Overview
(/study/app
ai-
hl/sid-
132-
cid-
761618/ov

Teacher view

Table of
contents

Notebook

Glossary

Reading
assistance

4. Probability and statistics / 4.14 Variance

The big picture

Mathematical models are extremely important and useful in our lives. A few examples of how these models can be used include predicting the weather, analysing the flow of traffic within a city, and interpreting the possible effects of a company's latest earnings release. However, building a model to represent a real-life event is not easy. Consider the following video.

This weather forecasting model is actually accurate | Lloyd Treinish | ...
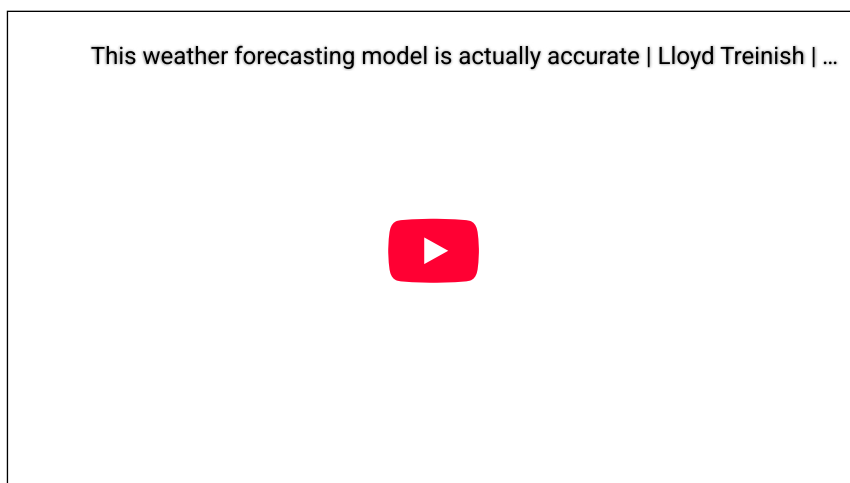
When building a model, mathematicians must consider the factors that might have an effect on the outcome. The more possible factors they take into account, the more likely it is that the model will make accurate predictions, but the harder it will be to build and the more unwieldy to use. For example, consider the possible factors that might influence a weather or climate model, as shown by the arrows in the diagram below.

Student
view

👁 More information
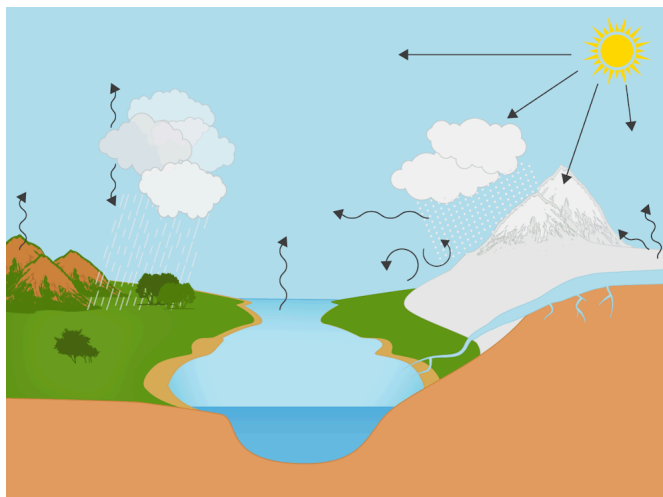
The diagram illustrates various environmental factors influencing weather and climate. On the left, there are mountains and vegetation with upwards arrows indicating evaporation. Nearby clouds are depicted releasing rain, signifying precipitation.

In the center, a body of water is shown, suggesting a lake or river contributing to local humidity and evaporation.

On the right, a snow-capped mountain is displayed with dotted arrows coming from the sun, indicating solar radiation and melting effects. Curved and straight arrows show wind patterns and air flow. Clouds above the mountain suggest different phases of cloud formation and precipitation. This comprehensive arrangement outlines the interacting components involved in climate models.

[Generated by AI]

After deciding on the factors to include, each of them is then worked into an equation that attempts to model the real-life outcomes of the weather.

🔑 **Concept**

Mathematicians analyse possible contributing factors to help find structure in seemingly random events. This analysis can be used for **modelling** that can help predict future events.

4. Probability and statistics / 4.14 Variance

Linear transformations of a single random variable

Linear transformations and expectation

🔗 **Making connections**

Recall from section 4.7.2 (/study/app/math-ai-hl/sid-132-cid-761618/book/expected-value-of-a-discrete-random-variable-id-26111/) that the notation $\mathrm{E}\left(X\right)$ refers to the expected value of a random discrete variable. This expectation is equal to the mean for a discrete random variable and can be calculated using the formula $\mathrm{E}\left(X\right)=\sum x\,\mathrm{P}\left(X=x\right).$

