|  |  |
| --- | --- |
| 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 | Nominal |
| 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 | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

**Ans:** Total outcomes = 23 = 8

Probability = Total number of outcomes **(HHT, HTH, and THH) /** Number of favorable outcomes

Probability = 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:**  a) Probability is 0.

b) Combinations of dice rolls that result in a sum less than or equal to 4: (1,1), (1,2), (2,1), (1,3), (3,1), (2,2).

Total possible outcomes when two dice are rolled: 6×6=36

Probability = 6/36 = 1/6

c) The sums divisible by both 2 and 3 are 6 and 12. The combinations for sum 6 are (1,5), (2,4), (3,3), (4,2), (5,1), and for sum 12 is (6,6).

Total possible outcomes when two dice are rolled: 6×6=36

Probability = 6/36 = 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 number of balls = 2+3+2=7

Total number of non-blue balls is 2+3=5

P (non-blue on first draw) = 5/7​  
P (Non-blue on second draw) = 4/6​

P (Both non-blue) = 5/7​×4/6​ = 20/42​ = **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:** The formula for the expected value of a discrete random variable is:

E[X]=∑i ​xi​⋅pi​

Given the probabilities and candy counts for each child:

E[X]= (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

**=3.075​**

**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**

**Ans:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.596563 | 3.21725 | 17.84875 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 17.02 |
| Variance | 0.285881 | 0.957379 | 3.193166 |
| Standard Deviation | 0.534679 | 0.978457 | 1.786943 |
| Range | 2.17 | 3.911 | 8.4 |

The points dataset is relatively symmetrically distributed with moderate variability around the mean.

The score dataset shows moderate variability around the mean with no dominant value.

The weight dataset shows moderate variability around the mean with no dominant value, similar to the score dataset

**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:** Formula:

E[X] = (∑ni=1 ​xi​)/ n

E[X]= (108+110+123+134+135+145+167+187+199​) / 9

= 1308/9 = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Speed**

Skewness = E[(x-μ/σ)]3 and Kurtosis=E [(x-μ/σ)]4-3

Where μ (speed) =15.4 And σ (speed) = 5.234501

Skewness (speed) = -0.117510

Kurtosis (speed) = -0.508994

**Distance**

Again μ (dist) =42.98 And σ (dist) = 25.51038

Skewness (dist) = 0.806895

Kurtosis (dist) = 0.405053

**SP and Weight (WT)**

**Use Q9\_b.csv**

**SP**

Skewness = E[(x-μ/σ)]3 and Kurtosis=E [(x-μ/σ)]4-3

Where μ (SP) = 121.5402722 And σ (SP) = 14.09362

Skewness (SP) = 1.61145

Kurtosis (SP) = 2.977329

**WT**

Skewness = E[(x-μ/σ)]3 and Kurtosis=E [(x-μ/σ)]4-3

Where μ (WT) = 32.41257691 And σ (WT) = 7.446417

Skewness (WT) = -0.61475

Kurtosis (WT) = 0.950291

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





**Ans:** According to Histogram the data is concentrated in left side so the histogram is positively skewed and mean is in between 50 to 100.

According to box plot the data is concentrated downwards and Q2 - Q1 < Q3 - Q2 which shows positive skewness and has some outliers at which are greater than Q4.

**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:** Here 2000 is greater than 30 so the sample is normal and it is a simple random sample.

Sample statistic is 200.

Confidence level is 94%, 98%, 96%.

