Transformations Exercise

CMSC320

Consider data for variable $\mathbf{x} = x_1, x_2, \ldots, x_n$. We use $\overline{x}$ to denote the sample mean of $\mathbf{x}$, and s_x is the sample standard deviation of $\mathbf{x}$.

Part I

For each of the following three transformations derive (a) the sample mean $\overline{z}$, and (b) the sample standard deviation s_z.

1. Centering

$$z_i = (x_i - \overline{x})$$

2. Scaling

$$z_i = \frac{x_i}{s_x}$$

3. Centering and scaling (standardizing)

$$z_i = \frac{(x_i - \overline{x})}{s_x}$$

Part II

4. Consider transformation $z_i = \log x_i$. Show that the sample mean $\overline{z}$ equals the logarithm of the geometric sample mean of the original data x_i.

Note: The sample mean we use most commonly is the *arithmetic* mean ($\overline{x} = \frac{1}{n} \sum_i x_i$). For strictly positive data, especially where there is skew, the *geometric* mean is a better summary of central trend. It is defined as:

$$\mathrm{gm}(\mathbf{x}) = \left(\prod_i x_i \right)^{1/n}$$

So, your problem is to show that $\overline{z} = \log \mathrm{gm}(\mathbf{x})$.