 i.e5C2= 5\*4/2=**10**

So,by formula= 10/21,so the probability that the ball is not blue is **10/21=0.47**

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

Ans: Expected no of candies for randomly selected will be given by:

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

i.e: 0.015+0.8+1.95+0.025+0.06+0.24=3.090

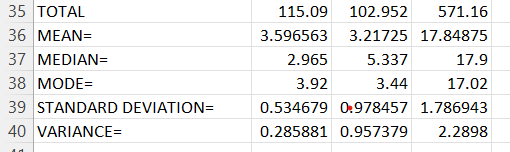
= **Expected number of candies for a randomly selected child  = 3.09**

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.

ANS:



|  |  |  |  |
| --- | --- | --- | --- |
| TOTAL | 115.09 | 102.952 | 571.16 |
| MEAN= | 3.596563 | 3.21725 | 17.84875 |
| MEDIAN= | 2.965 | 5.337 | 17.9 |
| MODE= | 3.92 | 3.44 | 17.02 |
| STANDARD DEVIATION= | 0.534679 | 0.978457 | 1.786943 |
| VARIANCE= | 0.285881 | 0.957379 | 2.2898 |
| MIN= | 2.76 | 1.513 | 14.5 |
| MAX= | 4.93 | 5.424 | 22.9 |
| RANGE= | 2.17 | 3.911 | 8.4 |

**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?

Ans: Expected value= Summation(prob\*value)

There are 9 patients, so probability of selecting one patient=1/9

So,

(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 is 145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**ANS:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | SPEED | DISTANCE | |
| SKEWNESS= | -0.11751 | 0.806895 |  |
| KURTOSIS= | -0.50899 | 0.405053 |  |

**SP and Weight(WT)**

**Use Q9\_b.csv**

**ANS:**

|  |  |  |
| --- | --- | --- |
|  | SP | WEIGHT |
|  |  |  |
| SKEWNESS | 1.61145 | -0.61475 |
| KURTOSIS | 2.977329 | 0.950291 |

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



ANS: The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.



ANS: 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?

stats**.**norm**.**interval(0.94,200,30**/**(2000**\*\***0.5))=198.738,201.2626

stats**.**norm**.**interval(0.98,200,30**/**(2000**\*\***0.5))=198.4394,201.5605

stats**.**norm**.**interval(0.96,200,30**/**(2000**\*\***0.5))=198.6223,201.3776

**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?

ANS:MOST OF THE MARKS OCCUR BETWEEN 36-42

|  |  |
| --- | --- |
| TOTAL | 738 |
| MEAN | 41 |
| MEDIAN | 40.5 |
| MODE | 41 |
| VARIANCE | 25.52941 |
| STD DEV | 5.052664 |

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

ANS: **ZERO SKEWNESS**

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

ANS**: POSITIVELY SKEWED**

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

**ANS: NEGATIVELY SKEWED**

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

ANS: **IT MEANS ITS DISTRIBUTION IS PEAK AND THICK TAIL**

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

ANS: **WIDER PEAK AND THINNER TAIL**

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



What can we say about the distribution of the data?

ANS: MOST OF THE DATA IS ON THE LEFT SIDE OF THE BOX. MEDIAN IS ABOUT 15.

What is nature of skewness of the data?

ANS: **POSITIVELY SKEWED**

What will be the IQR of the data (approximately)?   
  
ANS: **18-10=8(UQ-LQ)**

Q19) Comment on the below Boxplot visualizations?



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

ANS: here are no outliers and 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)

Ans: Probability of MPG>38 is 0.4074

* 1. P(MPG<40)

Ans: Probability of MPG<40 is 0.7530

* 1. P (20<MPG<50)

Ans: 0.89899

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: Yes it is 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

Ans: For Waist- yes, it is normally distributed

For AT: It is right skewed and does not follow normal distribution

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

94% confidence interval, 60% confidence interval

**Ans**: for 90%= 1.6448

For 94%=1.8807

For 60%=0.8416

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

Ans: T-score for sample size of 25-95% 2.0638

96%- 2.1715

99%-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

t**=**(260**-**270)**/**(90**/**18**\*\***0.5)=-0.4714

p\_value**=**1**-**stats**.**t**.**cdf(abs(**-**0.4714),df**=**17)=0.32167