A blue and white logo

Description automatically generated **San Francisco Bay University**

**MATH208 - Probability and Statistics**

**2023 Fall Homework #1**

**Due day: 9/30/2023**

**Instruction:**

1. **Homework answer sheet should contain the original questions and corresponding answers.**
2. **Answer sheet must be in PDF file format with Github links for the programming questions, but MS Word file can’t be accepted. As follows is the answer sheet name format.**

***<course\_id>\_week<week\_number>\_StudentID\_FirstName\_LastName.pdf***

1. **The program name in Github must follow the format like   
   *<course\_id>\_week<week\_number>\_q<question\_number>\_StudentID\_FirstName\_LastName***
2. **If the calculation in Excel is needed, the original file must be provided.**
3. **Show screenshot of all running results, including the system date/time.**
4. **The calculation process must be printed if needed, handwriting can’t be accepted.**
5. **Only accept homework submission uploaded via Canvas.**
6. **Overdue homework submission can’t be accepted.**

**3. Takes academic honesty and integrity seriously (Zero Tolerance of Cheating & Plagiarism)**

**For the students in Engineering School**

1. Write the program in any computer language, Python preferred to create 500 random numbers from -20 to +20 in uniform distribution and find the mean, median and standard deviation. After that, plot the histogram with 10 bins. Notice that the only user defined function can be used to calculate the mean, median and standard deviation, don’t directly call existing function from Python library.

1. Similar to the above, write the program to create 500 random numbers with mean = 10 and standard deviation = 0.5 in Gaussian distribution and find the mean, median and standard deviation. After that, plot the histogram with 10 bins. Notice that the only user defined function can be used to calculate the mean, median and standard deviation, don’t directly call existing function from Python library.
2. The 10-leading causes of death in the United States during 2006 were listed on the Centers for Disease Control and Prevention website. There are a total of 1,855,610 deaths recorded. Plot the Pareto chart in Python or Excel and explain your results.

|  |  |
| --- | --- |
| Cause of Death | Number (x 10,000) |
| Alzheimer’s | 7.2 |
| Chronic Respiratory Disease | 12.5 |
| Diabetes | 7.2 |
| Heart Disease | 63.2 |
| Influenza/Pneumonia | 5.6 |
| Malignant Neoplasms | 56.0 |
| Accidents | 12.2 |
| Nephritis/Nephrosis | 4.5 |
| Septicemia | 3.4 |
| Stroke | 13.7 |

1. The following data are the ages of 118 known offenders who committed an auto theft last year in Garden City, Michigan. Write the program to find the median, the mode, Q1 and Q3, P10 and P95.

11 14 15 15 16 16 17 18 19 21 25 36

12 14 15 15 16 16 17 18 19 21 25 39

13 14 15 15 16 17 17 18 20 22 26 43

13 14 15 15 16 17 17 18 20 22 26 46

13 14 15 16 16 17 17 18 20 22 27 50

13 14 15 16 16 17 17 19 20 23 27 54

13 14 15 16 16 17 18 19 20 23 29 59

13 15 15 16 16 17 18 19 20 23 30 67

14 15 15 16 16 17 18 19 21 24 31

14 15 15 16 16 17 18 19 21 24 34

**For the students in Business School**

1. Given the following dataset, try to find the mean, median and standard deviation in Excel and plot the histogram with 10 bins. The Excel lab manual is available as reference.

