|  |  |
| --- | --- |
| 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) | Nominal |

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 | Interval |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Ordinal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

**Ans**. { HHH,HHT,HTH,HTT,THH,THT,TTH,TTT}

Two heads=3

Total=8

3/8=0.375 \*100=**37.5%**

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-**

1. **0**
2. **16.6 % (1/6)**
3. **66.66 % (24/36)**

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 or 47.61 %**

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\*0.120

=0.015+0.8+1.95+0.025+0.06

**=3.090**

Hence, Expected number of candies for a randomly selected child is **3.090.**

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.

**Ans.**

**1.Mean**- Points = 3.596563

Score = 3.217250

Weigh = 17.848750

**2.Median**- Points = 3.695

Score = 3.325

Weigh = 17.710

**3.Mode** Points 3.92

Score Na

Weigh 18.90

**4.Variance** Points 0.285881

Score 0.957379

Weigh 3.193166

**5.Standard Deviation**- Points 0.534679

Score 0.978457

Weigh 1.786943

**Use Q7.csv file**

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**. Expected value= ∑ (P(x) ̽Weight of Patient

= ∑ [1/9\*(108+110+123+134+135+145+167+187+199)]

**= 145.33**

There are 9 patients hence; Probability of X is 1/9. i.e P(X)= 1/9 for every patient.

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

Cars speed and distance

**Use Q9\_a.csv**

**1. Skewness-**

A. Cars speed = -0.11750986144663393

B. Cars Distance = 0.8068949601674215

**2. Kurtosis-**

A. Cars speed = -0.5089944204057617

B. Cars Distance = 0.4050525816795765

**SP and Weight (WT)**

**Use Q9\_b.csv**

1. **Skewness-**

**A.** SP = 1.611450

**B.** WT = -0.614753

1. **Kurtosis-**

**A.** SP = 2.977329

**B.** WT = 0.950291

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



Ans. About Histogram:

1. Chicken Weight Data is positively skewed

2. More than 50 % chicken weight is between 50 to 150

3. Mostly weight of the chicken is between 50 to 100

About Box plot:

1. The data is Right Skewed.

2. There are outliers at upper 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.**Given Data,

Sample mean of µ= 200

Sample standard deviation of S = 30.

Sample size of n= 2000

Interval= µ±t Wherea,t=critical value for two tailed confidence

t=1.8916 for 94% Confidence

t=2.3452 for 98% Confidence

t=2.0673 for 96% Confidence

**1. For 94%,**

Interval = 200 + 1.8916 = **201.27**

Interval = 200 - 1.8916 = **198.73**

**2. For, 98%**

Interval = 200 + 2.3452 = **201.58**

Interval = 200 - 2.3452 = **198.44**

**2. For, 96%**

Interval = 200 + 2.0673 = **198.62**

Interval = 200 - 2.0673 = **201.38**

**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?

**Ans.**

1. Mean=41, Median=40.5, variance=25.52, Std. Deviation=5.052
2. We can say that mean of marks of student is 41 which is slightly greater than median.

Most of the students got marks in between 41-42; there are two outlier 49, 56.

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

**Ans.** If a frequency distribution graph has a symmetrical frequency curve i.e which is Normally Distributed Curve, then mean, median and mode will be equal.

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

**Ans.**  The distribution is positively skewed.

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

**Ans.**  The distribution is negatively skewed.

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

**Ans.**It indicates that distributions is peaked and possess thick tails.

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

**Ans**.It indicates that a 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.** The data is distributed in De-assigned format

What is nature of skewness of the data?

**Ans.** Left side skewed

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

**Ans**. IQR = Q3-Q1

= 18-10

= **8 is IQR**

Q19) Comment on the below Boxplot visualizations?



