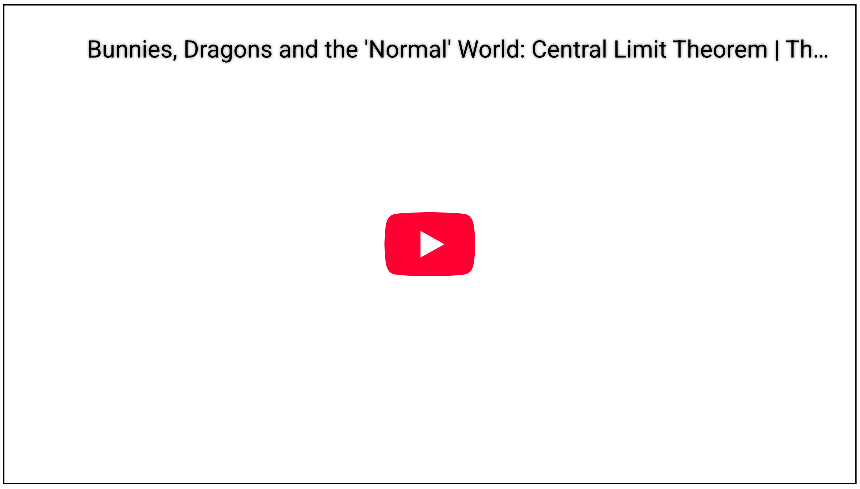
Index

The big picture

The word ‘normal’ can be misused and care should be taken to apply it properly. With over seven billion people in the world, finding the normal can be difficult. In fact, what does it even mean for something to be ‘normal’? Consider this question as you watch the following video.

You have encountered the normal distribution within your maths class, and perhaps in biology too. Examples of mass-produced items or things grown in nature are often used to demonstrate the power of the normal distribution. Could the normal distribution have applications beyond describing items that are produced in large quantities? Could you use it to find some normality in a seemingly random world?



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

Concept

Different probability distributions provide a representation of the relationship between the theory and reality. The central limit theorem allows us to use these distributions to make predictions about what might happen.

Theory of Knowledge

Central limit theorem provides for the consideration that phenomena that appear randomly distributed are actually normally distributed. This certainly has implications in a variety of contexts such as economics, psychology and sociology and brings up key issues in perception — if we perceive the world one way but upon further consideration through an alternative method discover the world is actually another way, what does that say about our initial perception and conclusion?

Knowledge Question: Does mathematical knowledge provide a complete understanding of phenomena?

4. Probability and statistics / 4.15 The central limit theorem

Linear combinations of normally distributed random variables

Section Student... (0/0) Feedback Print (/study/app/math-ai-hl/sid-132-cid-761618/book/linear-combinations-of-normally-distributed-random-id-27546/print/) Assign

Combinations of normally distributed variables

Making connections

Recall from subtopic 4.9 (/study/app/math-ai-hl/sid-132-cid-761618/book/the-big-picture-id-26125/) that $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$ denotes a random variable whose outcome follows a normal distribution defined by the mean, μ , and variance, σ^2 .

In section 4.14.2 (/study/app/math-ai-hl/sid-132-cid-761618/book/linear-combinations-of-n-random-variables-id-27541/) you studied linear combinations of independent random variables. Let us now explore how that concept can be applied to normally distributed random variables. Use the applet shown below to investigate what happens when two independent random variables, $X_1 \sim \mathrm{N}\left(\mu_1, {\sigma_1}^2\right)$ and $X_2 \sim N\left(\mu_2, {\sigma_2}^2\right)$, are added together.

Click the box next to ‘Collect random data’ to have the applet begin finding the sum of simultaneously collected random data from both the blue and orange data set. After resetting the applet, click on the ‘Sum’ button to have the applet find the difference of the data rather than the sum.

Student
view

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

Interactive 1. Investigating the Addition of Two Random Variables.

More information for interactive 1

This interactive tool allows users to explore how the sum or difference of two normally distributed random variables affects the resulting distribution's mean and standard deviation.

On the x-axis (ranging from 0 to 35), two independent normal distributions are shown:

- Yellow curve (X_1) with mean μ_1 and standard deviation σ_1.
- Blue curve (X_2) with mean μ_2 and standard deviation σ_2.

Users can modify the mean and standard deviation of these distributions by dragging the corresponding colored circles and squares along the axis. As changes are made, the graph and parameter values update dynamically.

When the "Collect random data" checkbox is selected, a simulation begins:

- A purple curve gradually forms above the original curves, representing the distribution of the sum $X_1 + X_2$ (or difference, depending on mode) with mean μ_3 and standard deviation σ_3.
- Users can adjust the speed of data collection using the "Slower" or "Faster" buttons.
- The "Sum" button toggles between summing and subtracting the distributions.
- The "New Data" button generates a new set of random values.
- The "Reset" button restores the tool to its default state.

This simulation illustrates key principles of combining normal distributions:

- Addition: The means add $\mu_3 = \mu_1 + \mu_2$, and the variances add $\sigma_3^2 = \sigma_1^2 + \sigma_2^2$.
- Subtraction: The means subtract $\mu_3 = \mu_1 - \mu_2$, but the variances still add.

Over time, users can observe how the empirical distribution converges to the theoretical normal curve, reinforcing understanding of how linear combinations of random variables behave—a foundational idea in statistical inference, measurement error analysis, and hypothesis testing.

Student view

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

How do the mean and standard deviation (see section 4.9.1 (/study/app/math-ai-hl/sid-132-cid-761618/book/the-normal-distribution-id-26126/)) of the new curve relate to the means and standard deviations of the original two random variables? How does the new curve change when you find the difference instead of the sum?

