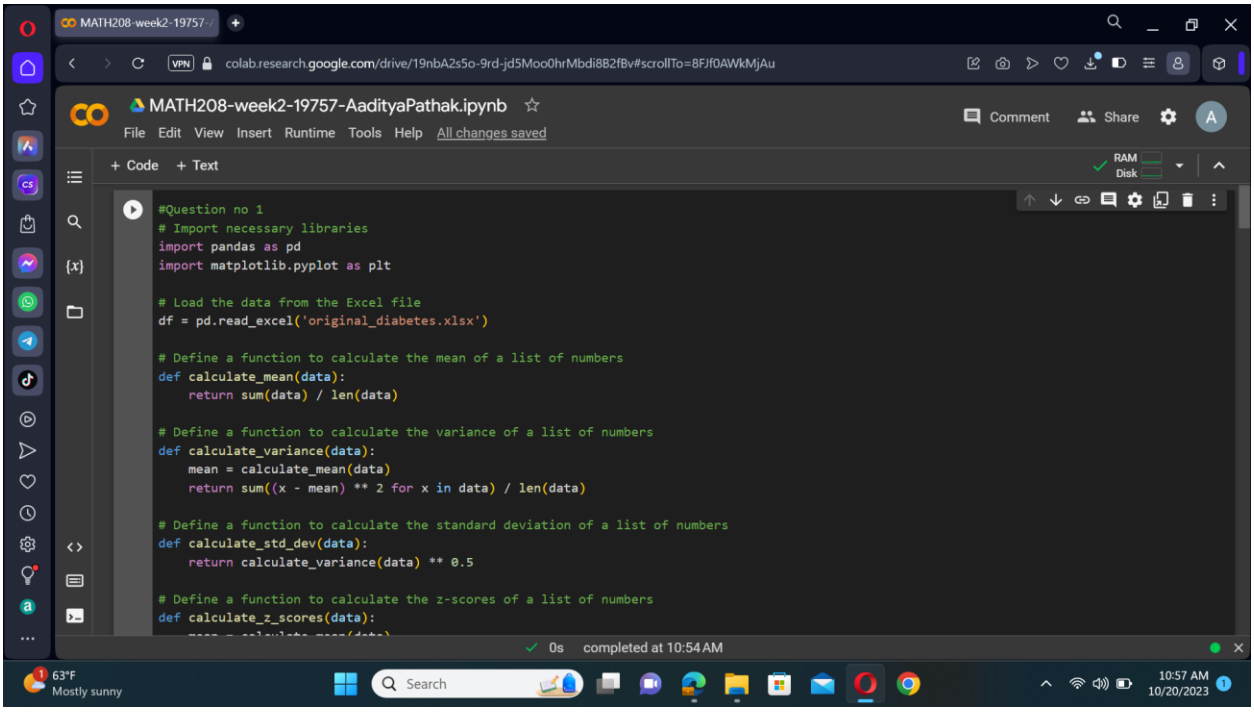
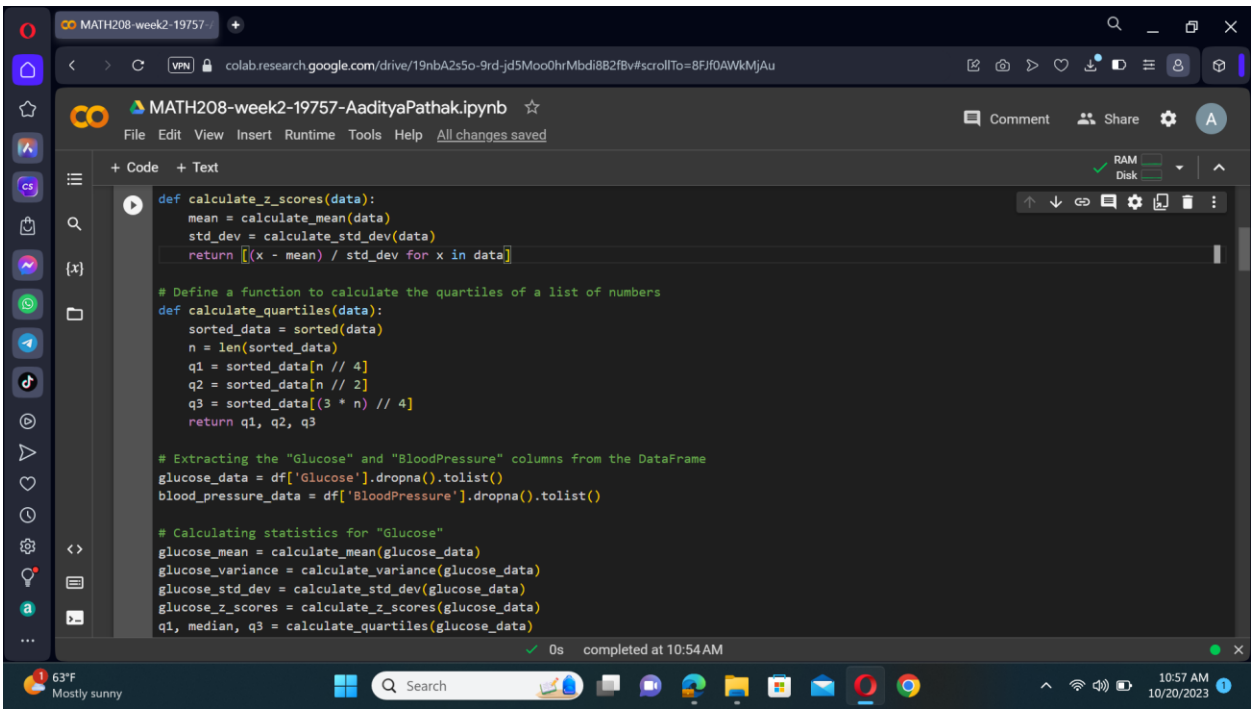
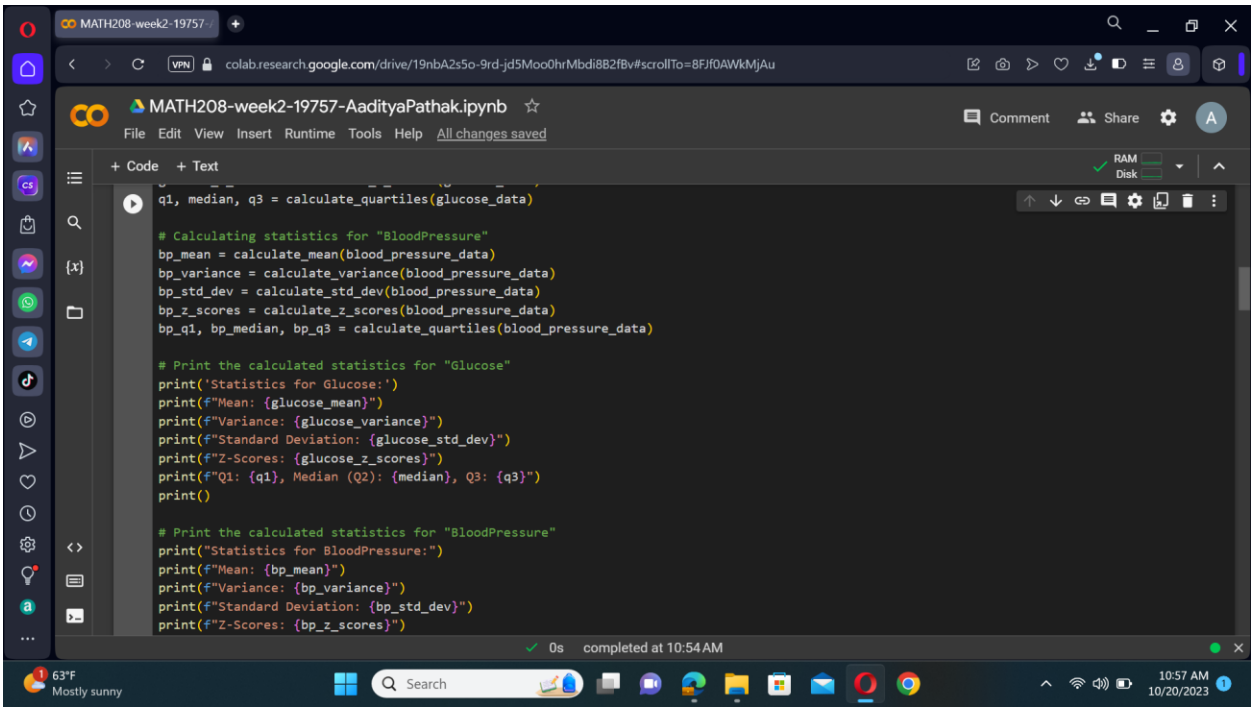
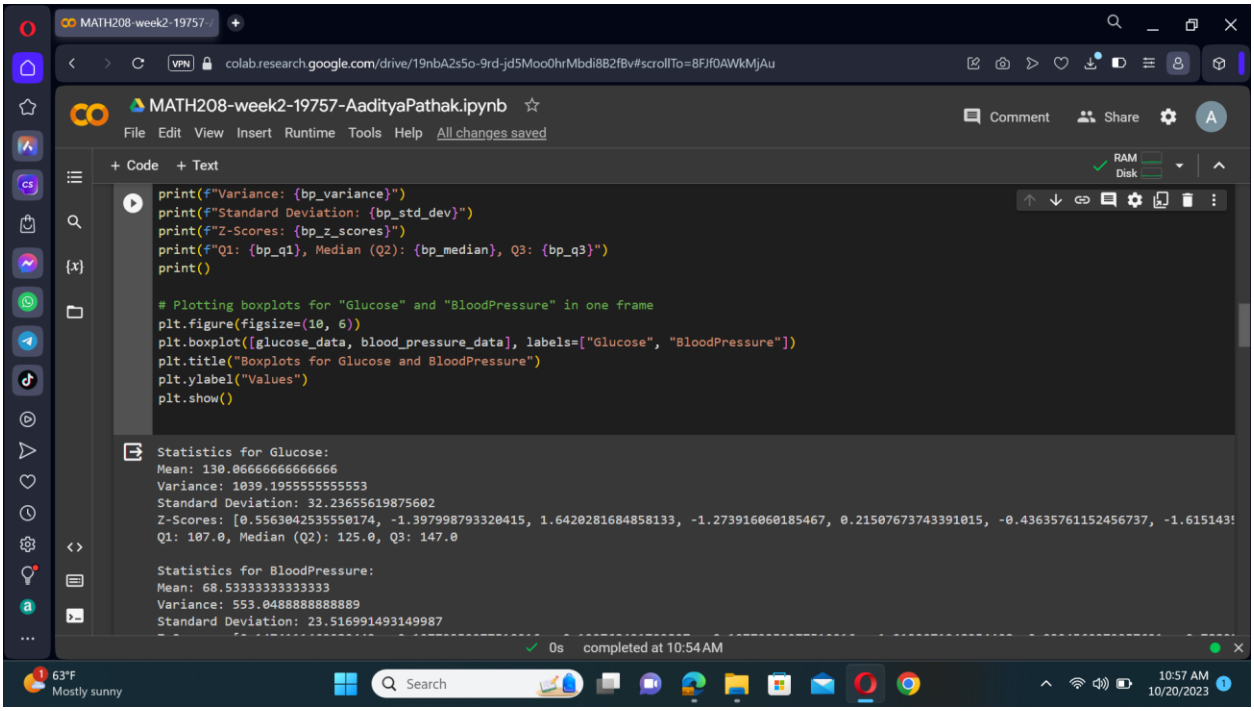
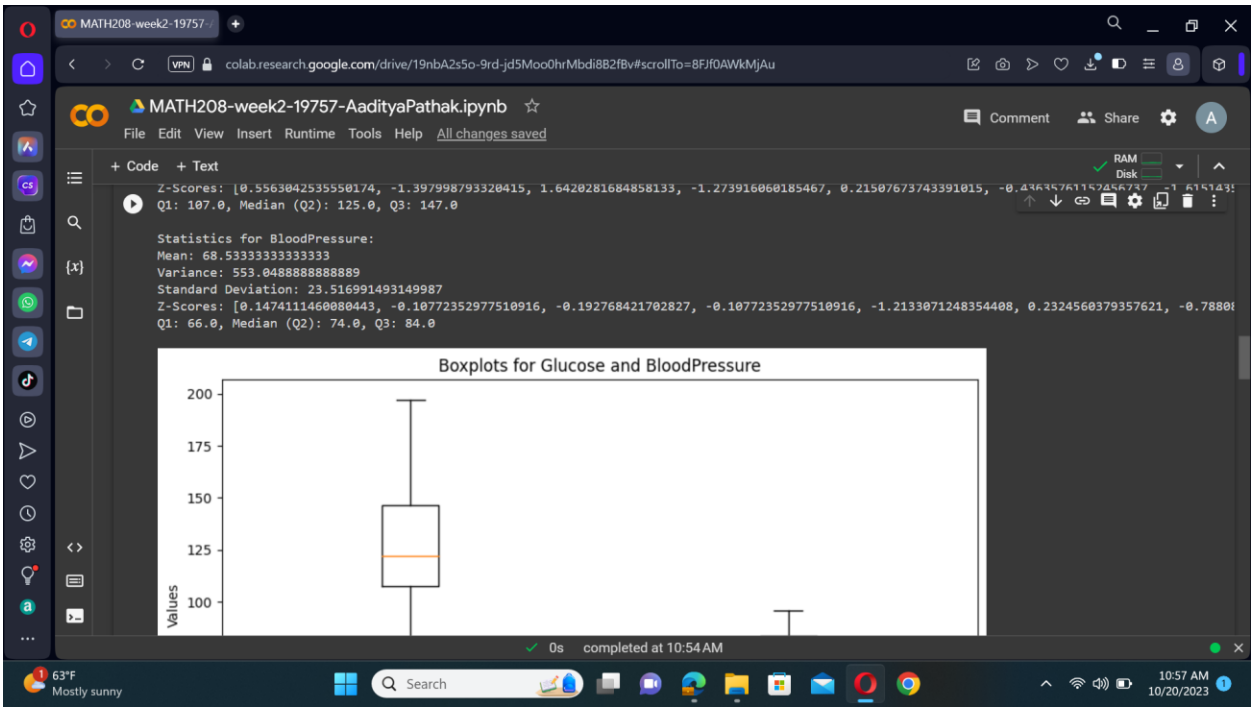
1.Write the program in any computer language to read-in the data from the attached file "original_diabetes.xlsx" partially coming from Pima Indian Diabetes in the National Institute of Diabetes and Digestive and Kidney Diseases1. Find the means, variances,standard deviations, z scores and the values of $Q1$, x(median) and $Q3$ of "Glucose" and "BloodPressure" by user-defined functions rather than calling existing functions in the libraries. After that, please plot the boxplots of both variables in one frame.

