Chapter 4. Variability

Likan Zhan
Beijing Language and Culture University

2021-09-27

https://likan.info
zhanlikan@blcu.edu.cn

Table of Contents

Table of Contents

Introduction of Variability

Introduction of Variability

- **Variability** (变异性) provides a quantitative measure of the differences between scores in a distribution and describes the degree to which the scores are spread out or clustered together.

Introduction of Variability

- **Variability** (变异性) provides a quantitative measure of the differences between scores in a distribution and describes the degree to which the scores are spread out or clustered together.
- In simple terms, if the scores in a distribution are all the same, then there is no variability.

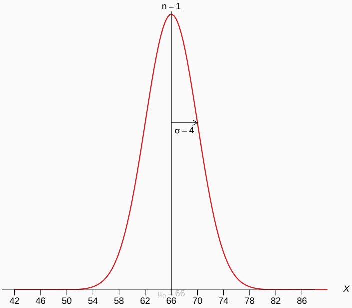
Introduction of Variability

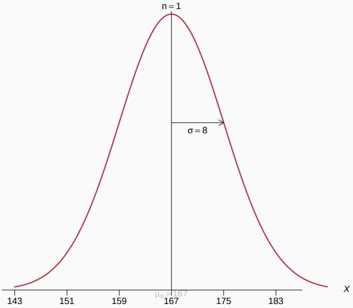
- **Variability** (变异性) provides a quantitative measure of the differences between scores in a distribution and describes the degree to which the scores are spread out or clustered together.
- In simple terms, if the scores in a distribution are all the same, then there is no variability.
- If there are small differences between scores, then the variability is small, and

Introduction of Variability

- **Variability** (变异性) provides a quantitative measure of the differences between scores in a distribution and describes the degree to which the scores are spread out or clustered together.
- In simple terms, if the scores in a distribution are all the same, then there is no variability.
- If there are small differences between scores, then the variability is small, and
- if there are large differences between scores, then the variability is large.

Adult male heights and weights





The two purposes of a good measure of variability

The two purposes of a good measure of variability

1. Variability describes the distribution.

The two purposes of a good measure of variability

1. Variability describes the distribution.
   - Usually, variability is defined in terms of *distance*.

The two purposes of a good measure of variability

1. Variability describes the distribution.
   - Usually, variability is defined in terms of *distance*.
   - It tells whether the scores are clustered close together or are spread out over a large distance.

The two purposes of a good measure of variability

The two purposes of a good measure of variability

2. Variability measures how well an individual score (or group of scores) represents the entire distribution.

The two purposes of a good measure of variability

2. Variability measures how well an individual score (or group of scores) represents the entire distribution.
   - This aspect of variability is very important for inferential statistics, in which relatively small samples are used to answer questions about populations.

The two purposes of a good measure of variability

2. Variability measures how well an individual score (or group of scores) represents the entire distribution.
   - This aspect of variability is very important for inferential statistics, in which relatively small samples are used to answer questions about populations.
   - Variability provides information about how much error to expect if you are using a sample to represent a population.

Common measures of variability

Common measures of variability

- Variance and Standard Deviation

Common measures of variability

- Variance and Standard Deviation
- Range

Common measures of variability

- Variance and Standard Deviation
- Range
- Interquartile range

Table of Contents

Defining Standard Deviation and Variance

Defining Standard Deviation and Variance

- The standard deviation is the most commonly used and the most important measure of variability.

Defining Standard Deviation and Variance

- The standard deviation is the most commonly used and the most important measure of variability.
- Standard deviation uses the mean of the distribution as a reference point and measures variability by considering the distance between each score and the mean.

Defining Standard Deviation and Variance

- The standard deviation is the most commonly used and the most important measure of variability.
- Standard deviation uses the mean of the distribution as a reference point and measures variability by considering the distance between each score and the mean.
- In simple terms, the standard deviation provides a measure of the standard, or *average distance* from the mean, and describes whether the scores are clustered closely around the mean or are widely scattered.

The calculation of variance and standard deviation

The calculation of variance and standard deviation

- Given the following $N = 5$ scores

X
1
9
5
8
7

The calculation of variance and standard deviation

The calculation of variance and standard deviation

- Although the concept of standard deviation is straightforward, the actual equations tend to be more complex.

The calculation of variance and standard deviation

- Although the concept of standard deviation is straightforward, the actual equations tend to be more complex.
- Therefore, we begin by looking at the logic that leads to these equations.

The calculation of variance and standard deviation

- Although the concept of standard deviation is straightforward, the actual equations tend to be more complex.
- Therefore, we begin by looking at the logic that leads to these equations.
- If you remember that our goal is to measure the standard, or typical, distance from the mean, then this logic and the equations that follow should be easier to remember.