As you saw in The big picture (/study/app/math-ai-hl/sid-132-cid-761618/book/the-big-picture-id-27539/) for this subtopic, random variables are a part of many of the models used to represent real-life phenomena. When analysing these variables, mathematicians need to understand how changes to these variables will affect the overall model.
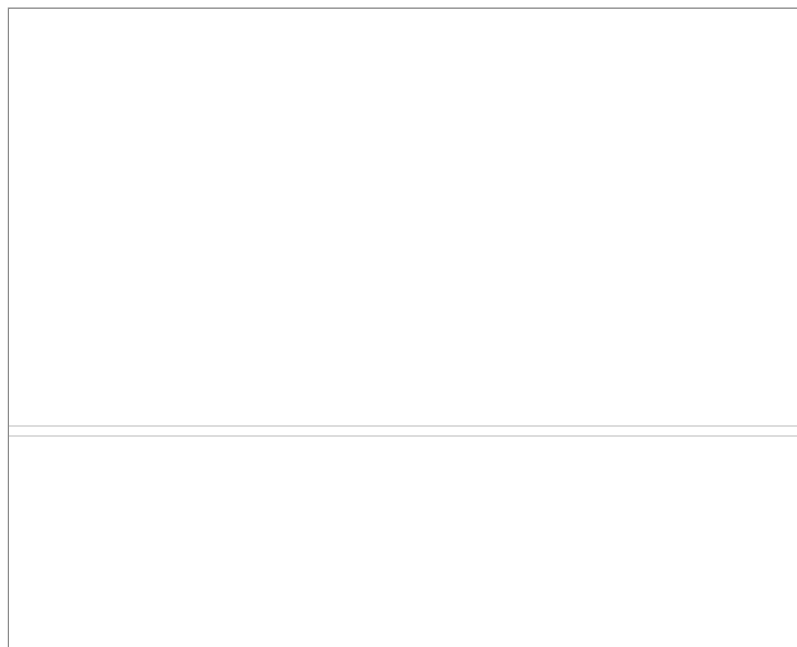
Consider, using the applet below, the simple situation of rolling a dice. Use the sliders to change the values of a and b. This will change the numbers written on the faces of the dice.

- Using a you can multiply the usual numbers by the same factor.
- Using b you can add the same amount to all numbers.

The formula that gives the numbers is $ax + b$, where x ranges from 1 to 6.

Interactive 1. Linear Transformations.

Credit: GeoGebra [↗](https://www.geogebra.org/m/JHg7VJUk) Juan Carlos Ponce Campuzano

👁 More information

This interactive probability simulator offers an engaging way to explore how linear transformations affect random variables through a virtual six-sided die experiment.

The screen is divided into two halves. The top half displays a graph of the xy-axis, with the x-axis representing the average and the y-axis representing the number of rolls ranging from 0 to 160. A blue dashed line is shown on the graph, representing $E(X)$. A checkbox labeled "values" highlights an individual die roll's coordinates on the graph.

On the bottom half, a "roll die" button enables users to simulate a single die roll and updates the graph. The "simulation" button enables users to run multiple die rolls automatically and updates the graph continuously. Users can also speed up the simulation or stop it with their respective buttons. The "New sample" button allows users to reset the simulation, clearing the graph and starting a new set of die rolls.

Users can manipulate two key parameters using horizontal sliders: a multiplicative factor (a) ranging from 1 to 10 and an additive constant (b) adjustable between 0 and 20, which transform the die's outcomes according to the equation $ax + b$. As users conduct simulated rolls, the tool dynamically plots the running average against the number of rolls, visually demonstrating how sample means converge to the theoretical expected value. The display updates in real-time to show how changing these parameters affects the distribution - for instance, setting $a = 3$ and $b = 4$ would produce face values of 7, 10, 13, 16, 19, and 22 with a corresponding expected value of $3(3.5) + 4 = 14.5$. Through this hands-on exploration, users gain concrete understanding of fundamental probability concepts including expectation, linear transformations ($E(aX + b) = aE(X) + b$), and the law of large numbers, making abstract statistical principles tangible and intuitive. The interactive nature of adjusting parameters and immediately observing their effects on both the numerical outcomes and graphical representation creates a powerful learning experience that reinforces theoretical knowledge through practical experimentation.

Section   Student...   (0/0)   📢 Feedback   🖨 Print (/study/app/math-ai-hl/sid-132-cid-761618/book/the-big-picture-id-27539/print/)   Assign ▼

Note that the expectation of the rolling of a regular six-sided dice is 3.5, or $\mathrm{E}(X) = 3.5$.

Run the simulation and see how the overall average of the rolls gets closer to this value as more rolls are completed.

The variable b in the applet can be changed to add the same amount to each side of the dice.

⚙ **Activity**

Change the value of b and take note of what happens to $\mathrm{E}(X)$.

Can you express the change in terms of $\mathrm{E}(X)$ and b?

How can you verify that the value given for $\mathrm{E}(X)$ in the applet is correct?

Example 1

★☆☆

Calculate the value of $\mathrm{E}(X)$ for one six-sided dice with values $2, 3, 4, 5, 6$ and 7, i.e. $b = 1$.