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

**Ans** .The box plot 1 designed with range = 3, the second one range is = 1.5

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)
  2. P(MPG<40)

c. P (20<MPG<50)

**Ans.** A. 0.3475939251582705

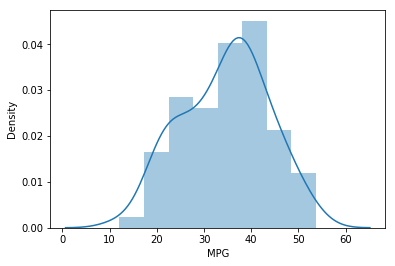
B. 0.7293498762151616

C. 1.2430968797327613

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

**Dataset: Cars.csv**

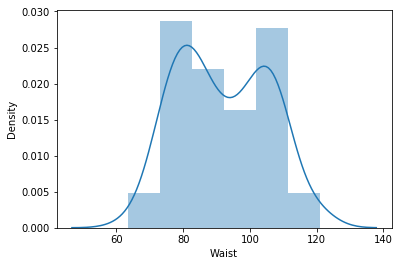


1. Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution.

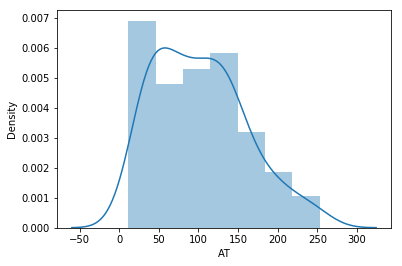
**Dataset: wc-at.csv**

**Ans.**

**1. Waist-**



**2. AT-**



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

**Ans.**

1. Given,

CL=90%

A=

==0.95

This value lies between the z-score column=0.08& Row=1.2

So,Z-Score for 90% CL is **1.28**

**2.** A=

==0.97

This value lies between the z-score column=0.05 & Row=1.5

=1.89

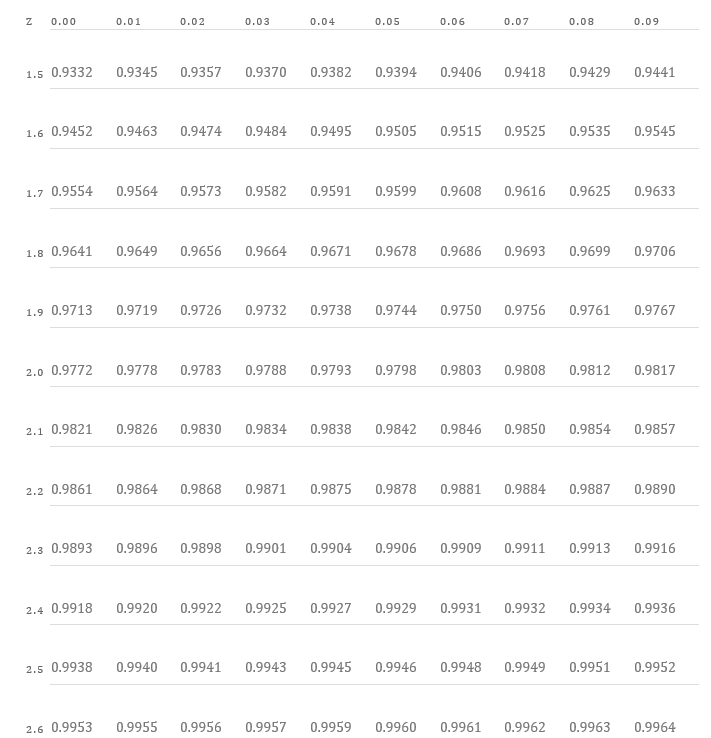
So,Z-Score for 94 % CL is **1.55**

**3.** A=

==0.80

This value lies between the z-score column=0.05 & Row=0.2

So,Z-Score for 94 % CL is **0.25**

For reference use below table

Also, I used following scripting Part-

stats.norm.ppf(CL)

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

**Ans**. 1. T-Score with CL 95% = 2.0638985616280205

2. T-Score with CL 96%= 2.1715446760080677

3. T-Score with CL 99% = 2.796939504772804

**Refered this Scripting Part-stats.t.ppf(0.995,24)**

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**. Don’t Understand.Kindly Help me regarding the same.