|  |  |
| --- | --- |
| 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 | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| 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 | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| 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 | Interval |
| SAT Scores | Interval |
| Years of Education | Interval |

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

Number of total event=2^3=8.

Number of event {2h,1t}={h,h,t},{t,h,h},{h,t,h}

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

a)p(x=1)=0

b)Total number of possibilities are=36

required possibilities=(p(x=2)=1/36+p(x=3)=2/36+p(x=4)=3/36)=6/36=1/6

c)only one number is divisible by 2 and 3=6,12

number of occurs of p(x=6) p(x=12)=5,1

probability of getting sum divisible by 2 and 3 is =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?

P(x=!2)=5/7\*4/6=20/42

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 number of candies for randomly selected child =1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120=3

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.



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?

Sol)Expected value of weight of that patient=145

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

**Cars speed and distance**

|  |
| --- |
|  |
|  |
|  |



**Sol:cardistance**

1. Skewness:1.290763 positively skewed rightly skewed and the data points more on the left side there some outliers on the right side.
2. Kurtosis:2.464546 positive kurtosis as skinner tail and broader peak because is close to value 3 which is normal distribution value

Speed:

1)skewness:-0.89542 it is left skewed and tail is at right side

2)kurtosis:0.249561positive kurtosis as skinner tail and broader peak

|  |
| --- |
|  |

**SP and Weight(WT)**



skewness(SP)

-0.407692

kurtosis(SP)

2.086731

Negatively skewed, platykurtic

skewness(WT)

-1.287359

kurtosis(WT)

3.818812

Negatively skewed, nearly mesokurtic

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



Sol: Histogram:

\*Right skewed

\*mode is from 50-100

Boxplot:

\*outliers are there in this data set

\*data right skewed

**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 ?

Sample mean=200, sample sd=30, n=2000, N=3,000,000

SE (mean)=30/sqrt(2000)= 0.6708

94% CI=200±1.88\*0.6708=(198.73,201.26)

96% CI=200±2.05\*0.6708=(198.62,201.38)

98% CI=200±2.32\*0.6708=(198.44,201.55)

**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.

Sol:Mean:41

Median:40.5

Variance:25.52

Standard deviation:5.052

1. What can we say about the student marks?

Sol:The average of the students is 41 and variability of the marks in between the students is low the students are performing in the range of 34-56=22

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

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

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

Q16) What does positive kurtosis value indicates for a data ?More values are at center and skinner tails and broader tails

Q17) What does negative kurtosis value indicates for a data?More values are at tails with broader tails and thinner peaks

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



What can we say about the distribution of the data?

Not symmetric distribution it is left skewed distribustion

What is nature of skewness of the data?negative skewness and left 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.

Boxplot1 has shorter range compared to boxplot 2

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)
  4. mean(Cars$MPG)
  5. [1] 34.42208
  6. &gt; sd(Cars$MPG)
  7. [1] 9.131445
  8. 1-pnorm(0.3921)
  9. [1] 0.3474922
  10. pnorm(0.6112)
  11. [1] 0.7294664
  12. x&lt;-pnorm(20, mean = 34.42, sd = 9.13, lower.tail = TRUE, log.p = FALSE)
  13. &gt; y&lt;-pnorm(50, mean = 34.42, sd = 9.13, lower.tail = TRUE, log.p = FALSE)
  14. &gt; y-x
  15. [1] 0.8989178

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

shapiro.test (Cars$MPG)

Shapiro-Wilk normality test

data: Cars$MPG

W = 0.97797, p-value = 0.1764

MPG follows Normal Distribution.

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

Dataset: wc-at.csv

shapiro.test (wt$Waist)

Shapiro-Wilk normality test

data: wt$Waist

W = 0.95586, p-value = 0.00117

&gt; shapiro.test (wt$AT)

Shapiro-Wilk normality test

data: wt$AT

W = 0.95234, p-value = 0.000654

Both AT and Waist do not follow normal distribution. P value less than 0.05.

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

1.64, 1.88, 0.84

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

#For getting Student&#39;s t value

&gt; qt(0.975,24)

[1] 2.063899

&gt; qt(0.98,24)

[1] 2.171545

&gt; qt(0.995,24)

[1] 2.79694

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

Pop mean 270, n=18, xbar=260, sd=90,

tscore= (260-270)/90/sqrt(18)

tscore&lt;-(260-270)/(90/sqrt(18))

&gt; df&lt;-17

&gt; pt(tscore,df)

[1] 0.3216725