Lecture 4: Statistics ctd.

MA208 Quantitative Techniques for Business

Ilia Pirashvili, PhD

With special thanks to Dr. Kirsten Pfeiffer

School of Mathematical and Statistical Sciences
University of Galway

January 23, 2024

Lecture 4 - Outline

- Standard deviation
- Chebyshev's Inequality
- Coefficient of variation

Lecture 4 - Outline

- Standard deviation
- Chebyshev's Inequality
- Coefficient of variation

Lecture 4 - Outline

- Standard deviation
- Chebyshev's Inequality
- Coefficient of variation

Recall from last lecture...

We talked about the following types of means:

- (Arithmetic) mean, simply called mean in statistics;
- Harmonic mean (useful for ratios)
- Geometric mean (the product version of arithmetic mean)
- Weighted mean (generalises harmonic mean)

We also talked about the mode (most frequent items).
Lastly we mentioned the **limitations of** only using the **means** or median or other averages and aggregates (we need to understand the **variance** of the dataset).

Recall from last lecture...

We talked about the following types of means:

- (Arithmetic) mean, simply called mean in statistics;
- Harmonic mean (useful for ratios)
- Geometric mean (the product version of arithmetic mean)
- Weighted mean (generalises harmonic mean)

We also talked about the mode (most frequent items).
Lastly we mentioned the **limitations of** only using the **means** or median or other averages and aggregates (we need to understand the **variance** of the dataset).

Recall from last lecture...

We talked about the following types of means:

- (Arithmetic) mean, simply called mean in statistics;
- Harmonic mean (useful for ratios)
- Geometric mean (the product version of arithmetic mean)
- Weighted mean (generalises harmonic mean)

We also talked about the mode (most frequent items).
Lastly we mentioned the **limitations of** only using the **means** or median or other averages and aggregates (we need to understand the **variance** of the dataset).

Recall from last lecture...

We talked about the following types of means:

- (Arithmetic) mean, simply called mean in statistics;
- Harmonic mean (useful for ratios)
- Geometric mean (the product version of arithmetic mean)
- Weighted mean (generalises harmonic mean)

We also talked about the mode (most frequent items).
Lastly we mentioned the **limitations of** only using the **means** or median or other averages and aggregates (we need to understand the **variance** of the dataset).

Recall from last lecture...

We talked about the following types of means:

- (Arithmetic) mean, simply called mean in statistics;
- Harmonic mean (useful for ratios)
- Geometric mean (the product version of arithmetic mean)
- Weighted mean (generalises harmonic mean)

We also talked about the mode (most frequent items).
Lastly we mentioned the **limitations of** only using the **means** or median or other averages and aggregates (we need to understand the **variance** of the dataset).

Recall from last lecture...

We talked about the following types of means:

- (Arithmetic) mean, simply called mean in statistics;
- Harmonic mean (useful for ratios)
- Geometric mean (the product version of arithmetic mean)
- Weighted mean (generalises harmonic mean)

We also talked about the mode (most frequent items).

Lastly we mentioned the **limitations of** only using the **means** or median or other averages and aggregates (we need to understand the **variance** of the dataset).

Recall from last lecture...

We talked about the following types of means:

- (Arithmetic) mean, simply called mean in statistics;
- Harmonic mean (useful for ratios)
- Geometric mean (the product version of arithmetic mean)
- Weighted mean (generalises harmonic mean)

We also talked about the mode (most frequent items).
Lastly we mentioned the **limitations of** only using the **means** or median or other averages and aggregates (we need to understand the **variance** of the dataset).

Recall the Standard Deviation

A ore powerful measure of spread than IQR is the standard deviation which takes into account how far each data is from the mean.

Given a set of data values $x_1, x_2, x_3, ..., x_n$, the standard deviation is

Standard Deviation

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + ... + (x_n - \bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

Recall the Standard Deviation

A ore powerful measure of spread than IQR is the standard deviation which takes into account how far each data is from the mean.

Given a set of data values $x_1, x_2, x_3, ..., x_n$, the standard deviation is

Standard Deviation

$$\begin{aligned} \sigma &= \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + ... + (x_n - \overline{x})^2}{n-1}} \\ &= \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}} \end{aligned}$$

Standard Deviation

Exercise

A hospital records the following data of blood samples:

$$\{8, 11, 7, 13, 10, 11, 7, 9\}$$

Find the mean $\overline{x}$ and the standard deviation σ.

