|  |  |
| --- | --- |
| 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) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| 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?

Sol) Total no of events= (HHH,HHT,HTH,HTT,THH,THT,TTT,TTH,HTT) =8

Events that has 2 heads and one tail = (HHT,THH,HTH)=3

Probability =

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

Sol) Total no of outcomes when two dice are rolled = 6X6=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)

1. Equal to 1 = 0 probability
2. Less than or equal to 4=(1,2),(3,1),(2,2),(2,1),(1,3),(1,1)=6/36=1/6
3. Sum is divisible by 2 and 3

( 2 3 4 5 6 7

3 4 5 6 7 8

4 5 6 7 8 9

5 6 7 8 9 10

6 7 8 9 10 11

7 8 9 10 11 12 )

Probability =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?

Sol) total no of events=nCr = 7C2 = =21

No of events that none of the balls drawn is blue = 5C2 = =10

Probability that none of the balls is blue=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

Sol) Expectation =

=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**

**For Points**

**>>** mean(Q7$Points)

3.596563

**>>**median(Q7$Points)

3.695

**>>** var(Q7$Points)

0.2858814

**>>** sd(Q7$Points)

0.5346787

**>>** range(Q7$Points)

2.76 4.93

**For Score**

**>>** mean(Q7$Score)

3.21725

**>>** median(Q7$Score)

3.325

**>>** var(Q7$Score)

0.957379

**>>** sd(Q7$Score)

0.9784574

**>>** range(Q7$Score)

1.513 5.424

**For Weigh**

**>>** mean(Q7$Weigh)

17.84875

**>>** median(Q7$Weigh)

17.71

**>>** var(Q7$Weigh)

3.193166

**>>** sd(Q7$Weigh)

1.786943

**>>** range(Q7$Weigh)

14.5 22.9

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?

Sol) Expectation =

= \*108+\*110+\*123+\*134+\*135+\*145+\*167+\*187+\*199 \

=

= 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

kurtosis(Q9\_a$speed)

[1] 2.422853

> kurtosis(Q9\_a$dist)

[1] 3.248019

> skewness(Q9\_a$speed)

[1] -0.1139548

> skewness(Q9\_a$dist)

[1] 0.7824835

**SP and Weight(WT)**

**Use Q9\_b.csv**

> kurtosis(Q9\_b$SP)

[1] 5.723521

> kurtosis(Q9\_b$WT)

[1] 3.819466

> skewness(Q9\_b$SP)

[1] 1.581454

> skewness(Q9\_b$WT)

[1] -0.6033099

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



Sol) most of that chick Weight are lies between 50-100 and its has positively skewed



Sol) its has positively

**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?

Sol) stats.t.interval(0.94,1999,200,30/np.sqrt(2000))

>> (198.7376089443071, 201.2623910556929)

stats.t.interval(0.98,1999,200,30/np.sqrt(2000))

>> (198.4381860483216, 201.5618139516784)

stats.t.interval(0.96,1999,200,30/np.sqrt(2000))

>> (198.6214037429732, 201.3785962570268)

**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.

x=c(34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56)

mean(x)

median(x)

var(x)

sd(x)

sol) mean(x)

[1] 41

> median(x)

[1] 40.5

> var(x)

[1] 25.52941

> sd(x)

[1] 5.052664

1. What can we say about the student marks?

Sol) mean> median, so that distribution is skewed towards right.no outliers are present

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

Sol) Mean=median, symmetric and skewness=0

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

Sol) right skewed, skewness=positive

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

Sol) left skewed , skewness=negative

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

Sol) High peak at center of data

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

Sol) Wide peak at center of data

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



What can we say about the distribution of the data?

Sol) Its not a Normal Distribution

What is nature of skewness of the data?

Sol) It left skewed

What will be the IQR of the data (approximately)?   
Sol) IQR is 8

Q19) Comment on the below Boxplot visualizations?



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

Sol) Both boxplots have same median, doesn’t have outliers

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)**

from scipy import stats

np.round(1-stats.norm.cdf (38, df.MPG.mean(), df.MPG.std()),3)

**0.348**

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

np.round(stats.norm.cdf(40, df.MPG.mean(),df.MPG.std()),3)

**0.729**

**c. P (20<MPG<50)**

stats.norm.cdf (50, df.MPG.mean(), df.MPG.std ()) -(1-stats.norm.cdf (20, df.MPG.mean(), df.MPG.std()))

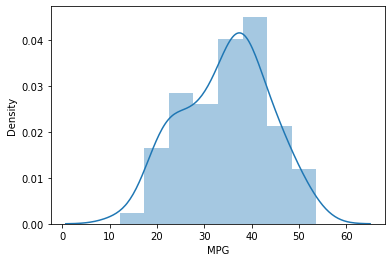
**0.013116469610523374**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

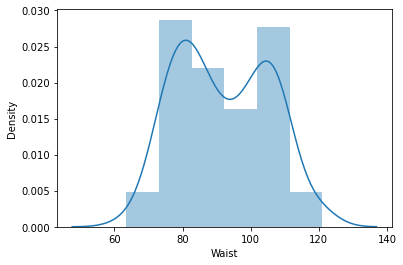
Sol) MPG is not follows Normal Distribution

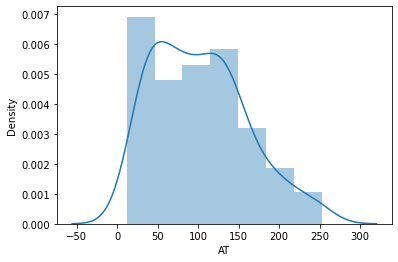


1. Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Sol) Adipose Tissue (AT) and Waist Circumference (Waist) doesn’t follow Normal Distribution





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

Sol) import scipy import stats

From scipy.stats import norm

90%= stats.norm.ppf(0.95)

1.6448536269514722

94%= stats.norm.ppf(0.97)

1.8807936081512509

96%= stats.norm.ppf(0.8)

0.8416212335729143

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

Sol) 95% = stats.t.ppf(0.97,24) =1.971

96% = stats.t.ppf(0.98,24) =2.171

99% = stats.t.ppf(0.995,24) =2.796

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

sol) µ=270, =260, SD=90, n=18, df= 17

t-score= = -0.4714

degree of freedom = 17

pt(tscore, df) = pt(-0.4714,17) = 0.32