|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuous / Ratio |
| Weight of Gold | Continuous / Ratio |
| Distance between two places | Continuous / Ratio |
| Length of a leaf | Continuous / Ratio |
| Dog's weight | Continuous / Ratio |
| Blue Color | Nominal |
| 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/ Nominal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal/Interval |
| 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?

Answer: P(H H T) + P(H T H) + P(T H H)

=1/8 +1/8+1/8

= 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

Answer :

a) There is no outcomes which corresponds sum is equal to one. i.e.

0/36. Probability is 0.

b) (1,3) (2,2) (3,1) = 3 outcomes, 3/36 i.e. 1/12

c) 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:

P (2R, 3G, 2B)

P (5/7, 4/6) = 20/42 i.e. 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:

Expected number of candies for randomly selected child =

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**

Answer: Refer Jupyter notebook: [Basic Stats L1\Q7\q7.pdf](Basic%20Stats%20L1/Q7/q7.pdf)

Points:

Mean =3.596563, Median= 3.695, Mode= “numeric”, Variance=0.2858814, Standard deviation= 0.5346787, Range = 2.17

Score:

Mean= 3.21725, Median= 3.325, Mode= “numeric”,

Variance= 0.957379, Standard deviation= 0.9784574

Note: Mean value are closer for both ‘Point’ and ‘Score’.

Range = 3.91

Weight:

Mean= 17.84875, Median= 17.71, Mode= “numeric”,

Variance= 3.193166, Standard deviation= 1.786943

Range = 8.39

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:

Expected Value  =  ∑ ( probability  \* Value )

 ∑ P(x).E(x)

there are 9 patients

Probability of selecting each patient = 1/9

Ex  108, 110, 123, 134, 135, 145, 167, 187, 199

P(x)  1/9  1/9   1/9  1/9   1/9   1/9   1/9   1/9  1/9

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**

**Answer:** refer notebook [Basic Stats L1\Q9\q9a.pdf](Basic%20Stats%20L1/Q9/q9a.pdf)

Skewness

1. Speed distribution is left skewed (negative skewness)
2. Distance distribution is right skewed (positive skewness)

kurtosis

1. Speed distribution is platykurtic (negative kurtosis i.e. flatter than normal distribution)
2. Distance distribution is leptokurtic (positive kurtosis i.e. peaked than normal distribution)

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Answer:** refer notebook[Basic Stats L1\Q9\q9b.pdf](Basic%20Stats%20L1/Q9/q9b.pdf)

Skewness:

1. SP distribution is Right skewed (Positive skewness)
2. WT distribution is Left skewed (Negative skewness)

Kurtosis:

Both the SP and WT distributions are leptokurtic (have positive kurtosis i.e. Peaked than normal distribution)

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



Answer:

The most of the data points are concentrated in the range 50-100 with frequency 200.

And least range of weight is 400 somewhere around 0-10.

So, the expected value the above distribution is 75.

In case of Skewness-we can notice a long tail towards right so it is heavily right skewed



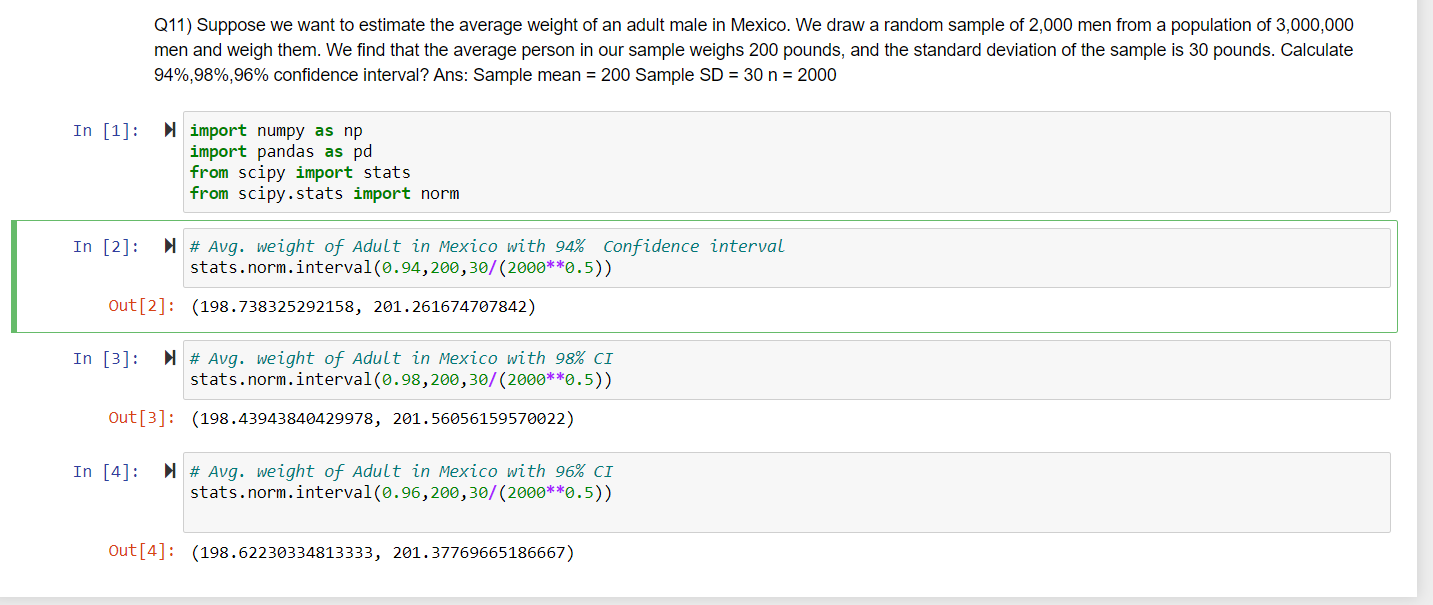
Answer:

