Standard Deviation

Lecture 18
Section 5.3.4

Robb T. Koether

Hampden-Sydney College

Mon, Sep 29, 2008

Outline

Exercise 5.12, page 333.

The five-number summary for the distribution of income (in \$1000s) for the 200 households in your neighborhood is provided below.

$$\$25, \$37, \$67, \$100, \$250$$

Exercise 5.12, page 333.

(a) Draw a basic boxplot for the income distribution in your neighborhood.

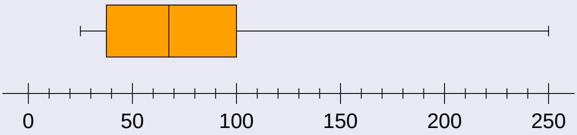
(b) Suppose that your household income is \$56,000. What can you say about the percentage of households that have a higher income than you?

(c) If the lowest 25% of the households will be classified as poor, what is the minimum household income that would lead to being classified as not poor?

Solution

(a)

- First, do not find a five-number summary for these data. These numbers are the five-number summary.
- The boxplot:



Solution

(b) \$56,000 is between the first quartile and the median, so we can say that at least half the neighborhood, but no more than three-quarters, have a higher income.

(c) You must have an income of at least \$37,000 not to be classified as poor.

Variability

- Our ability to estimate a parameter accurately depends on the variability of the population.
- What do we mean by "variability" in the population?
- How do we measure it?

Variability

- Our ability to estimate a parameter accurately depends on the variability of the population.
- What do we mean by "variability" in the population?
- How do we measure it?

Variability

- Our ability to estimate a parameter accurately depends on the variability of the population.
- What do we mean by "variability" in the population?
- How do we measure it?

Deviations from the Mean

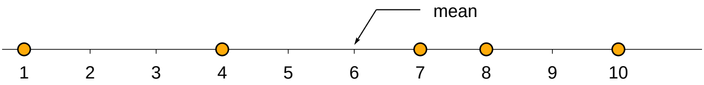
Definition (Deviation)

The deviation of an observation x is the difference between x and the sample mean $\bar{x}$.

$$\text{deviation of } x = x - \bar{x}.$$

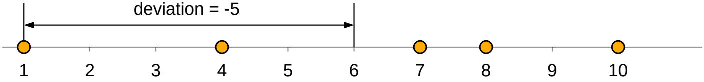
Deviations from the Mean

- Deviations from the mean.



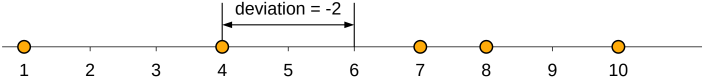
Deviations from the Mean

- Deviations from the mean.



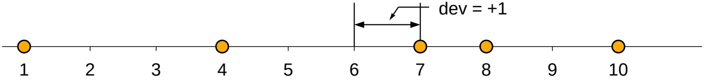
Deviations from the Mean

- Deviations from the mean.



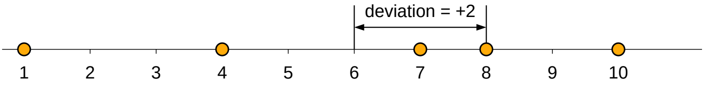
Deviations from the Mean

- Deviations from the mean.



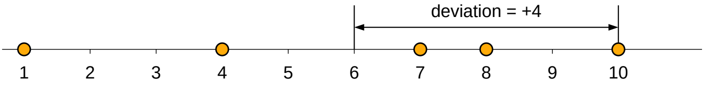
Deviations from the Mean

- Deviations from the mean.



Deviations from the Mean

- Deviations from the mean.



Deviations from the Mean

- How do we obtain one number that is representative of the whole set of individual deviations?
- Normally we use an average to summarize a set of numbers.
- Why will the average not work in this case?

Sum of Squared Deviations

- Rather than average the deviations, we will average their squares. That way, there will be no canceling.
- So we compute the sum of the squared deviations.

Definition (Sum of squared deviations)

The sum of squared deviations, denoted SSX, of a set of numbers is the sum of the squares of their deviations from the mean of the set.

$$\text{SSX} = \sum (x - \overline{x})^2.$$

Sum of Squared Deviations

To find SSX

- Find the average: $\bar{x} = \frac{\sum x}{n}$.
- Find the deviations from the average: $x - \bar{x}$.
- Square the deviations: $(x - \bar{x})^2$.
- Add them up:

$$\text{SSX} = \sum (x - \bar{x})^2.$$

Sum of Squared Deviations

Example (Calculating SSX)

- Let the sample be $\{1, 4, 7, 8, 10\}$.
- Then

$$\begin{aligned} \text{SSX} &= (1-6)^2 + (4-6)^2 + (7-6)^2 \\ &\quad +(8-6)^2 + (10-6)^2 \\ &= (-5)^2 + (-2)^2 + (1)^2 + (2)^2 + (4)^2 \\ &= 25 + 4 + 1 + 4 + 16 \\ &= 50. \end{aligned}$$

Sum of Squared Deviations

Practice

- Let the sample be $\{1, 3, 4, 6, 9, 11, 15\}$.
- Calculate
  - The sample mean.
  - The deviations.
  - The squared deviations.
  - The sum of the squared deviations.

The Population Variance

Definition (Variance of a population)

The variance of a population, denoted σ^2, is the average of the squared deviations of the members of the population.

$$\sigma^2 = \frac{\sum (x-\mu)^2}{N}.$$

Definition (Standard deviation of a population)

The standard deviation of a population, denoted σ, is the square root of the population variance.