Standard Deviation

Solution:

A hospital records the following data of blood samples:

$$\{8, 11, 7, 13, 10, 11, 7, 9\}$$

Find the mean $\overline{x}$ and the standard deviation σ.

$$\begin{aligned} \overline{x} &= \frac{\sum_{i=1}^{n} x_i}{n} \\ &= \frac{8+11+7+13+10+11+7+9}{8} \\ &= \frac{76}{8} = 9.5 \end{aligned}$$

Standard Deviation

Solution:

A hospital records the following data of blood samples:

$$\{8, 11, 7, 13, 10, 11, 7, 9\}$$

Find the mean $\overline{x}$ and the standard deviation σ.

$$\begin{aligned} \overline{x} &= \frac{\sum_{i=1}^{n} x_i}{n} \\ &= \frac{8 + 11 + 7 + 13 + 10 + 11 + 7 + 9}{8} \\ &= \frac{76}{8} = 9.5 \end{aligned}$$

Standard Deviation

Solution:

A hospital records the following data of blood samples:

$$\{8, 11, 7, 13, 10, 11, 7, 9\}$$

Find the mean $\overline{x}$ and the standard deviation σ.

$$\begin{aligned} \overline{x} &= \frac{\sum_{i=1}^{n} x_i}{n} \\ &= \frac{8 + 11 + 7 + 13 + 10 + 11 + 7 + 9}{8} \\ &= \frac{76}{8} = 9.5 \end{aligned}$$

Standard Deviation

Solution:

A hospital records the following data of blood samples:

$$\{8, 11, 7, 13, 10, 11, 7, 9\}$$

Find the mean $\overline{x}$ and the standard deviation σ.

$\overline{x} = 9.5$

$$\begin{aligned}
\overline{\sigma} &= \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n-1}} \\
&= \sqrt{\frac{(8-9.5)^2 + (11-9.5)^2 + \ldots (7-9.5)^2 + (9-9.5)^2}{7}} \\
&= \sqrt{\frac{(-1.5)^2 + 1.5^2 + (-2.5)^2 + 3.5^2 + 0.5^2 + 1.5^2 + (-2.5)^2 + (-0.5)^2}{7}} \\
&= \sqrt{\frac{2.25 + 2.25 + 6.25 + 12.25 + 0.25 + 2.25 + 6.25 + 0.25}{7}} \\
&= \sqrt{\frac{32}{7}} \simeq 2.17
\end{aligned}$$

Standard Deviation

Solution:

A hospital records the following data of blood samples:

$$\{8, 11, 7, 13, 10, 11, 7, 9\}$$

Find the mean $\overline{x}$ and the standard deviation σ.

$\overline{x} = 9.5$

$$\begin{aligned}
\overline{\sigma} &= \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n-1}} \\
&= \sqrt{\frac{(8-9.5)^2 + (11-9.5)^2 + \ldots (7-9.5)^2 + (9-9.5)^2}{7}} \\
&= \sqrt{\frac{(-1.5)^2 + 1.5^2 + (-2.5)^2 + 3.5^2 + 0.5^2 + 1.5^2 + (-2.5)^2 + (-0.5)^2}{7}} \\
&= \sqrt{\frac{2.25 + 2.25 + 6.25 + 12.25 + 0.25 + 2.25 + 6.25 + 0.25}{7}} \\
&= \sqrt{\frac{32}{7}} \simeq 2.17
\end{aligned}$$

Standard Deviation

Solution:

A hospital records the following data of blood samples:

$$\{8, 11, 7, 13, 10, 11, 7, 9\}$$

Find the mean $\overline{x}$ and the standard deviation σ.

$\overline{x} = 9.5$

$$\begin{aligned}
\overline{\sigma} &= \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n-1}} \\
&= \sqrt{\frac{(8-9.5)^2 + (11-9.5)^2 + \ldots (7-9.5)^2 + (9-9.5)^2}{7}} \\
&= \sqrt{\frac{(-1.5)^2 + 1.5^2 + (-2.5)^2 + 3.5^2 + 0.5^2 + 1.5^2 + (-2.5)^2 + (-0.5)^2}{7}} \\
&= \sqrt{\frac{2.25 + 2.25 + 6.25 + 12.25 + 0.25 + 2.25 + 6.25 + 0.25}{7}} \\
&= \sqrt{\frac{32}{7}} \simeq 2.17
\end{aligned}$$

Standard Deviation

Solution:

A hospital records the following data of blood samples:

$$\{8, 11, 7, 13, 10, 11, 7, 9\}$$

Find the mean $\overline{x}$ and the standard deviation σ.

