|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | **COUNT** |
| Results of rolling a dice | **COUNT** |
| 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 | **CATEGORICAL** |
| Number of kids | **COUNT** |
| Number of tickets in Indian railways | **COUNT** |
| Number of times married | **COUNT** |
| Gender (Male or Female) | **CATEGORICAL** |

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 | **ORDINAL** |
| Type of living accommodation | **ORDINAL** |
| Level of Agreement | **INTERVAL** |
| IQ(Intelligence Scale) | **RATIO** |
| Sales Figures | **ORDINAL** |
| Blood Group | **NOMINAL** |
| Time Of Day | **NOMINAL** |
| Time on a Clock with Hands | **NOMINAL** |
| Number of Children | **ORDINAL** |
| Religious Preference | **NOMINAL** |
| Barometer Pressure | **RATIO** |
| SAT Scores | **RATIO** |
| Years of Education | **ORDINAL** |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained? **Ans: 3/8**

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

1. Equal to 1 **Ans:0**
2. Less than or equal to 4 **Ans:1/6**
3. Sum is divisible by 2 and 3 **Ans: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 : 5C2 /7C2 =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

**Ans: E(x) =∑ x.p(x) = (1\*0.015) +(4\*.2) +(3\*0.65) +(5\*0.005) + (6\*0.01)+(2\*0.12)**

**=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: Mean Median Mode variance std.dev Range**

**Points 3.5965 3.695 3.07 &3.92 0.2858 0.5347 2.17**

**Score 3.2172 3.325 3.44 0.9574 0.9784 3.911**

**Weigh 17.8487 17.71 17.02&18.9 3.1932 1.7869 8.3999**

* **As all three have values of mean, median and mode nearly equal values we can say they are almost normally distributed.**
* **Since means of points and Score are nearly equal but as standard deviation of Score is high the data points of score are more dispersed compared to points data.**

**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: (108\*(1/8)) + (110\*(1/8) + (123\*(1/8)) + (134\*(1/8)) + (135\*(1/8))+ (145\*(1/8)) + (167\*(1/8)) + (187\*(1/8)) + (199\*(1/8)) = 163.5 pounds**

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

**Cars speed and distance**

**Use Q9\_a.csv**

For speed : skewness = -0.1175 kurtosis = -0.5089

For distance : skewness = 0.8069 kurtosis = 0.405

So from skewness values we can conclude distance data is right skewed and speed data is left skewed.

From kurtosis values we can conclude probability distribution curve for distance has sharper peak and speed has comparatively wider peak.

**SP and Weight(WT)**

**Use Q9\_b.csv**

For SP : skewness = 1.6114 kurtosis = 2.9773

For WT: skewness = -0.6147 kurtosis = 0.9503

So data for SP is right skewed and for WT it is left skewed.

Probability distribution curve for SP has a sharper peak compared to that of WT.

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



* **The data set given for chick weight is right skewed as can be seen from histogram and also from the longer right side whisker of box plot .**
* **It also has outlier which are present above the upper fence.**
* **From histogram a rough estimate of mean for data set can be taken as 75.**

**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?

**ANS: For 94% confidence interval: [198.738, 201.262]**

**For 98% confidence interval: [198.438, 201.562]**

**For 96% confidence interval:[198.621, 201.379]**

**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: Mean = 41 , median = 40.5 , variance =25.529,**

**standard deviation=5.053**

* **Since the median is 40.5 and minimum value is 34 and maximum value is 56, we see data is more dispersed towards right thus making it right skewed.**

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

**No skewness**

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

**Right Skewed**

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

**Left skewed**

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

**Sharper peak of probability distribution curve**

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

**Wider peak of probability distribution curve**

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



What can we say about the distribution of the data? **Left skewed data**

What is nature of skewness of the data? **Negatively skewed**

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

**Range of Dataset 2 is greater than dataset 1.**

**Median of both the data sets are same.**

**Both the datasets are normally distributed.**

**Dataset 1 is less dispersed than dataset2 and thus has lower variance.**

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:33/81**
  2. P(MPG<40) **Ans:61/81**

c. P (20<MPG<50) **Ans:69/81**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Mean =34.42 median = 35.15 skewness = -0.178 kurtosis = -0.612**

**As can be see from above values it is not normally distributed but since the deviations are very less it can be considered as normally distributed for statistical analysis.**

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

Dataset: wc-at.csv

**AT: Mean= 101.89, Median = 96.54 (Right Skewed)**

**Waist: Mean = 91.9 , Median = 90.8 ( mean very close to median so normally distributed)**

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

**90% - 1.645**

**94% - 1.881**

**60% - 0.842**

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

**95% - 2.064**

**96% - 2.171**

**99% - 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

**Ans: n=18 sample mean =260 sample std dev(s) =90**

**External standard = 260**

**Ho : Average life < 260 days**

**H1: Average life > 260 days**

**t = (260-270)/(90/(18)^0.5) = -0.471**

**p=0.3216**

**Thus probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 0.3216.**