|  |  |
| --- | --- |
| 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 | Ordinal |
| 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 | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

ANS:- 1) When three coins are tossed the total number of possible

Combinations are 2^3=8

Probability of getting two head p(H)=2

Probability of getting one tail P(T)=1

the probability that two heads and one tail =P(H)+P(T)/8

=2+1/8

***=0.375***

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

1. Equal to 1
2. than Less or equal to 4
3. Sum is divisible by 2 and 3

**ANS:-** If two dice are rolled then n(s)=36

**Probability:**- a) Equal to 1=p(1)/n(s)

=0/36=***0***

b) than Less or equal to 4

probability= [(1,1),(1,2)(2,2)(3,1)(1,3)]

**probability=5/36=0.1388888**

1. Sum is divisible by 2 and 3:-

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?

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 number of candies for a randomly selected child

=  1 \* 0.015  + 4\*0.20  + 3 \*0.65  + 5\*0.005  + 6 \*0.01  + 2 \* 0.12

= 0.015 + 0.8  + 1.95 + 0.025 + 0.06 + 0.24

= 3.09

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

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range &

For Points,Score,Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

***ANS:-***

**TABLE:-1**

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameters** | MEAN | MODE | MEDIAN |
| Points | 3.596563 | 3.92 | 3.695 |
| Score | 3.21725 | 3.44 | 3.325 |
| Weigh | 17.84875 | 17.02 | 17.71 |

**TABLE:-2**

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameters** | VARIANCE | SD | RANGE**(MAX-MIN**) |
| Points | 0.285881 | 0.534679 | 2.17 |
| Score | 0.957379 | 0.978457 | 3.911 |
| Weigh | 3.193166 | 1.786943 | 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  =  ∑ ( probability  \* Value )

**= ∑ P(x).E(x)**

Probability of selecting each patient = 1/9

Ex  108, 110, 123, 134, 135, 145, 167, 187, 199

P(x)  1/9  1/9   1/9  1/9   1/9   1/9   1/9   1/9  1/9

Expected Value  =  (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 = 145.33**

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

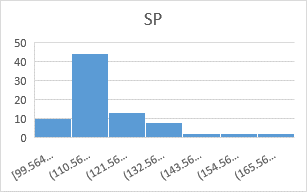
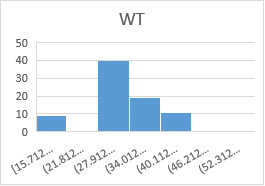
**Cars speed and distance**

**Use Q9\_a.csv**

**ANS:-**

**Table:-1**

|  |  |  |
| --- | --- | --- |
| **PARAMETERS** | **SKEWNESS** | **KURTOSIS** |
| **SP** | 1.61145 | 2.977329 |
| **WT** | -0.61475 | 0.950291 |

**SP Weight(WT**

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



**ANS:- Histogram is positively skewed it means oyliers at higher side.**



**ANS:- Box plot shows there are an outliers above upper whisker!**

**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:- Confidence Interval = x̄ ± z \* ơ / √n**

**Where, x̄ =mean, Z=value of z-stat at confidence level**

**Ơ=standard deviation, n=no of sample**

**a)AT 94%:-**

**=200±1.96\*30/√2000**

**=200 ± 1.314**

**=**[**198.686, 201.314]**

**b)AT 98%:-**

**=200±2.236\*30/√2000**

**=200 ± 1.562**

**=**[**198.438, 201.562]**

**c)AT 94%:-**

**=200±2.04\*30/√2000**

**=200 ± 1.375**

=**[198.625, 201.375]**

**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:-**

**a)Mean=sum of marks/total no. of marks**

**=34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56/18**

**=36.22**

**b)Median=40+41/2**

**=40.5**

**C)variance=σ2 = ∑ f (m − x̅)2 / n**

**=4.92+0.0484+0.0484+3.1684+3.1684+14.28+14.28+22.84+22.84+22.84+33.40+33.40+77.08+95.64+391.24/18**

**=**25.46484

**d)standard deviation=**sq.root of variance

= sq.root(25.46484)

**=5.04627**

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

**ANS:-** it means, the distribution is**symmetric**

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

**ANS:-** the distribution is**positively skewed**

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

**ANS:-** distribution is**positively skewed**

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

**ANS:-** **distribution has heavier tails than the normal distribution**

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

**ANS:-** **distribution has lighter tails 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:- Data is not normally distributed as it tilted towards right and it has high tail.**

What is nature of skewness of the data?

**ANS:- Data is positively skew.**

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

**ANS:-** IQR  = Q3 – Q1

IQR=18-10=8

**IQR=8**

Q19) Comment on the below Boxplot visualizations?



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

ANS:- **data is normally distributed in boxplot 1 as compare to boxplot2**

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.

**ANS:-**

MPG <- Cars$MPG

* 1. **P(MPG>38):-From phyton jypter notebook**

**# P(MPG>38)**

**stats.norm.cdf(38,cars.MPG.mean(),cars.MPG.std())**

**=0.347593925158270**

* 1. **P(MPG<40):-**

***# P(MPG<40)***

**stats.norm.cdf(40,cars.MPG.mean(),cars.MPG.std())**

**= 0.7293498762151616**

* 1. **P (20<MPG<50):- 20<MPG<50)**

**stats.norm.cdf(0.50,cars.MPG.mean(),cars.MPG.std())-stats.norm.cdf(0.20,cars.MPG.mean(),cars.MPG.std())**

**= 1.2430968797327613e-05**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**ANS:-**From histogram Data is some how **normaly 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:- Data in case of AT is not normally distributed it is positively skewed.

For waist it is normally distributed.

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

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

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