**NAME: J.SUNIL KUMAR**

**MAIL ID:sunilkumar.jayamoorthy@gmail.com**

**BASIC STATISTICS**

**ASSIGNMENT-1**

Q1) Identify the Data type for the Following

|  |  |
| --- | --- |
| 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 |

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

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

Answer:

3 coins are tossed=8 outcomes, In that 2Head & 1 Tail =3 outcomes

Expected outcome =3/8

Total no of outcomes

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. NO value
2. 6/36=1/6
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 balls=8

Two balls drawn random=2(not a blue ball)

For first ball, the probability=5/7

For second ball, the probability=4/6

TOTAL: 5/7\*4/6=20/42=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

Answer:

1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120=1.875

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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | Mean | Median | Mode | Variance | SD | Range |
| Points | 3.59 | 3.69 | 3.44 | 0.28 | 0.53 | 2.17 |
| Score | 3.21 | 3.25 | 3.07&3.98 | 0.95 | 0.97 | 8.4 |
| Weighs | 17.8 | 17.71 | 17.02 &18.9 | 3.19 | 1.78 | 3.91 |

**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?

Answer=108+110+123+134+135+145+167+187+199=1308/9=145.33

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

**Cars speed and distance**

|  |  |  |
| --- | --- | --- |
| **Cars** | **Skewness** | **Kurtosis** |
| **Speed** | **-0.1139548** | **2.422853** |
| **Distance** | **0.7824835** | **3.248019** |
| **Index** | **0.0000000** | **1.799040** |