Steps	Explanation
Begin with the formula for $\mathrm{E}(X)$.	$\mathrm{E}(X) = \sum x\, \mathrm{P}(X = x)$
Substitute the new values for the sides into the formula. Note that the theoretical probability for each side does not change and is still $\dfrac{1}{6}$.	$\mathrm{E}(X) = 2 \times \dfrac{1}{6} + 3 \times \dfrac{1}{6} + 4 \times \dfrac{1}{6} + 5 \times \dfrac{1}{6} + 6 \times \dfrac{1}{6} + 7 \times \dfrac{1}{6}$ $\qquad = 4.5$ Therefore the value shown in the applet is correct.

Note that as b increases, $\mathrm{E}(X)$ increases by the same amount.

This can be represented by the formula $\mathrm{E}(X + b) = \mathrm{E}(X) + b$.

Note that $\mathrm{E}(X + b)$ means that expectation when the value of b is being added to **every possible outcome** of X.

The variable a in the applet gives you the ability to increase the values of the sides of the dice through multiplication.

⚙ **Activity**

Reset the applet and then change the value of a to 2.

Take note of what happens to $\mathrm{E}(X)$. Can you express the change in terms of $\mathrm{E}(X)$ and a?

Again, how can you verify that the value given for $\mathrm{E}(X)$ in the applet is correct?

Example 2

★☆☆

Calculate the value of $E(X)$ for a dice with side values 2, 4, 6, 8, 10, and 12, i.e. $a = 2$.

Steps	Explanation
Begin with the formula for $E(X)$.	$E(X) = \sum x \, P(X = x)$
Substitute the new values for the sides into the formula. Note that the theoretical probability for each side does not change and is still $\frac{1}{6}$.	$E(X) = 2 \times \dfrac{1}{6} + 4 \times \dfrac{1}{6} + 6 \times \dfrac{1}{6} + 8 \times \dfrac{1}{6} + 10 \times \dfrac{1}{6} + 12 \times \dfrac{1}{6}$ $= 7$ Therefore, the value shown in the applet when $a = 2$ is correct.

Note that as you increase the value of a, the value of $E(X)$ increases by a factor of a. Therefore, the formula $E(aX) = aE(X)$ can be used to demonstrate the effects of changing the value of a.

✓ **Important**

The expectation of a single random variable after a linear transformations   can be found using the formula $E(aX + b) = aE(X) + b$.

Linear transformations and variance

🔗 **Making connections**

In section 4.3.2 (/study/app/math-ai-hl/sid-132-cid-761618/book/measures-of-dispersion-id-26074/) you saw how variance , σ^2 , the square of the standard deviation is used to measure the spread of data in statistics.

The concept of the variance of a random variable is similar to the variance of a data set. It measures how likely it is for the random variable to be close to the mean. The following is the formal definition.

$$\mathrm{Var}(X) = E\left(\left(X - E(X)\right)^2\right)$$

ⓘ **Exam tip**

You will not need to know the formulae for variance for the exams.

✓ **Important**

When working with random variables, the notation $\text{Var}(X)$ is used for variance rather than σ^2. $\text{Var}(X)$ measures the variability of the random variable X.

In the following example the formula is used to find the variance. On exams you will not be expected to carry out this calculation, but it can be helpful for the understanding of the concept.

Example 3

★★☆

Find the variance of the outcomes of rolling a standard six-sided dice.

Steps	Explanation
The first step is to find the expected value.	$\text{E}(X) = \sum x\,\text{P}(X = x)$
Substitute the values for the sides into the formula. Note that the theoretical probability for each side is $\dfrac{1}{6}$.	$\text{E}(X) = 1 \times \dfrac{1}{6} + 2 \times \dfrac{1}{6} + 3 \times \dfrac{1}{6} + 4 \times \dfrac{1}{6} + 5 \times \dfrac{1}{6} + 6 \times \dfrac{1}{6}$ $= 3.5$
In the variance formula you need the values of the random variable $\left(X - \text{E}(X)\right)^2 = \left(X - 3.5\right)^2$	$\begin{array}{c\|cccccc} X & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline \left(X-3.5\right)^2 & 6.25 & 2.25 & 0.25 & 0.25 & 2.25 & 6.25 \end{array}$
The variance of X is the expected value of $\left(X - 3.5\right)^2$	$\text{Var}(X) = 6.25 \times \dfrac{1}{6} + 2.25 \times \dfrac{1}{6} + 0.25 \times \dfrac{1}{6}$ $\qquad + 0.25 \times \dfrac{1}{6} + 2.25 \times \dfrac{1}{6} + 6.25 \times \dfrac{1}{6}$ ≈ 2.92

Consider the similarity of the formula between the variance of a data set and the variance of a discrete random variable:

$$\sigma^2 = \frac{\sum f_i(x_i - \mu)^2}{n} = \sum \frac{f_i}{n}(x_i - \mu)^2$$
$$\text{Var}(X) = \sum p_i\left(x_i - \text{E}(X)\right)^2$$

This similarity allows you to use the calculator to find $\text{Var}(X)$ for a discrete random variable X.

