|  |  |
| --- | --- |
| 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 | **Ratio** |
| Blood Group | **Nominal** |
| Time Of Day | **Interval** |
| Time on a Clock with Hands | **Interval** |
| Number of Children | **Nominal** |
| Religious Preference | **Nominal** |
| 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: Total no. of events=**

**{HHH,HHT,HTT,TTT,TTH,THH,HTH,THT}=8**

**Interasted events=3**

**Probability=3/8=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: 0 % (Never possible)**

**Ans: 6/36=1/6=0.1667**

**Ans: 6/36=1/6=0.1667**

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: s.s=7choose2=21**

**5choose2=10**

**P=10/21=0.4762**

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]=1\*.015+4\*.20+3\*.65+5\*.005+6\*.01+2\*.120=3.09/count**

**=309/100count**

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: Mean Median Mode**

**Points : 3.5966 3.695 3.07,3.92(Bimode)**

**Score : 3.2173 3.325 3.44(Unimode)**

**Weight:17.8488 17.710 17.02,18.90(Bimode)**

**Variance Std Dev Range**

**Points :0.2859 0.5347 2.1700**

**Score :0.9574 0.9785 3.9110**

**Weight:3.1932 1.7469 8.3999**

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:**

**E(X)=SUM(x/n)**

**=(108\*.11+110\*.11+123\*.11+134\*.11+135\*.11+145\*.11+167\*.11+187\*.11+199\*.11)**

**=145.33**

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

Cars speed and distance

Use Q9\_a.csv

**Ans: Skewness Kurtosis**

**Speed: -0.1175 -0.5090**

**Dist : 0.8069 0.4050**

**Note: 1) mass of distribution of speed is concentrated on right side also wider pick and thinner tails.**

**2) mass of distribution of distance is concentrated on lift side also high pick and thick head.**

SP and Weight(WT)

Use Q9\_b.csv

**Ans: Skewness Kurtosis**

**SP : 1.6115 2.9773**

**WT: -0.6148 0.9503**

**Note: 1) mass of the distribution of SP is concentrated on left side also high pick.**

**2) mass of the distribution of WT is concentrated on right side also wider pick and thinner tails.**

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



**Ans: Positive skewness (mass concentrated on left side)**

**Positive kurtosis(high pick)**



**Ans: It shows the distribution has lots of outliers towards upper extreme.**

**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%**

**stats.norm.interval(.94+0.06/2,loc=200,scale=30)**

**[134.8972886724632, 265.1027113275368]**

**For 98%**

**stats.norm.interval(.98+0.02/2,loc=200,scale=30)**

**[122.72512089353299, 277.274879106467]**

**For 96%**

**stats.norm.interval(.96+0.04/2,loc=200,scale=30)**

**[130.2095637787748, 269.7904362212252]**

**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 = 41.0**

**Median=40.5**

**Mode =41.0**

**Var =25.5294**

**Std dev=5.0527**

**2) Most of the student got average marks.**

**Here mean>median so distribution is slightly Right-skewed**

**And no outliers are present.**

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

**Ans: Symmetric distribution with two modes, Normalized Skewnenn (unimode or bimode) or no skewness.**

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

**Ans: Positive Skewness=Right-skewed (data concentrated on left side)**

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

**Ans: Negative Skewness=left-skewed (data concentrated on right side)**

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

**Ans: positive kurtosis value indicates sharp peakness and thik head**,**(more peaked distribution than the normal distribution) less variation**

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

**Ans: negative kurtosis value indicates wider peakness and thinner tail,(a flatter distribution than the normal distribution) more variance**

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



What can we say about the distribution of the data?

**Ans: The portion of the box to the left of the median is longer than the portion**

**to the right of the median, and the left whisker is longer than the right**

**whisker, so the data are left-skewed**

**(it is not Normal distribution)**

What is nature of skewness of the data?

**Ans: Negative Skewness(Lift-skewed)**

What will be the IQR of the data (approximately)?   
**Ans: (Q3-Q1)=18-10**

**=8**

Q19) Comment on the below Boxplot visualizations?



**Ans: Data is symmetrically distributed.**

**mean,median and mode of boxplot1 and box plot2 are same.**

**Medians are appro =260**

**Boxplot are NO skewed in +ve and –ve.**

**And no outliers.**

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

**Ans:Boxplot1 : small variance**

**Boxplot2 : high variance**

**No outliers in both of data sets**

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) first we find mean and std dev of MPG**

**Mean=34.4221, Std dev=9.1314**

**from scipy import stats**

**stats.norm.cdf(38,loc=34.4221, scale=9.1314)**

**= 1- 0.6524**

**=0.3476**

1. **stats.norm.cdf(40,loc=34.4221, scale=9.1314)**

**=0.7293**

1. **stats.norm.cdf(50,loc=34.4221, scale=9.1314) -stats.norm.cdf(20,loc=34.4221, scale=9.1314)**

**=0.8989**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans: Follows Normal distribution as indicated by qq-plot.**

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 follows Normal Distribution from the below QQ-plot**

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

**Ans: for 90% CI**

**stats.norm.ppf(0.90+0.10/2) = 1.6449**

**for 94% CI**

**stats.norm.ppf(0.94+0.06/2) =1.8808**

**for 60% CI**

**stats.norm.ppf(0.60+0.40/2) =0.8816**

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

**Ans:Given n=25**

**for 95% CI**

**stats.t.ppf((0.95+0.05/2),df=24) = 2.0639**

**for 96% CI**

**stats.t.ppf((0.96+0.04/2),df=24) =2.1715**

**for 99% CI**

**stats.t.ppf((0.99+0.01/2),df=24) =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

**Ans: x\_bar=260days, s=90days, n=18bulbs, mu=270days**

**Hypothesis testing**

**H**o**=average life of bulbs more than 260days**

**H**a**= average life of bulbs no more than 260days**

**(Sigma is unknown so use t-test)**

**t=-0.4714**

**stats.t.cdf(-0.4714,df=17)**

**=0.3217(p-value)**

**Which is greater than alpha(0.05)**

**So fail to reject null hypothesis**

**The 18 randomly selected bulbs would have an average life of more than 260 days.**

**And probability is =32.17%**