The calculation of variance and standard deviation

The calculation of variance and standard deviation

- **Deviation** (离差) is distance from the mean:

$$\text{deviation score} = X - \mu$$

The calculation of variance and standard deviation

- **Deviation** (离差) is distance from the mean:

$$\text{deviation score} = X - \mu$$

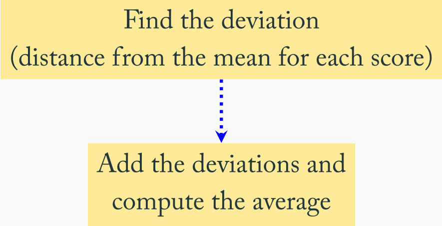
- The deviation for the previous scores are

X	Mean	Deviation
1	6	-5
9	6	3
5	6	-1
8	6	2
7	6	1

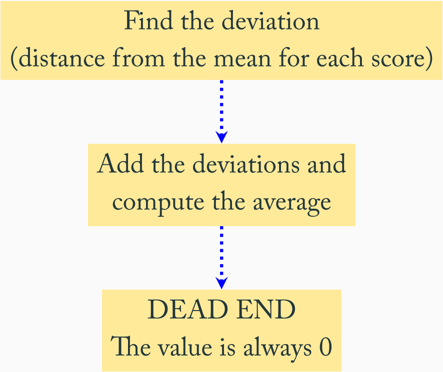
The calculation of variance and standard deviation

Find the deviation
(distance from the mean for each score)

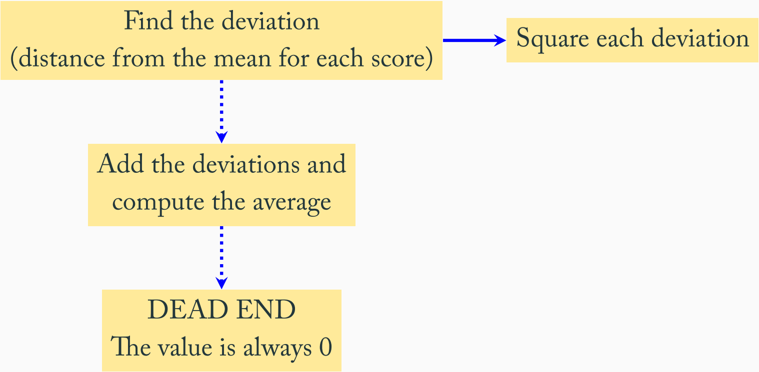
The calculation of variance and standard deviation



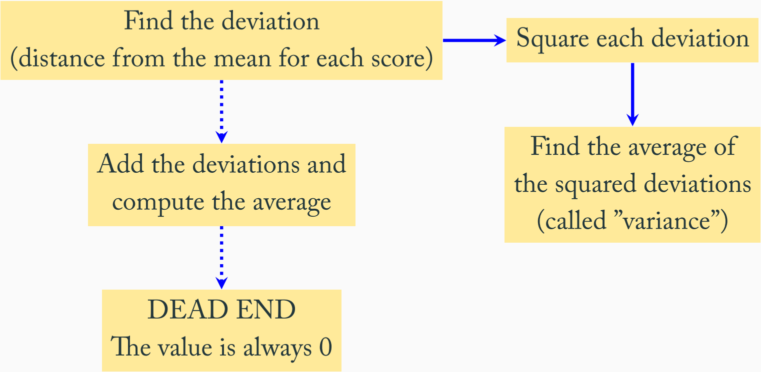
The calculation of variance and standard deviation



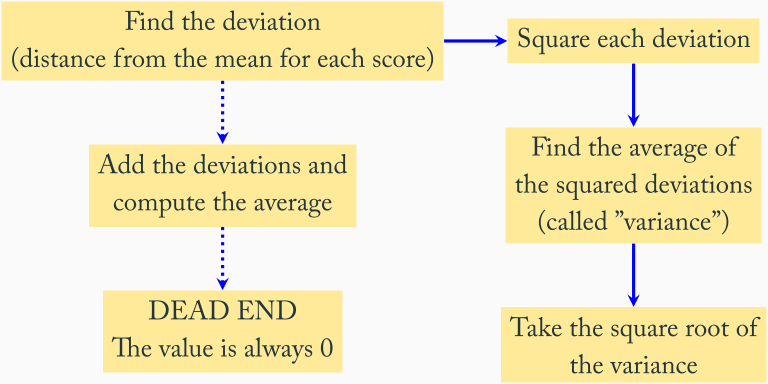
The calculation of variance and standard deviation



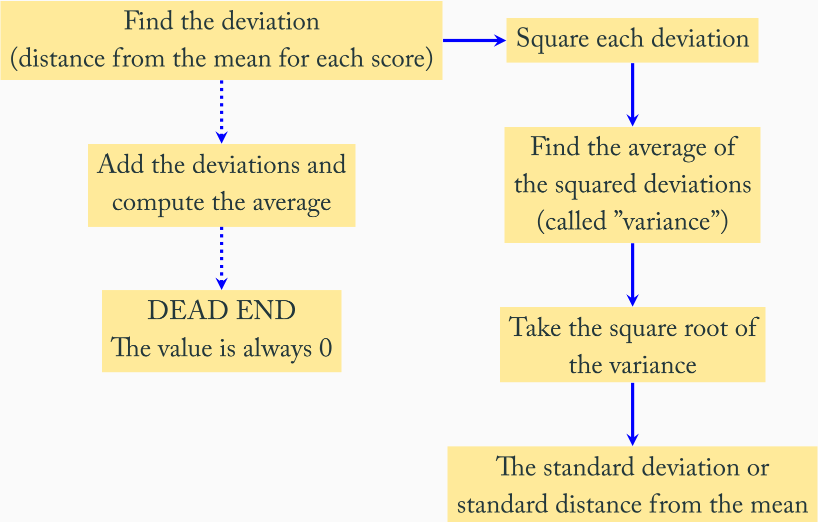
The calculation of variance and standard deviation



The calculation of variance and standard deviation



The calculation of variance and standard deviation



The calculation of variance and standard deviation

The calculation of variance and standard deviation

- **Variance** (方差) equals the mean of the squared deviations. Variance is the average squared distance from the mean.