✓ **Important**

If you store the possible outcomes and the corresponding probabilities in the memory of your calculator and run the one-variable statistics application, the calculator will tell you both the expected value and the standard deviation of a discrete random variable.

Student view

Example 4

★☆☆

Use your graphic display calculator to find the variance of the outcomes of rolling a six-sided dice with side values of $2, 3$, $4, 5, 6,$ and $7.$

Steps	Explanation
Enter the values 2, 3, 4, 5, 6, and 7 in a list and enter $\dfrac{1}{6}$ for the corresponding probabilities. Use the one-variable statistics application with the probabilities as frequencies.	$\sigma X = 1.70782\ldots$ $\mathrm{Var}\,(X) = \sigma^2 \approx 2.917$

Note that the variance for this transformation was the same value as the original variance found in **Example 3**. This can be generalised by saying that $\mathrm{Var}\,(X + b) = \mathrm{Var}\,(X)$.

Example 5

★☆☆

Use your graphic display calculator to find the variance of the outcomes of rolling a six-sided dice with side values of $2, 4$, $6, 8, 10,$ and $12.$

Steps	Explanation
Enter the values 2, 4, 6, 8, 10, and 12 in a list and enter $\dfrac{1}{6}$ for the corresponding probabilities. Use the one-variable statistics application with the probabilities as frequencies.	$\sigma X = 3.415650255$ $\mathrm{Var}\,(X) = \sigma^2 \approx 11.67$

Note that the answer from **Example 5** is four times larger than the variance for the dice with the original side values. In fact, this change in the variance can be summarised as $\mathrm{Var}\,(aX) = a^2 \mathrm{Var}\,(X)$.

✓ **Important**

The two linear transformations explored above for variance can be combined into one formula that is given in the formula booklet:

$$\mathrm{Var}\,(aX + b) = a^2 \mathrm{Var}\,(X)$$

4 section questions ⌄
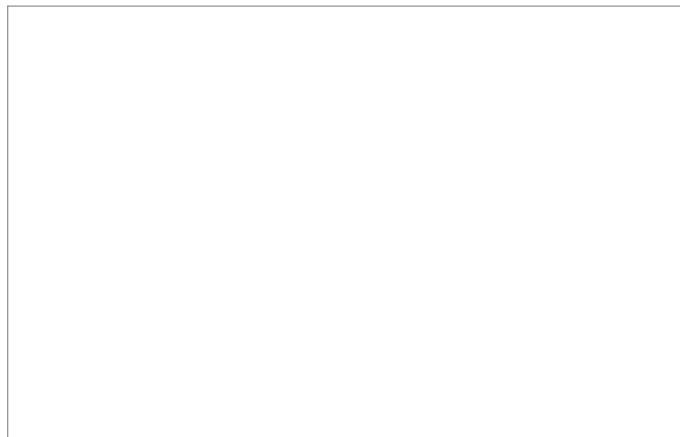
4. Probability and statistics / 4.14 Variance

Linear combinations of n random variables

⚙️ **Activity**

Consider the dice-rolling applet below.

Interactive 1. Linear Combinations of Random Variables in Dice Rolling.

👁️ More information for interactive 1

This interactive of probability exploration activity allows users to investigate the statistical properties of rolling four six-sided dice X_1, X_2, X_3, and X_4 through hands-on experimentation. The interface includes individual "Roll" buttons for each die, enabling users to roll each die independently. Participants can generate multiple data sets by rolling the dice and recording their sums, either as complete sets or individual results.

The equation below $(X_1 + X_2 + X_3 + X_4)$ is written dynamically and shows one possible outcome where the sum of all four dice is displayed.

Through this flexible experimentation, users discover how statistical measures behave and learn to predict outcomes using probability rules while gaining a practical understanding of the relationship between empirical results and theoretical expectations.

For example:
- Click "Roll" for X_1, and you might get 3.
- Click "Roll" for X_2, and you might get 5.
- Click "Roll" for X_3, and you might get 2.
- Click "Roll" for X_4, and you might get 4.
- The equation $X_1 + X_2 + X_3 + X_4 = 3 + 5 + 2 + 4 = 14$ shows one possible outcome where the sum of all four dice is 14.

Experiment by rolling the dice multiple times to see different sums and observe how random outcomes vary. This hands-on approach helps build intuition about probability and chance.

1. Roll each of the four dice shown above. Write down the sum. Repeat this until you have four sums written down.

2. Find the mean and variance of your four sums.

3. Repeat Steps 1 and 2.

4. Consider the following questions:

   a) How did the mean and variance of the first four sums compare to those of the second four sums?

   b) If you were to collect another four sums, what do you think the mean and

variance would be?

c) Can you think of a way to predict the sum of four dice?

What was your answer to the last question from the above activity? One possible solution would be to roll the four dice many times, collect many sums, and then calculate the mean and variance for your large data set. However, is there a more efficient way to find these results?