✓ **Important**

The linear combination of two independent, normally distributed random variables also follows a normal distribution.

Example 1

★☆☆

The radius of a wire, in cm, follows the distribution $X_1 \sim \mathrm{N}\left(0.5, 0.009\right)$. The wire is wrapped in a thermoplastic layer whose thickness, in cm, follows the distribution $X_2 \sim \mathrm{N}\left(0.2, 0.006\right)$.

Calculate the probability that the radius of a wrapped wire will be greater than 0.75 cm.

Steps	Explanation
$\begin{aligned}\mathrm{E}\left(X_1+X_2\right) &= \mathrm{E}\left(X_1\right)+\mathrm{E}\left(X_2\right)\\ &= \mu_{X_1}+\mu_{X_2}\\ &= 0.5+0.2\\ &= 0.7\end{aligned}$ $\begin{aligned}\mathrm{Var}\left(X_1+X_2\right) &= \mathrm{Var}\left(X_1\right)+\mathrm{Var}\left(X_2\right)\\ &= 0.009+0.006\\ &= 0.015\end{aligned}$ $\therefore$ t he radius of the wrapped wire follows the distribution $X \sim \mathrm{N}\left(0.7,\ 0.015\right)$	The radius of the wrapped wire is a linear combination of the independent X_1 and X_2 .
$\begin{aligned}\mathrm{P}\left(X>0.75\right) &= 0.341545\ldots\\ &\approx 34.2\%\end{aligned}$	Use your calculator to find the probability that the radius of the wrapped wire is greater than 0.75 cm.

Making connections

Instructions for finding probabilities with a graphic display calculator (GDC) are given in section 4.9.2 (/study/app/math-ai-hl/sid-132-cid-761618/book/the-normal-distribution-and-calculator-functions-id-26127/) .

Making connections

Recall from section 4.14.2 (/study/app/math-ai-hl/sid-132-cid-761618/book/linear-combinations-of-n-random-variables-id-27541/) that the expectation and variance of a linear combination of independent random variables can be found using the formulae:

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

Student view

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

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

Using these equations, you can now see the result of a linear combination of the two variables from the above applet. Given $X_1 \sim \text{N}\left(\mu_1, {\sigma_1}^2\right)$ and $X_2 \sim N\left(\mu_2, {\sigma_2}^2\right)$:

$$\begin{aligned} \text{E}\left(aX_1 \pm bX_2\right) &= a\text{E}\left(X_1\right) \pm b\text{E}\left(X_2\right) \\ &= a\mu_1 \pm b\mu_2 \end{aligned}$$

$$\begin{aligned} \text{Var}\left(aX_1 \pm bX_2\right) &= a^2\text{Var}\left(X_1\right) + b^2\text{Var}\left(X_2\right) \\ &= a^2{\sigma_1}^2 + b^2{\sigma_2}^2 \end{aligned}$$

✓ **Important**

Given two independent normal random variables $X_1 \sim \text{N}\left(\mu_1,\ {\sigma_1}^2\right)$ and $X_2 \sim N\left(\mu_2, {\sigma_2}^2\right)$, the distribution of their linear combination can be found using the following rule:

$$aX_1 \pm bX_2 \sim \text{N}\left(a\mu \pm b\mu,\ a^2{\sigma_1}^2 + b^2{\sigma_2}^2\right)$$

Example 2

★★☆

Let $X_1 \sim \text{N}\left(3.5, 1.44\right)$ and $X_2 \sim \text{N}\left(5, 1.21\right)$ be independent random variables.

Find $\text{P}\left(3X_1 + 2X_2 > 30\right)$.

Steps	Explanation
$aX_1 \pm bX_2 \sim \text{N}\left(a\mu \pm b\mu,\ a^2{\sigma_1}^2 + b^2{\sigma_2}^2\right)$ $3X_1 + 2X_2 \sim \text{N}\left(3 \times 3.5 + 2 \times 5,\ 3^2 \times 1.44 + 2^2 \times 1.21\right)$ $\sim \text{N}\left(20.5,\ 17.8\right)$	Begin by finding the distribution of the linear combination of the independent X_1 and X_2 .
$\text{P}\left(3X_1 + 2X_2 > 30\right) = 0.0121700\ldots$ $\approx 1.22\%$	Use your calculator to find the probability that an outcome of this in combination is greater than 30.

3 section questions ⌄

4. Probability and statistics / 4.15 The central limit theorem

Sampling a normally distributed random variable

Section Student... (0/0) Feedback Print (/study/app/math-ai-hl/sid-132-cid-761618/book/sampling-a-normally-distributed-random-variable-id-27547/print/) Assign

Student view

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

In the previous section you learned about how to work with linear combinations of one or more normally distributed random variables. For now, let us go back to considering only one random variable.

Consider the random variable $X_1 \sim \mathrm{N}\left(50, 100\right)$. If you were to collect samples of this variable, how would the mean of the samples compare with the mean of the random variable? Would they be the same? What about the variance? What is the range of sample means that you would expect?

Activity

Explore the questions from above with the applet below. To begin, note that the sample size is currently set at 4.

1. Observe the variety of values that appear for the sample mean and sample standard deviation as the simulation runs.
2. Let the simulation run for a few minutes. What shape is the distribution of the sample means beginning to look like?
3. Press pause and clear the sampling data. Change the sample size to 9 and press play. Have the variety of values for the mean and standard deviation changed from the original sample size?
4. Repeat the process from the previous step for sample sizes of 16, 25, and 100.
5. How does the sample size affect the mean and standard deviation of the samples? How can you explain these changes?