The calculation of variance and standard deviation

- **Variance** (方差) equals the mean of the squared deviations. Variance is the average squared distance from the mean.
- **Standard deviation** (标准差) is the square root of the variance and provides a measure of the standard, or average distance from the mean.

$$\text{Standard deviation} = \sqrt{\text{variance}}$$

The calculation of variance and standard deviation

The calculation of variance and standard deviation

- Return to the previous $N = 5$ scores

X	Mean	Deviation	Squared_Deviation
1	6	-5	25
9	6	3	9
5	6	-1	1
8	6	2	4
7	6	1	1

The calculation of variance and standard deviation

The calculation of variance and standard deviation

- The sum of the squared deviation

$$\sum(X-\mu)^2 = 25+9+1+4+1 = 40$$

The calculation of variance and standard deviation

- The sum of the squared deviation

$$\sum (X-\mu)^2 = 25 + 9 + 1 + 4 + 1 = 40$$

- The variance is

$$\begin{aligned} \text{Variance} &= \frac{\text{Sum of Squared Deviation}}{N} \\ &= \frac{40}{5} = 8 \end{aligned}$$

The calculation of variance and standard deviation

- The sum of the squared deviation

$$\sum(X-\mu)^2 = 25+9+1+4+1 = 40$$

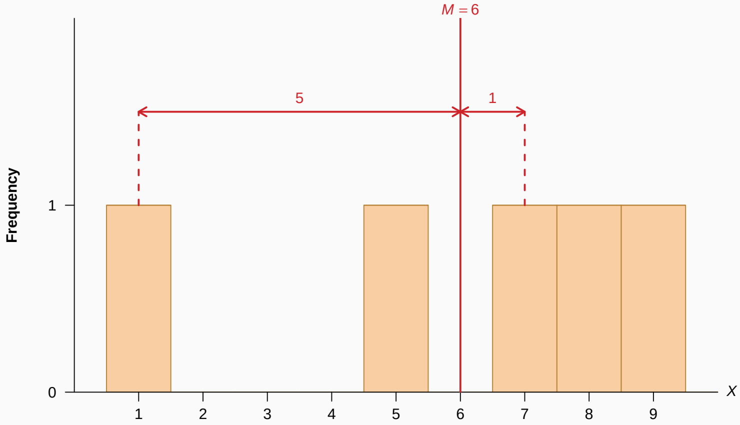
- The variance is

$$\begin{aligned}\text{Variance} &= \frac{\text{Sum of Squared Deviation}}{N}\\ &= \frac{40}{5} = 8\end{aligned}$$

- The standard deviation is

$$\begin{aligned}\text{Standard Deviation} &= \sqrt{\text{Variance}}\\ &= \sqrt{8} = 2.828\end{aligned}$$

The calculation of variance and standard deviation



The Sum of Squared Deviations (SS)

The Sum of Squared Deviations (SS)

- Variance is the mean of the squared deviations.

$$\begin{aligned}\text{Variance} &= \text{mean squared deviation} \\ &= \frac{\text{sum of squared deviations}}{\text{number of scores}}\end{aligned}$$

The Sum of Squared Deviations (SS)

- Variance is the mean of the squared deviations.

$$\begin{aligned}\text{Variance} &= \text{mean squared deviation} \\ &= \frac{\text{sum of squared deviations}}{\text{number of scores}}\end{aligned}$$

- **SS**, or **sum of squares** (离差平方和), is the sum of the squared deviation scores.

The Sum of Squared Deviations (SS)

The Sum of Squared Deviations (SS)

- Definitional formula

$$SS = \sum (X - \mu)^2 \qquad (1)$$

The Sum of Squared Deviations (SS)

- Definitional formula

$$SS = \sum (X - \mu)^2 \tag{1}$$

- Comoutational formula

$$SS = \sum X^2 - \frac{(\sum X)^2}{N} \tag{2}$$

From Definitional formula to computational formula

$$
\begin{aligned}
SS &= \sum (X - \mu)^2 \\
&= \sum (X^2 - 2 \cdot X \cdot \mu + \mu^2) \\
&= \sum X^2 - 2 \cdot \mu \cdot \sum X + N \cdot \mu^2 \\
&= \sum X^2 - 2 \cdot \frac{\sum X}{N} \cdot \sum X + N \cdot \frac{(\sum X)^2}{N^2} \\
&= \sum X^2 - 2 \cdot \frac{(\sum X)^2}{N} + \frac{(\sum X)^2}{N} \\
&= \sum X^2 - \frac{(\sum X)^2}{N}
\end{aligned}
$$

Sum of squares: Definitional formula

Sum of squares: Definitional formula

- Suppose a data set of $N = 4$ scores

```
X <- c(1, 0, 6, 1)
Mean <- mean(X)
Deviation <- X - Mean
Squared_Deviation <- Deviation ^ 2
SS <- sum(Squared_Deviation)
SS

## [1] 22
```

Sum of squares: Comoutational formula

Sum of squares: Comoutational formula

- Return to the same data set

```
X <- c(1, 0, 6, 1)
Sum_Sqr_X <- sum(X ^ 2)
Sqr_Sum_X <- (sum(X)) ^ 2
N <- length(X)
SS <- Sum_Sqr_X - Sqr_Sum_X / N
SS
```

```
## [1] 22
```

Final Formulas and Notation

Final Formulas and Notation

- **Population variance** (总体方差) is represented by the symbol σ^2 and equals the mean squared distance from the mean.

