|  |  |
| --- | --- |
| 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 | Interval |
| 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 | Interval |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Interval |

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

Ans. Total No. of outcome =2^3=8

(HHH,HHT,HTH,THH,TTH,THT,HTT,TTT)

Req. 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

Ans. Total Outcomes =36 (11,12,13,14,15,16,21,22,23,24,25,26,31,32,33,34,35,36,

41,42,43,44,45,46,51,52,53,54,55,56,61,62,63,64,65,66)

Prob. Of any Outcome= fav. Outcome % Total Outcome

Ans. A) 0

B) Fav. Outcome=11,12, 13,21,22,31

=6/36

=1/6

C) Fav. Outcome i.e Sum is divisible by 2 and 3 = 15, 24, 33,42,51,66

= 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. Number of ways of selecting 2 balls out of 7 is 7C2= 7! / (7!-2!)2!

=**21** (total Outcome)

Now as per condition selecting non blue balls is 5C2= 5!/2!(5!-2!)=10

Req. Probability = **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. The basic expected value formula is the probability of an event multiplied by the amount of times the event happens: ∑ (P (E) \* n).

So, Req. expected number is =(1\*0.015 + 4\*0.20 + 3\*0.65 + 5\*0.005 + 6\*0.01 +2\*.120) =.015+.80+1.95+.025+.06+.240)= **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.

|  |  |  |  |
| --- | --- | --- | --- |
|  | points | score | weigh |
| median | 3.5965 | 3.325 | 17.71 |
| mean | 3.6950000 | 3.217250 | 17.848750 |
| Mode | 3.07 & 3.92 | 3.44 | 17.02 & 18.90 |
| var | 0.285881350 | 0.9573 | 3.19316 |
| SD | 0.534679 | 0.978457 | 1.786943 |
| Range | 2.17 | 3.907 | 8.40 |
|  |  |  |  |

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.** The basic expected value formula is the probability of an event multiplied by the amount of times the event happens: ∑ (P (E) \* n).

Probability of selecting each patient = 1/9.And Here n is 108, 110, 123, 134, 135, 145, 167, 187, 199.

Expected Value = (1/9)(108) + (1/9)110  + (1/9)123 + (1/9)134 + (1/9)135 + (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

**Expected Value of the Weight of that patient = 145.33**

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

**Cars speed and distance**

Use Q9\_a.csv

Skewness= speed -0.117510 left skewed

dist 0.806895

Kurtosis = speed -0.508994

dist 0.405053

SP and Weight(WT)

Use Q9\_b.csv

Skewness =SP 1.611450 right skewed

WT -0.614753

Kurtosis= SP 2.977329 distribution is leptokurtik

WT 0.950291

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



This is a right-skewed distribution, indicating that there are a number of values greater than the mode.



When the median is closer to the bottom of the box, and if the whisker is shorter on the lower end of the box, then the distribution is positively skewed .

**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**. Here **X̅= 200, n= 2000, s(SD)=30.** Since sample size is large we use z Table instead of t table. So, formula **= X̅+z**val**\*s/n^.5** .

Z values for 94 ,98, 96 CI is 1.88, 2.33, 2.06. So by using above formula we get following values=

**#CI=94%** 200+1.88\*.670=201.261 { 30 /2000^.5=.670 }

200-1.88\*.670= 198.7388

**#CI=98%**

200+2.33\*.670=201.561

200-2.33\*.670=198.438

**#CI=96%**

200+2.06\*.670=201.380

200+2.06\*.670=198.619

Using Python:

stats.norm.interval(.94,200,30/math.sqrt(2000)) #CI=94%

(198.738325292158, 201.261674707842)

stats.norm.interval(.98,200,30/math.sqrt(2000)) #CI=98%

198.43943840429978, 201.56056159570022)

stats.norm.interval(.96,200,30/math.sqrt(2000)) #CI=96%

(198.62230334813333, 201.37769665186667)

**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?

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

Ans. Skewness is zero as distribution is normal if mean =median.

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

Ans. Positively skewed

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

Ans. Negatively skewed

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

Ans. Distribution is peaked and possess thick tails.

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

Ans. Distribution is flat and has thin tails.

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



What can we say about the distribution of the data?

Ans. asymmetrical

What is nature of skewness of the data?

Ans. Left skewed

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

Ans. 8

Q19) Comment on the below Boxplot visualizations?



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

Ans. Data in 2nd boxplot has more IQR , hence more spread the data less reliable it becomes.

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. cars[cars['MPG']>38].count() =33

P(MPG>38)= 33/81

* 1. P(MPG<40)= 61/81

c. P (20<MPG<50)=69/81

Q 21) Check whether the data follows normal distribution

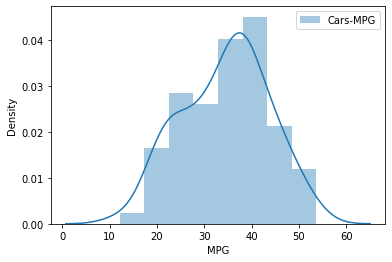
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Mean= 34.422

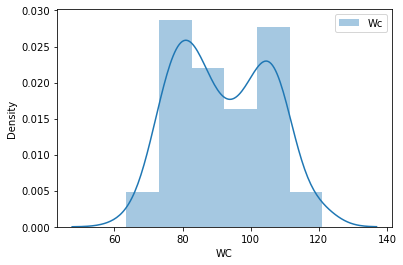
Median= 35.152

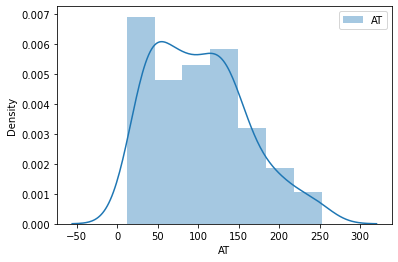
For normal Dist. Median=mean.So MPG doesn’t follows Normal Dist.



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

Dataset: wc-at.csv





Both doesn’t follow distribution

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

**Ans.**

from scipy import stats

from scipy.stats import norm

#z-score of 90% confidence interval

stats.norm.ppf(.95)

1.6448536269514722

#z-score of 94% confidence interval

stats.norm.ppf(.97)

1.8807936081512509

#z-score of 60% confidence interval

stats.norm.ppf(.80)

0.8416212335729143

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

**Ans.** N=25 Given

from scipy import stats

from scipy.stats import norm

#t-score of 95% confidence interval

stats.t.ppf(.975,25)

2.059538552753294

#t-score of 96% confidence interval

stats.t.ppf(.98,25)

2.1665866344527562

#t-score of 99% confidence interval

stats.t.ppf(.995,25)

2.787435813675851

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= n-1 =17

Ans. t - statistics for the data is given as follows:

https://tex.z-dn.net/?f=t%3D%5Cdfrac%7Bx-%5Cmu%7D%7B%5Cfrac%7Bs%7D%7B%5Csqrt%20n%7D%7D

x = mean of the sample of bulbs =  260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

So, t = (260-270)/90sqrt90

t = -0.471

now by t- score calculator with df= 17.

**Req. probability= .321**