Rolling four dice

Credit: pixhook GettyImages

Consider the situation again. The roll of dice can be considered random and the roll of each dice is independent of the other three. Therefore, you have four independent random variables. The last question of the activity is essentially asking you to find the expected sum and variance of the rolls of four dice. This can be written as $\mathrm{E}\left(X_1 + X_2 + X_3 + X_4\right)$ and $\mathrm{Var}\left(X_1 + X_2 + X_3 + X_4\right)$.

✓   **Important**

The following can be used to find the expectation and variance of a linear combination of independent random variables:

$$\mathrm{E}\left(X_1 \pm X_2 \pm \ldots \pm X_n\right) = \mathrm{E}\left(X_1\right) \pm \mathrm{E}\left(X_2\right) \pm \ldots \pm \mathrm{E}\left(X_n\right)$$

$$\mathrm{Var}\left(X_1 \pm X_2 \pm \ldots \pm X_n\right) = \mathrm{Var}\left(X_1\right) + \mathrm{Var}\left(X_2\right) + \ldots + \mathrm{Var}\left(X_n\right)$$

Example 1

★☆☆

Find the mean and variance for the sum rolled by four unbiased six-sided dice.

⊗

Student
view

Using the formula for expectation shown above.

$$\mathrm{E}\left(X_1 + X_2 + X_3 + X_4\right) = \mathrm{E}\left(X_1\right) + \mathrm{E}\left(X_2\right) + \mathrm{E}\left(X_3\right) + \mathrm{E}\left(X_4\right)$$
$$= 3.5 + 3.5 + 3.5 + 3.5$$
$$= 14$$

Using the formula for variance.

$$\mathrm{Var}\left(X_1 + X_2 + X_3 + X_4\right) = \mathrm{Var}\left(X_1\right) + \mathrm{Var}\left(X_2\right) + \mathrm{Var}\left(X_3\right) + \mathrm{Var}\left(X_4\right)$$
$$= 2.91\bar{7} + 2.91\bar{7} + 2.91\bar{7} + 2.91\bar{7}$$
$$= 11.\bar{6}$$

🔗 Making connections

Recall the following two formulae from the previous section.

$$\mathrm{E}\left(aX + b\right) = a\mathrm{E}\left(X\right) + b \text{ and } \mathrm{Var}\left(aX + b\right) = a^2\mathrm{Var}\left(X\right)$$

Using these two formulae, you can now extend the formulae for linear combinations of random variables to the following:

✓ Important

For linear combinations of n independent random variables:

$$\mathrm{E}\left(a_1X_1 \pm a_2X_2 \pm \ldots \pm a_nX_n\right) = a_1\mathrm{E}\left(X_1\right) \pm a_2\mathrm{E}\left(X_2\right) \pm \ldots \pm a_n\mathrm{E}\left(X_n\right)$$

$$\mathrm{Var}\left(a_1X_1 \pm a_2X_2 \pm \ldots \pm a_nX_n\right) = a_1{}^2\mathrm{Var}\left(X_1\right) + a_2{}^2\mathrm{Var}\left(X_2\right) + \ldots + a_n{}^2\mathrm{Var}\left(X_n\right)$$

⚠ Be aware

The above formulae are only valid when the individual random variables are independent of each other.

Example 2

★☆☆

Riku creates a game based on rolling dice.

A player will roll a dice three times and their overall score will be calculated using the formula

$$3 \times 1\text{st roll} - 2 \times 2\text{nd roll} + 3\text{rd roll}$$

Calculate a player's expected score and the standard deviation of their score.

Using the expectation formula for a linear combination of random variables.

$$E\left(a_1 X_1 \pm a_2 X_2 \pm \ldots \pm a_n X_n\right) = a_1 E\left(X_1\right) \pm a_2 E\left(X_2\right) \pm \ldots \pm a_n E\left(X_n\right)$$

$$E\left(3X_1 - 2X_2 + X_3\right) = 3E\left(X_1\right) - 2E\left(X_2\right) + E\left(X_3\right)$$

$$= 3 \times 3.5 - 2 \times 3.5 + 3.5$$

$$= 7$$

Then using the variance formula.

$$\mathrm{Var}\left(a_1 X_1 \pm a_2 X_2 \pm \ldots \pm a_n X_n\right) = a_1{}^2 \mathrm{Var}\left(X_1\right) + a_2{}^2 \mathrm{Var}\left(X_2\right) + \ldots + a_n{}^2 \mathrm{Var}\left(X_n\right)$$

$$\mathrm{Var}\left(3X_1 - 2X_2 + X_3\right) = 3^2 \mathrm{Var}\left(X_1\right) + (-2)^2 \mathrm{Var}\left(X_2\right) + 1^2 \mathrm{Var}\left(X_3\right)$$

$$= 9 \times 2.91\bar{6} + 4 \times 2.91\bar{6} + 2.91\bar{6}$$

$$= 40.8\bar{3}$$

Then taking the square root to find the standard deviation:

$$\mathrm{Standard\ deviation} = \sqrt{\mathrm{Var}\left(3X_1 - 2X_2 + X_3\right)}$$