Interactive 1. Sampling a Normally Distributed Random Variable.
Credit: GeoGebra (https://www.geogebra.org/m/kUmJeEwx) Steve Phelps

More information for interactive 1

This interactive allows users to demonstrate fundamental statistical concepts through sampling experiments.

The graph displayed is a histogram with xy axes, x-axis ranging from 0 to 80 and y-axis ranging from 0 to 40, overlaid with a normal distribution curve.

On the graph, there are purple dots in a bell shaped curve representing the shape of a normal distribution. Users can explore how sample statistics behave when drawn from a normal distribution $\mathrm{N}\left(50, 10\right)$ by adjusting the sample size. The tool allows users to set different sample sizes (from small to large) to understand how this affects the distribution of sample means. A play button on the bottom left corner allows users to run a real time simulation. As users run simulations, red dots emerge from the x-axis representing the histogram of the sampled data. Users understand three key phenomena: the sample means form their own normal distribution centered at the population mean (50), the variability of these means decreases as sample size increases, and the shape of the distribution becomes more tightly clustered around the true mean.

Student view

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

The "Clear Sampling Data" button lets users reset experiments to compare different scenarios, while the live updating display shows exactly how many samples have been collected. This dynamic visualization makes abstract statistical principles like the Central Limit Theorem concrete by showing how larger samples yield more precise estimates of population parameters.

Important

As the sample means are plotted in the applet, you should notice that they too are normally distributed. As the sample size increases the distribution becomes narrower.

Making connections

From section 4.14.1 (/study/app/math-ai-hl/sid-132-cid-761618/book/linear-transformations-of-a-single-random-variable-id-27540/), recall that

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

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

From section 4.15.1 (/study/app/math-ai-hl/sid-132-cid-761618/book/linear-combinations-of-normally-distributed-random-id-27546/), recall that if X is normally distributed then

$$\underbrace{X+X+\cdots+X}_{n}\sim\mathrm{N}\left(n\mu,n\sigma^2\right).$$

Consider again a random variable, X, that is normally distributed. Since X is normally distributed you know that $\mathrm{E}\left(X\right)=\mu$ and $\mathrm{Var}\left(X\right)=\sigma^2$.

To understand the results of the interactive activity above, we will need to consider how to find the expectation of an average sample, $\mathrm{E}\left(\bar{X}\right)$, and the expected variance of an average sample, $\mathrm{Var}\left(\bar{X}\right)$.

Consider a sample that contains n random selections of the variable. The mean value of those selections, $\bar{X}$, can be calculated as follows:

$$\bar{X}=\frac{X_1+X_2+\ldots+X_n}{n}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

We can now substitute this value into $\mathrm{E}\left(\bar{X}\right)$:

Substituting for $\bar{X}$:	$\mathrm{E}\left(\bar{X}\right)=\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)$
Using $\mathrm{E}\left(aX+b\right)=a\mathrm{E}\left(X\right)+b$:	$\mathrm{E}\left(\bar{X}\right)=\frac{1}{n}\mathrm{E}\left(\sum_{i=1}^{n}X_i\right)$
Using $\mathrm{E(X+\cdots+X)=E(X)+\cdots+E(X)}$:	$\mathrm{E}\left(\bar{X}\right)=\frac{1}{n}\times n\mu=\mu$
Simplifying :	$\mathrm{E}\left(\bar{X}\right)=\mu$

Student view

Therefore, the expected mean of a sample is the same as the overall mean of the random variable. Take note that since n cancels out, this will be the case no matter what size the sample is; you saw this in your trials with the applet.

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

Following the same process, you can find a similar formula for $\mathrm{Var}\left(\bar{X}\right)$, with one important difference!

$$\mathrm{Var}\left(\bar{X}\right) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \times n\sigma^2 = \frac{\sigma^2}{n}$$

Note how the result for $\mathrm{Var}\left(\bar{X}\right)$ is dependent on the sample size, n. What happens to $\mathrm{Var}\left(\bar{X}\right)$ as the value of n increases? How does this relate to your experience using the applet above? As $\mathrm{Var}\left(\bar{X}\right)$ is related to how far off the sample mean could be from the actual mean of the random variable, what does this tell you about the importance of the sample size?

✓ **Important**

The expected mean and variance of an n- sized sample from a normally distributed random variable can be found using the following formulae:

$$\mathrm{E}\left(\bar{X}\right) = \mu \text{ and } \mathrm{Var}\left(\bar{X}\right) = \frac{\sigma^2}{n}$$

Exam tip

You can take the square root of $\mathrm{Var}\left(\bar{X}\right)$ to find the standard error of the sample mean. This value, $\frac{\sigma}{\sqrt{n}}$, can be used to measure the spread of the sample means around the actual population mean. As n increases, the standard error decreases.

Example 1

★★☆

A farmer is selling peaches at a market. The average weight of his peaches is 175 g and the standard deviation is 5 g. To transport and display his fruits, he fills boxes with 18 randomly chosen peaches each. A customer buys one of the boxes.

Find the probability that the customer gets more than 3.096 kg of fruit.

Steps	Explanation
$\bar{X}_{\text{minimum}} = \frac{3096}{18} = 172 \text{ grams}$	Begin by finding the minimum average needed to guarantee a total weight of 3.096 kg of fruit.
$\bar{X} \sim \mathrm{N}\left(\mu,\ \frac{\sigma^2}{n}\right) \sim \mathrm{N}\left(175,\ \frac{25}{18}\right)$	Find the distribution of an average sample of 18 peaches.