$\overline{x} = 9.5$

$$\begin{aligned}
\overline{\sigma} &= \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n-1}} \\
&= \sqrt{\frac{(8-9.5)^2 + (11-9.5)^2 + \ldots (7-9.5)^2 + (9-9.5)^2}{7}} \\
&= \sqrt{\frac{(-1.5)^2 + 1.5^2 + (-2.5)^2 + 3.5^2 + 0.5^2 + 1.5^2 + (-2.5)^2 + (-0.5)^2}{7}} \\
&= \sqrt{\frac{2.25 + 2.25 + 6.25 + 12.25 + 0.25 + 2.25 + 6.25 + 0.25}{7}} \\
&= \sqrt{\frac{32}{7}} \simeq 2.17
\end{aligned}$$

Standard Deviation

Solution:

A hospital records the following data of blood samples:

$$\{8, 11, 7, 13, 10, 11, 7, 9\}$$

Find the mean $\overline{x}$ and the standard deviation σ.

$\overline{x} = 9.5$

$$\begin{aligned}
\overline{\sigma} &= \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n-1}} \\
&= \sqrt{\frac{(8-9.5)^2 + (11-9.5)^2 + \ldots (7-9.5)^2 + (9-9.5)^2}{7}} \\
&= \sqrt{\frac{(-1.5)^2 + 1.5^2 + (-2.5)^2 + 3.5^2 + 0.5^2 + 1.5^2 + (-2.5)^2 + (-0.5)^2}{7}} \\
&= \sqrt{\frac{2.25 + 2.25 + 6.25 + 12.25 + 0.25 + 2.25 + 6.25 + 0.25}{7}} \\
&= \sqrt{\frac{32}{7}} \simeq 2.17
\end{aligned}$$

Using Standard Deviation

One crude but simple tool for understanding standard deviation is the range rule of thumb, which is based on the principle that for many data sets, the vast majority (such as 95%) of sample values lie within two standard variations of the mean.

Range Rule of Thumb

We informally define *usual* values in a data set to be those that are typical and not too extreme. If the standard deviation of a data set is known, use it to find rough estimates of the minimum and maximum *usual* sample values as follows:

$$\begin{aligned} \text{minimum “usual” value} &= (\text{mean}) - 2 \times (\text{standard deviation}) \\ \text{maximum “usual” value} &= (\text{mean}) + 2 \times (\text{standard deviation}) \end{aligned}$$

Using Standard Deviation

Example

The Wechsler Adult Intelligence Scale involves an IQ test designed so that the mean score is 100 and the standard deviation is 15. Use the range rule of thumb to find the minimum and maximum “usual” IQ scores. Then determine whether an IQ score of 135 would be considered “unusual”.

Using Standard Deviation

Example

The Wechsler Adult Intelligence Scale involves an IQ test designed so that the mean score is 100 and the standard deviation is 15. Use the range rule of thumb to find the minimum and maximum "usual" IQ scores. Then determine whether an IQ score of 135 would be considered "unusual".

Solution

Minimum usual value: $\mu - 2 \times \sigma = 100 - 30$.

Maximum usual value: $\mu + 2 \times \sigma = 100 + 30$.

Hence, typical range is: $[70, 130]$ and so, 135 is outside.

Using Standard Deviation

Example

The Wechsler Adult Intelligence Scale involves an IQ test designed so that the mean score is 100 and the standard deviation is 15. Use the range rule of thumb to find the minimum and maximum "usual" IQ scores. Then determine whether an IQ score of 135 would be considered "unusual".

Solution

Minimum usual value: $\mu - 2 \times \sigma = 100 - 30$.

Maximum usual value: $\mu + 2 \times \sigma = 100 + 30$.

Hence, typical range is: $[70, 130]$ and so, 135 is outside.

Using Standard Deviation

Note:

- The standard deviation measures the *variation* among data values.
- Values close together have a small standard deviation, but values with much more variation have a larger standard deviation.
- For many data sets, a value is *unusual* if it differs from the mean by more than two standard deviations.

