PSY 201: Statistics in Psychology

Lecture 06

Variability

How to make IQ scores look good.

Greg Francis

Purdue University

Fall 2019

DESCRIPTION

- central tendency gives an indication of where most, many, or average, scores are
- also want some idea of how much variability exists in a distribution of scores
  - range
  - mean deviation
  - variance
  - standard deviation

RANGE

- Highest score - lowest score

Name	Sex	Score
Greg	Male	95
Ian	Male	89
Aimeé	Female	94
Jim	Male	92

- $95 - 89 = 6$

PROBLEM

- range is **very** sensitive to “extreme” scores

Name	Sex	Score
Greg	Male	95
Ian	Male	89
Aimeé	Female	94
Jim	Male	92
Bob	Male	32

- 95 - 32 = 63
- one score makes a big difference!

MEAN DEVIATION

- we can decrease sensitivity to extreme scores by considering deviations from a measure of central tendency
- a deviation score is

$$x_i = X_i - \overline{X}$$

- we define the mean deviation as:

$$\text{MD} = \frac{\Sigma|X_i - \overline{X}|}{n} = \frac{\Sigma|x_i|}{n}$$

- where $|x_i|$ means: "absolute value of x_i"
- why do we take the absolute value instead of just summing deviations?

VARIANCE

- mean deviation turns out to be mathematically messy
- squaring also removes minus signs!
- sum of squares

$$SS = \Sigma(X_i - \overline{X})^2 = \Sigma(x_i)^2$$

- variance is the average sum of squares
- calculation depends on whether scores are from a population or a sample

POPULATION

- a population includes **all** members of a specified group
- variance is defined as:

$$\sigma^2 = \frac{SS}{N} = \frac{\Sigma(X_i - \mu)^2}{N} = \frac{\Sigma(x_i)^2}{N}$$

- where
  - μ is the mean of the population
  - N is the number of scores in the population

SAMPLE

- a sample includes a **subset** of scores from a population
- variance is defined as:

$$s^2 = \frac{SS}{n-1} = \frac{\Sigma(X_i - \overline{X})^2}{n-1} = \frac{\Sigma(x_i)^2}{n-1}$$

- where
  - $\overline{X}$ is the mean of the sample
  - n is the number of scores in the sample
- why the differences? Don't worry for now. Just know the calculations.

SAMPLE VARIANCE

- deviation formula:

$$s^2 = \frac{\Sigma(x_i)^2}{n-1}$$

- alternative (but equivalent) calculation is the raw score formula

$$s^2 = \frac{SS}{n-1} = \frac{\Sigma(X_i)^2 - \left[(\Sigma X_i)^2 / n\right]}{n-1}$$

- use whichever formula is simpler!

EXAMPLE

Name	Sex	Score
Greg	Male	95
Ian	Male	89
Aimeé	Female	94
Jim	Male	92

- since we have the raw scores, we use the raw score formula (we assume a sample)

$$s^2 = \frac{SS}{n-1} = \frac{\Sigma(X_i)^2 - \left[(\Sigma X_i)^2 / n\right]}{n-1}$$

$$\Sigma X_i^2 = (95)^2 + (89)^2 + (94)^2 + (92)^2 = 34246$$

$$(\Sigma X_i)^2/n = (95+89+94+92)^2/4 = \frac{(370)^2}{4} = \frac{136900}{4} = 34225$$

- so,

$$s^2 = \frac{34246 - 34225}{3} = \frac{21}{3} = 7$$

SUM OF SQUARES

- earlier we calculated the squared deviation from the mean

$$\sum x_i^2 = \sum (X_i - \overline{X})^2$$

$$= (95 - 92.5)^2 + (89 - 92.5)^2 + (94 - 92.5)^2 + (92 - 92.5)^2$$

$$= (2.5)^2 + (-3.5)^2 + (1.5)^2 + (-0.5)^2 = 0$$

$$= 6.25 + 12.25 + 2.25 + 0.25 = 21.0$$

- we can use that to calculate variance with the deviation score formula:

$$s^2 = \frac{\Sigma x_i^2}{n - 1} = \frac{21}{3} = 7$$

- Same as before!
- Note! variance cannot be negative

STANDARD DEVIATION

- variance is in **squared** units of measurement
  - distance: squared meters
  - weight: squared kilograms
  - temperature: squared degrees
  - ...
- standard deviation is in the same units as the scores!
- square root of variance