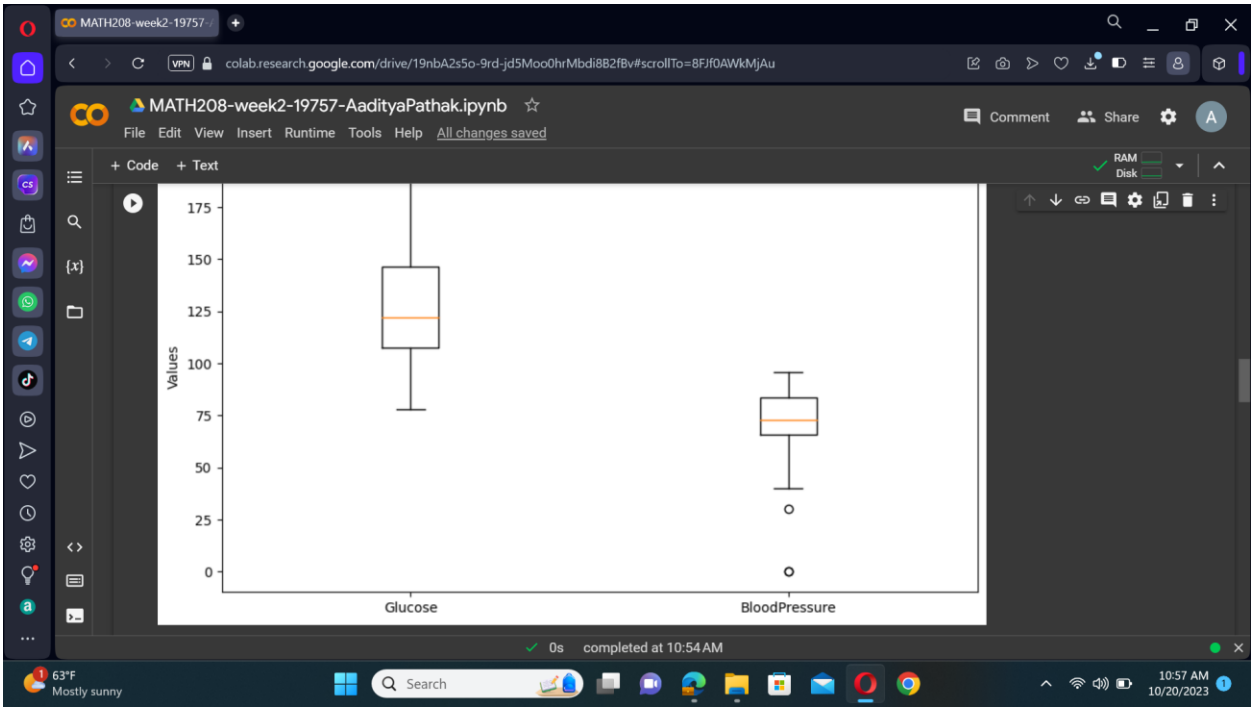





MATH208-week2-19757-
colab.research.google.com/drive/19nbA2s5o-9rd-jd5Moo0hrMbdi8B2fBv#scrollTo=8FJf0AWkMjAu
MATH208-week2-19757-AadityaPathak.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
q1, median, q3 = calculate_quartiles(glucose_data)
# Calculating statistics for "BloodPressure"
bp_mean = calculate_mean(blood_pressure_data)
bp_variance = calculate_variance(blood_pressure_data)
bp_std_dev = calculate_std_dev(blood_pressure_data)
bp_z_scores = calculate_z_scores(blood_pressure_data)
bp_q1, bp_median, bp_q3 = calculate_quartiles(blood_pressure_data)
# Print the calculated statistics for "Glucose"
print('Statistics for Glucose:')
print(f"Mean: {glucose_mean}")
print(f"Variance: {glucose_variance}")
print(f"Standard Deviation: {glucose_std_dev}")
print(f"Z-Scores: {glucose_z_scores}")
print(f"Q1: {q1}, Median (Q2): {median}, Q3: {q3}")
print()
# Print the calculated statistics for "BloodPressure"
print("Statistics for BloodPressure:")
print(f"Mean: {bp_mean}")
print(f"Variance: {bp_variance}")
print(f"Standard Deviation: {bp_std_dev}")
print(f"Z-Scores: {bp_z_scores}")
0s completed at 10:54 AM
63°F
Mostly sunny
Search
10:57 AM
10/20/2023


MATH208-week2-19757-
colab.research.google.com/drive/19nbA2s5o-9rd-jd5Moo0hrMbdi8B2fBv#scrollTo=8FJf0AWkMjAu
MATH208-week2-19757-AadityaPathak.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
print(f"Variance: {bp_variance}")
print(f"Standard Deviation: {bp_std_dev}")
print(f"Z-Scores: {bp_z_scores}")
print(f"Q1: {bp_q1}, Median (Q2): {bp_median}, Q3: {bp_q3}")
print()
# Plotting boxplots for "Glucose" and "BloodPressure" in one frame
plt.figure(figsize=(10, 6))
plt.boxplot([glucose_data, blood_pressure_data], labels=["Glucose", "BloodPressure"])
plt.title("Boxplots for Glucose and BloodPressure")
plt.ylabel("Values")
plt.show()
Statistics for Glucose:
Mean: 130.06666666666666
Variance: 1039.1955555555553
Standard Deviation: 32.23655619875602
Z-Scores: [0.5563042535550174, -1.397998793320415, 1.6420281684858133, -1.273916060185467, 0.21507673743391015, -0.43635761152456737, -1.615143
Q1: 107.0, Median (Q2): 125.0, Q3: 147.0
Statistics for BloodPressure:
Mean: 68.53333333333333
Variance: 553.0488888888889
Standard Deviation: 23.516991493149987
0s completed at 10:54 AM
63°F
Mostly sunny
Search
10:57 AM
10/20/2023


MATH208-week2-19757-AadityaPathak.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
Z-Scores: [0.5563042535550174, -1.3979987933204 15, 1.6420281684858133, -1.273916060185467, 0.21507673743391015,
Q1: 107.0, Median (Q2): 125.0, Q3: 147.0
Statistics for BloodPressure:
Mean: 68.53333333333333
Variance: 553.0488888888889
Standard Deviation: 23.516991493149987
Z-Scores: [0.1474111460080443, -0.10772352977510916, -0.192768421702827, -0.10772352977510916, -1.2133071248354408, 0.2324560379357621,
Q1: 66.0, Median (Q2): 74.0, Q3: 84.0
Boxplots for Glucose and BloodPressure
Values
200
175
150
125
100
0s completed at 10:54 AM
63°F Mostly sunny
Search
10:57 AM 10/20/2023


MATH208-week2-19757-AadityaPathak.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
Values
175
150
125
100
75
50
25
0
Glucose
BloodPressure
0s completed at 10:54 AM
63°F Mostly sunny
Search
10:57 AM 10/20/2023

2.Write the program to verify Chebyshev's inequity as follows by 50 random numbers generated from a normal distribution with mean μ = 10 and standard deviation σ =0.5.