Using Standard Deviation

Note:

- The standard deviation measures the *variation* among data values.
- Values close together have a small standard deviation, but values with much more variation have a larger standard deviation.
- For many data sets, a value is *unusual* if it differs from the mean by more than two standard deviations.

Using Standard Deviation

Note:

- The standard deviation measures the *variation* among data values.
- Values close together have a small standard deviation, but values with much more variation have a larger standard deviation.
- For many data sets, a value is *unusual* if it differs from the mean by more than two standard deviations.

Chebyshev's Inequality

Another concept that is helpful in understanding or interpreting a value of a standard deviation is Chebyshev's Inequality, which applies to *any* data set.

Chebyshev's Inequality

At least $1 - \frac{1}{k^2}$ of the distribution's values are within k standard deviations of the mean, where k is any positive number greater than 1.

For $k = 2$ and $k = 3$ we get:

- At least $\frac{3}{4}$ (or 75 %) of all values lie within 2 standard deviations of the mean.
- At least $\frac{8}{9}$ (or 89 %) of all values lie within 3 standard deviations of the mean.

Chebyshev's Inequality

Another concept that is helpful in understanding or interpreting a value of a standard deviation is Chebyshev's Inequality, which applies to *any* data set.

Chebyshev's Inequality

At least $1 - \frac{1}{k^2}$ of the distribution's values are within k standard deviations of the mean, where k is any positive number greater than 1.

For $k = 2$ and $k = 3$ we get:

- At least $\frac{3}{4}$ (or 75 %) of all values lie within 2 standard deviations of the mean.
- At least $\frac{8}{9}$ (or 89 %) of all values lie within 3 standard deviations of the mean.

Chebyshev's Inequality

Another concept that is helpful in understanding or interpreting a value of a standard deviation is Chebyshev's Inequality, which applies to *any* data set.

Chebyshev's Inequality

At least $1 - \frac{1}{k^2}$ of the distribution's values are within k standard deviations of the mean, where k is any positive number greater than 1.

For $k = 2$ and $k = 3$ we get:

- At least $\frac{3}{4}$ (or 75 %) of all values lie within 2 standard deviations of the mean.
- At least $\frac{8}{9}$ (or 89 %) of all values lie within 3 standard deviations of the mean.

Chebyshev's Inequality – Examples

Example

IQ scores have a mean of 100 and a standard deviation of 15. What can we conclude from Chebyshev's Inequality? ($k = 2$, $k = 3$)

Solution

$k = 2$:

$\mu - 2\sigma = 100 - 2 \times 15 = 70$
$\mu + 2\sigma = 100 + 2 \times 15 = 130$
$\Rightarrow$ At least 75% of IQs are within 2 standard derivations of the mean.

$k = 3$:

$\mu - 3\sigma = 100 - 3 \times 15 = 55$
$\mu + 3\sigma = 100 + 3 \times 15 = 145$
$\Rightarrow$ At least 89% of IQs are within 3 standard derivations of the mean.

Chebyshev's Inequality – Examples

Example

IQ scores have a mean of 100 and a standard deviation of 15. What can we conclude from Chebyshev's Inequality? ($k = 2$, $k = 3$)

Solution

$\mathbf{k = 2}$:

$\mu - 2\sigma = 100 - 2 \times 15 = 70$
$\mu + 2\sigma = 100 + 2 \times 15 = 130$
$\Rightarrow$ At least 75% of IQs are within 2 standard derivations of the mean.

$\mathbf{k = 3}$:

$\mu - 3\sigma = 100 - 3 \times 15 = 55$
$\mu + 3\sigma = 100 + 3 \times 15 = 145$
$\Rightarrow$ At least 89% of IQs are within 3 standard derivations of the mean.

Question 1(b) from 2018 exam paper

Computers from a particular company are found to last on average for three years without any hardware malfunction, with a standard deviation of two months. At least what percent of the computers last between 31 months and 41 months?

Solution

Mean months without malfunction: $\mu = 3 \times 12 = 36$.
Standard deviation in months: $\sigma = 2$.

$$\underline{31} = 36 - 5 = 36 - (2.5 \times 2) \Rightarrow 31 = \underline{\mu - 2.5\sigma}$$

$$\underline{41} = 36 + 5 = 36 + (2.5 \times 2) \Rightarrow 31 = \underline{\mu + 2.5\sigma}$$

By Chebyshev's Inequality, at least $1 - \frac{1}{2.5^2} = 0.84 = 84\%$ of the hardware last between 31 and 41 months.