Final Formulas and Notation

- **Population variance** (总体方差) is represented by the symbol $\boldsymbol{\sigma}^2$ and equals the mean squared distance from the mean.
- Population variance is obtained by dividing the sum of squares by N.

$$\text{population variance} = \boldsymbol{\sigma}^2 = \frac{SS}{N} \tag{3}$$

Final Formulas and Notation

Final Formulas and Notation

- **Population standard deviation** (总体标准差) is represented by the symbol $\boldsymbol{\sigma}$ and equals the square root of the population variance.

$$\begin{aligned} &\text{population standard deviation} \\ =&\boldsymbol{\sigma} = \sqrt{\boldsymbol{\sigma}^2} \\ =&\sqrt{\frac{SS}{N}} \end{aligned} \tag{4}$$

Table of Contents

Formulas for Sample Sum of Squares

Formulas for Sample Sum of Squares

- The goal of inferential statistics is to use the limited information from samples to draw general conclusions about populations.

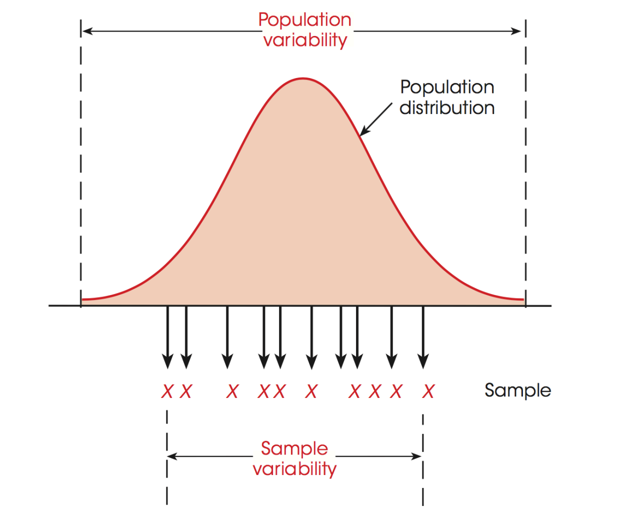
Formulas for Sample Sum of Squares

- The goal of inferential statistics is to use the limited information from samples to draw general conclusions about populations.
- The basic assumption of this process is that samples should be representative of the populations from which they come.

Formulas for Sample Sum of Squares

- The goal of inferential statistics is to use the limited information from samples to draw general conclusions about populations.
- The basic assumption of this process is that samples should be representative of the populations from which they come.
- This assumption poses a special problem for variability because samples consistently tend to be less variable than their populations.

The Problem with Sample Variability



Formulas for Sample Sum of Squares

Formulas for Sample Sum of Squares

- Notice that a few extreme scores in the population tend to make the population variability relatively large.

Formulas for Sample Sum of Squares

- Notice that a few extreme scores in the population tend to make the population variability relatively large.
- However, these extreme values are unlikely to be obtained when you are selecting a sample, which means that the sample variability is relatively small.

Formulas for Sample Sum of Squares

- Notice that a few extreme scores in the population tend to make the population variability relatively large.
- However, these extreme values are unlikely to be obtained when you are selecting a sample, which means that the sample variability is relatively small.
- The fact that a sample tends to be less variable than its population means that sample variability gives a **biased** estimate of population variability.

Formulas for Sample Sum of Squares

- Notice that a few extreme scores in the population tend to make the population variability relatively large.
- However, these extreme values are unlikely to be obtained when you are selecting a sample, which means that the sample variability is relatively small.
- The fact that a sample tends to be less variable than its population means that sample variability gives a **biased** estimate of population variability.
- Fortunately, the bias in sample variability is consistent and predictable, which means it can be corrected.

Formulas for Sample Sum of Squares

Formulas for Sample Sum of Squares

- The definitional formula for SS for a sample is

$$SS = \sum (X - M)^2 \tag{5}$$

Formulas for Sample Sum of Squares

- The definitional formula for SS for a sample is

$$SS = \sum (X - M)^2 \tag{5}$$

- The computational formula for SS for a sample is

$$SS = \sum X^2 - \frac{(\sum X)^2}{n} \tag{6}$$

Sample Variance and Standard Deviation

Sample Variance and Standard Deviation

- **Sample variance** (样本方差) is represented by the symbol $\mathbf{s}^2$ and equals the mean squared distance from the mean.