$P(-k\sigma < X - \mu < k\sigma) \geq (1 - 1/k^2) or\ P(|X - \mu| \geq k\sigma) \leq 1/k^2$

For instance,

if k = 1, $P(-\sigma < X - \mu < \sigma) = P(\mu - \sigma < X < \mu + \sigma)$

$= how\ many\ random\ numbers\ are\ within\ the\ range\ [\mu - \sigma, \mu + \sigma]$

$Total\ random\ numbers$

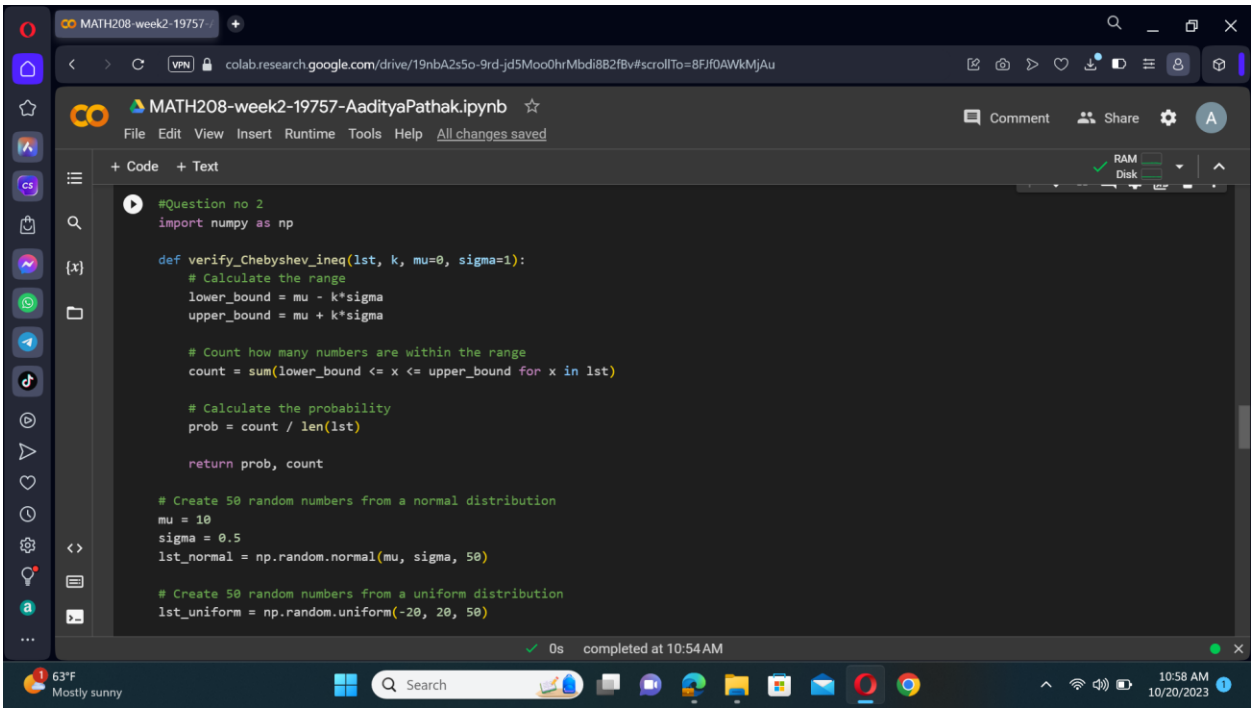
if k = 2, $P(-2\sigma < X - \mu < 2\sigma) = P(\mu - 2\sigma < X < \mu + 2\sigma)$