STANDARD DEVIATION

- deviation score formula:

$$s = \sqrt{s^2} = \sqrt{\frac{SS}{(n-1)}} = \sqrt{\frac{\Sigma(x_i)^2}{(n-1)}}$$

- raw score formula:

$$s = \sqrt{s^2} = \sqrt{\frac{SS}{(n-1)}} = \sqrt{\frac{\Sigma(X_i)^2 - \left[\left(\Sigma X_i\right)^2/n\right]}{n-1}}$$

EXAMPLE

Name	Sex	Score
Greg	Male	95
Ian	Male	89
Aimeé	Female	94
Jim	Male	92

- since we have the raw scores, we use the raw score formula to calculate variance

$$s^2 = \frac{SS}{n-1} = \frac{\Sigma(X_i)^2 - \left[(\Sigma X_i)^2 / n\right]}{n-1}$$

- we calculated earlier that the variance equals:

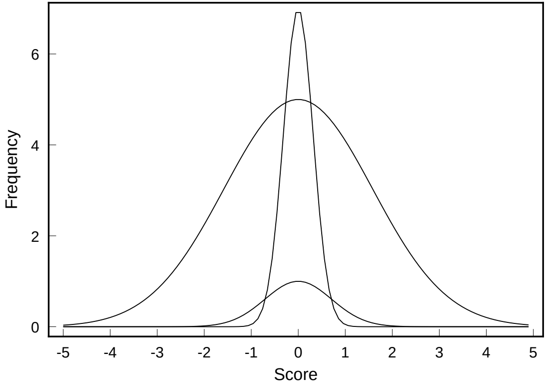
$$s^2 = \frac{34246 - 34225}{3} = \frac{21}{3} = 7$$

- and then the standard deviation equals:

$$s = \sqrt{s^2} = \sqrt{7} = 2.646$$

WHY BOTHER?

- the value of the standard deviation gives us an idea of how spread out scores are
- larger standard deviations indicate that scores are more spread out



WHY BOTHER?

- we will use standard deviation to let us estimate how different a score is relative to the central tendency of the distribution
- we can then compare (in a certain sense) **across** distributions!

STANDARD SCORE

- also called z-score

$$\text{Standard score} = \frac{\text{raw score} - \text{mean}}{\text{standard deviation}}$$

$$z = \frac{X - \overline{X}}{s}$$

- indicates the number of standard deviations a raw score is above or below the mean

EXAMPLE

- if

$$\overline{X} = 26$$

- and

$$s = 4$$

- and you have (among others) the scores $X_1 = 16$, $X_2 = 32$, $X_3 = 28$
- then

$$z_1 = \frac{X_1 - \overline{X}}{s} = \frac{16 - 26}{4} = -2.5$$

$$z_2 = \frac{X_2 - \overline{X}}{s} = \frac{32 - 26}{4} = 1.5$$

$$z_3 = \frac{X_3 - \overline{X}}{s} = \frac{28 - 26}{4} = 0.5$$

PROPERTIES

- when a raw score is **above** the mean, its z-score is positive
- when a raw score is **below** the mean, its z-score is negative
- when a raw score **equals** the mean, its z-score is zero
- absolute size of the z-score indicates how far from the mean a raw score is

UNITS

- z-scores work in units of standard deviation
- new numbers for same information!
- just like converting units for other familiar measures
  - length: feet into meters, miles into kilometers
  - weight: pounds into kilograms
  - temperature: fahrenheit into celsius
  - data: raw score units into standard deviation units
- trick!: standard deviation units depend on your particular set of data!