Sample Variance and Standard Deviation

- **Sample variance** (样本方差) is represented by the symbol $\mathbf{s}^2$ and equals the mean squared distance from the mean.
- Sample variance is obtained by dividing the sum of squares by $n - 1$.

$$\text{sample variance} = s^2 = \frac{SS}{n-1} \tag{7}$$

Sample Variance and Standard Deviation

- **Sample variance** (样本方差) is represented by the symbol **s**^{2} and equals the mean squared distance from the mean.
- Sample variance is obtained by dividing the sum of squares by $n - 1$.

$$\text{sample variance} = s^2 = \frac{SS}{n-1} \tag{7}$$

- **Sample standard deviation** (样本标准差) is represented by the symbol **s**^{2} and equal the square root of the sample variance.

$$\begin{aligned} &\text{sample standard deviation} \\ &= s = \sqrt{s^2} = \sqrt{\frac{SS}{n-1}} \end{aligned} \tag{8}$$

Sample Variance and Standard Deviation

Sample Variance and Standard Deviation

- Remember that the formulas for sample variance and standard deviation were constructed so that the sample variability would provide a good estimate of population variability.

Sample Variance and Standard Deviation

- Remember that the formulas for sample variance and standard deviation were constructed so that the sample variability would provide a good estimate of population variability.
- For this reason, the sample variance is often called estimated population variance (整体方差的估计), and the sample standard deviation is called estimated population standard deviation (整体标准差的估计).

Sample Variance and Standard Deviation

- Remember that the formulas for sample variance and standard deviation were constructed so that the sample variability would provide a good estimate of population variability.
- For this reason, the sample variance is often called estimated population variance (整体方差的估计), and the sample standard deviation is called estimated population standard deviation (整体标准差的估计).
- When you have only a sample to work with, the variance and standard deviation for the sample provide the best possible estimates of the population variability.

Sample Variability and Degrees of Freedom

Sample Variability and Degrees of Freedom

- Although the concept of a deviation score and the calculation SS are almost exactly the same for samples and populations, the minor differences in notation are really very important.

Sample Variability and Degrees of Freedom

- Although the concept of a deviation score and the calculation SS are almost exactly the same for samples and populations, the minor differences in notation are really very important.
- Specifically, with a population, you find the deviation for each score by measuring its distance from the population mean.

Sample Variability and Degrees of Freedom

Sample Variability and Degrees of Freedom

- With a sample, on the other hand, the value of μ is unknown and you must measure distances from the sample mean.

Sample Variability and Degrees of Freedom

- With a sample, on the other hand, the value of μ is unknown and you must measure distances from the sample mean.
- Because the value of the sample mean varies from one sample to another, you must first compute the sample mean before you can begin to compute deviations.

Sample Variability and Degrees of Freedom

- With a sample, on the other hand, the value of μ is unknown and you must measure distances from the sample mean.
- Because the value of the sample mean varies from one sample to another, you must first compute the sample mean before you can begin to compute deviations.
- However, calculating the value of M places a restriction on the variability of the scores in the sample.

Sample Variability and Degrees of Freedom

- Suppose we select a sample of $n = 3$ scores and compute a mean of $M = 5$.

X	A sample of $n = 3$ scores with a mean of $M = 5$
2	
9	
___	$\leftarrow$ What is this third score?

Sample Variability and Degrees of Freedom

Sample Variability and Degrees of Freedom

- For a sample of n scores, the **degrees of freedom** (自由度), or **df**, for the sample variance are defined as $df = n - 1$.

Sample Variability and Degrees of Freedom

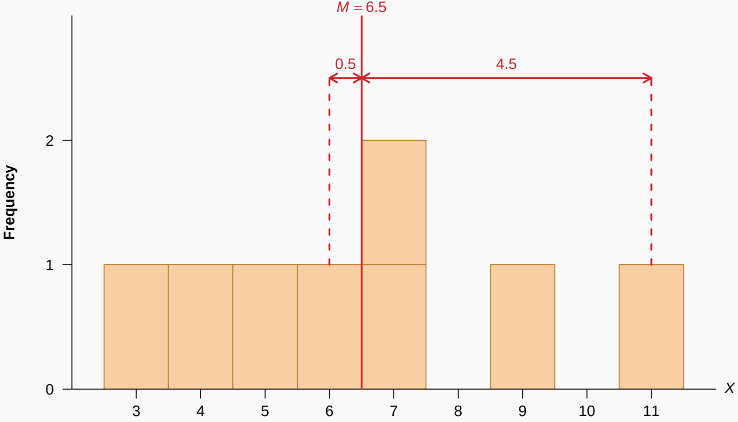
- For a sample of n scores, the **degrees of freedom** (自由度), or **df**, for the sample variance are defined as $df = n - 1$.
- The degrees of freedom determine the number of scores in the sample that are independent and free to vary.

