Topic: Exploratory Data Analysis (EDA)

Linear Transformations

School of Mathematics and Applied Statistics

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

Linear Transformations

Ethics

Nature of the question to be answered

(Transforms – data from different perspectives)

Context/ Expertise

Design: Experiments Vs Observation

  Sampling

  Measurement

Description and Analysis

Conclusions & Decision Making

VARIATION

Activity: Transformation of Measurement Units

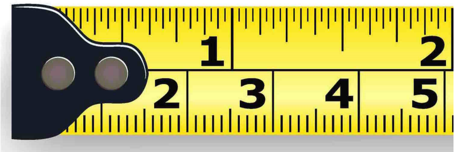
A sample of data was collected from students: an excerpt is shown

Height	Shoe Size	Sex
153	8	f
6’	11	m
180~~cm~~	15	m
170~~cm~~	9.5	m

Discuss:

- What do you notice? 6' = 6 feet.
- How could you fix the problem? convert 6' (feet) into cm.
- What do you need to know to fix the problem? convert feet → inches. inches → cm

Activity: Transforming Height



We know: 1 inch = 2.54 cm and 1 foot = 12 inches

6 feet = 6 x 12 = 72 inches

= 72 x 2.54 = 182.9 cm

This is an example of a linear transformation or a rescaling of measurements

Transforming Data: Mean

Discuss: What happens to the mean if each data value x_i is rescaled by a **linear transformation** $a + bx_i$?

Example:

- Values of x: $\{1, 2, 3\}$, then $\bar{x} = 2$
- Let $y_i = a + bx_i$ with $a = 10$ and $b = 3$

  then values of y:
  $\{10 + 3 \times 1, 10 + 3 \times 2, 10 + 3 \times 3\} = \{13, 16, 19\}$,

  and $\bar{y} = \dfrac{13 + 16 + 19}{3} = \dfrac{48}{3} = 16.$

$$\bar{y} = a + b\bar{x} = 10 + 3 \times 2 = 10 + 6 = \underline{\underline{16}}$$

So, the mean is rescaled in the same way as the values of x

Rescaling Data: Prove mathematically for mean

If each data value x_i, $i = 1 \dots n$ is **rescaled by a linear transformation** such that $y_i = a + bx_i$, show that $\bar{y} = a + b\bar{x}$

$y_i = a + bx_i$

start $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$

$= \frac{1}{n} \sum_{i=1}^{n} (a + bx_i)$

$= \frac{1}{n} \left[na + b \sum_{i=1}^{n} x_i \right]$

$= a + b \times \boxed{\frac{1}{n} \sum_{i=1}^{n} x_i}$

$\bar{y} = a + b\bar{x}$ ✓

Rescaling Data: Properties of standard deviation

- What happens to the standard deviation when a constant is added to or subtracted from each data value?
- Eg. the standard deviation of $\{2, 4, 6\}$ compared with the standard deviation of $\{1, 3, 5\}$ or $\{8, 10, 12\}$.

Rescaling Data: Properties of standard deviation

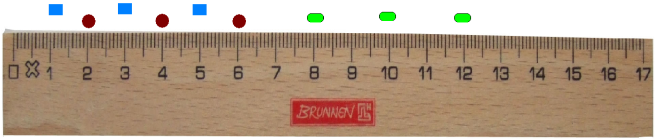
- What happens to the standard deviation when a constant is added to or subtracted from each data value?
- Eg. the standard deviation of $\{2, 4, 6\}$ compared with the standard deviation of $\{1, 3, 5\}$ or $\{8, 10, 12\}$.

Rescaling Data: Properties of standard deviation

- What happens to the standard deviation when a constant is added to or subtracted from each data value?
- Eg. the standard deviation of $\{2, 4, 6\}$ compared with the standard deviation of $\{1, 3, 5\}$ or $\{8, 10, 12\}$.

Rescaling Data: Properties of standard deviation

- What happens to the standard deviation when a constant is added to or subtracted from each data value?
- Eg. the standard deviation of $\{2, 4, 6\}$ compared with the standard deviation of $\{1, 3, 5\}$ or $\{8, 10, 12\}$.



Rescaling Data: Properties of standard deviation

- What happens to the standard deviation when a constant is added to or subtracted from each data value?
- Eg. the standard deviation of $\{2, 4, 6\}$ compared with the standard deviation of $\{1, 3, 5\}$ or $\{8, 10, 12\}$.

- The standard deviation of $\{2, 4, 6\}$ is the same as the standard deviation of $\{1, 3, 5\}$ or $\{8, 10, 12\}$.
- Standard deviation is unaffected when a constant is added to or subtracted from each data value.
- Prove mathematically...

Rescaling Data: Properties of standard deviation

- What happens to the standard deviation when each data value is multiplied by a constant c?
- Eg. if data values $\{1, 3, 5\}$ are multiplied by 3: $\{3, 9, 15\}$.

Rescaling Data: Properties of standard deviation

- What happens to the standard deviation when each data value is multiplied by a constant c?
- Eg. if data values $\{1, 3, 5\}$ are multiplied by 3: $\{3, 9, 15\}$.

Rescaling Data: Properties of standard deviation

- What happens to the standard deviation when each data value is multiplied by a constant c?
- Eg. if data values $\{1, 3, 5\}$ are multiplied by 3: $\{3, 9, 15\}$.

Rescaling Data: Properties of standard deviation

- What happens to the standard deviation when each data value is multiplied by a constant c?
- Eg. if data values $\{1, 3, 5\}$ are multiplied by 3: $\{3, 9, 15\}$.

- For example, the standard deviation of $\{1, 3, 5\}$ is 2, the standard deviation of $\{3, 9, 15\}$ is 6.

Rescaling Data: Properties of standard deviation

- What happens to the standard deviation when each data value is multiplied by a constant c?
- Eg. if data values $\{1, 3, 5\}$ are multiplied by 3: $\{3, 9, 15\}$.

- For example, the standard deviation of $\{1, 3, 5\}$ is 2, the standard deviation of $\{3, 9, 15\}$ is 6.
- In fact, the standard deviation is multiplied by the constant $|c|$

$$s_{\text{new}} = |c| \times s_{\text{old}}$$

- Prove mathematically...

Rescaling Data: Prove mathematically for sd

If each data value x_i, $i = 1 \ldots n$ is **rescaled by a linear transformation** such that $y_i = a + bx_i$, show that $s_y = \sqrt{b^2 s_x^2} = |b| s_x$

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad \bar{y} = a + b\bar{x}$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left[\cancel{a} + bx_i - \cancel{a} - b\bar{x}\right]^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left[b(x_i - \bar{x})\right]^2$$

$$= b^2 \times \boxed{\frac{1}{n-1}\sum(x_i - \bar{x})^2}$$

$$s_y^2 = b^2 \times s_x^2$$

$$s_y = +\sqrt{b^2 s_x^2}$$

$$= |b| \times s_x$$

✓ as required.

$s_x \geq 0$

$s_y \geq 0$.