Student view

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

Steps	Explanation
$\mathrm{P}\left(\bar{X} > 172\right) = 0.994545227\ldots$ ≈ 0.995	Use your calculator to find the probability that the average is greater than 172 grams. Besides the lower bound, you will also need to enter a large number as upper bound, the mean, 175, and the standard deviation, $\sqrt{\frac{25}{18}}$.

3 section questions

4. Probability and statistics / 4.15 The central limit theorem

The central limit theorem

Section Student... (0/0) Feedback Print (/study/app/math-ai-hl/sid-132-cid-761618/book/the-central-limit-theorem-id-27548/print/) Assign

Making connections

In the previous section you saw that multiple n- sized samples of a normally distributed variable will also be normally distributed.

What do you think would happen if you took multiple samples of a variable that was not normally distributed? Consider the applet shown below:

Interactive 1. Different Distributions of Random Variable.

More information for interactive 1

This interactive simulation demonstrates the powerful concept of the Central Limit Theorem that allows users to explore how the distribution of sample means behaves for different sample sizes and different distributions.

Student view

The pink-colored histogram represents the frequency of data values across a range from 0 to 100. Users can select from various non-normal population distributions by clicking their respective checkboxes and understand how the sampling distribution of the mean changes as users adjust the sample size using a horizontal slider, ranging from 2 to 50. The “Draw distribution of sample means” button generates a sampling distribution of means based on the original data and the selected sample size in a blue histogram emerging from the x-axis. The users can generate a new set of data and update the histogram by clicking on the “New data” button in the top right corner.

When users begin with a sample size of 2, the distribution of sample means perfectly matches the user’s chosen non-normal population distribution. However, as the user increases the sample size using the slider, a remarkable transformation - the sampling distribution gradually becomes more symmetric and bell-shaped, regardless of the original population's distribution. The tool displays key parameters, mean and standard deviation and generates fresh random samples with each click of the "Draw distribution of sample means" button. As users set larger samples, the results tend to group more closely around the true average, which means the data gets more consistent and precise with bigger samples.

To the bottom left of the histogram is a row of distribution icons, each representing a different type of population distribution, a flat distribution, right-skewed distribution, normal distribution, left-skewed distribution, and bimodal distribution, where the flat distribution is currently selected.

This interactive helps users to understand why the Central Limit Theorem is so fundamental in statistics - it explains how we can make reliable inferences about population means even when dealing with non-normal data, as long as our sample sizes are sufficiently large. Try experimenting with different skewed or irregular distributions to see how the theorem holds true across various scenarios.

In the applet above, you can select the actual distribution of the random variable (bottom left). Choose one of the distributions that is not a normal distribution. Start with a sample size of 1 and click on ‘Draw’. What is the shape of the plot of the sample means? It is probably very close to the original distribution, as you might expect since the mean of a sample size of 1 will be the variable itself.

Now choose a sample size of 5 and click ‘Draw’. Does the plot look different? Explore and see what happens as you increase the sample size. Experiment with the different distributions for the random variable.

✓ **Important**

The central limit theorem tells us that for any random variable the distribution of the sample mean is approximately normal if the sample size is large enough.

Furthermore, if the mean, μ , and the variance, σ^2 , of the population are known, then the distribution of the sample mean can be modelled as $\bar{X} \sim \mathrm{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$.

Example 1

★☆☆

Let $X_1, X_2, \ldots, X_{45}$ be a random sample taken from a population with a mean of 27 and a standard deviation of 10.

Find $\mathrm{P}\left(25 < \bar{X} < 29\right)$, giving your answer in decimal form correct to 3 significant figures.

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

Steps	Explanation
$\bar{X} \sim \mathrm{N}\left(\mu,\ \frac{\sigma^2}{n}\right)$ $\sim \mathrm{N}\left(27,\ \frac{10^2}{45}\right)$	Since you are using a sample size of 45, the average sample can be modelled as follows.
$\mathrm{P}\left(25 < \bar{X} < 29\right) = 0.820287\ldots$ ≈ 0.820	Use your calculator to find the probability of the sample mean being between 25 and 29.

Example 2

★★☆

For the population in **Example 1** (mean 27, standard deviation 10), find the minimum value of n needed to guarantee that $\mathrm{P}\left(25 < \bar{X} < 29\right) > 0.9$.

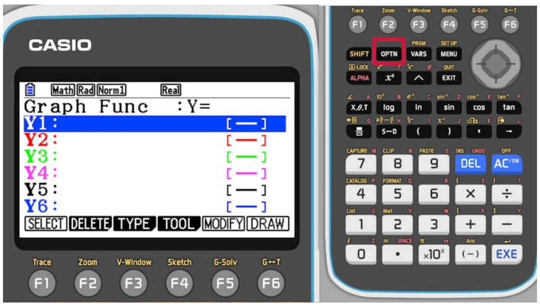
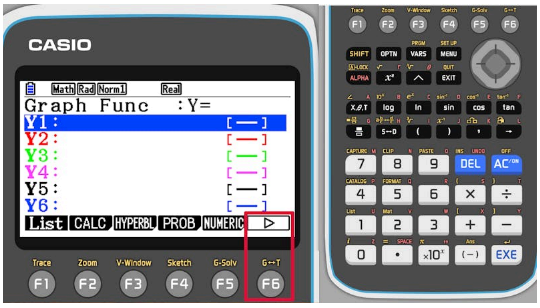
Making connections

For **Example 2**, you will need to use your GDC's Normal CDF function.