Question 1(b) from 2018 exam paper

Computers from a particular company are found to last on average for three years without any hardware malfunction, with a standard deviation of two months. At least what percent of the computers last between 31 months and 41 months?

Solution

Mean months without malfunction: $\mu = 3 \times 12 = 36$.
Standard deviation in months: $\sigma = 2$.

$$\underline{31} = 36 - 5 = 36 - (2.5 \times 2) \Rightarrow 31 = \underline{\mu - 2.5\sigma}$$

$$\underline{41} = 36 + 5 = 36 + (2.5 \times 2) \Rightarrow 31 = \underline{\mu + 2.5\sigma}$$

By Chebyshev's Inequality, at least $1 - \frac{1}{2.5^2} = 0.84 = 84\%$ of the hardware last between 31 and 41 months.

Question 1(b) from 2018 exam paper

Computers from a particular company are found to last on average for three years without any hardware malfunction, with a standard deviation of two months. At least what percent of the computers last between 31 months and 41 months?

Solution

Mean months without malfunction: $\mu = 3 \times 12 = 36$.
Standard deviation in months: $\sigma = 2$.

$$\underline{31} = 36 - 5 = 36 - (2.5 \times 2) \Rightarrow 31 = \underline{\mu - 2.5\sigma}$$

$$\underline{41} = 36 + 5 = 36 + (2.5 \times 2) \Rightarrow 31 = \underline{\mu + 2.5\sigma}$$

By Chebyshev's Inequality, at least $1 - \frac{1}{2.5^2} = 0.84 = 84\%$ of the hardware last between 31 and 41 months.

Question 1(b) from 2018 exam paper

Computers from a particular company are found to last on average for three years without any hardware malfunction, with a standard deviation of two months. At least what percent of the computers last between 31 months and 41 months?

Solution

Mean months without malfunction: $\mu = 3 \times 12 = 36$.
Standard deviation in months: $\sigma = 2$.

$$\underline{31} = 36 - 5 = 36 - (2.5 \times 2) \Rightarrow 31 = \underline{\mu - 2.5\sigma}$$

$$\underline{41} = 36 + 5 = 36 + (2.5 \times 2) \Rightarrow 31 = \underline{\mu + 2.5\sigma}$$

By Chebyshev's Inequality, at least $1 - \frac{1}{2.5^2} = 0.84 = 84\%$ of the hardware last between 31 and 41 months.

Coefficient of Variation

When comparing variation in two different sets of data, the standard deviations should be compared only if:

- the two sets of data use the same scale and units,
- they have approximately the same mean.

If the means are substantially different, or if the sample uses different scales or measurement units, we can use the coefficient of variation (cv).
This measure describes the standard deviation relative to the mean and is usually expressed as a percentage:

Coefficient of Variation

$$cv = \frac{\sigma}{\bar{x}} \times 100\%$$

When two distributions are compared, the distribution with the largest coefficient of variation has the greatest dispersion.
Note: The coefficient of variation is often used as a measure for economic inequality.

Coefficient of Variation

When comparing variation in two different sets of data, the standard deviations should be compared only if:

- the two sets of data use the same scale and units,
- they have approximately the same mean.

If the means are substantially different, or if the sample uses different scales or measurement units, we can use the coefficient of variation (cv).
This measure describes the standard deviation relative to the mean and is usually expressed as a percentage:

Coefficient of Variation

$$cv = \frac{\sigma}{\bar{x}} \times 100\%$$

When two distributions are compared, the distribution with the largest coefficient of variation has the greatest dispersion.
Note: The coefficient of variation is often used as a measure for economic inequality.

Coefficient of Variation

When comparing variation in two different sets of data, the standard deviations should be compared only if:

- the two sets of data use the same scale and units,
- they have approximately the same mean.

If the means are substantially different, or if the sample uses different scales or measurement units, we can use the coefficient of variation (cv).

This measure describes the standard deviation relative to the mean and is usually expressed as a percentage:

Coefficient of Variation

$$cv = \frac{\sigma}{\bar{x}} \times 100\%$$

When two distributions are compared, the distribution with the largest coefficient of variation has the greatest dispersion.

Note: The coefficient of variation is often used as a measure for economic inequality.

