|  |  |
| --- | --- |
| 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 | Interval |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Interval |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| 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 | Ordinal |

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

Answer: Total number of events= {hhh, hht, htt, ttt, tth, thh, hth, tht} =8

Interested events=3

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

Answer:

Total number of outcomes when two dice are rolled=6\*6=36.

(1, 1)(1, 2)(1, 3)(1, 4)(1, 5)(1, 6)

(2, 1)(2, 2)(2, 3)(2, 4)(2, 5)(2, 6)

(3, 1)(3, 2)(3, 3)(3, 4)(3, 5)(3, 6)

(4, 1)(4, 2)(4, 3)(4, 4)(4, 5)(4, 6)

(5, 1)(5, 2)(5, 3)(5, 4)(5, 5)(5, 6)

(6, 1)(6, 2)(6, 3)(6, 4)(6, 5)(6, 6)

a) Equal to 1 = 0% probability

b) Less than or equal to 4= 6/36 = 1/6

c) Sum is divisible by 2 and 3= 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?

Answer: Total number of events=7c2=7!/2!\*5!=21

Interested events =5c2=5!/2!\*3!=10

Probability that none of the balls is blue=10/21=0.47

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

Answer: Expected number = E(x) ==1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120= **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** ; ***For points:***

Q7.points.mean()

3.596563

Q7.points.median()

3.695

Q7.points.mode()

0 3.07

1 3.92

Q7.points.var()

0.2858814

Q7.points.std()

0.5346787

Q7.points.max() , Q7.points.min()

Range= max 4.93, min 2.76

***For score:***

Q7.score.mean()

3.217250

Q7.score.median()

3.325

Q7.score.mode()

0 3.44

Q7.score.var()

0.9573789677419356

Q7.score.std()

0.978457

Q7.score.max() , Q7.score.min()

Range= max 5.42, min 1.513

***For weigh:***

Q7.weigh.mean()

17.848750

Q7.weigh.median()

17.71

Q7.weigh.mode()

0 17.02

1 18.90

Q7.weigh.var()

3.193166129032258

Q7.weigh.std()

1.786943

Q7.weigh.max() , Q7.weigh.min()

Range= max 22.9, min 14.5

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?

Answer: EV=Σx/n ==145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

Answer**: Q9\_a.skew()**

**skewness of speed = -0.11750986144663393**

**Skewness of distance = 0.8068949601674215**

**Kurtosis of speed= -0.508994**

**Kurtosis of distance= 0.405053**

***Inferences: for speed the value of skewness is negative so it is main orientation on right side and longer tail on left side . For distance value is positive so its main orientation on left side and from left to right peaks are start to decreasing with longer tail on right.***

**SP and Weight(WT)**

**Use Q9\_b.csv**

Answer: **: Q9\_b.skew()**

**skewness of SP= 1.611450**

**Skewness of weight = -0.614753**

**Kurtosis of SP =** **2.977329**

**Kurtosis of Weight = 0.950291**

***Inferences: for Weight the value of skewness is negative so it is main distribution/orientation on right with longer tail on left and peaks start to decrease from right to left . For SP, main orientation on left side as the skewness value is positive with longer tail on right .***

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



**Answer*: In this graph the main distribution is on left side and as we move from left to right peaks starts to decreasing with longer/flatter tail on left so the skewness is on positive side.***

**

Answer: ***The above boxplot suggests that 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?

Answer: ***we have***

***x̅=200 s=30 n= 2000***

***since we don’t have µ, we have s , so t distribution is required***

***For 94%: t.interval(.94,2000,loc=240,scale=30)= (183.54417173267188, 296.4558282673281)***

***For 98%: t.interval(0.98,2000,loc=240,scale=30)= (170.1535847418068, 309.8464152581932)***

***For 96%=t.interval(.96,2000,loc=240,scale=30)= (178.34732124381935, 301.65267875618065)***

**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.

Answer: stats.mean(stud)= 41

Stats.variance(stud)= 25.529411764705884

Stats.median(stud)= 40.5

Stats.mode(stud)= 41

Stats.stdev(stud)= 5.05266382858645

1. What can we say about the student marks?

Answer: Everyone scores decent marks and no outlier is present

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

Answer: Symmetrical

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: ***Sharp peak , less variation and distribution is peaked***

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

Ans: ***wider peaks and distribution is flat.***

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



What can we say about the distribution of the data?

Ans: ***it is not a Normal Distribution and main orientation on right side***

What is nature of skewness of the data?

Ans***: It is left/negative skewed***.

What will be the IQR of the data (approximately)?   
Ans: ***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.

Answer: ***\* Median is same in both boxplots that is approx. 260***

***\* Distribution is symmetrical***

***\* No outliers are also not present in both boxplots***

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)

***a=stats.norm.cdf(53,34.42,9.13)= 0.9790770430687412=97%***

***b=stats.norm.cdf(38,34.42,9.13)= 0.6525129749869594=65%***

***c=a-b=97-65=32%***

Answer: ***32% probability***

* 1. P(MPG<40)

Answer: ***stats.norm.cdf(40,34.42,9.13)= 0.7294571279557076=72.9% probability***

* 1. P (20<MPG<50)

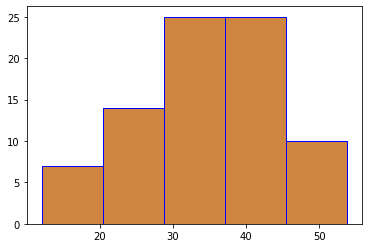
Answer: ***stats.norm.cdf(50,34.42,9.13)- stats.norm.cdf(20,34.42,9.13)= 0.9560389719783978- 0.05712118952347562= 0.8989177824549222=89% probability***

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer: ***Follows Normal distribution as indicated by bar graph***

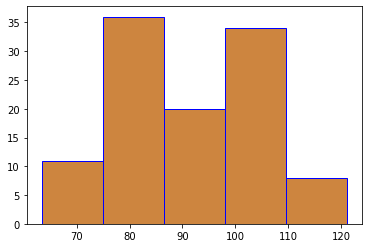
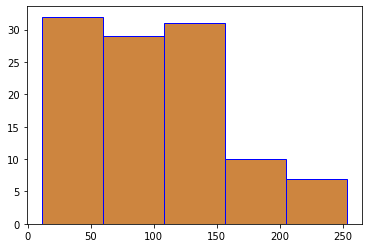
******

***Skewness is -0.1 which is almost symmetric***

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

Dataset: wc-at.csv

Answer: ***Both AT and waist not follows normal distribution.***

******

***AT Waist***

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

Answer***: 90%= stats.norm.ppf(0.95) = 1.6448536269514722***

***94%= stats.norm.ppf(0.97) = 1.8807936081512509***

***60%= stats.norm.ppf(0.80) = 0.8416212335729143***

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

**Answer:** ***95%=stats.t.ppf(0.975,df=24)= 2.0638985616280205***

***96%=stats.t.ppf(0.98,df=24)= 2.1715446760080677***

***99%=stats.t.ppf(0.995,df=24)= 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

Answer:

***µ=270, =260, SD=90, n=18, df=n-1=18-1= 17***

***tscore= = = -10/21.23= -0.47***

***> pt(-0.47,17)= 0.3221639***

***Required probability = 0.32=32%***