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}}.$$

The Sample Variance

Definition (Variance of a sample)

The variance of a sample, denoted s^2, is the sum of the squared deviations of the members of the sample, divided by 1 less than the sample size.

$$s^2 = \frac{\sum (x - \overline{x})^2}{n - 1}.$$

Definition (Standard deviation of a sample)

The standard deviation of a sample, denoted s, is the square root of the sample variance.

$$s = \sqrt{\frac{\sum (x - \overline{x})^2}{n - 1}}.$$

The Sample Variance

- Theory shows that if we divide $\sum (x - \overline{x})^2$ by $n - 1$ instead of n, then s^2 will be a better estimator of σ^2.
- Otherwise, s^2 will systematically underestimate σ^2.
- Therefore, we do it.

Example

Example (Calculating s)

- For the sample $\{1, 4, 7, 8, 10\}$, we found that

$$\text{SSX} = 50.$$

- Therefore,

$$s^2 = \frac{50}{4} = 12.5$$

and so

$$s = \sqrt{12.5} = 3.536.$$

Sum of Squared Deviations

Practice

- Let the sample be $\{1, 3, 4, 6, 9, 11, 15\}$.
- Calculate s^2 and s.

Example

- How does s compare to the individual deviations?
- We will interpret s as being representative of the deviations in the sample.
- Does that seem reasonable for the previous examples?

Alternate Formula for SSX

- An alternate formula to compute SSX is

$$\text{SSX} = \sum x^2 - \frac{\left(\sum x\right)^2}{n}.$$

- Then, as before

$$s^2 = \frac{\text{SSX}}{n-1}$$

and

$$s = \sqrt{\frac{\text{SSX}}{n-1}}.$$

Example

Example (Alternate formula for SSX)

- Let the sample be $\{1, 4, 7, 8, 10\}$.
- Then

$$\sum x = 30$$

and

$$\sum x^2 = 1 + 16 + 49 + 64 + 100 = 230.$$

- So

$$\begin{aligned} \text{SSX} &= 230 - \frac{30^2}{5} \\ &= 230 - 180 \\ &= 50. \end{aligned}$$

Sum of Squared Deviations

Practice

- Let the sample be $\{1, 3, 4, 6, 9, 11, 15\}$.
- Find $\sum x$.
- Find $\sum x^2$.
- Use the alternate formula to find SSC, s^2, and s.

TI-83 - Standard Deviations

TI-83 Standard Deviations

- Follow the procedure for computing the mean.
- The display shows `Sx` and σ`x`.
- `Sx` is the sample standard deviation.
- σ`x` is the population standard deviation.

Example

Example

TI-83 Standard Deviations

- Let the sample be $\{1, 4, 7, 8, 10\}$.
- We get
  - $\mathrm{Sx} = 3.535533906$.
  - $\sigma\mathrm{x} = 3.16227766$.

Sum of Squared Deviations

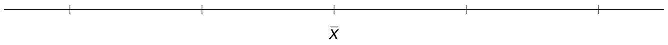
Practice

- Let the sample be $\{1, 3, 4, 6, 9, 11, 15\}$.
- Use the TI-83 to find s and s^2.
- What are the values of $\sum x$ and $\sum x^2$?

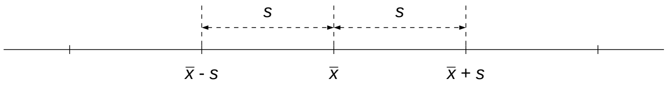
Interpreting the Standard Deviation

- Observations that deviate from $\bar{x}$ by much more than s are unusually far from the mean.
- Observations that deviate from $\bar{x}$ by much less than s are unusually close to the mean.

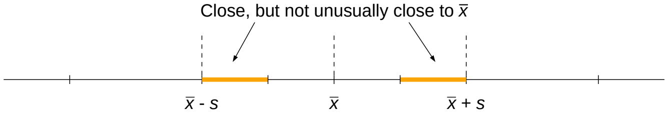
Interpreting the Standard Deviation



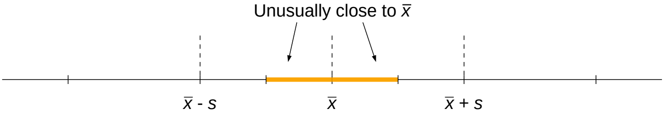
Interpreting the Standard Deviation



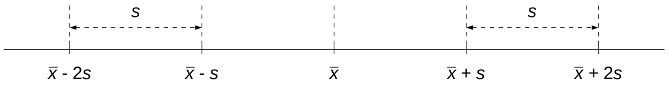
Interpreting the Standard Deviation



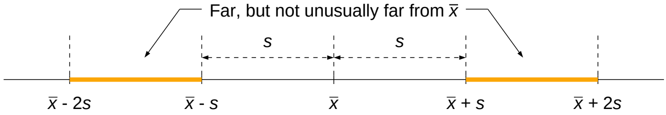
Interpreting the Standard Deviation



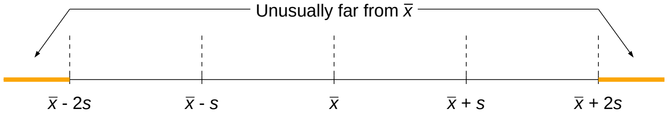
Interpreting the Standard Deviation



Interpreting the Standard Deviation



Interpreting the Standard Deviation



Assignment

Homework

- Read Section 5.3.4, pages 326 - 333.
- Let's Do It! 5.13, 5.14, 5.15.
- Page 333, exercises 10, 11, 14, 16 - 18, 20, 21.
- Chapter 5 review, p. 345, exercises 29 - 32, 36 - 40, 42 - 44, 47, 52, 53, 55.