EE3801 LAB 1

AY 2023/24

Submitted by

Gulati Shobhit (A0244507H)

Q.1 a) The output for Q.1a is the following:

```
Q.1a The mean, median and sum of the fat present in the body parts for each individual is :

 ID  Mean  Median   Sum
  0 50.37   36.75 503.7
  1 51.08   37.90 510.8
  2 51.00   36.45 510.0
 47 48.22   36.05 482.2
 48 47.79   34.30 477.9
 49 45.13   34.20 451.3
```

Q.1 b) The output for Q.1b is the following:

```
Q.1b The mean, median, and sum of fat for each body part is:

Features    Mean  Median    Sum
    neck  38.016   38.10 1900.8
   chest 101.128  101.10 5056.4
 abdomen  92.080   89.15 4604.0
     hip 102.058  100.45 5102.9
   thigh  61.506   61.60 3075.3
    knee  38.828   38.70 1941.4
   ankle  23.586   23.10 1179.3
  biceps  32.798   32.45 1639.9
 forearm  28.852   29.00 1442.6
   wrist  18.118   18.20  905.9
```

Q.1 c) The following output shows three different types of mean:

```
Q.1c The Results of 3 different types of mean across different bodyparts

Features  Arithmetic Mean  Geometric Mean  Harmonic Mean
    neck          38.016       37.883069      37.755200
   chest         101.128      100.606458     100.106007
 abdomen          92.080       91.065145      90.135062
     hip         102.058      101.577150     101.131966
   thigh          61.506       61.161311      60.827484
    knee          38.828       38.725157      38.625040
   ankle          23.586       23.490255      23.403851
  biceps          32.798       32.598184      32.403743
 forearm          28.852       28.769638      28.685546
   wrist          18.118       18.084399      18.051573
```

Each has its own characteristics and use cases:

1. Arithmetic Mean:
    a. Calculated by adding up all values in a dataset and dividing by the number of values.
    b. This is the most commonly used measure of central tendency.
2. Geometric Mean:
    a. Calculated by taking the N-th root of the product of N values.
    b. It's useful for dealing with values that are subject to exponential growth or decay.
3. Harmonic Mean:
    a. Calculated by taking the reciprocal of the arithmetic mean of the reciprocals of a set of values.
    b. Used in situations where rates or ratios need to be averaged, such as calculating average speeds or average times.

In our case where the three means are relatively close to each other, it may not matter much which one you choose.
However, in our case the dataset is a symmetric distribution and dataset is neither a ratios nor showed exponential growth or decay ; so we are using Arithmetic Mean.

Q2. a) In the following output, the Max Value shows the maximum value in the specified column, and Max ID displays the respective ID of that maximum values. Similarity for the Min Value and Min ID.

```
Q2.a The individuals that have the maximum and minimum fat

 Features  Max Value  Max ID  Min Value  Min ID
  density     1.0911      25     1.0101      35
  bodyfat    40.1000      35     3.7000      25
     neck    51.2000      38    31.5000      44
    chest   136.2000      38    83.4000      49
  abdomen   148.1000      38    70.4000      49
      hip   147.7000      38    85.3000      26
    thigh    87.3000      38    50.0000      44
     knee    49.1000      38    34.4000      49
    ankle    33.9000      30    20.6000      48
   biceps    45.0000      38    26.1000      44
  forearm    32.8000      21    23.1000      44
    wrist    21.4000      38    16.1000      44
```

Q2. b) In the following output, I display the ID of individuals who are appearing more than once under Max ID and Min ID together with their corresponding part of the body.

```
Q.2b The Individuals with Max Id and their features are:

Max ID: 38 - Features : neck, chest, abdomen, hip, thigh, knee, biceps, wrist

Q.2b The Individuals with Min Id and their features are:

Min ID: 44 - Features : neck, thigh, biceps, forearm, wrist
Min ID: 49 - Features : chest, abdomen, knee
```

Q. 3 The following is the output for Q.3, number of individuals in each feature that fall within 10% of standard deviation from its respective mean and median:

```
Q.3 number of individuals in each feature that fall within 10% of standard deviation from its respective mean and median :

Feature  Within 10% of Mean  Within 10% of Median
density                   1                     3
bodyfat                   1                     3
    age                   3                     4
 weight                   7                     8
 height                   3                     5
   neck                   4                     6
  chest                   7                     7
abdomen                   4                     6
    hip                   4                     6
  thigh                   2                     2
   knee                   6                     5
  ankle                   5                     7
 biceps                   5                     6
forearm                   4                     3
  wrist                   5                     7
```