Coefficient of Variation

When comparing variation in two different sets of data, the standard deviations should be compared only if:

- the two sets of data use the same scale and units,
- they have approximately the same mean.

If the means are substantially different, or if the sample uses different scales or measurement units, we can use the coefficient of variation (cv).

This measure describes the standard deviation relative to the mean and is usually expressed as a percentage:

Coefficient of Variation

$$\text{cv} = \frac{\sigma}{\overline{x}} \times 100\%$$

When two distributions are compared, the distribution with the largest coefficient of variation has the greatest dispersion.

Note: The coefficient of variation is often used as a measure for economic inequality.

Coefficient of Variation

When comparing variation in two different sets of data, the standard deviations should be compared only if:

- the two sets of data use the same scale and units,
- they have approximately the same mean.

If the means are substantially different, or if the sample uses different scales or measurement units, we can use the coefficient of variation (cv).
This measure describes the standard deviation relative to the mean and is usually expressed as a percentage:

Coefficient of Variation

$$\text{cv} = \frac{\sigma}{\overline{x}} \times 100\%$$

When two distributions are compared, the distribution with the largest coefficient of variation has the greatest dispersion.
Note: The coefficient of variation is often used as a measure for economic inequality.

Coefficient of Variation

Exercises

Calculate the coefficients of variation for the following data sets.

1. A data set of $\{100, 100, 100\}$ has constant values. Its standard deviation is 0 and its mean is 100.
   $\text{cv} = \frac{\sigma}{\mu} \times 100\% = \frac{0}{100} \times 100\% = 0\%$.
2. A data set of $\{90, 100, 110\}$ has more variability. Its standard deviation is 8.16 and its mean is 100.
   $\text{cv} = \frac{8.16}{100} \times 100\% = 8.16\%$.
3. A data set of $\{1, 5, 6, 8, 10, 40, 65, 88\}$ has more variability again. Its standard deviation is 30.78 and its average is 27.875.
   $\text{cv} = \frac{30.78}{27.875} \times 100\% = 110\%$.

Coefficient of Variation

Exercises

Calculate the coefficients of variation for the following data sets.

1. A data set of $\{100, 100, 100\}$ has constant values. Its standard deviation is 0 and its mean is 100.
   $\text{cv} = \frac{\sigma}{\mu} \times 100\% = \frac{0}{100} \times 100\% = 0\%.$
2. A data set of $\{90, 100, 110\}$ has more variability. Its standard deviation is 8.16 and its mean is 100.
   $\text{cv} = \frac{8.16}{100} \times 100\% = 8.16\%.$
3. A data set of $\{1, 5, 6, 8, 10, 40, 65, 88\}$ has more variability again. Its standard deviation is 30.78 and its average is 27.875.
   $\text{cv} = \frac{30.78}{27.875} \times 100\% = 110\%.$

Coefficient of Variation

Exercises

Calculate the coefficients of variation for the following data sets.

1. A data set of $\{100, 100, 100\}$ has constant values. Its standard deviation is 0 and its mean is 100.
   $\text{cv} = \frac{\sigma}{\mu} \times 100\% = \frac{0}{100} \times 100\% = 0\%$.
2. A data set of $\{90, 100, 110\}$ has more variability. Its standard deviation is 8.16 and its mean is 100.
   $\text{cv} = \frac{8.16}{100} \times 100\% = 8.16\%$.
3. A data set of $\{1, 5, 6, 8, 10, 40, 65, 88\}$ has more variability again. Its standard deviation is 30.78 and its average is 27.875.
   $\text{cv} = \frac{30.78}{27.875} \times 100\% = 110\%$.

Coefficient of Variation

Exercises

Calculate the coefficients of variation for the following data sets.

1. A data set of $\{100, 100, 100\}$ has constant values. Its standard deviation is 0 and its mean is 100.
   $\text{cv} = \frac{\sigma}{\mu} \times 100\% = \frac{0}{100} \times 100\% = 0\%.$
2. A data set of $\{90, 100, 110\}$ has more variability. Its standard deviation is 8.16 and its mean is 100.
   $\text{cv} = \frac{8.16}{100} \times 100\% = 8.16\%.$
3. A data set of $\{1, 5, 6, 8, 10, 40, 65, 88\}$ has more variability again. Its standard deviation is 30.78 and its average is 27.875.
   $\text{cv} = \frac{30.78}{27.875} \times 100\% = 110\%.$

Coefficient of Variation

Exercises

Calculate the coefficients of variation for the following data sets.

