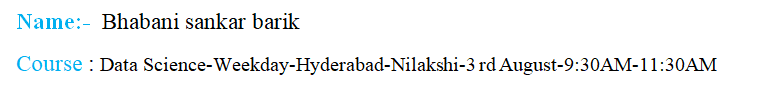
|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuous |
| Weight of Gold | Continuous |
| Distance between two places | Continuous |
| Length of a leaf | Continuous |
| Dog's weight | Continuous |
| 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) | Ratio |
| 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 | Ordinal |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?  
**ANS:-`**Probability(p) of two heads and one tail =

*P*= Number of favorable outcomes​ / Total number of possible outcomes  
 =two heads(2)+one tail(1) / Three coins (8 faces)  
 =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:  
   (a)Probability(p) that sum is equal to 1=  
    Number of ways to get a sum of 1/Total number of possible outcomes  
   = 0/36

(b)Less than or equal to 4 =   
No of ways to get sum of 4/total number of possible outcomes  
= (1,1),(1,2),(2,1),(1,3),(2,2),(3,1) **/** 36  
= 6/36  
=1/6  
(c) sum is divisible by 2 and 3=  
Number of ways to get sum is divisible by 2 and 3 **/** Total no of possible outcomes  
= (1,5),(2,4),(3,3),(4,2),(5,1),(6,6) / 36  
= 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:- Total no of balls = (2+3+2) =7  
 Let S be the sample space .  
then, n(S) =Number of ways of drawing 2 balls out of 7  
=7 c2  
=(7\*6) / (2\*1)  
=21  
Let E = Event of drawing 2 balls, none of which is blue.  
n(E)= Number of ways of drawing 2 balls out of (2+3)balls .  
=5 c2  
=(5\*4) / (2\*1)  
=10  
 i,e P(E) = N(E) / N(S) = 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 Expected number of candies for a randomly selected child   
let X = Candies count

(x\*P(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***> mean(Q7$Points)

[1] 3.596563

> median(Q7$Points)

[1] 3.695

> mfv(Q7$Points)

[1] 3.07 3.92

> var(Q7$Points)

[1] 0.2858814

> sd(Q7$Points)

[1] 0.5346787

> range(Q7$Points)

[1] 2.76 4.93  
  
**Inferences**

The mean and median are close, indicating a relatively symmetric distribution.  
The mode (mfv) has two values (3.07 and 3.92), suggesting some biomodality or variability in the central tendency.  
The low variance and standard deviation suggest that the values are clustered around the mean .  
***#for score***> mean(Q7$Score)

[1] 3.21725

> median(Q7$Score)

[1] 3.325

> mfv(Q7$Score)

[1] 3.44

> var(Q7$Score)

[1] 0.957379

> sd(Q7$Score)

[1] 0.9784574

> range(Q7$Score)

[1] 1.513 5.424  
 **Inferences**

\*The mean and the median are relatively close, indicating a moderately symmetric distribution.  
\*The mode (mfv) is 3.44 indicating a prominent value in the data set.  
\*The higher variance and standard deviation suggest more spread in the data compared to the points variable .

***#for weigh***> mean(Q7$Weigh)

[1] 17.84875

> median(Q7$Weigh)

[1] 17.71

> mfv(Q7$Weigh)

[1] 17.02 18.90

> var(Q7$Weigh)

[1] 3.193166

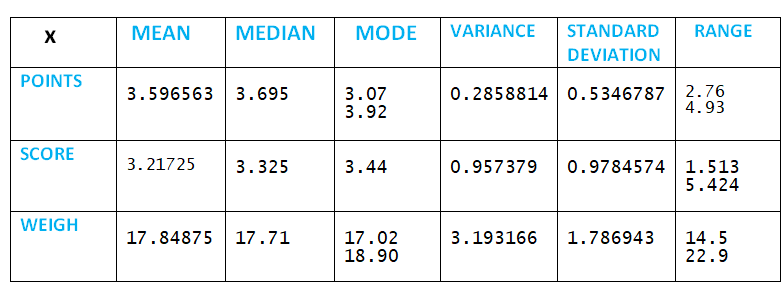
> sd(Q7$Weigh)

[1] 1.786943

> range(Q7$Weigh)

[1] 14.5 22.9

**Inferences**\*The mean and median are close, indicating a relatively symmetric distribution.  
\*The mode (mfv) has two values (17.02 and 18.90), suggesting some bimodality or variability in the central tendency.  
\*The higher variance and standard deviation compared to the Points variable suggest more spread in the data.

****

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=sum(X\*probability(X))

={[(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)]}  
=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv  
ANS:-   
for speed**> kurtosis(Q9\_a$speed)

**[1] 2.422853**

> skewness(Q9\_a$speed)

[1] -0.1139548

**for distance**> kurtosis(Q9\_a$dist)

[1] 3.248019

> skewness(Q9\_a$dist)

[1] 0.7824835

**inferences**

\*For the speed variable, there is a moderate amount of variability with potentially more extreme values on the lower end  
\* For the distance variable, there is a higher amount of variability with potential presence of outliers on the higher end .  
\*Both variable have distributions that deviate from a normal distribution, with heavy tails and skewness .