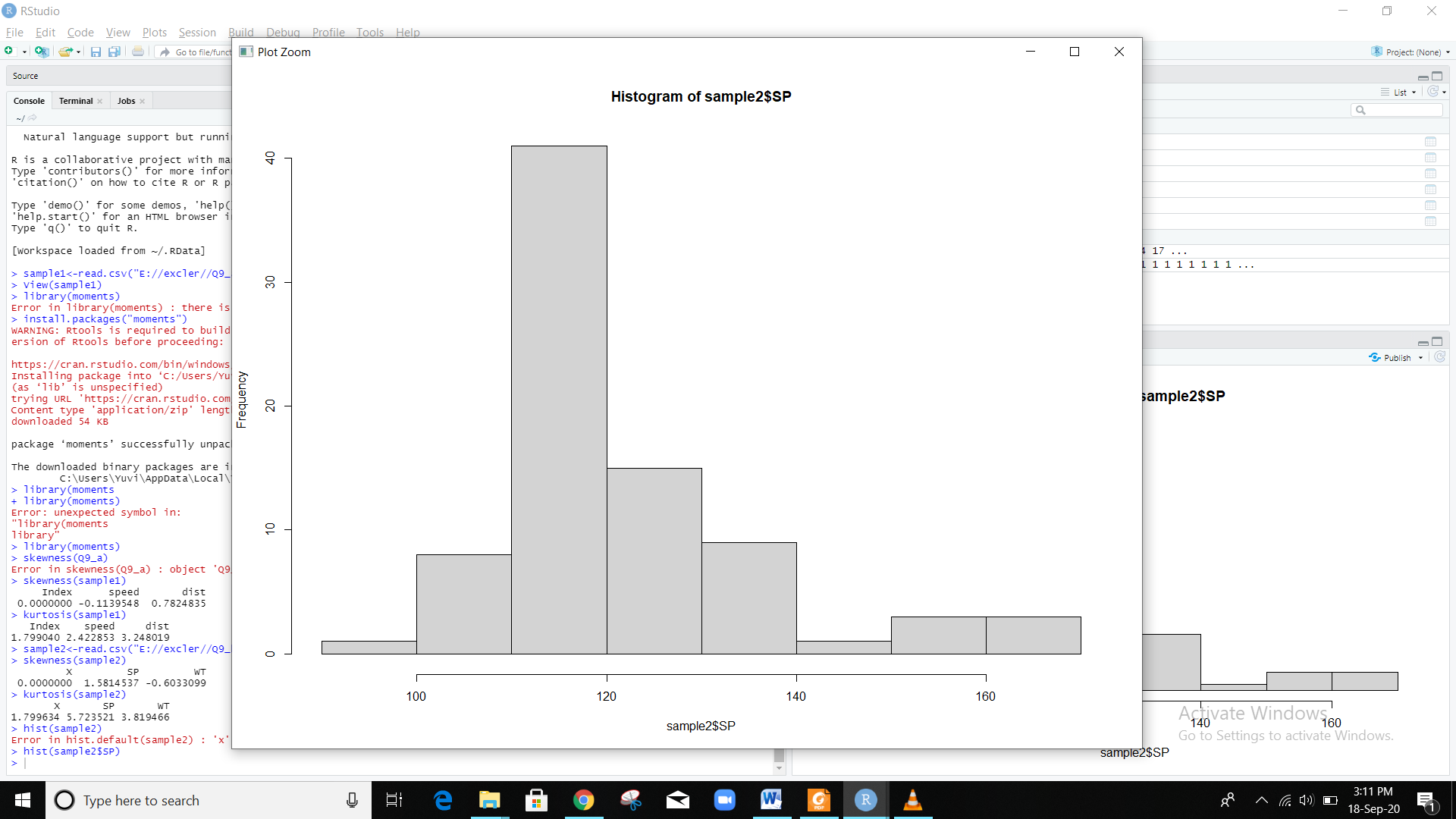
**Use Q9\_a.csv**

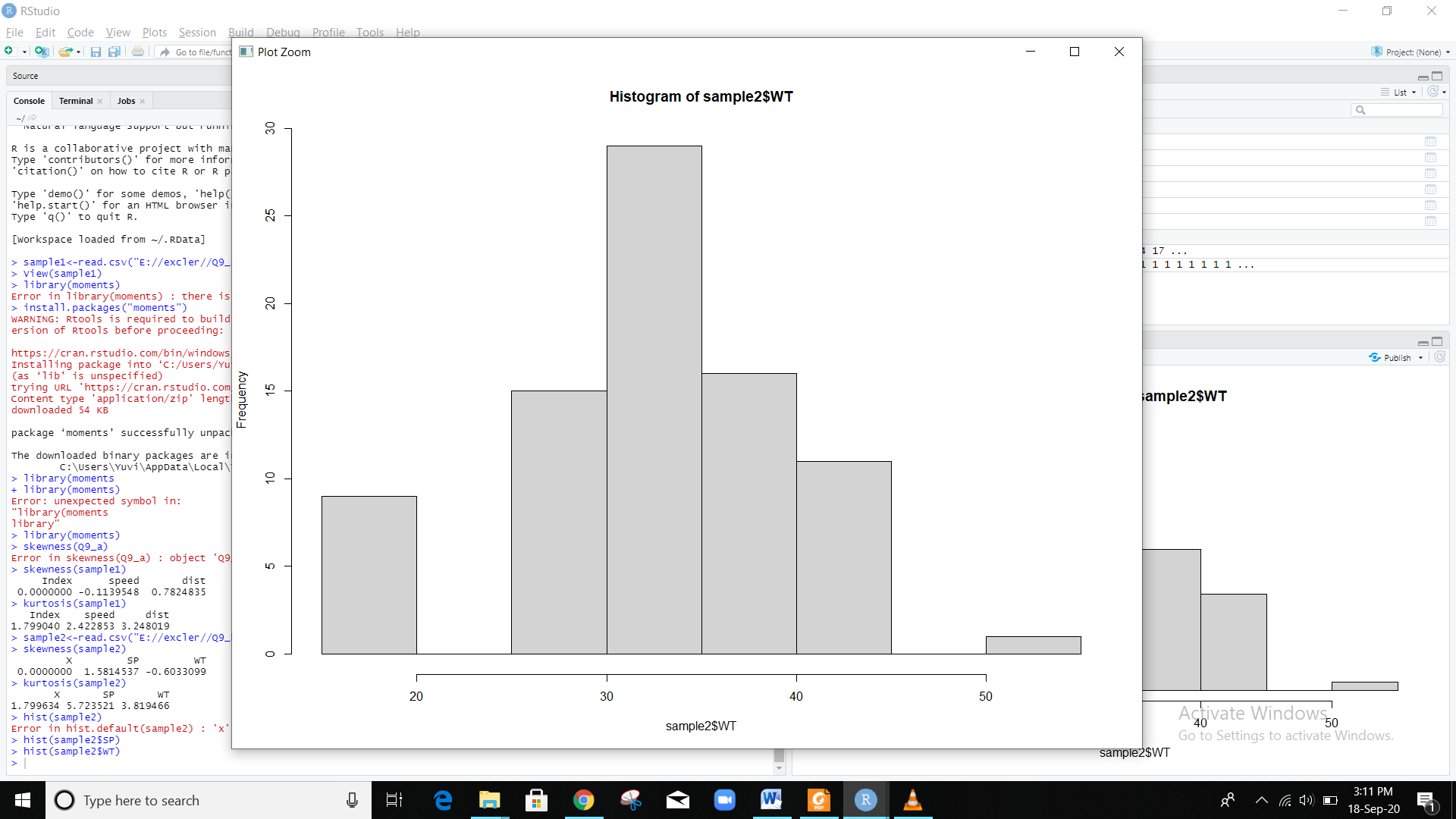
**SP and Weight(WT)**