1. A data set of $\{100, 100, 100\}$ has constant values. Its standard deviation is 0 and its mean is 100.
   $\text{cv} = \frac{\sigma}{\mu} \times 100\% = \frac{0}{100} \times 100\% = 0\%$.
2. A data set of $\{90, 100, 110\}$ has more variability. Its standard deviation is 8.16 and its mean is 100.
   $\text{cv} = \frac{8.16}{100} \times 100\% = 8.16\%$.
3. A data set of $\{1, 5, 6, 8, 10, 40, 65, 88\}$ has more variability again. Its standard deviation is 30.78 and its average is 27.875.
   $\text{cv} = \frac{30.78}{27.875} \times 100\% = 110\%$.

Coefficient of Variation

Exercises

Calculate the coefficients of variation for the following data sets.

1. A data set of $\{100, 100, 100\}$ has constant values. Its standard deviation is 0 and its mean is 100.
   $\text{cv} = \frac{\sigma}{\mu} \times 100\% = \frac{0}{100} \times 100\% = 0\%.$
2. A data set of $\{90, 100, 110\}$ has more variability. Its standard deviation is 8.16 and its mean is 100.
   $\text{cv} = \frac{8.16}{100} \times 100\% = 8.16\%.$
3. A data set of $\{1, 5, 6, 8, 10, 40, 65, 88\}$ has more variability again. Its standard deviation is 30.78 and its average is 27.875.
   $\text{cv} = \frac{30.78}{27.875} \times 100\% = 110\%.$

Coefficient of Variation

Exercises

Calculate the coefficients of variation for the following data sets.

1. A data set of $\{100, 100, 100\}$ has constant values. Its standard deviation is 0 and its mean is 100.
   $\text{cv} = \frac{\sigma}{\mu} \times 100\% = \frac{0}{100} \times 100\% = 0\%$.
2. A data set of $\{90, 100, 110\}$ has more variability. Its standard deviation is 8.16 and its mean is 100.
   $\text{cv} = \frac{8.16}{100} \times 100\% = 8.16\%$.
3. A data set of $\{1, 5, 6, 8, 10, 40, 65, 88\}$ has more variability again. Its standard deviation is 30.78 and its average is 27.875.
   $\text{cv} = \frac{30.78}{27.875} \times 100\% = 110\%$.

Coefficient of Variation

Exercise

Given the following data set:

$$\{20, 20, 21, 24, 25, 28\}$$

Find the standard deviation and the coefficient of variation.

Solution

$$\mu = \frac{20+20+21+24+25+28}{6} = 23$$

$$\sigma = \sqrt{\frac{(20-23)^2+(20-23)^2+(21-23)^2+(24-23)^2+(25-23)^2+(28-23)^2}{5}}$$

$$= \sqrt{\frac{9+9+4+1+4+25}{5}} = \sqrt{\frac{52}{5}} \simeq 3.22$$

$$cv = \frac{\sigma}{\mu} \times 100\% = \frac{3.22}{23} \times 100\% = 14\%.$$

Coefficient of Variation

Exercise

Given the following data set:

$$\{20, 20, 21, 24, 25, 28\}$$

Find the standard deviation and the coefficient of variation.

Solution

$$\mu = \frac{20+20+21+24+25+28}{6} = 23$$

$$\sigma = \sqrt{\frac{(20-23)^2+(20-23)^2+(21-23)^2+(24-23)^2+(25-23)^2+(28-23)^2}{5}}$$

$$= \sqrt{\frac{9+9+4+1+4+25}{5}} = \sqrt{\frac{52}{5}} \simeq 3.22$$

$$\text{cv} = \frac{\sigma}{\mu} \times 100\% = \frac{3.22}{23} \times 100\% = 14\%.$$

Coefficient of Variation

Exercise

Given the following data set:

$$\{20, 20, 21, 24, 25, 28\}$$

Find the standard deviation and the coefficient of variation.