Median is less than mean right skewed and we have outlier on the upper

side of box plot and there are less data points between Q1 and bottom point.

**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?

Answer:



**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?

Answer:

Mean= 41, Median= 40.5, variance= 25.52,

Standard deviation= 5.52

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

Answer: Symmetrical

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

Answer: Right Skewed

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

Answer: Left Skewed

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

Answer : The data is normally distributed and Kurtosis value is 0.

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

Answer :

It indicates that the distribution is flat and has thin tails. Platykurtic distributions have negative kurtosis values.

A platykurtic distribution is flatter (less peaked) when compared with the normal distribution, with fewer values in its shorter (i.e. lighter and thinner) tails.

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



What can we say about the distribution of the data?

Answer:

Let’s assume above box plot is about age’s of the students in a school.

50% of the people are above 10 years old and remaining are less.

And students who’s age is above 15 are approximately 40%

What is nature of skewness of the data?

Answer:

It is Left skewed; median is greater than mean.

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

Answer: approx. 8 (Upper quartile 18 and lower quartile is approx. 10, hence 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.

Answer:

By observing both the plots whisker’s level is high in boxplot 2, mean and

median are equal hence distribution is symmetrical.

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)

**Answer:**

1-pnorm(38,34.422,9.13144)= 0.3475908

* 1. P(MPG<40)

**Answer:**

pnorm(40,34.422,9.13144)= 0.7293527

* 1. P (20<MPG<50)

**Answer:**

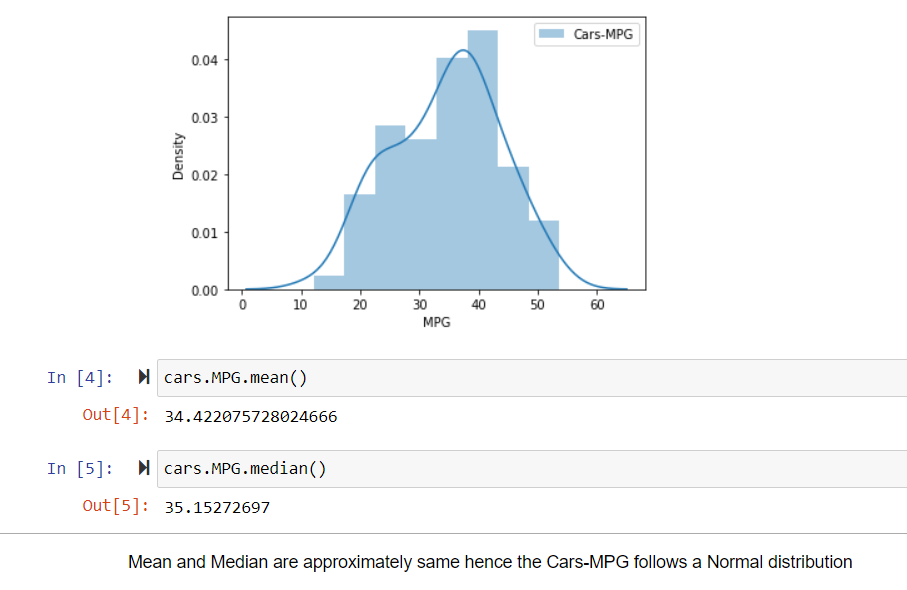
pnorm(50,34.422,9.13144)-(1-pnorm(20,34.422,9.13144))= 0.01311818