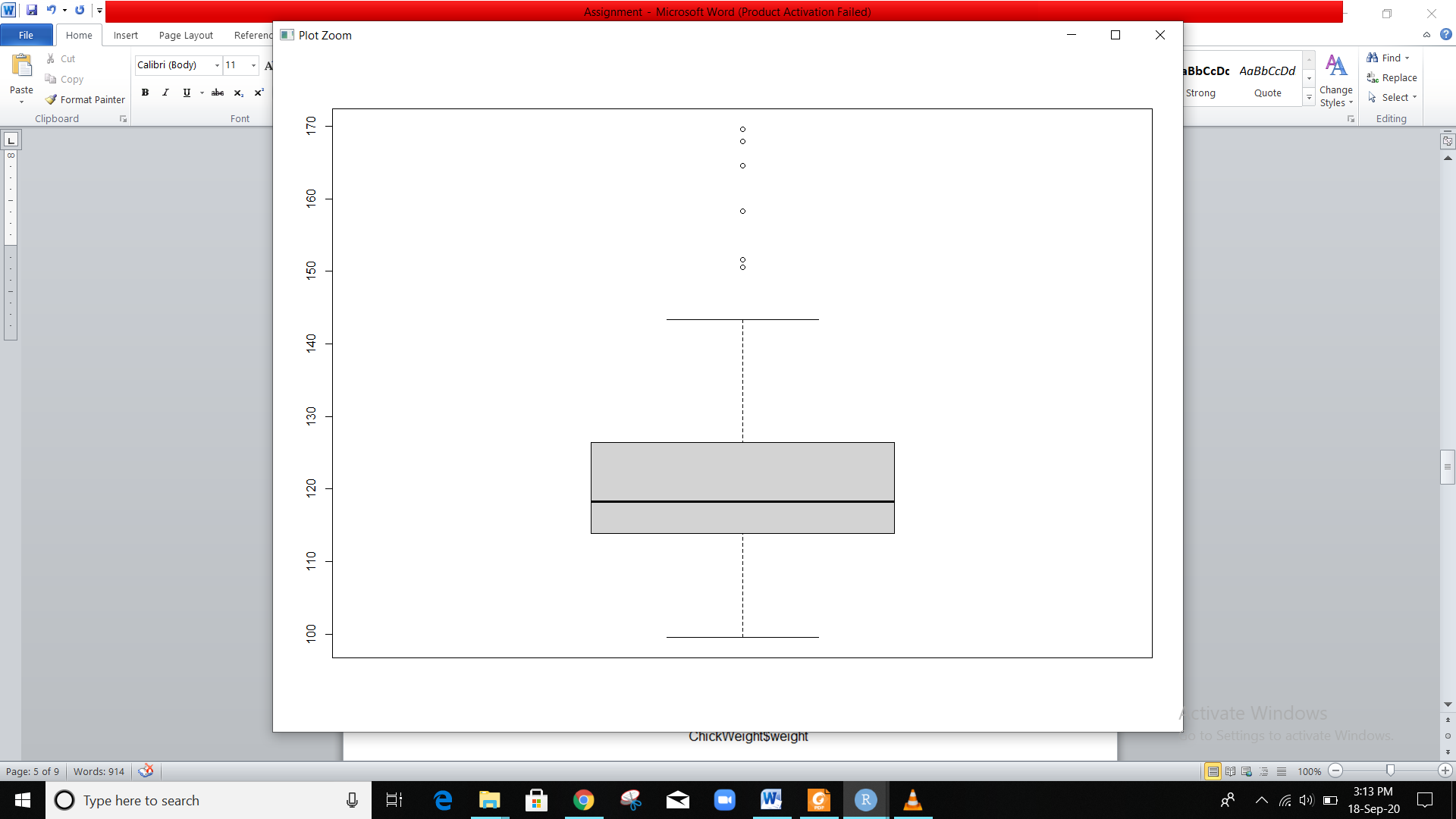
**Use Q9\_b.csv**

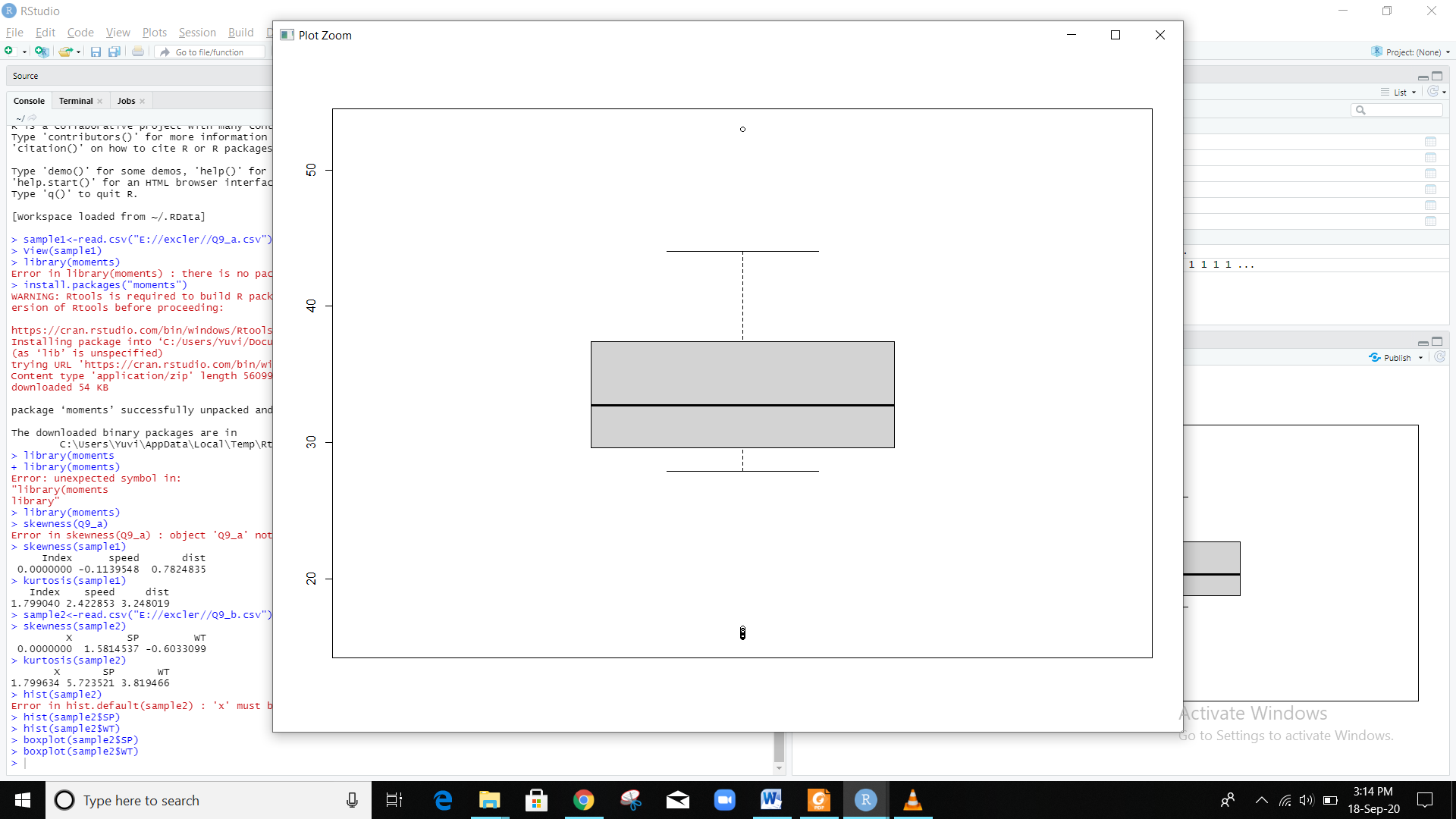
|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **SP** | **1.5814537** | **5.723521** |
| **WT** | **-0.6033099** | **3.819466** |