Q. 4 The following output has the number of missing values in every feature:

```
Q.4 Number of Missing Values in Each Feature:

Feature  Missing Values Count
density                     0
bodyfat                     4
    age                     0
 weight                     7
 height                     2
   neck                     3
  chest                     1
abdomen                     0
    hip                     6
  thigh                     3
   knee                     1
  ankle                     2
 biceps                     4
forearm                     0
  wrist                     2
```

Q. 5 a) The output in absolute differences in Mean Values:

```
Q.5a Absolute Differences in Mean Values (bodyfat3b vs. bodyfat2):

density    0.000000
bodyfat    0.536348
age        0.000000
weight     1.001302
height     0.036667
neck       0.139404
chest      0.017796
abdomen    0.000000
hip        0.410182
thigh      0.193234
knee       0.074939
ankle      0.034833
biceps     0.158522
forearm    0.000000
wrist      0.013250
```

Q.5 b) The absolute differences in Median Values:

```
Q.5b Absolute Differences in Median Values (bodyfat3c vs. bodyfat2):

density    0.000
bodyfat    0.850
age        0.000
weight     0.875
height     0.000
neck       0.100
chest      0.200
abdomen    0.000
hip        1.100
thigh      1.500
knee       0.000
ankle      0.000
biceps     0.050
forearm    0.000
wrist      0.000
```

Q.5 c) The Mean Difference vs Median Difference:

Q.5c The Mean Difference vs The Median Difference:

Feature	Mean Difference	Median Difference
density	0.000000	0.000
bodyfat	0.536348	0.850
age	0.000000	0.000
weight	1.001302	0.875
height	0.036667	0.000
neck	0.139404	0.100
chest	0.017796	0.200
abdomen	0.000000	0.000
hip	0.410182	1.100
thigh	0.193234	1.500
knee	0.074939	0.000
ankle	0.034833	0.000
biceps	0.158522	0.050
forearm	0.000000	0.000
wrist	0.013250	0.000

From the above table, we can infer that Mean Difference is less than ~0.5 for almost all features except *weight* column. This means replacing Mean to null values, doesn't affect the Mean that much of the dataset.
The Median Difference have some features that have 0 difference and it is accurate for those features. But for features greater than 1, it will be less accurate.

Q.6 a) The output of following question with top 3 and bottom 3 rows are:

```
Q.6a The top 3 and bottom 3 rows of normalised bodyfat dataframe:
     density   bodyfat       age    weight    height      neck     chest  \
0   0.561613 -0.567159 -1.268721 -0.720556 -0.893031 -0.552493 -0.756337
1   1.217281 -1.197111 -1.387738 -0.248598  0.795473  0.147250 -0.709231
2  -0.767810  0.753706 -1.387738 -0.726766 -1.455866 -1.221814 -0.501964
47  1.271543 -1.247913  0.635551 -0.863386  0.420250 -1.039272 -1.067238
48  0.425958 -0.435073  1.349653 -1.180095 -0.611614 -1.586898 -0.831707
49  1.443374 -1.410481  1.587687 -1.385024 -1.268254 -1.221814 -1.670197

     abdomen       hip     thigh      knee     ankle    biceps   forearm  \
0  -0.474192 -0.722223 -0.373159 -0.527376 -0.742220 -0.215020 -0.666231
1  -0.625823 -0.320882 -0.417831 -0.527376 -0.081882 -0.619192  0.022024
2  -0.288099 -0.273103 -0.283815  0.024850  0.182253 -1.077254 -1.675671
47 -0.867054 -0.894227 -1.311268 -0.458348 -0.742220 -1.077254 -0.941533
48 -0.598254 -1.114009 -1.415503 -1.045088 -1.314513 -1.077254 -1.538020
49 -1.494255 -1.419793 -1.623971 -1.528286 -0.742220 -1.616150 -1.400369

       wrist
0  -0.901998
1   0.072656
2  -1.345022
47 -0.193159
48 -1.610837
49 -1.167813
```

Q.6 b) Number of individuals greater than mean:

```
Q.6b Number of Individuals Greater Than the Respective Feature's Mean:

density    27
bodyfat    23
age        21
weight     25
height     23
neck       26
chest      25
abdomen    21
hip        22
thigh      25
knee       23
ankle      21
biceps     22
forearm    27
wrist      27
dtype: int64
```