$$= \sqrt{40.8\bar{3}}$$

$$\approx 6.39$$

5 section questions ⌄

4. Probability and statistics / 4.14 Variance

Unbiased estimates of the mean and variance of a population

Estimating mean and variance

🔗 **Making connections**

Recall from section 4.1.1 (/study/app/math-ai-hl/sid-132-cid-761618/book/data-types-and-sources-id-26062/) that in statistics population refers to the entire group you are investigating. It is often difficult to collect data from the entire population, so you collect a sample of data instead. Also, recall the methods you studied in sections 4.1.2 (/study/app/math-ai-hl/sid-132-cid-761618/book/sampling-id-26063/) and 4.12.1 (/study/app/math-ai-hl/sid-132-cid-761618/book/valid-data-collection-and-analysis-id-27560/) to remove bias from the process of collecting a sample.

Consider a theoretical situation in which you have successfully removed all bias from the method you used to collect a sample. Could bias still be present within the estimates you make about the population using the sample data?

To explore this question, let us again consider the case of rolling two dice. However, for now, imagine that you do not know how many sides are on the dice. They could be 6-sided dice, 8-sided dice, 12-sided dice, etc. All you can see is the two values that are face up when the two dice are rolled. How well do you think the mean of those two values shown will estimate the actual mean of all of the sides of the dice?

Value of 1st dice

The mean of the two rolled dice

👁 More information

The diagram is a table showing the means of all possible pairs that can be rolled using two 6-sided dice. The rows represent the value of the second dice ranging from 1 to 6, while the columns represent the value of the first dice, also ranging from 1 to 6. Each cell in the table contains the mean of the two dice values for that pair. For instance, the mean for rolling a 5 on the first dice and a 4 on the second dice is highlighted as 4.5. This table helps visualize how the averages are calculated based only on the values rolled, without considering the total number of sides on each dice.

[Generated by AI]

The diagram above shows the means of all of the possible pairs that could be rolled if the two dice were 6-sided. In the situation circled, rolling a 5 on the first dice and a 4 on the second dice would give you a mean of 4.5. Note that arriving at this mean **only depends on the two values rolled**, not on how many sides there actually are on the dice.

So, how well does this value of 4.5 estimate the actual mean of rolling two 6-sided dice? (Remember, the actual mean of a 6-sided dice is 3.5.)

You may not think 4.5 is a very good estimate of 3.5. However, what if you considered all of the possible pairs that could be rolled?

Example 1

★☆☆

Calculate the mean of the means of rolling the two dice from the diagram above.

Using the mean formula.

$$\bar{x} = \sum_{i=1}^{k} \frac{f_i x_i}{n}$$

$$= \frac{1 \times 1 + 2 \times 1.5 + 3 \times 2 + 4 \times 2.5 + 5 \times 3 + 6 \times 3.5 + 5 \times 4 + 4 \times 4.5 + 3 \times 5 + 2 \times 5.5 + 1 \times 6}{36}$$

$$= 3.5$$

Note that if you were to collect all possible samples with two pieces of data, their average would be the actual mean of the population. The significance of this is that the sample mean is unbiased, meaning that it is a good estimate of the population mean. This is also true for any other sized sample.

✓ **Important**

The mean of a sample, $\bar{x}$, is an unbiased estimate of the mean of the population, μ , and is calculated using the formula $\bar{x} = \sum_{i=1}^{n} \dfrac{x_i}{n}$.

You have now seen that the mean of a sample of data is unbiased. However, what about the variance? Consider the information shown in the diagram below. The numbers in the table give the variance of the corresponding two element data set. For example, the circled value is the variance of the data set $\{4, 5\}$.

🚫 More information

The image is a table representing the variance outcomes for two rolled dice, with the x-axis labeled 'Value of 1st dice' and the y-axis labeled 'Value of 2nd dice.' Both axes range from 1 to 6, indicating possible dice values. Each cell in the table contains a numerical value representing the variance resulting from rolling a pair of dice showing the corresponding numbers. For example, the variance for the pair (1, 1) is 0, while the pair (4, 5) circled in the table has a variance of 0.25. The text on the right side of the table reads "The variance of the two rolled dice," pointing out that the numbers in the table are variance values for the respective dice combinations.

[Generated by AI]

Recall from section 4.14.1 (/study/app/math-ai-hl/sid-132-cid-761618/book/linear-transformations-of-a-single-random-variable-id-27540/) that the variance of the data set $\{1, 2, 3, 4, 5, 6\}$ is $2.91\overline{6}$. In the situation circled, rolling a 5 on the first dice and a 4 on the second would give you a variance of 0.25. This does not seem like a very good estimate of the actual variance. But again, what if you considered all of the possible pairs that could be rolled?

Example 2

★☆☆

Calculate the average variance from the two rolled dice in the diagram above.