Sample Variability and Degrees of Freedom

- To calculate sample variance (mean squared deviation), we find the sum of the squared deviations (SS) and divide by the number of scores that are free to vary. This number is $n - 1 = df$.
- Thus, the formula for sample variance is

$$s^2 = \frac{\text{sum of squared deviation}}{\text{numbers of scores free to vary}} = \frac{SS}{df} = \frac{SS}{n-1}$$

Presenting Standard Deviation in Graphs



The Sample Variance and Standard Deviation

```
X <- c(4, 6, 5, 11, 7, 9, 7, 3)
```

The Sample Variance and Standard Deviation

```r
X <- c(4, 6, 5, 11, 7, 9, 7, 3)
```

```r
SS <- sum((X - mean(X))^2)
SS <- sum(X^2) - (sum(X))^2 / length(X)
```

The Sample Variance and Standard Deviation

```r
X <- c(4, 6, 5, 11, 7, 9, 7, 3)
```

```r
SS <- sum((X - mean(X))^2)
SS <- sum(X^2) - (sum(X))^2 / length(X)
```

```r
s_sqr <- SS / (length(X) - 1)
s <- sqrt(s_sqr)
```

The Sample Variance and Standard Deviation

```
X <- c(4, 6, 5, 11, 7, 9, 7, 3)
```

```
SS <- sum((X - mean(X))^2)
SS <- sum(X^2) - (sum(X))^2 / length(X)
```

```
s_sqr <- SS / (length(X) - 1)
s <- sqrt(s_sqr)
```

```
## [1] 6.857143
## [1] 2.618615
```

The Sample Variance and Standard Deviation

The Sample Variance and Standard Deviation

- Use the `var()` function in R

```
var(X)

## [1] 6.857143
```

The Sample Variance and Standard Deviation

- Use the `var()` function in R

```
var(X)

## [1] 6.857143
```

- Use the `sd()` function in R

```
sd(X)

## [1] 2.618615
```

Biased and Unbiased Statistics

Biased and Unbiased Statistics

- A sample statistic **unbiased** (无偏) if the average value of the statistic is equal to the population parameter.

Biased and Unbiased Statistics

- A sample statistic **unbiased** (无偏) if the average value of the statistic is equal to the population parameter.
- The average value of the statistic is obtained from all the possible samples for a specific sample size, n.

Biased and Unbiased Statistics

- A sample statistic **unbiased** (无偏) if the average value of the statistic is equal to the population parameter.
- The average value of the statistic is obtained from all the possible samples for a specific sample size, n.
- A sample statistic is **biased** (有偏的) if the average value of the statistic either underestimates or overestimates the corresponding population parameter.

Biased and Unbiased Statistics

Biased and Unbiased Statistics

- Suppose a population that consists of exactly $N = 3$ scores

```
X <- c(0, 3, 9)
```

Biased and Unbiased Statistics

- Suppose a population that consists of exactly $N = 3$ scores

```
X <- c(0, 3, 9)
```

- This population has a mean of $\mu = 4$ and,

Biased and Unbiased Statistics

- Suppose a population that consists of exactly $N = 3$ scores

```
X <- c(0, 3, 9)
```

- This population has a mean of $\mu = 4$ and,
- This population has a variance of $\sigma^2 = 14$.

Biased and Unbiased Statistics

- Suppose a population that consists of exactly $N = 3$ scores

```r
X <- c(0, 3, 9)
```

- This population has a mean of $\mu = 4$ and,
- This population has a variance of $\sigma^2 = 14$.
- Next, we select samples of $n = 2$ scores from this population.

Biased and Unbiased Statistics

Biased and Unbiased Statistics

- In fact, we obtain every single possible sample with $n = 2$.

	Score_1	Score_2	Mean
1	0	0	0.00
2	3	0	1.50
3	9	0	4.50
4	0	3	1.50
5	3	3	3.00
6	9	3	6.00
7	0	9	4.50
8	3	9	6.00
9	9	9	9.00

Biased and Unbiased Statistics

Biased and Unbiased Statistics

- The biased and unbiased variances for each sample are:

	Score_1	Score_2	Mean	Biased	Unbiased
1	0	0	0.00	0.00	0.00
2	3	0	1.50	2.25	4.50
3	9	0	4.50	20.25	40.50
4	0	3	1.50	2.25	4.50
5	3	3	3.00	0.00	0.00
6	9	3	6.00	9.00	18.00
7	0	9	4.50	20.25	40.50
8	3	9	6.00	9.00	18.00
9	9	9	9.00	0.00	0.00

Biased and Unbiased Statistics

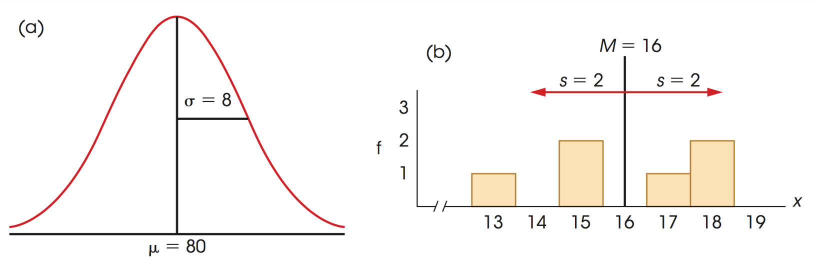
Biased and Unbiased Statistics

- The mean of sample mean and of the two variances are

```
## Mean
##    4
## Biased
##      7
## Unbiased
##       14
```

Table of Contents

Presenting the Mean and Standard Deviation



Transformations of Scale

Transformations of Scale

- Adding a constant to each score does not change the standard deviation.

Transformations of Scale

- Adding a constant to each score does not change the standard deviation.
- Multiplying each score by a constant causes the standard deviation to be multiplied by the same constant.

Reporting the Standard Deviation

Reporting the Standard Deviation

- In reporting the results of a study, the researcher often provides descriptive information for both central tendency and variability.

