Yeo Meng Han

A0251772A

EE3801 Lab 1

**Question 1a**

A screenshot of a black and white screen

Description automatically generated

**Question 1b**

A screenshot of a computer screen

Description automatically generated

**Question 1c**

A screenshot of a computer screen

Description automatically generated

Arithmetic mean: Mean represents the average value of a set of data points. It gives us a sense of the central tendency of the data

Geometric mean: Geometric mean is useful when dealing with quantities that are multiplicatively related, such as growth rates, ratios or compound interest. It is less significantly affected by extreme values compared to the arithmetic mean.

Harmonic mean: Harmonic mean is useful when dealing with rates, such as speed or efficiency. It’s the reciprocal of the arithmetic mean of the reciprocals of the data. The harmonic mean is influenced more by smaller values, skewing the harmonic mean to be lower than arithmetic mean.

Working with body measurements, which might involve physical sizes and proportions. Considering that body measurements often have multiplicative relationships (for example, doubling a body part might imply a different significance compared to a linear increase) and our objective is to study the relative changes or growth rates, the geometric mean could be a suitable choice.

However, if we’re interested in understanding the overall distribution of body measurements and hope to consider extreme values (which naturally occurs), the arithmetic mean will be the most suitable.

**Question 2a**

A screenshot of a computer screen

Description automatically generated

**Question 2b**

**A screenshot of a computer

Description automatically generated**

**Question 3**

A screenshot of a computer

Description automatically generated

**Question 4**

A computer screen with white text

Description automatically generated

**Question 5a**

**A black screen with white text

Description automatically generated**

**Question 5b**

A screen shot of a computer

Description automatically generated

**Question 5c**

The absolute difference values give us insights into how much the imputed values differ from the original values in the bodyfat2 dataset. A lower absolute difference indicates that the imputed values are closely aligned with the original values in terms of their central tendency – accurately estimating the missing value, resulting in more representative approximation.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mean | Median | Comparison |
| Density | 0.000000 | 0.000 | No difference |
| Bodyfat | 0.536348 | 0.850 | Mean |
| Age | 0.000000 | 0.000 | No difference |
| Weight | 1.001302 | 0.875 | Median |
| Height | 0.036667 | 0.000 | Not significant difference |
| Neck | 0.139404 | 0.100 | Median |
| Chest | 0.017796 | 0.200 | Mean |
| Abdomen | 0.000000 | 0.000 | No difference |
| Hip | 0.410182 | 1.100 | Mean |
| Thigh | 0.193234 | 1.500 | Mean |
| Knee | 0.074939 | 0.000 | Not significant difference |
| Ankle | 0.034833 | 0.000 | Not significant difference |
| Biceps | 0.158522 | 0.050 | Median |
| Forearm | 0.000000 | 0.000 | No difference |
| Wrist | 0.013250 | 0.000 | Not significant difference |

From the table above, imputation using means appears to be more accurate and closer to original results. It can also be inferred that median imputation is more useful and provide more accuracy for features where data distribution may be skewed or has extreme outliers – this may tell us that within the dataset, there’s significant outliers for weight, neck and biceps. We can also infer that the data of bodyfat, chest, hip and thighs generally follow normal distribution and there’s no significant outliers as imputed means provided a closer means of the feature’s data to the original results.

**Question 6a**

A black background with white text

Description automatically generated

**Question 6b**

A screenshot of a computer

Description automatically generated