-4.18989434, -16.39625703, -3.64263082, 1.07083339,

10.19357629, 6.36652527, 6.69708917, 15.10301644,

-15.37275945, 0.96502973, -7.66820899, -10.83614889,

-11.18807869, 6.98246507, -8.59731822, 10.81493784,

-5.03634343, -15.44139705, -4.69402281, -8.45440955,

-10.50903958, -14.75280713, -12.04994154, -14.41702196,

-4.90869898, -2.51146482, 16.69386058, 16.10521476,

-8.0111506 , -0.52341855, 16.7660631 , 8.81572916,

-14.69540149, -16.91979764, -19.443883 , 18.55418315,

17.38571045, 2.73447136, -13.85297921, 15.67017262,

3.14875289, 3.79776751, -0.35373847, 11.7425206 ,

-13.56059579, -16.49553125, -17.7148762 , -2.09348509,

-16.89094231, 3.06873144, -6.97146231, 19.73717365,

-3.97369338, 3.56939124, -2.26756128, -15.28950097,

-2.30605174, 11.87653503, 0.63952388, 11.64489851,

17.02832979, 19.15737457, 13.31233543, 11.9539493 ,

-13.49808171, 11.2898073 , -8.90318926, 8.9881454 ,

-18.23322783, 0.59265736, 7.09487351, -9.66020789,

19.58989269, -4.83636985, -16.75161561, -14.12835361,

0.30150355, 2.87156167, 5.60830776, -14.36093008,

17.54046228, -8.05631402, 18.34368584, -1.77490521,

16.07313921, 14.12119943, 0.13228049, 1.92983826,

14.85573007, -17.72362047, 16.14025917, -19.47611883,

-2.04534322, 19.82706782, -11.02116676, -14.48463378,

19.46822451, -5.07537173, 8.16836663, 19.39447016

1. Similar to the above, try to find the mean, median and standard deviation in Excel for the following dataset and plot the histogram with 10 bins. The Excel lab manual is available as reference.

9.4213798 , 10.2301027 , 9.76534516, 11.36808598, 10.0779665 ,

10.1714552 , 9.79452853, 9.28774604, 10.20323265, 9.93234335,

9.61064104, 10.42219556, 10.02687956, 9.62984905, 10.08933524,

9.64629745, 9.97147114, 9.92061038, 10.01794133, 9.50242332,

9.0496859 , 9.73396068, 9.83737981, 9.98355996, 10.26274408,

10.28893084, 10.39075768, 9.92847436, 11.01890017, 9.28424041,

10.45679792, 10.19666782, 10.86064153, 10.49657478, 10.03656705,

10.02326243, 11.1044741 , 10.14146821, 10.60541338, 9.00568628,

10.10099188, 9.93017714, 9.14624405, 9.16308994, 9.55799541,

9.97748609, 10.12256364, 9.53857323, 10.2434414 , 10.60253704,

10.15445592, 9.10046222, 11.02864311, 10.02058428, 10.54058978,

10.52428045, 9.99690432, 9.65926343, 9.46544324, 10.38825349,

9.40089265, 10.00759389, 10.86663891, 9.59797556, 10.66093283,

9.77758696, 8.67864203, 10.72460343, 9.73610607, 10.3375458 ,

9.50523926, 10.19263198, 10.5701187 , 9.08132523, 9.48877511,

9.78204876, 10.02626177, 9.9532261 , 10.06037811, 9.97986134,

9.78875599, 10.33006754, 10.05705824, 10.55695896, 10.59390069,

9.53997908, 9.46284674, 9.80203652, 10.99484118, 9.82861927,

10.02241117, 8.66347508, 9.65586799, 9.26305778, 9.3854242 ,

10.70342391, 9.91188565, 9.91219612, 10.4657285 , 10.24439263

1. The 10-leading causes of death in the United States during 2006 were listed on the Centers for Disease Control and Prevention website. There are a total of 1,855,610 deaths recorded. Plot the Pareto chart in Excel and explain your results.

|  |  |
| --- | --- |
| Cause of Death | Number (x 10,000) |
| Alzheimer’s | 7.2 |
| Chronic Respiratory Disease | 12.5 |
| Diabetes | 7.2 |
| Heart Disease | 63.2 |
| Influenza/Pneumonia | 5.6 |
| Malignant Neoplasms | 56.0 |
| Accidents | 12.2 |
| Nephritis/Nephrosis | 4.5 |
| Septicemia | 3.4 |
| Stroke | 13.7 |

1. The following data are the ages of 118 known offenders who committed an auto theft last year in Garden City, Michigan. Find the median, the mode, Q1 and Q3, P10 and P95 in Excel.

11 14 15 15 16 16 17 18 19 21 25 36

12 14 15 15 16 16 17 18 19 21 25 39

13 14 15 15 16 17 17 18 20 22 26 43

13 14 15 15 16 17 17 18 20 22 26 46

13 14 15 16 16 17 17 18 20 22 27 50

13 14 15 16 16 17 17 19 20 23 27 54

13 14 15 16 16 17 18 19 20 23 29 59

13 15 15 16 16 17 18 19 20 23 30 67

14 15 15 16 16 17 18 19 21 24 31

14 15 15 16 16 17 18 19 21 24 34