Using the mean formula.

$$\bar{x} = \sum_{i=1}^{k} \frac{f_i x_i}{n}$$

$$= \frac{6 \times 0 + 10 \times 0.25 + 8 \times 1 + 6 \times 2.25 + 4 \times 4 + 2 \times 6.25}{36}$$

$$= 1.458\bar{3}$$

How does your result to **Example 2** compare with the actual variance of $2.91\bar{6}$ for an unbiased 6-sided dice? Note that the average variance of the samples is exactly half of the actual variance of an unbiased 6-sided dice! This inherent bias will occur any time you try to estimate the variance of a population with samples containing only two pieces of data.

What do you think would happen if you collected samples containing three pieces of data or more?

Section   Student... (0/0)   📢 Feedback   🖨 Print (/study/app/math-ai-hl/sid-132-cid-761618/book/linear-transformations-of-a-single-random-variable-id-27540/print/)   Assign

Interactive 1. Estimating Population Mean and Variance Using Dice Rolls.

Credit: GeoGebra ☑ (https://www.geogebra.org/m/JHg7VJUk) Juan Carlos Ponce Campuzano

👁 More information for interactive 1

This interactive simulation lets you explore statistical estimation through a hands-on dice experiment.

The screen is divided into two halves, on the top half a graph is presented with XY axes, with x-axis ranging from 0 to 160 representing the number of dice rolled and y-axis ranging from 0 to 5 representing average.

The interface features three primary controls: A 'Roll all 3 dice' button allows users to generate new rolls simultaneously, with each roll resulting in a red point being plotted on the graph. As the user continues clicking the button, additional points are added to the graph, which are connected by a line to form a curve. A "Simulation" button for automated repeated sampling, and a "New Sample" button to reset your experiment. Users can roll three dice together using the "Roll all 3 dice" button.

The tool tracks and displays two key aspects of your experiment: the running average of your dice rolls (showing how sample means estimate the population mean) and the variance across your rolls (demonstrating sample variability). As you conduct more rolls, a dynamic graph updates in real-time, illustrating the convergence of sample statistics toward the true population parameters (with the theoretical variance of 2.9167 displayed as a reference point for standard 6-sided dice).

❌ Section   Student... (0/0)   📢 Feedback   🖨 Print (/study/app/math-ai-hl/sid-132-cid-761618/book/linear-combinations-of-n-random-variables-id-27541/print/)   Assign

Student
view

Additionally, users can toggle the 'Values' checkbox on the top, to display individual roll results on the graph. After clicking the checkbox, a horizontal

slider appears to the right, which can be dragged to view the average for each roll, displaying both the number of rolls (n) and the corresponding average value."

This engaging setup transforms abstract statistical theory into concrete understanding by showing immediate results of your sampling experiments. The combination of manual rolling options and automated simulation features provides multiple ways to investigate the relationship between sample statistics and population parameters.

In the applet above, Note that average variance is trending towards a value of $1.9\bar{4}$. In fact, this value is two-thirds of the actual value the actual variance of $2.91\bar{6}$. Interestingly enough as you increase the number of dice being rolled, a pattern emerges which can be seen in the table.

Sample size	2	3	4	5	n
$\dfrac{\text{sample variance}}{\text{population variance}}$	$\dfrac{1}{2}$	$\dfrac{2}{3}$	$\dfrac{3}{4}$	$\dfrac{4}{5}$	$\dfrac{n-1}{n}$

So, the bias gets smaller as the sample size increases. However, how can you fully remove the bias from the sample variance? Let us use the information from the table:

$$\frac{\text{sample variance}}{\text{population variance}} = \frac{n-1}{n}$$
$$\text{sample variance} = \frac{n-1}{n} \times \text{population variance}$$
$$\frac{n}{n-1} \times \text{sample variance} = \text{population variance}$$
$$\frac{n}{n-1} s_n^2 = \sigma^2$$

Therefore, the value of $\dfrac{n}{n-1} s_n{}^2$ is an unbiased estimate of the population variance.

ⓘ **Exam tip**

The notation σ^2 is used when discussing the variance of a population and the notation $s_n{}^2$ is used when discussing the variance of a sample.

The notations σ and s_n are used similarly for standard deviation.

As the value for $\dfrac{n}{n-1} s_n{}^2$ is obtained by using a sample, we set it equal to $s_{n-1}{}^2$ instead of σ^2.

✓ **Important**

The value of $s_{n-1}{}^2$ is an unbiased estimate of the variance of a population and is calculated using the formula $s_{n-1}{}^2 = \dfrac{n}{n-1} s_n{}^2$.

Example 3

★☆☆

Aimee collected a sample with 50 pieces of data and found that the variance of the sample was 12.3954.

Use this information to calculate an unbiased estimate for the variance of the population.

Consider the information given.

$$n = 50$$
$$s_n{}^2 = 12.3954$$

Use the formula for $s_{n-1}{}^2$ to estimate the variance of the population.

$$s_{n-1}{}^2 = \frac{n}{n-1} s_n{}^2$$
$$= \frac{50}{49} \times 12.3954$$
$$= 12.6483\ldots$$

✓ **Important**

When given data in a frequency table the formula $s_{n-1}{}^2 = \sum_{i=1}^{n} \frac{f_i(x_i - \bar{x})^2}{n-1}$ can be used to calculate an estimate for the variance of the population, where $\bar{x}$ is the sample mean.