Reporting the Standard Deviation

- In reporting the results of a study, the researcher often provides descriptive information for both central tendency and variability.
- The dependent variables in psychology research are often numerical values obtained from measurements on interval or ratio scales.

Reporting the Standard Deviation

- In reporting the results of a study, the researcher often provides descriptive information for both central tendency and variability.
- The dependent variables in psychology research are often numerical values obtained from measurements on interval or ratio scales.
- With numerical scores the most common descriptive statistics are the mean (central tendency) and the standard deviation (variability), which are usually reported together.

Reporting the Standard Deviation

Reporting the Standard Deviation

- In many journals, especially those following APA style, the symbol SD is used for the sample standard deviation.

Reporting the Standard Deviation

- In many journals, especially those following APA style, the symbol SD is used for the sample standard deviation.

  Children who viewed the violent cartoon displayed more aggressive responses ($M = 12.45$, $SD = 3.7$) than those who viewed the control cartoon ($M = 4.22$, $SD = 1.04$).

Reporting the Standard Deviation

Reporting the Standard Deviation

- When reporting the descriptive measures for several groups, the findings may be summarized in a table.

	Type of Video Game	
	Violent	Control
Males	M=7.72 SD=2.43	M=4.34 SD=2.16
Females	M=2.47 SD=0.92	M=1.61 SD=0.68

Reporting the Standard Deviation

- When reporting the descriptive measures for several groups, the findings may be summarized in a table.

	Type of Video Game	
	Violent	Control
Males	M=7.72 SD=2.43	M=4.34 SD=2.16
Females	M=2.47 SD=0.92	M=1.61 SD=0.68

- Sometimes the table also indicates the sample size, n, for each group.

Standard Deviation and Descriptive Statistics

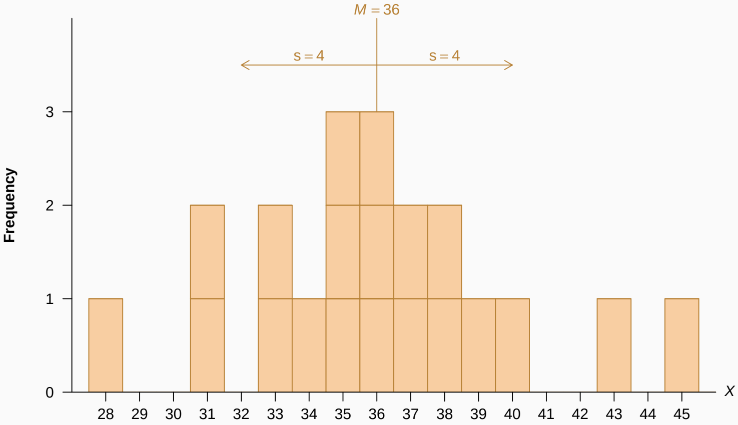
Standard Deviation and Descriptive Statistics

- **Describing an Entire Distribution.** Because the mean identifies the center of a distribution and the standard deviation describes the average distance from the mean, these two values should allow you to create a reasonably accurate image of the entire distribution.

Standard Deviation and Descriptive Statistics

- **Describing an Entire Distribution.** Because the mean identifies the center of a distribution and the standard deviation describes the average distance from the mean, these two values should allow you to create a reasonably accurate image of the entire distribution.
- **Describing the Location of Individual Scores.** Knowing the mean and standard deviation should also allow you to describe the relative location of any individual score within the distribution.

Standard Deviation and Descriptive Statistics



Variance and Inferential Statistics

Variance and Inferential Statistics

- In very general terms, the goal of inferential statistics is to detect meaningful and significant patterns in research results.

Variance and Inferential Statistics

- In very general terms, the goal of inferential statistics is to detect meaningful and significant patterns in research results.
- The basic question is whether the patterns observed in the sample data reflect corresponding patterns that exist in the population, or are simply random fluctuations that occur by chance.

Variance and Inferential Statistics

- In very general terms, the goal of inferential statistics is to detect meaningful and significant patterns in research results.
- The basic question is whether the patterns observed in the sample data reflect corresponding patterns that exist in the population, or are simply random fluctuations that occur by chance.
- Variability plays an important role in the inferential process because the variability in the data influences how easy it is to see patterns.

Variance and Inferential Statistics

- In very general terms, the goal of inferential statistics is to detect meaningful and significant patterns in research results.
- The basic question is whether the patterns observed in the sample data reflect corresponding patterns that exist in the population, or are simply random fluctuations that occur by chance.
- Variability plays an important role in the inferential process because the variability in the data influences how easy it is to see patterns.
- In general, low variability means that existing patterns can be seen clearly, whereas high variability tends to obscure any patterns that might exist.