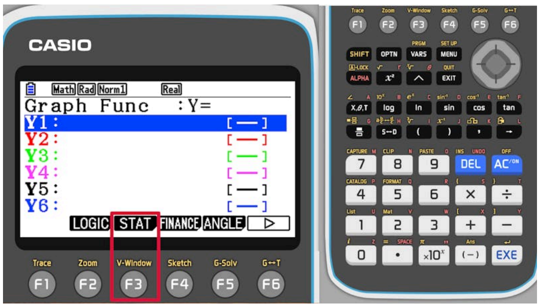
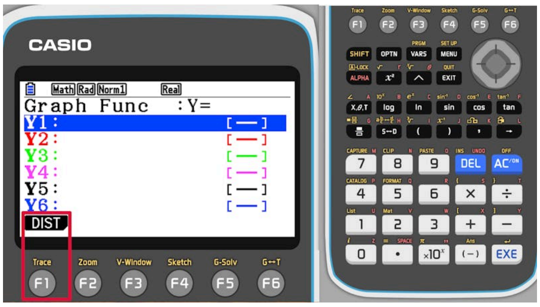
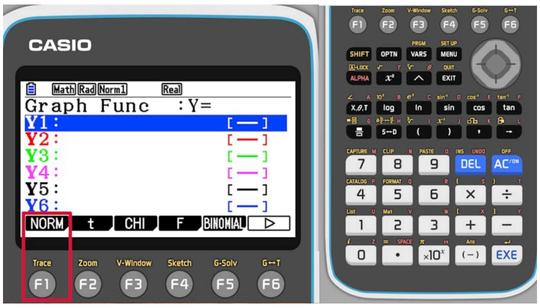
Instructions for using this function are given in Section 4.9.2 (/study/app/math-ai-hl/sid-132-cid-761618/book/the-normal-distribution-and-calculator-functions-id-26127/).

Steps	Explanation
$\bar{X} \sim \mathrm{N}\left(\mu,\ \frac{\sigma^2}{n}\right)$ $\sim \mathrm{N}\left(27,\ \frac{10^2}{n}\right)$	As in **Example 1**, begin by finding the model for the average sample.
$\mathrm{P}\left(25 < \bar{X} < 29\right) \geq 0.9$	Set up an inequality that can be solved in your calculator.
As the function crosses the line at $x = 67.638574$ the minimum value of n needed to guarantee that $\mathrm{P}\left(25 < \bar{X} < 29\right) > 0.9$ is $n = 68$.	Use your calculator's graphing function to find where graph crosses the line . See the instructions below for help on how to do this on your calculator

Student view

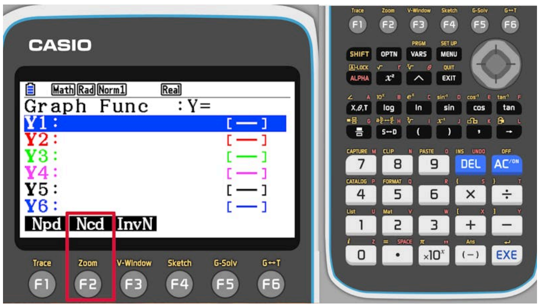
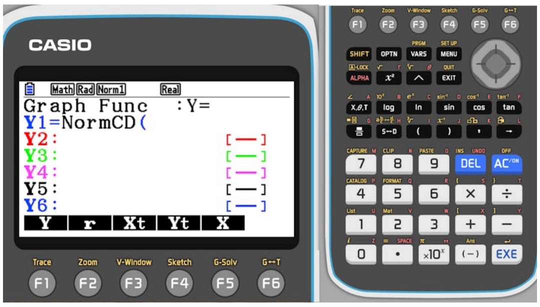
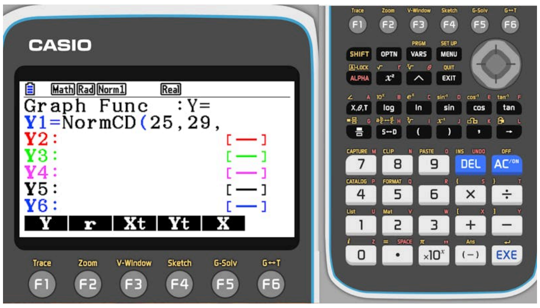
Steps	Explanation
In these instructions you will see how to graph the probability $$\mathrm{P}\left(25 < \bar{X} < 29\right)$$ as a function of the unknown size of the sample. To start, open the graph mode.	
You need to enter the probability as a function, so start looking for the tool. Press OPTN to open the options ...	
... press F6 to scroll to the options not on the first screen ...	

