|  |  |
| --- | --- |
| 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 | **Continuous-Interval** |
| Weight | **Continuous-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 | **Ordinal** |
| Time on a Clock with Hands | **Interval** |
| Number of Children | **Nominal** |
| Religious Preference | **Nominal** |
| Barometer Pressure | **Interval** |
| SAT Scores | **Interval** |
| Years of Education | **Ordinal** |

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

**Ans. P(H H T) + P(H T H) + P(T H H) =1/8 +1/8+1/8 = 3/8**

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

1. Equal to 1

**Ans. There is no outcomes which corresponds sum is equal to one. i.e. 0/36. Probability is 0.**

1. Less than or equal to 4

**Ans. P (Sum is less than or equal to 4) = N (Event (Sum is less than or equal to 4)) / N**

**(Event (Two dice rolled))**

**= 6 / 36 = 1/6 = 0.166 = 16.66%**

1. Divisible by 2 and 3

**Ans. P (Sum is divisible by 2 and 3) = N (Event (Sum is divisible by 2 and 3)) / N**

**(Event (Two dice rolled))**

**= 6 / 36 = 1/6 = 0.16 = 16.66%**

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 10/21**

**P (2R, 3G, 2B) P (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.**

**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**

**=       3.090**

**=  3.09**

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

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(point,score,weight) = 3.59 , 2.21 ,17.84**

**median(point,score,weight) = 3.695 , 3.325 ,17.71**

**Mode(weight)= (17.02, 18.9)**

**Mode(score)= 3.44**

**Mode(points)= 3.92, 3.07**







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.**

**There are 9 tatents**

**Probability of selecting eachpatient=1/9**

**Expectetvalue=(1/9)108+(1/9)+110+(1/9)123+(1/9)134+(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**

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

Cars speed and distance

**Use Q9\_a.csv**

**Refer to( Assignment 1 basic statics level Q9\_a .ipynb)**

**Ans:**

**Skewness(index,speed,dist)=0.0000,-0.1175,0.8068**

**Kurtosis(index,speed,dist)=-1.2000,-0.5089,0.4050**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Refer to( Assignment 1 basic statics level Q9\_a .ipynb)**

**Ans:**

**Skewness(SP,WT)= 1.611450, -0.614753**

**Kurtosis(SP,WT)=** **2.977329,0.950291**

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



**Ans: The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.**



**Ans: The boxplot has outliers on the maximum side.**

**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.**

**Degrees of freedom= 2000-1= 1999**

**Confidence interval= 94% (1- σ/2)= 1-0.03) =0.97**

**Confidene interval for 94% is 1.882**

**Confidence interval for 98%= 2.33**

**Confidence interval for 96% = 2.05**

**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.

**Mean= 41, Median= 40, variance= 24.111, Standard deviation= 4.910**

1. What can we say about the student marks?

**Ans: we don’t have outliers and the data is slightly skewed towards right because mean is**

**greater than median.**

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

**Ans. Symetrical**

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. The data is notmally distributed and kurtosis value is 0.**

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

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

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



Q. What can we say about the distribution of the data?

**Ans. Let’s assume above box plot is about age’s of the students in a school. 50% of the people**

**are above 10 yrs old and remainig are less. And students who’s age is above 15 are approx 40%.**

Q. What is nature of skewness of the data?

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

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

**Ans. Approximately= -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 are**

**equal hence distribution is symetrical.**

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)

**Ans: 0.348**

* 1. P(MPG<40)

**Ans: 0.729**

* 1. P (20<MPG<50)

**Ans: 0.013000000000000012**

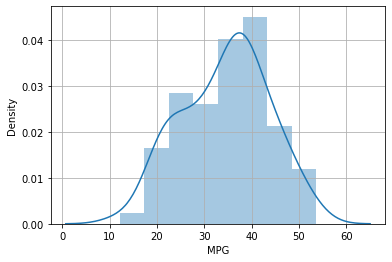
**Refer to( Assignment 1 basic statics level Q20 .ipynb)**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

**Dataset: Cars.csv**

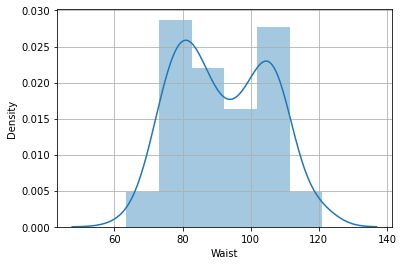
**Refer to( Assignment 1 basic statics level Q21 .ipynb)**

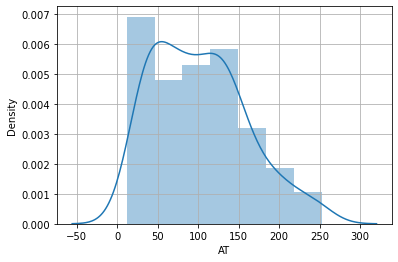


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

**Dataset: wc-at.csv**

**Refer to( Assignment 1 basic statics level Q21 .ipynb)**





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

confidence interval

**Ans:**

**# z value for 90% confidence interval**

**print('Z score for 60% Conifidence Intervla =',np.round(stats.norm.ppf(.05),4))**

**Z score for 60% Conifidence Intervla = -1.6449**

**# z value for 94% confidence interval**

**print('Z score for 60% Conifidence Intervla =',np.round(stats.norm.ppf(.03),4))**

**Z score for 60% Conifidence Intervla = -1.8808**

**# z value for 60% confidence interval**

**print('Z score for 60% Conifidence Intervla =',np.round(stats.norm.ppf(.2),4))**

**Z score for 60% Conifidence Intervla = -0.8416**

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99%

confidence interval for sample size of 25

**Ans:**

**# t score for 95% confidence interval**

**print('T score for 95% Confidence Interval =',np.round(stats.t.ppf(0.025,df=24),4))**

**T score for 95% Confidence Interval = -2.0639**

**# t value for 94% confidence interval**

**print('T score for 94% Confidence Inteval =',np.round(stats.t.ppf(0.03,df=24),4))**

**T score for 94% Confidence Inteval = -1.974**

**# t value for 99% Confidence Interval**

**print('T score for 95% Confidence Interval =',np.round(stats.t.ppf(0.005,df=24),4))**

**T score for 95% Confidence Interval = -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=260**

**Mean=270**

**Sample standard daviation=90**

**Square root of sample size=18**

**t\_score = (260-270)/90/np.sqrt(18))**

**t\_score = -0.471**

**stats.t.cdf(t\_score, df = 17)**

**0.32 = 32%**