|  |  |
| --- | --- |
| 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 | Ratio |
| Number of Children | Ratio |
| 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= **3coins tossed then total outcomes are 2\*2\*2=8**

**Out of which (HHT, HTH, THH) =3**

**P=3/8=0.375**

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

1. **Equal to 1 =**

**If two dices were rolled, then total possible cases =36**

**Total Favorable cases (Having sum =1) = 0**

**As minimum sum is 2 for outcome (1,1).**

**Hence, probability is 0.**

1. **Less than or equal to 4 =**

**(1,1) (1,2) (1,3) (2,2) total 4 out of 36**

**P = 4/36 =1/9**

1. **Sum is divisible by 2 and 3 = 1/6**

**The only numbers between 2–12 which are divisible by both 2 and 3 are 6 and 12.**

**While 12 can only be made 1 way (double 6) 6 can be made 5 ways (1,5 2,4 3,3 4,2 5,1) so the probability of throwing a sum of either 12 or 6 is (number of successful outcomes)/(number of all possible outcomes) = (1+5)/36 = 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- R=2, G=3, B=2 Total 7 balls**

**No ball should blue so 3G and 2R total 5 out of 2**

**P = 5c2 / 7c2**

**P =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] = (1 \* 0.015) + (4 \* 0.20) + (3 \* 0.65) + (5 \* 0.005) + (6 \* 0.01) + (2 \* 0.120)**

**E[X] = 0.015 + 0.80 + 1.95 + 0.025 + 0.06 + 0.24**

**E[X] = 3.095**

**So, the expected number of candies for a randomly selected child is 3.095 candies.**

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**

**count 32.000000 32.000000 32.000000**

**mean 3.596563 3.217250 17.848750**

**std 0.534679 0.978457 1.786943**

**min 2.760000 1.513000 14.500000**

**25% 3.080000 2.581250 16.892500**

**50% 3.695000 3.325000 17.710000**

**75% 3.920000 3.610000 18.900000**

**max 4.930000 5.424000 22.900000**

**#### Distribution of data shows that score and weight contains outsiders while point give double peak curve**

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 X = [108, 110, 123, 134, 135, 145, 167, 187, 199]**

**Ex = [108+110+ 123+ 134+135+145+167+187+199]/9**

**= 145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
|  | **speed** | **distance** |
| Skewness | -0.117510 | 0.806895 |
| Kurtosis | -0.508994 | 0.405053 |

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | **SP** | **WT** |
| **Skewness** | 1.611450 | **-** 0.614753 |
| **Kurtosis** | 2.977329 | 0.950291 |

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



**Distribution shows positive skewness (Right skew) i.e data concentrated more toward left.**

**it contains outliers.**



**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=

n=2000 ; m=200; σ=30

|  |  |
| --- | --- |
| % | confidence interval |
| 94 | **(143.57619175546247, 256.42380824453755)** |
| 98 | **(130.2095637787748, 269.7904362212252)** |
| 96 | **(138.38753268104531, 261.61246731895466** |

**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?

|  |  |
| --- | --- |
| mean | **41.0** |
| median | **40.5** |
| variance | **25.529411764705884** |
| standard deviation | **5.05266382858645** |

**We can say it is a uniform distribution as mean and median close to each other, 56 acts as outlier**

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

**ANS= distribution is normal, no skewness present.**

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

**ANS= concentration of distribution is towards right, -ve skew**

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

**ANS= concentration of distribution is towards left, +ve skew**

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

**ANS = sharp peak**

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

**ANS=Wider peak**

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



What can we say about the distribution of the data?

**ANS=concentrated toward right**

What is nature of skewness of the data?

**Ans=from fig mean > median i.e. –ve skewness**

What will be the IQR of the data (approximately)?   
  
**ANS=7**

Q19) Comment on the below Boxplot visualizations?



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

ANS=**2ND data is widely spread from its mean as compare to 1st data**

**Both data have appox. Same mean.**

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

ANS=

* 1. **P(MPG>38) = 0.34759392515827137**
  2. **P(MPG<40) = 0.7293498762151609**

**c. P(20<MPG<50) = 1.2430968797327491e-05**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv=-ve skew

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

Dataset: wc-at.csv

Waist\_mean 91.90183486238531

Waist\_median 90.8

AT\_mean 101.89403669724771

AT\_median 96.54

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

Ans=

|  |  |  |
| --- | --- | --- |
| **60%** | **90%** | **94%** |
| **0.2533471031357997** | **1.2815515655446004** | **1.5547735945968535** |

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

|  |  |  |
| --- | --- | --- |
| **95%** | **96%** | **99%** |
| **2.0638985616280205** | **2.1715446760080677** | **2.796939504772804** |

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 = 0.32167253567098364 32.16%

Assignment 1 = Assignment\_1.ipynb