Solution

$$\mu = \frac{20+20+21+24+25+28}{6} = 23$$

$$\sigma = \sqrt{\frac{(20-23)^2+(20-23)^2+(21-23)^2+(24-23)^2+(25-23)^2+(28-23)^2}{5}}$$

$$= \sqrt{\frac{9+9+4+1+4+25}{5}} = \sqrt{\frac{52}{5}} \simeq 3.22$$

$$\text{cv} = \frac{\sigma}{\mu} \times 100\% = \frac{3.22}{23} \times 100\% = 14\%.$$

Coefficient of Variation

Exercise

Given the following data set:

$$\{20, 20, 21, 24, 25, 28\}$$

Find the standard deviation and the coefficient of variation.

Solution

$$\mu = \frac{20+20+21+24+25+28}{6} = 23$$

$$\sigma = \sqrt{\frac{(20-23)^2+(20-23)^2+(21-23)^2+(24-23)^2+(25-23)^2+(28-23)^2}{5}}$$

$$= \sqrt{\frac{9+9+4+1+4+25}{5}} = \sqrt{\frac{52}{5}} \simeq 3.22$$

$$\text{cv} = \frac{\sigma}{\mu} \times 100\% = \frac{3.22}{23} \times 100\% = 14\%.$$

Coefficient of Variation

Exercise

Given the following data set:

$$\{20, 20, 21, 24, 25, 28\}$$

Find the standard deviation and the coefficient of variation.

Solution

$$\mu = \frac{20+20+21+24+25+28}{6} = 23$$

$$\sigma = \sqrt{\frac{(20-23)^2+(20-23)^2+(21-23)^2+(24-23)^2+(25-23)^2+(28-23)^2}{5}}$$

$$= \sqrt{\frac{9+9+4+1+4+25}{5}} = \sqrt{\frac{52}{5}} \simeq 3.22$$

$$\text{cv} = \frac{\sigma}{\mu} \times 100\% = \frac{3.22}{23} \times 100\% = 14\%.$$

Question from 2008 exam paper

The scores of twenty-one exam students are

72	93	97	52	57	69	69
84	49	72	68	68	98	51
65	84	79	55	64	71	78

(i) Present the data using a stem-and-leaf display.

(ii) Present the data using a histogram with classes $0-9$, $10-19$, ..., $90-99$.

(iii) What is the median exam score? What is the upper quartile score and the lower quartile score?

(iv) Consider the sample consisting of the first five exam scores. What is the sample mean, what is the standard deviation and what is the coefficient of variation?

Question from 2008 exam paper

The scores of twenty-one exam students are

72	93	97	52	57	69	69
84	49	72	68	68	98	51
65	84	79	55	64	71	78

(i) Present the data using a stem-and-leaf display.

(ii) Present the data using a histogram with classes $0-9$, $10-19$, ..., $90-99$.

(iii) What is the median exam score? What is the upper quartile score and the lower quartile score?

(iv) Consider the sample consisting of the first five exam scores. What is the sample mean, what is the standard deviation and what is the coefficient of variation?

Question from 2008 exam paper

The scores of twenty-one exam students are

72	93	97	52	57	69	69
84	49	72	68	68	98	51
65	84	79	55	64	71	78

(i) Present the data using a stem-and-leaf display.

(ii) Present the data using a histogram with classes $0-9$, $10-19$, ..., $90-99$.

(iii) What is the median exam score? What is the upper quartile score and the lower quartile score?

(iv) Consider the sample consisting of the first five exam scores. What is the sample mean, what is the standard deviation and what is the coefficient of variation?

Question from 2008 exam paper

The scores of twenty-one exam students are

72	93	97	52	57	69	69
84	49	72	68	68	98	51
65	84	79	55	64	71	78

(i) Present the data using a stem-and-leaf display.

(ii) Present the data using a histogram with classes $0-9$, $10-19$, ..., $90-99$.

(iii) What is the median exam score? What is the upper quartile score and the lower quartile score?

(iv) Consider the sample consisting of the first five exam scores. What is the sample mean, what is the standard deviation and what is the coefficient of variation?

Question from 2008 exam paper

The scores of twenty-one exam students are

72	93	97	52	57	69	69
84	49	72	68	68	98	51
65	84	79	55	64	71	78

(i) Present the data using a stem-and-leaf display.

(ii) Present the data using a histogram with classes $0-9$, $10-19$, ..., $90-99$.

(iii) What is the median exam score? What is the upper quartile score and the lower quartile score?

(iv) Consider the sample consisting of the first five exam scores. What is the sample mean, what is the standard deviation and what is the coefficient of variation?