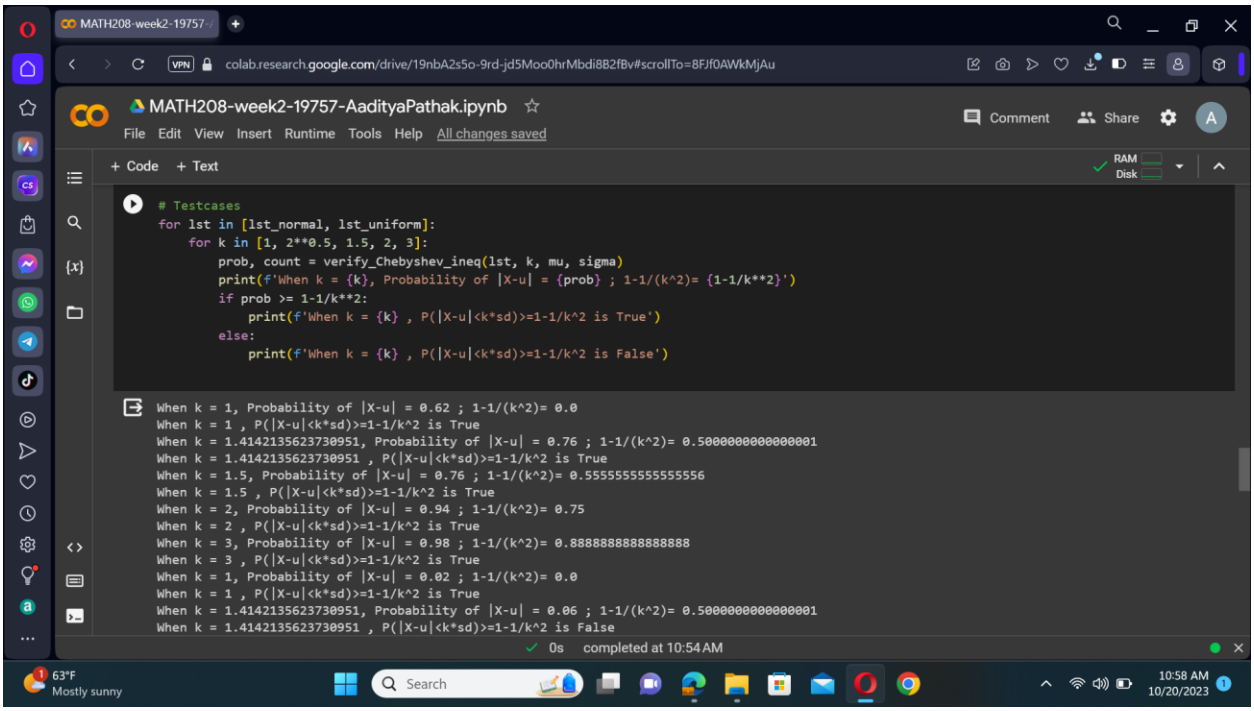
$= how\ many\ random\ numbers\ are\ within\ the\ range\ [\mu - 2\sigma, \mu + 2\sigma]$

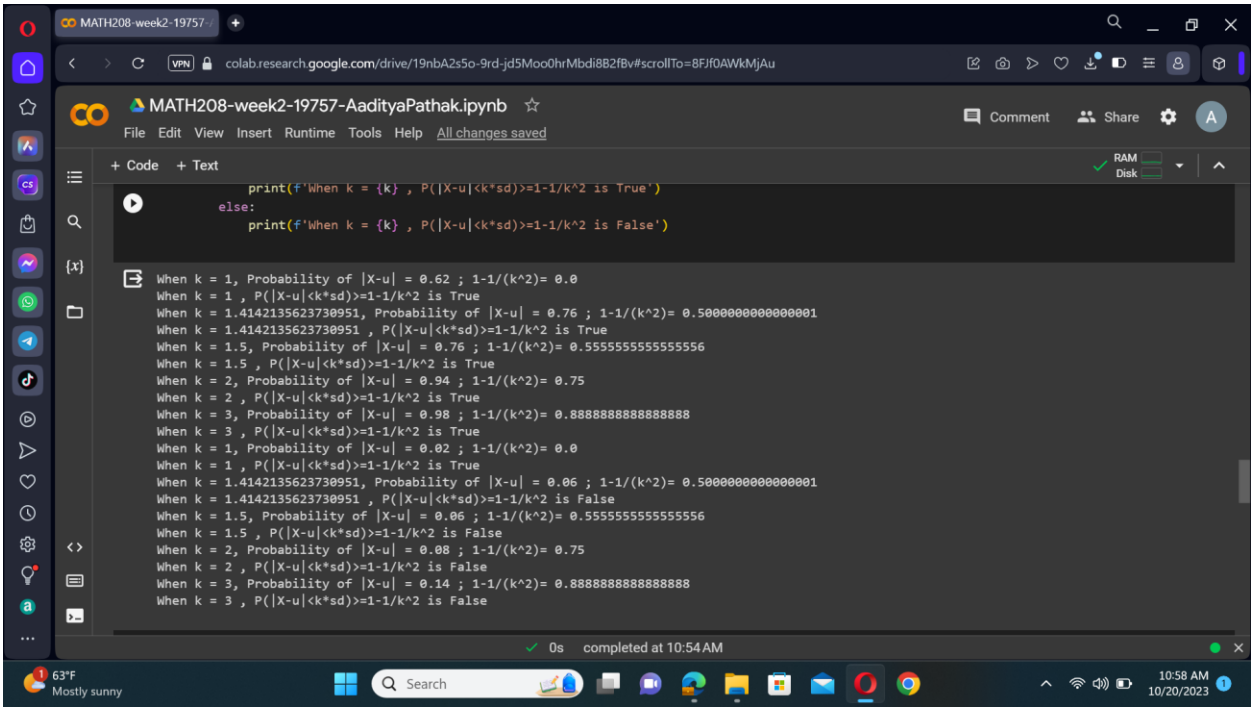
$Total\ random\ numbers$

Repeat the similar process to verify Chebyshev's inequity as well by 50 random

numbers within the range [-20, +20] generated from a uniform distribution.