Example 4

★★☆

Raymond surveys students at a school in New York to find out the number of siblings they have. Calculate an estimate for the variance of the population using his data in the table below.

Number of siblings	Frequency
0	13
1	22
2	16
3	5

To use the above formula for frequency tables, you must first find the sample mean.

$$\bar{x} = \sum_{i=1}^{k} \frac{f_i x_i}{n}$$
$$= \frac{13 \times 0 + 22 \times 1 + 16 \times 2 + 5 \times 3}{56}$$
$$= 1.232142\ldots$$

Now you can substitute into the formula for $s_{n-1}{}^2$.

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

$$= \frac{13 \times (0 - 1.23\ldots)^2 + 22 \times (1 - 1.23\ldots)^2 + 16 \times (2 - 1.23\ldots)^2 + 5 \times (3 - 1.23\ldots)^2}{56 - 1}$$

$$= 0.836038\ldots$$

$$\approx 0.836$$

⚠ **Be aware**

Note that, in the working for **Example 4**, no value was rounded until the final answer.

In examinations, you should never round numbers in the middle of a calculation unless the question instructs you to round.

⊕ **International Mindedness**

In **Example 4**, you used data related to the number of students' siblings at a school in New York. The average size of a family is a value that can vary for different parts of the world. Some countries have taken steps to limit the size of families, while others have worked to increase it. As the population of the world continues to increase, what role do you think governments should take in managing population? Should statisticians play a part in such decisions?

4 section questions ⌄

4. Probability and statistics / 4.14 Variance

Checklist

| Section | Student... | (0/0) | 📢 Feedback | 🖨 Print (/study/app/math-ai-hl/sid-132-cid-761618/book/checklist-id-27543/print/) | Assign ▼ |

🗐 **What you should know**

By the end of this subtopic you should be able to:

- calculate the expectation and variance of a random variable after a linear transformation using the formulae
$\mathrm{E}\,(aX + b) = a\mathrm{E}\,(X) + b$ and $\mathrm{Var}\,(aX + b) = a^2\mathrm{Var}\,(X)$

- calculate the expectation and variance of the linear combinations of n independent random variables,
$X_1,\ X_2,\ \ldots,\ X_n$ using the formulae
$\mathrm{E}\,(a_1X_1 \pm a_2X_2 \pm \ldots \pm a_nX_n) = a_1\mathrm{E}\,(X_1) \pm a_2\mathrm{E}\,(X_2) \pm \ldots \pm a_n\mathrm{E}\,(X_n)$
and
$\mathrm{Var}\,(a_1X_1 \pm a_2X_2 \pm \ldots \pm a_nX_n) = a_1{}^2\mathrm{Var}\,(X_1) + a_2{}^2\mathrm{Var}\,(X_2) + \ldots$
$\qquad\qquad\qquad + a_n{}^2\mathrm{Var}\,(X_n)$

- state that the sample mean, $\bar{x}$, is an unbiased estimate of the population mean, μ

- calculate an unbiased estimate for the variance of a population using the formula
$s_{n-1}{}^2 = \dfrac{n}{n-1}s_n{}^2 = \sum_{i=1}^{k} \dfrac{f_i(x_i - \bar{x})^2}{n-1}$.

4. Probability and statistics / 4.14 Variance

Investigation

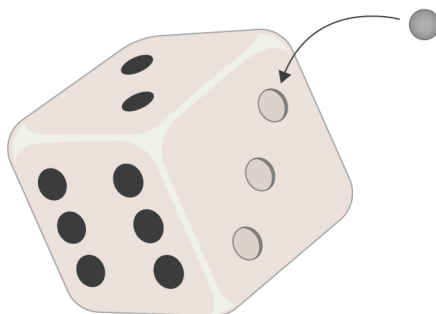Section   Student...   (0/0)   📢 Feedback   🖨 Print   (/study/app/math-ai-hl/sid-132-cid-761618/book/investigation-id-27544/print/)   Assign ▼

Loading dice to change the probability of the roll

In section 4.14.3 (/study/app/math-ai-hl/sid-132-cid-761618/book/unbiased-estimates-of-the-mean-and-variance-id-27542/) you learned about unbiased estimates of the variance of a population using calculations pertaining to a 6-sided dice.

Each side of a fair 6-sided dice has the same probability of facing upwards after being rolled.

However, do you think the rules for estimating the variance of a population hold true for a loaded dice with different probabilities?

1. Create your own probability distribution for a weighted dice and make the appropriate calculations to estimate the variance of the dice using a sample size of two rolls.
2. Devise a method to calculate the estimate using a sample size of three rolls.
3. Investigate how this may be extended to create the variance estimate using a sample size of n rolls.

Rate subtopic 4.14 Variance

Help us improve the content and user experience.

☆ ☆ ☆ ☆ ☆