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

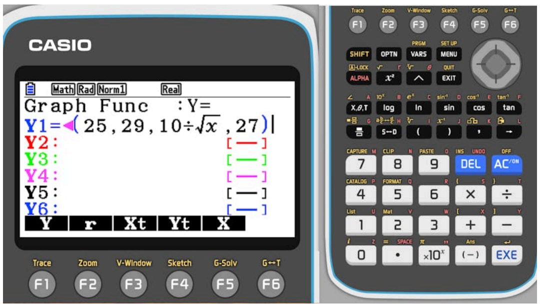
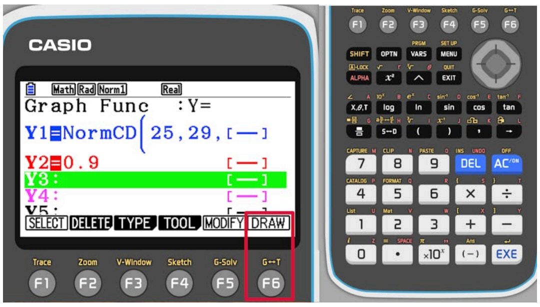
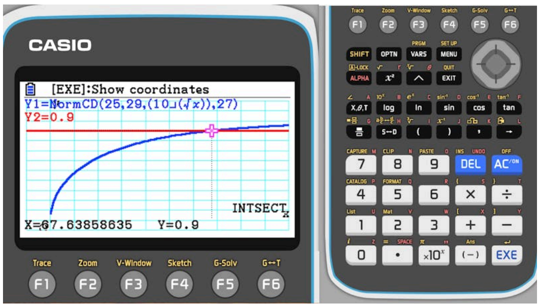
Steps	Explanation
... press F3 for the statistics related options ...	
... press F1 for the distributions ...	
... press F1 again for the normal distribution options ...	

Student
view

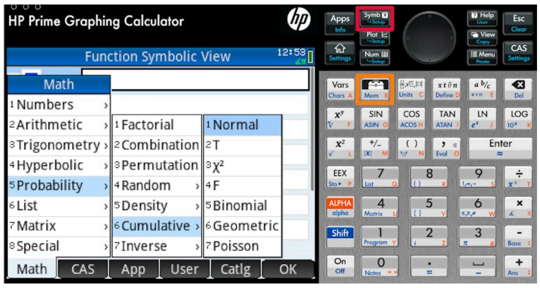
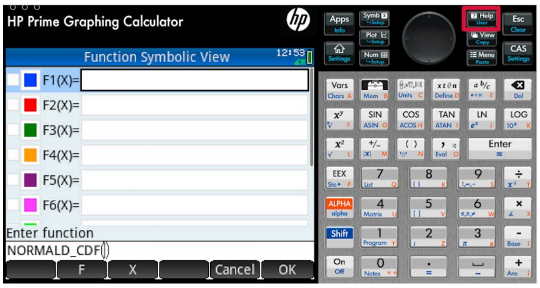
Overview
(/study/app/math-ai-hl/sid-132-cid-761618/ov

Steps	Explanation
... and finally, press F2 to access the cumulative distribution function of the normal distribution.	
At this point the calculator is waiting for the information about the specific question. There is no indication here of what the calculator is waiting for, so you need to remember.	
First you need to tell the lower and upper limits in the probability you are interested in $$\mathrm{P}\left(25 < \bar{X} < 29\right)$$ Separate the numbers by a comma.	

Student view

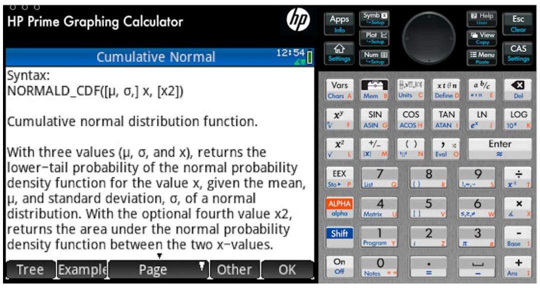
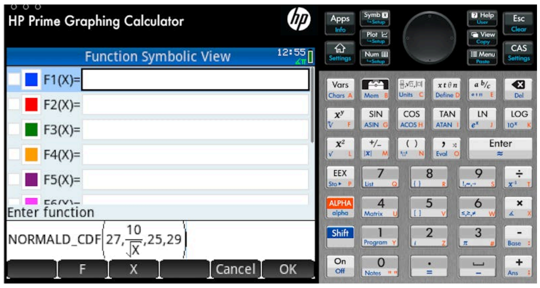
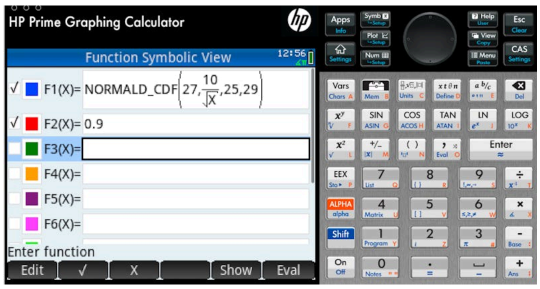
Steps	Explanation
Next, you need to tell the standard deviation and the mean of the normal distribution. Remember, the standard deviation depends on the sample size, which you don't know, so use the variable x here. The order is important 1. lower limit 2. upper limit 3. standard deviation 4. mean.	
Once the probability is entered as a function of the sample size, enter also the target probability in another function. Once these are done, press F6 to draw the graphs.	
You will need to adjust the window to see the intersection point. This is a good place to remind yourself on how to find the intersection point of two graphs.	

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

Steps	Explanation
In these instructions you will see how to graph the probability $$\mathrm{P}\left(25 < \bar{X} < 29\right)$$ as a function of the unknown size of the sample. To start, open the function application.	
You need to enter the probability as a function. In symbolic view open the toolbox and look for the cumulative distribution function of the normal distribution.	
At this point the calculator is waiting for the information about the specific question. You can open the help screen if you need to remind yourself on what the calculator is waiting for now.	