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

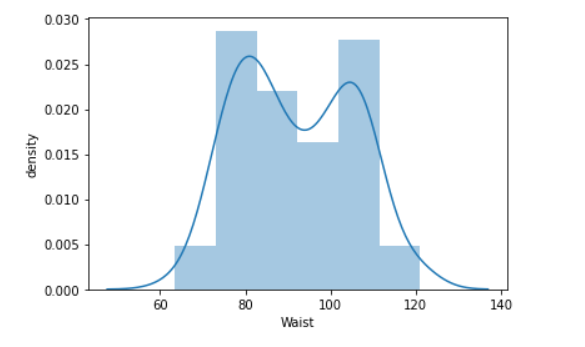
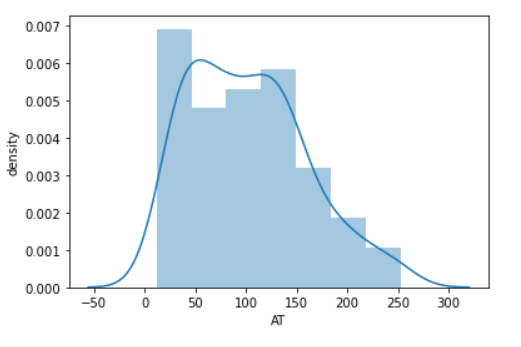
Answer: refer [Basic Stats L1\Q21\q21a.pdf](Basic%20Stats%20L1/Q21/q21a.pdf)



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

Dataset: wc-at.csv

Answer: refer [Basic Stats L1\Q21\q21b.pdf](Basic%20Stats%20L1/Q21/q21b.pdf)

Both the Adipose Tissue (AT) and Waist Circumference(Waist) data set do follow the normal distribution approximately (as mean and median of both the data are approximately same)

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

Answer: refer [Basic Stats L1\Q22\q22.pdf](Basic%20Stats%20L1/Q22/q22.pdf)

1. Z-score of 90% confidence interval : 1.64
2. Z-score of 94% confidence interval : 1.88
3. Z-score of 60% confidence interval: 0.84

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

Answer: refer [Basic Stats L1\Q23\q23.pdf](Basic%20Stats%20L1/Q23/q23.pdf)

1. t scores of 95% confidence interval for sample size of 25: 2.06
2. t scores of 96% confidence interval for sample size of 25: 2.17
3. t scores of 99% confidence interval for sample size of 25: 2.79

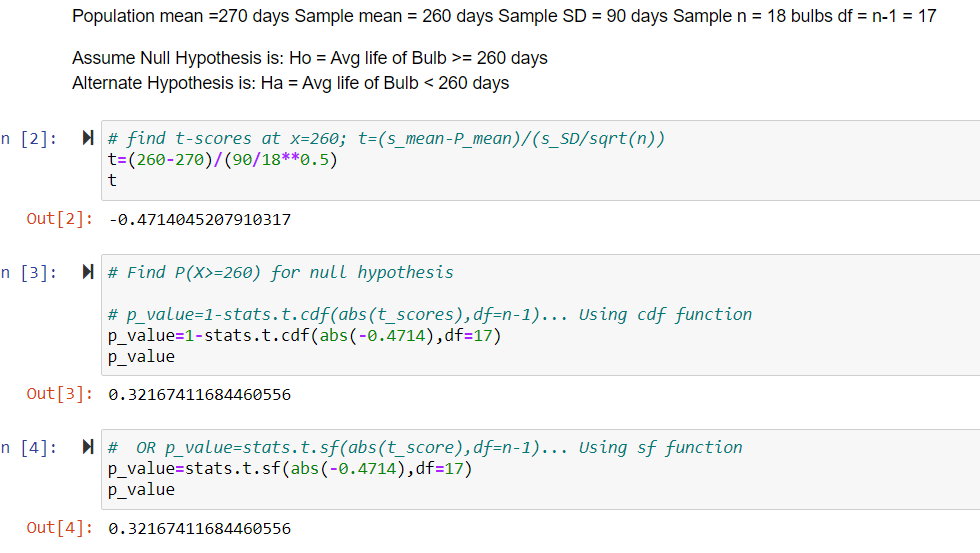
Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

Answer: refer [Basic Stats L1\Q24\q24.pdf](Basic%20Stats%20L1/Q24/q24.pdf)



Conclusion: Probability that 18 randomly selected bulbs would have an average life of no more than 260 days is **32.17%**  
Assuming significance value α = 0.05 (Standard Value)(If p\_value < α ; Reject H0 and accept Ha or vice-versa)  
Thus, as p-value > α ; Accept H0 i.e. The CEO claims are false and the avg life of bulb > 260 days