MATH208-week2-19757-
colab.research.google.com/drive/19nbA2s5o-9rd-jd5Moo0hrMbdi8B2fBv#scrollTo=8FJf0AWkMjAu
MATH208-week2-19757-AadityaPathak.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
# Testcases
for lst in [lst_normal, lst_uniform]:
for k in [1, 2**0.5, 1.5, 2, 3]:
prob, count = verify_Chebyshev_ineq(lst, k, mu, sigma)
print(f'When k = {k}, Probability of |X-u| = {prob} ; 1-1/(k^2)= {1-1/k**2}')
if prob >= 1-1/k**2:
print(f'When k = {k} , P(|X-u|<k*sd)>=1-1/k^2 is True')
else:
print(f'When k = {k} , P(|X-u|<k*sd)>=1-1/k^2 is False')
When k = 1, Probability of |X-u| = 0.62 ; 1-1/(k^2)= 0.0
When k = 1 , P(|X-u|<k*sd)>=1-1/k^2 is True
When k = 1.4142135623730951, Probability of |X-u| = 0.76 ; 1-1/(k^2)= 0.5000000000000001
When k = 1.4142135623730951 , P(|X-u|<k*sd)>=1-1/k^2 is True
When k = 1.5, Probability of |X-u| = 0.76 ; 1-1/(k^2)= 0.5555555555555556
When k = 1.5 , P(|X-u|<k*sd)>=1-1/k^2 is True
When k = 2, Probability of |X-u| = 0.94 ; 1-1/(k^2)= 0.75
When k = 2 , P(|X-u|<k*sd)>=1-1/k^2 is True
When k = 3, Probability of |X-u| = 0.98 ; 1-1/(k^2)= 0.8888888888888888
When k = 3 , P(|X-u|<k*sd)>=1-1/k^2 is True
When k = 1, Probability of |X-u| = 0.02 ; 1-1/(k^2)= 0.0
When k = 1 , P(|X-u|<k*sd)>=1-1/k^2 is True
When k = 1.4142135623730951, Probability of |X-u| = 0.06 ; 1-1/(k^2)= 0.5000000000000001
When k = 1.4142135623730951 , P(|X-u|<k*sd)>=1-1/k^2 is False
0s completed at 10:54 AM
63°F
Mostly sunny
Search
10:58 AM
10/20/2023


MATH208-week2-19757-
colab.research.google.com/drive/19nbA2s5o-9rd-jd5Moo0hrMbdi8B2fBv#scrollTo=8FJf0AWkMjAu
MATH208-week2-19757-AadityaPathak.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
print(f'When k = {k} , P(|X-u|<k*sd)>=1-1/k^2 is True')
else:
print(f'When k = {k} , P(|X-u|<k*sd)>=1-1/k^2 is False')
When k = 1, Probability of |X-u| = 0.62 ; 1-1/(k^2)= 0.0
When k = 1 , P(|X-u|<k*sd)>=1-1/k^2 is True
When k = 1.4142135623730951, Probability of |X-u| = 0.76 ; 1-1/(k^2)= 0.5000000000000001
When k = 1.4142135623730951 , P(|X-u|<k*sd)>=1-1/k^2 is True
When k = 1.5, Probability of |X-u| = 0.76 ; 1-1/(k^2)= 0.5555555555555556
When k = 1.5 , P(|X-u|<k*sd)>=1-1/k^2 is True
When k = 2, Probability of |X-u| = 0.94 ; 1-1/(k^2)= 0.75
When k = 2 , P(|X-u|<k*sd)>=1-1/k^2 is True
When k = 3, Probability of |X-u| = 0.98 ; 1-1/(k^2)= 0.8888888888888888
When k = 3 , P(|X-u|<k*sd)>=1-1/k^2 is True
When k = 1, Probability of |X-u| = 0.02 ; 1-1/(k^2)= 0.0
When k = 1 , P(|X-u|<k*sd)>=1-1/k^2 is True
When k = 1.4142135623730951, Probability of |X-u| = 0.06 ; 1-1/(k^2)= 0.5000000000000001
When k = 1.4142135623730951 , P(|X-u|<k*sd)>=1-1/k^2 is False
When k = 1.5, Probability of |X-u| = 0.06 ; 1-1/(k^2)= 0.5555555555555556
When k = 1.5 , P(|X-u|<k*sd)>=1-1/k^2 is False
When k = 2, Probability of |X-u| = 0.08 ; 1-1/(k^2)= 0.75
When k = 2 , P(|X-u|<k*sd)>=1-1/k^2 is False
When k = 3, Probability of |X-u| = 0.14 ; 1-1/(k^2)= 0.8888888888888888
When k = 3 , P(|X-u|<k*sd)>=1-1/k^2 is False
0s completed at 10:54 AM
63°F
Mostly sunny
Search
10:58 AM
10/20/2023

3.Given the following dataset, write the program to fit it by linear regression showing the values of $b1$, $b0$ and coefficient of linear correlation r. After that, please plot the curve of X vs Y and straight fitting line. Can we draw the conclusion that linear model is good for the dataset if the value of r is very close to +1? Suggest which fitting model should be better than linear based on the data visualization of the given dataset.