PROPERTIES

- z-scores are data
- we can find distributions, means, and standard deviations
- special properties of z-score distributions
  - The shape of the distribution of standard scores is identical to that of the original distribution of raw scores.
  - The mean of a distribution of z-scores will always equal 0.
  - The variance (and standard deviation) of a distribution of z-scores always equals 1.

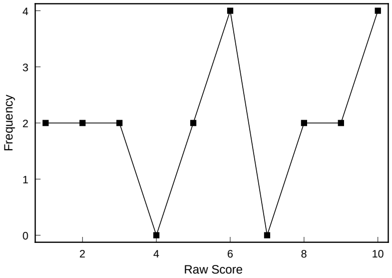
EXAMPLE

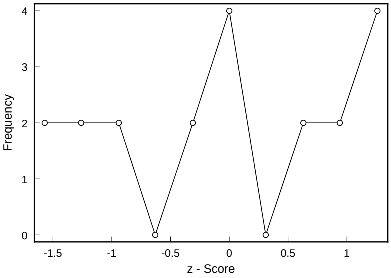
- A simple data set to play with
  - when a raw score is **above** the mean, its z-score is positive
  - when a raw score is **below** the mean, its z-score is negative
  - when a raw score **equals** the mean, its z-score is zero
  - absolute size of the z-score indicates how far from the mean a raw score is

Subject	Raw score	z-score
1	10	1.26
2	9	0.94
3	3	-0.94
4	10	1.26
5	9	0.94
6	2	-1.26
7	2	-1.26
8	10	1.26
9	5	-0.31
10	5	-0.31
11	1	-1.57
12	6	0.0
13	8	0.63
14	6	0.0
15	6	0.0
16	1	-1.57
17	3	-0.94
18	6	0.0
19	10	1.26
20	8	0.63
$n = 20$		
$\overline{X}$	6.0	0.0
s	3.18	1.0

EXAMPLE

- compare distributions of raw scores and z-scores
- shape is the same





USES

- suppose we want to compare the scores of a student in several classes
- we know the student's score, the mean score, the standard deviation, and the student's z-score

Subject	X	$\overline{X}$	s	z
Psychology	68	65	6	0.50
Mathematics	77	77	9	0.00
History	83	89	8	-0.75

- comparison of **raw scores** suggests that student did best in history, mathematics, then psychology
- comparison of **z-scores** suggests that student did best in psychology, mathematics, then history (relative to other students)

TRANSFORMED SCORES

- sometimes z-scores are unattractive
  - zero mean
  - negative values
- need to convert same information into a new distribution with a new mean and standard deviation

$$X' = (s')(z) + \overline{X}'$$

- where
  - X' = new or transformed score for a particular individual
  - s' = desired standard deviation of the distribution
  - z = standard score for a particular individual
  - $\overline{X}'$ = desired mean of the distribution

TRANSFORMED SCORES

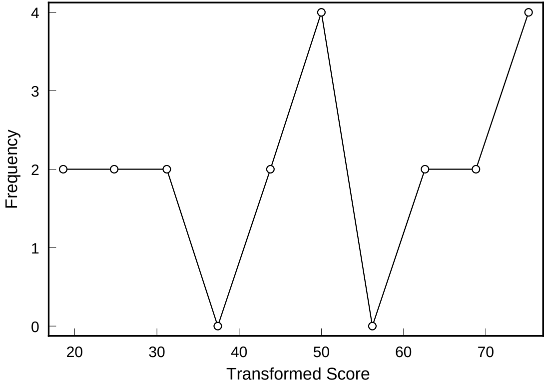
- GOAL: make data understandable; IQ scores, personality tests,...
- NOTE: you can change the mean and standard deviation all you want, but it does **not** change the information in the data
- shape remains the same!
  - conversion back to z-scores would produce the same z-scores!
  - a percentile maps to the corresponding transformed score

TRANSFORMED SCORES

- if we transform the scores from our earlier data set using

$$X' = 20X + 50$$

- we get



CONCLUSIONS

- variance
- standard deviation
- standard scores

NEXT TIME

- a very important distribution
- normal distribution

Describing everyone's height.