**SP and Weight(WT)**

**Use Q9\_b.csv  
ANS:-  
for SP**

> skewness(Q9\_b$SP)

[1] 1.552258

> kurtosis(Q9\_b$SP)

[1] 5.723521

**For Weight**> skewness(Q9\_b$WT)

[1] -0.5921721

> kurtosis(Q9\_b$WT)

[1] 3.819466

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



**ANS:-** **Histogram:**

\*Chick weight is a Right skewed or positively skewed

\*Fifty percent of the chick weight falls between 50 to 150

\*Highest Chick Weight is between 50-100

**BOX PLOT**  
\*It is Right skewed or positively skewed .



\*It has some outliers on the top 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:-**   
X+/-(Z1- α. σ/sqrt(n)  
Degrees of freedom= 2000-1= 1999   
Confidence interval= 94%   
(1- σ/2)= 1-0.03) =0.97   
for 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.**

**ANS**:-  
> score=c(34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56)

> mean(score)

[1] 41

> median(score)

[1] 40.5

> var(score)

[1] 25.52941

> sd(score)

[1] 5.052664

1. What can we say about the student marks?

ANS:- We can say that there is no outlier present in the data and It is slightly skewed towards the right side because the mean is greater than the median.

Q13) What is the nature of skewness when mean, median of data are equal?  
ANS:- No skewness Is present if mean and median are equal ,it is called normal distribution

Q14) What is the nature of skewness when mean > median ?  
ANS:- If the mean is greater than the median than the distribution is positively skewed.

Q15) What is the nature of skewness when median > mean?  
ANS:-If the median is greater than the mean than the distribution is negatively skewed .

Q16) What does positive kurtosis value indicates for a data ?  
ANS:-positive kurtosis means curve is more peaked and it is Leptokurtic .

Q17) What does negative kurtosis value indicates for a data?  
ANS:- Negative kurtosis indicates that the peak is wide , thinner tails and the curve is flatter .

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



What can we say about the distribution of the data?  
ANS:-The box plot is not normally distributed , median is towards right side. The box plot does not have any outliers. It is skewed towards left side.

What is nature of skewness of the data?  
ANS:-It is left skewed data .

What will be the IQR of the data (approximately)?   
ANS:-IQR=Upper quartile(Q3)-Lower quartile(Q1 )  
=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.  
ANS:-   
\*We can see that both the Boxplot doesn’t have an Outlier. Both the data have the same point as a Median of around 262.

\*They are normally distributed with no skew ness either at the minimum or maximum whisker range

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:-**P(MPG>38)= 1-pnorm(38,mean(Cars$MPG),sd(Cars$MPG))

=0.3475939

(here >38 is given so when we do normal distribution to 38 all values till 38 i.e (<38) is generated so for (>38) we should subtract it with 1.)

* 1. P(MPG<40)

**ANS:-**P(MPG<40)= pnorm(40,mean(Cars$MPG),sd(Cars$MPG))

= 0.7293499

c. P (20<MPG<50)

**ANS:-**  
P(20<MPG<50)= pnorm(50,mean(Cars$MPG),sd(Cars$MPG))-

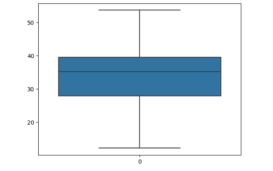
(1-pnorm(20,mean(Cars$MPG),sd(Cars$MPG)))

=0.01311647

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

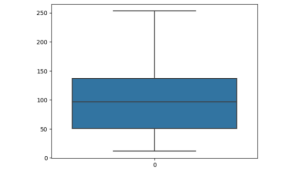
Dataset: Cars.csv  
**ANS:-**



The MPG of Cars follows Normal Distribution with Median is 35.

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

Dataset: wc-at.csv  
**ANS:-**



Adipose Tissue Does not follow Normal Distribution.

Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval   
**ANS:-** A=(1+CL)/2 ; z-score formula=qnorm(CI value)

For 90%

qnorm((1+0.90)/2)

[1] 1.644854

For 94%

> qnorm((1+0.94)/2)

[1] 1.880794

For 60%

> qnorm((1+0.60)/2)

[1] 0.8416212

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25  
**ANS:-** n=25 , df=n-1=25-1=24

t-score formula=qt(Confidence interval value,degree of freedom)

For 95%

qt((1+0.95)/2,24)

[1] 2.063899

For 96%

> qt((1+0.96)/2,24)

[1] 2.171545

For 99%

> qt((1+0.99)/2,24)

[1] 2.79694

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:-** Consider it is a normal distribution

X(mean)=260

µ=270

std=90

number of sample=18

t\_score=(x-µ)/(std/(number of sample^1/2))

(260-270)/(90/(18^(1/2)))

[1] -0.4714045

t\_score = -0.4714045

df=n=1=18=1=17

pt(t\_score,df)= pt(-0.4714045,17)

0.3216725

Probability is 32%