X Y

2 30

3 25

4 95

5 115

6 265

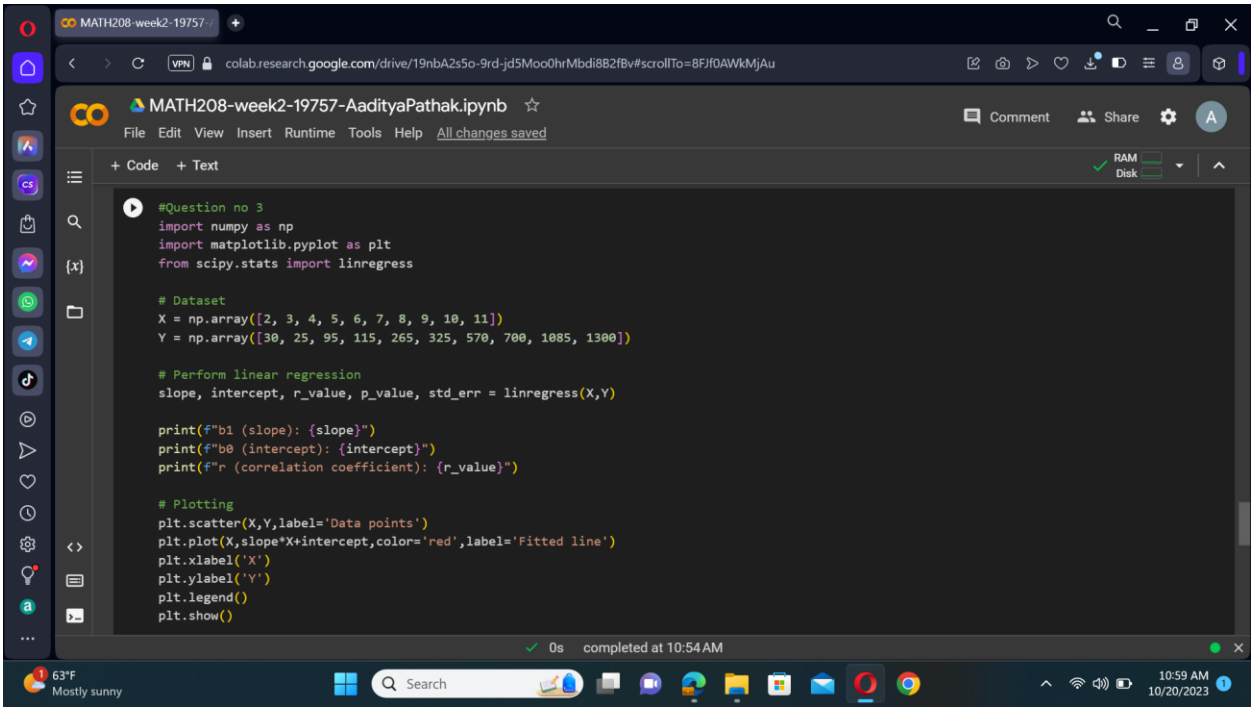
7 325

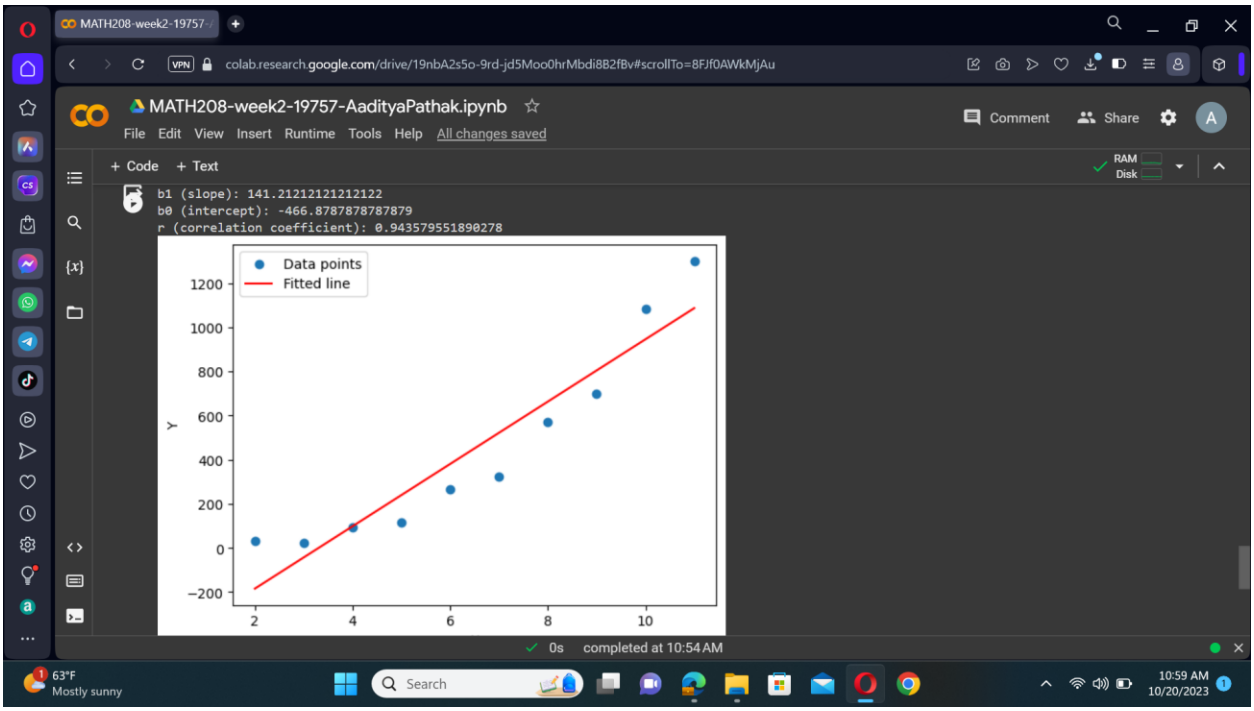
8 570

9 700

10 1085

11 1300




MATH208-week2-19757-AadityaPathak.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
b1 (slope): 141.21212121212122
b0 (intercept): -466.8787878787879
r (correlation coefficient): 0.943579551890278
Data points
Fitted line
y
1200
1000
800
600
400
200
0
−200
2
4
6
8
10
0s completed at 10:54 AM
63°F
Mostly sunny
Search
10:59 AM
10/20/2023