Standard error is 30/√2000 = 0.67082

Alpha (α) =100-confidence level=6%, 2% & 4%=0.06, 0.02 & 0.04 respectively

Critical Probability = 1 – α / 2 = 0.97, 0.99 & 0.98 respectively

Degree of freedom = n - 1 = 2000 - 1 = 1999

From t distribution critical values are = 1.882, 2.328 & 2.055 respectively Margin of error (94% confidence level) = critical value (94% confidence level) \* standard error = 1.882 \* 0.67082 = 1.262

Margin of error (98% confidence level) = critical value (98% confidence level) \* standard error = 2.328 \* 0.67082 = 1.561

Margin of error (96% confidence level) = critical value (96% confidence level) \* standard error = 2.055 \* 0.67082 = 1.378

**Confidence interval = Sample statistics ± Margin of error**

Confidence interval (94% confidence level) = 200 ± 1.262 =198.738 to 201.262

Confidence interval (98% confidence level) = 200 ± 1.561 =198.439 to 201.561

Confidence interval (96% confidence level) = 200 ± 1.378 = 198.622 to 201.378

**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: 1)** Mean (μ)= (34 + 36 + 36 + 38 + 38 + 39 + 39 + 40 + 40 + 41 + 41 + 41 + 41 + 42 + 42 + 45 + 49 + 56) / 18 = 738 / 18 = 41

Median= (40 + 41) / 2 = 40.5

Variance (σ2) = [(34-41) 2+ (36-41) 2+ (36-41) 2+ (38-41) 2+ (38-41) 2+ (39- 41) 2+ (39-41) 2+ (40-41) 2+ (40-41) 2+ (41-41) 2+ (41-41) 2+ (41-41) 2+ (41-41) 2+ (42-41) 2+ (45-41) 2+ (45-41) 2+ (49-41) 2+ (56-41) 2] / 18 = 434/18 = 24.111

Standard Deviation (σ)= (24.111)1/2= 4.910

**Ans: 2)** According to marks we can say that students who scored in between 34- 39 are 7 in count, students who scored in between 40-45 are 9 in count, student who scored in between 46-51 is 1 in count and student who scored in between 52-57 is 1 in count. Hence if we calculate skewness and kurtosis that will be +1.686 and +3.953 respectively and are positively skewed and kurtosis is also positive.

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

**Ans:** Symmetrical

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

**Ans**: Right Skewed

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

**Ans:** Left Skewed

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

**Ans:** Positive kurtosis indicates heavier tails and a more peaked distribution

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

**Ans:** The distribution of the data has lighter tails and a flatter peak than the normal distribution.

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



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

**Ans:**  Let’s assume above box plot is about ages of the students in a school.

50% of the people are above 10 yrs. old and remaining are less.

And students whose age is above 15 are approx. 40%.

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

**Ans:** Left skewed, median is greater than mean.

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

**Ans:** -8

Q19) Comment on the below Boxplot visualizations?



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

**Ans:** By observing both the plots whisker’s level is high in boxplot 2, mean and

median is equal hence distribution is symmetrical

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)

c. P (20<MPG<50)

**Ans:** a. P(MPG>38): 1-pnorm (38,34.422,9.13144) =0.3475908

b. P(MPG<40): pnorm (40,34.422,9.13144) =0.7293527

c. P (20<MPG<50): pnorm (50,34.422,9.13144)-(1-pnorm(20,34.422,9.13144) = 0.01311818

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 follows normal distribution

Because Mean = Median (very close) and skewness is close to 0.

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

Dataset: wc-at.csv

**Ans:** Waist and AT does not follow normal distribution.

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

**Ans:** Z score of 90% confidence interval is 1.65

Z score of 94% confidence interval is 1.55

Z score of 60% confidence interval is 0.85

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

**Ans:** For 95%= 1.96

For 96%= 2.5

For 99% = 2.47

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:** The probability that 18 randomly selected bulbs would have an average life of no more than 260 days can be calculated using the z-score formula and the central limit theorem.  
  
Firstly, calculating the standard deviation of the sample mean, which is the standard deviation divided by the square root of the sample size (90/√18 = 21.2).  
  
Next, calculating the z-score, which is the difference between the sample mean and the population mean divided by the standard deviation of the sample mean ((260-270)/21.2 = -0.47).  
  
Finally, the z-score in the z-score table to find the corresponding probability. A z-score of -0.47 corresponds to a probability of 0.3192 or 31.92%.  
  
Therefore, there is a 31.92% probability that 18 randomly selected bulbs would have an average life of no more than 260 days if the CEO's claim were true.