Variance and Inferential Statistics

Variance and Inferential Statistics

- Experiment A

Treatment_1	Treatment_2
35	39
34	40
36	41
35	40

Variance and Inferential Statistics

- Experiment A

Treatment_1	Treatment_2
35	39
34	40
36	41
35	40

- Mean (Treatment_1) = 35

Variance and Inferential Statistics

- Experiment A

Treatment_1	Treatment_2
35	39
34	40
36	41
35	40

- Mean (Treatment_1) = 35
- Mean (Treatment_2) = 40

- Experiment B

Treatment_1	Treatment_2
31	46
15	21
57	61
37	32

Variance and Inferential Statistics

- Experiment A

Treatment_1	Treatment_2
35	39
34	40
36	41
35	40

- Mean (Treatment_1) = 35
- Mean (Treatment_2) = 40

- Experiment B

Treatment_1	Treatment_2
31	46
15	21
57	61
37	32

- Mean (Treatment_1) = 35

Variance and Inferential Statistics

- Experiment A

Treatment_1	Treatment_2
35	39
34	40
36	41
35	40

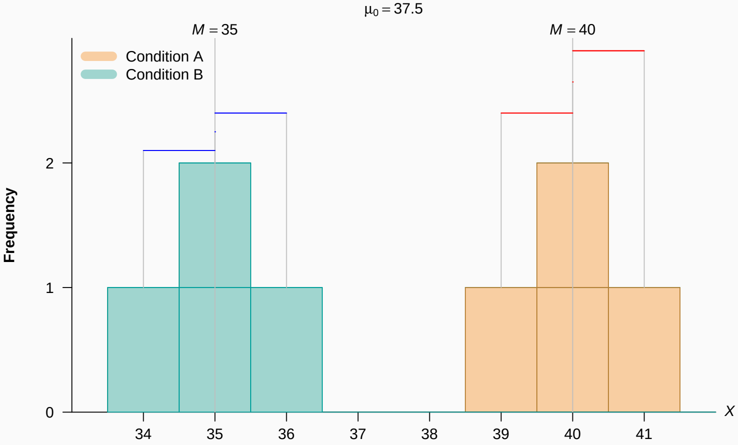
- Mean (Treatment_1) = 35
- Mean (Treatment_2) = 40

- Experiment B

Treatment_1	Treatment_2
31	46
15	21
57	61
37	32

- Mean (Treatment_1) = 35
- Mean (Treatment_2) = 40

Variance and Inferential Statistics



Variance and Inferential Statistics

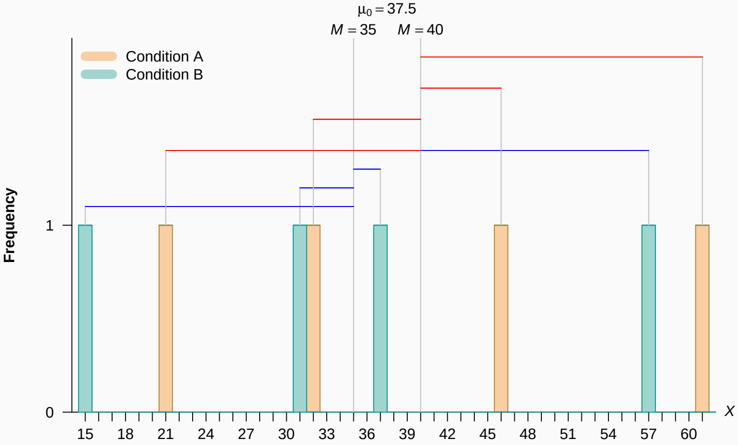


Table of Contents

The Range

The Range

- The **Range** (全距) is the distance covered by the scores in a distribution, from the smallest score to the largest score.

$$\text{Range} = X_{max} - X_{min}$$

The Range

- The **Range** (全距) is the distance covered by the scores in a distribution, from the smallest score to the largest score.

$$\text{Range} = X_{max} - X_{min}$$

- When the scores are measurements of a continuous variable, the range can be defined as the difference between real limits.

$$\text{Range} = \text{URL for} X_{max} - \text{LRL for} X_{min}$$

The Range

- The **Range** (全距) is the distance covered by the scores in a distribution, from the smallest score to the largest score.

$$\text{Range} = X_{max} - X_{min}$$

- When the scores are measurements of a continuous variable, the range can be defined as the difference between real limits.

$$\text{Range} = \text{URL for} X_{max} - \text{LRL for} X_{min}$$

- When the scores are whole numbers, the range can also be de ned as the number of measurement categories.

$$\text{Range} = X_{max} - X_{min} + 1$$

The Range function in R

```
(X <- seq(2, 17, by = 3))
range(X)
min(X)
max(X)

## [1]  2  5  8 11 14 17
## [1]  2 17
## [1] 2
## [1] 17
```

Interquartile range

Interquartile range

- Instead of looking at the entire range of the data, looking at the middle 50% will often prove to be a useful measure of spread, especially when the data are skewed.

Interquartile range

- Instead of looking at the entire range of the data, looking at the middle 50% will often prove to be a useful measure of spread, especially when the data are skewed.
- The **interquartile range (IQR)** (四分位距) is defined as $IQR = Q3 - Q1$.

Interquartile range

```
X

## [1]  2  5  8 11 14 17
```

Interquartile range

```
X
```

```
## [1]  2  5  8 11 14 17
```

```
quantile(X)
```

```
##    0%   25%   50%   75%  100%
##  2.00  5.75  9.50 13.25 17.00
```

Interquartile range

```
X
```

```
## [1]  2  5  8 11 14 17
```

```
quantile(X)
```

```
##    0%   25%   50%   75%  100%
##  2.00  5.75  9.50 13.25 17.00
```

```
IQR(X)
```

```
## [1] 7.5
```

Questions?