Linear Transformations

Notes

If x represents a variable, then a **linear transformation** is an equation of the form:

$$y = a + bx$$

where a, b are some numbers. For instance $y = x + 10$, or $y = 2x - 1$.

We think of y here as a *new variable*, and the equation tells us how to convert values of the one variable into values of the other.

Examples:

- x is a student's height measured in "inches from 5 feet". y is their actual height in inches. Then:
$$y = x + 60$$
- x is a student's height in inches, y is their height in meters. 1 meter is 39.37 inches. So:
$$y = x/39.37 = \frac{1}{39.37}x$$
- x is a room's temperatures measured in F, y is the same temperatures in C. The relation between Fahrenheit and Celsius degrees is:
$$F = 1.8C + 32$$
So:
$$y = \frac{x - 32}{1.8} = \frac{x}{1.8} - \frac{32}{1.8} = \frac{1}{1.8}x - \frac{32}{1.8}$$
A linear transformation between two variables tells us how individual values transform to each other.

So for instance if we had the temperature in Fahrenheit, $x = 56$, then we can find the corresponding temperature in Celsius:

$$y = \frac{1}{1.8} \times 56 - \frac{32}{1.8} = 13.333$$

But how measures of center or spread behave requires more thinking!

Behavior of variables under linear transformation

Assume $y = a + bx$. Then we can observe the following relation in the properties between x and y.

shape stays the same (modes, skewness, outliers)

center Follows the same transformation (mean, median do that)

e.g. $\bar{y} = a + b\bar{x}$

spread Only follows the multiplier (std. dev., IQR do that)

e.g. $s_y = bs_x$.

Practice: If some temperatures have a mean of 67 degrees F, and standard deviation of 5 degrees F, how would the corresponding temperatures in C behave?

A special case: Standardized scores

Standardized scores, also called z-scores, are given by the following linear transformation:

$$z = \frac{x - \bar{x}}{s_x}$$

Alternatively, they relate to x via:

$$x = \bar{x} + s_x z$$

Key properties:

- Think of z-score as saying "how many standard deviations away from the mean you are".
- z-scores are unitless.
- z-scores always have a mean of 0 and a standard deviation of 1.
- They are great for comparing quantities that are measured in different scales.

Practice In the Behavioral Survey data we have examined, let us consider as x the height variable. The variable has a mean $\bar{x} = 67.18$ and a standard deviation $s_x = 4.126$.

1. The maximum height was 93. What z-score corresponds to that?
2. We would consider a typical range of heights anything with a z-score between -2 and 2. What heights would that cover? Convert the zs to xs to answer this.