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









**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?

2000> 30, so it is normal

With 94% of confidence level;

Finding Standard Error=30/Square root(2000)

=0.670825

Finding Critical Probability=1-α/2

=1-0.06/2=0.97

Qnorm(0.97)=1.880794=critical Value

Margin Error=CriticalValue \*Standard Error=1.880794\*0.670825

=0.126

The 94% of confidence Level is (200-0.126)-(200+0.126)

(199.9874-200.126)

With 96% of confidence level;

Finding Standard Error=30/Square root(2000)

=0.670825

Finding Critical Probability=1-α/2

=1-0.04/2=0.98

Qnorm(0.98)= 2.053749=critical Value

Margin Error=CriticalValue \*Standard Error=2.053749\*0.670825

=1.377

The 96% of confidence Level is (200-1.377)-(200+1.377)

(198.623-201.377)

With 98% of confidence level;

Finding Standard Error=30/Square root(2000)

=0.670825

Finding Critical Probability=1-α/2

=1-0.02/2=0.99

Qnorm(0.99)= 2.326348=critical Value

Margin Error=CriticalValue \*Standard Error= 2.326348\*0.670825

=1.56

The 96% of confidence Level is (200-1.56)-(200+1.56)

(198.44-201.56)

**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.

Mean=41

Median=40.5

Variance=25.52941

Standard deviation=5.052664

1. What can we say about the student marks?

Normal distribution

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

When the mean= median it is Zero Skewness.  If the distribution is both symmetric and unimodal, then the mean = median = mode.

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

Positive Skewness

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

Negative Skewness

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

Positive Kurtosis implies high peak and taller(flatter) tails.

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

Negative kurtosis implies wider peak and thinner tails.

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



What can we say about the distribution of the data?

Asymmetric Distribution

What is nature of skewness of the data?

🡪Long tail is on Left Side

🡪Mass is on the Right side

🡪Skewness is Negative Skewness

What will be the IQR of the data (approximately)?   
  
 IQR=Q3-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.

|  |  |
| --- | --- |
| BOXPLOT 1 | BOX PLOT 2 |
| Min=237.5 | Min=200 |
| Max=287.5 | Max=350 |
| Q1=250 | Q1=225 |
| Q3=275 | Q3=300 |
| IQR=25 | IQR=75 |

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)=33/81
  2. P(MPG<40)=61/81
  3. 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

sam<-read.csv("E://excler//Cars.csv")

> View(sam)

> shapiro.test(sam$MPG)

Shapiro-Wilk normality test

data: sam$MPG

W = 0.97797, p-value = 0.1764

So, P value is high ,so it 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

shapiro.test(sam1$AT)

Shapiro-Wilk normality test

data: sam1$AT

W = 0.95234, p-value = 0.000654

P value is less than 0.05, it doesn’t follows Normal distribution

shapiro.test(sam1$Waist)

Shapiro-Wilk normality test

data: sam1$Waist

W = 0.95586, p-value = 0.00117

P value is less than 0.05, it doesn’t follows Normal distribution

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

Z90%=1.64

Z94%=1.88

Z60%=0.84

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

T score for 25 sample with 95% of confidence interval is qt(0.975,24)=2.06

T score for 25 sample with 96% of confidence interval is qt(0.98,24)=2.17

T score for 25 sample with 99% of confidence interval is qt(0.995,24)=2.79

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

t - statistics for the data is given as follows:

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

T=x-u/s/square root(n)

t = - 0.471

df(-0.471,17)=0.321814

here, t value is less than t value obtained with degrees of freedon 17 which is 0.321>-0.471

the probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of the bulbs is 300days