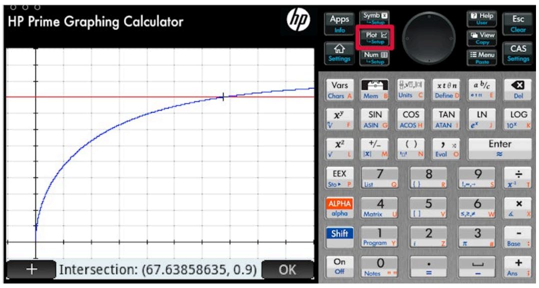
Student view

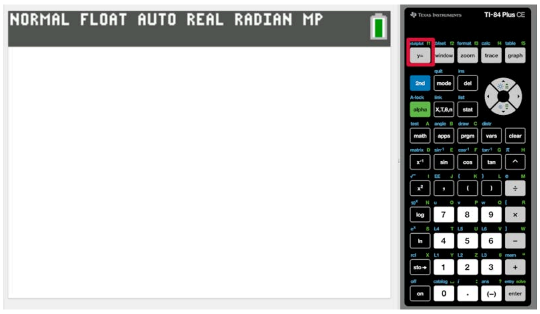
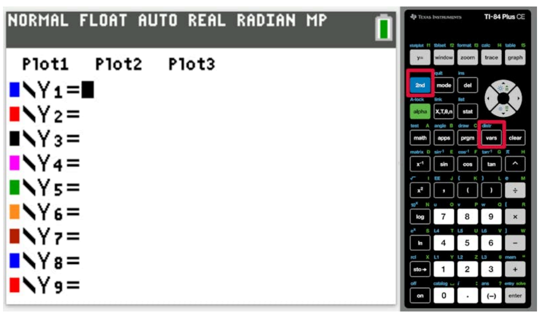
Overview (/study/ap ai-hl/sid-132-cid-761618/ov

Steps	Explanation
This is the help screen about the cumulative normal distribution function. Tap on OK when you have finished reading it.	
The first two numbers the calculator needs is the mean and the standard deviation of the normal distribution. Remember, the standard deviation depends on the sample size, which you don't know, so use the variable x here. Then you need to tell the lower and upper limits in the probability you are interested in $$\mathrm{P}\left(25 < \bar{X} < 29\right)$$	
Let's emphasise again, that the order of the numbers is important. You can always open the help screen if you forget. Once the probability is entered as a function of the sample size, enter also the target probability in another function.	

Student view

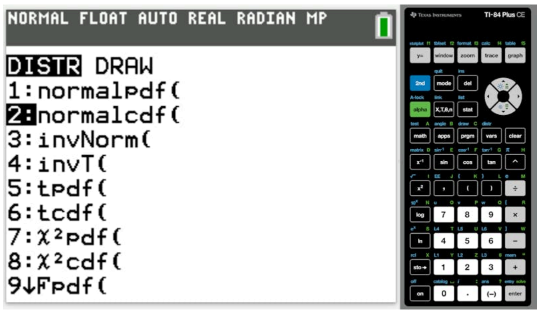
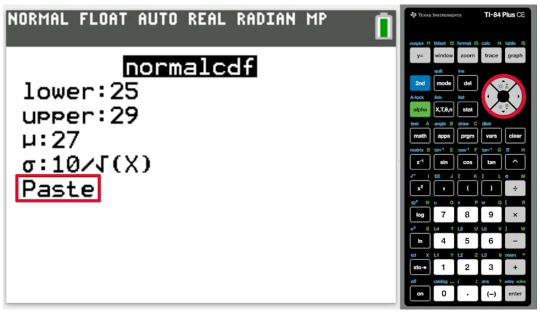
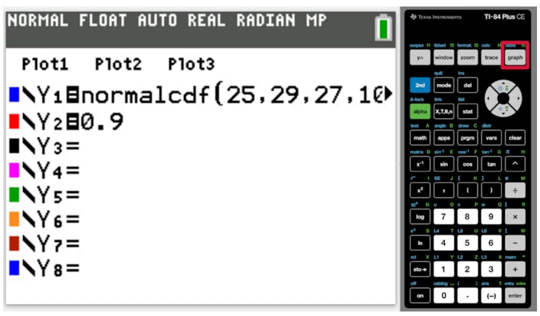
Overview
(/study/ap
ai-
hl/sid-
132-
cid-
761618/ov

Steps	Explanation
When you entered both functions, change to numeric view to see the graphs. You will need to adjust the window to see the intersection point. This is a good place to remind yourself on how to find the intersection point of two graphs.	

Steps	Explanation
In these instructions you will see how to graph the probability $\mathrm{P}\left(25 < \bar{X} < 29\right)$ as a function of the unknown size of the sample. To start, open the function editor menu.	
You need to enter the probability as a function, so start looking for the tool. Open the distribution menu ...	

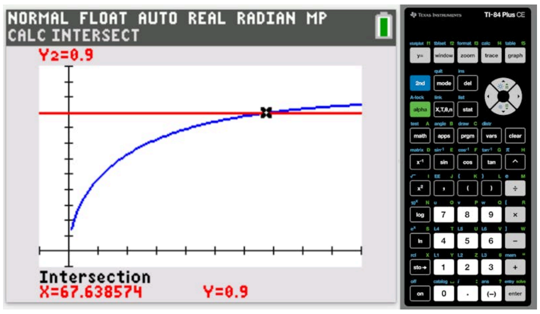
Student
view

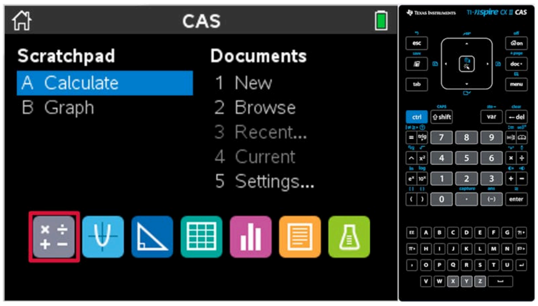
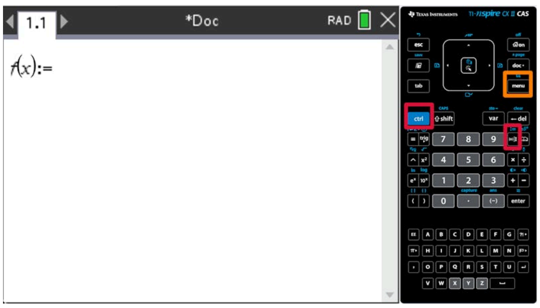
Overview
(/study/ap
ai-
hl/sid-
132-
cid-
761618/ov

Steps	Explanation
... and select the cumulative normal distribution function.	
At this point the calculator is waiting for the information about the specific question. You need to tell the lower and upper limits in the probability you are interested in $\mathrm{P}\left(25 < \bar{X} < 29\right)$ You also need to tell the mean and the standard deviation of the normal distribution. Remember, the standard deviation depends on the sample size, which you don't know, so use the variable x here. Once done, scroll down and paste the information.	
Once the probability is entered as a function of the sample size, enter also the target probability in another function. Once these are done, press the graph button to draw the graphs.	

Student
view

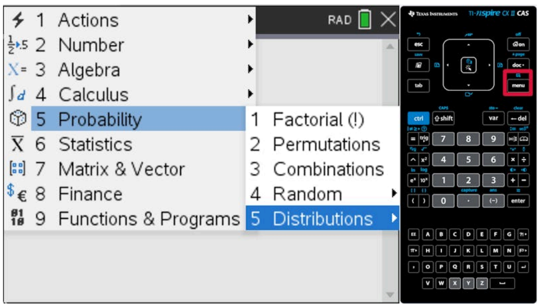
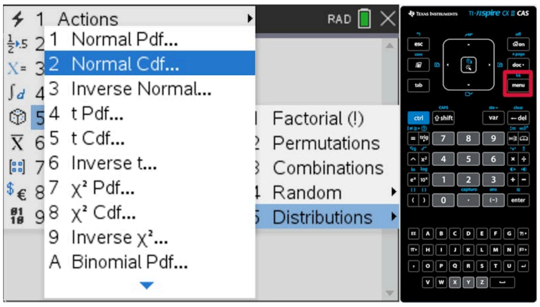
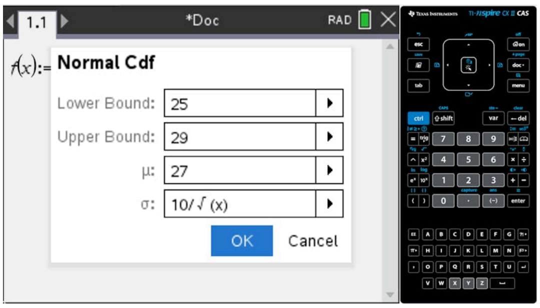
Overview (/study/ap ai-hl/sid-132-cid-761618/ov

Steps	Explanation
You will need to adjust the window to see the intersection point. This is a good place to remind yourself on how to find the intersection point of two graphs.	

Steps	Explanation
In these instructions you will see how to graph the probability $\mathrm{P}\left(25 < \bar{X} < 29\right)$ as a function of the unknown size of the sample. To start, open a calculator page.	
You need to enter the probability as a function. Remember to use the colon equal symbol for the definition. To start looking for the tool, open the menu ...	

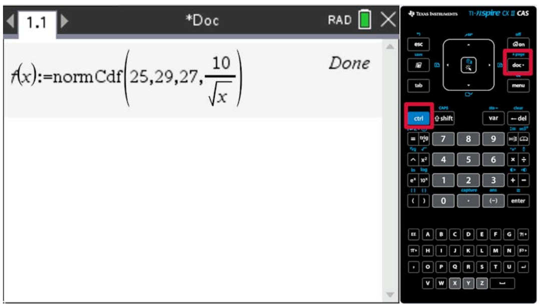
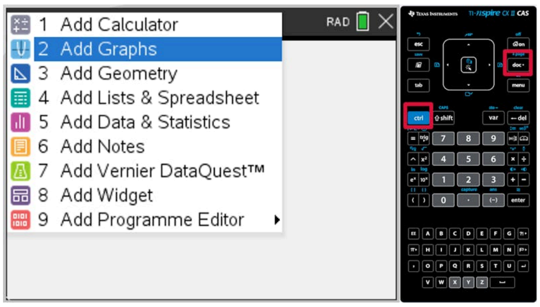
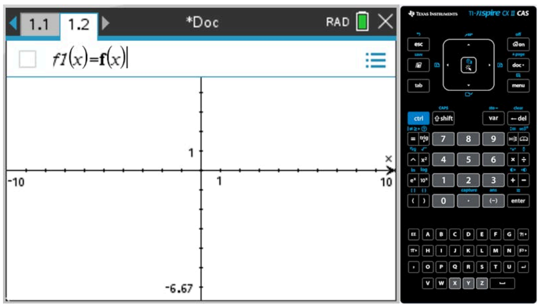
Student view

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

Steps	Explanation
... search for the probability distributions ...	
... and choose the cumulative distribution function of the normal distribution.	
At this point the calculator is waiting for the information about the specific question. You need to tell the lower and upper limits in the probability you are interested in $$\mathrm{P}\left(25<\bar{X}<29\right)$$ You also need to tell the mean and the standard deviation of the normal distribution. Remember, the standard deviation depends on the sample size, which you don't know, so use the variable x here.	

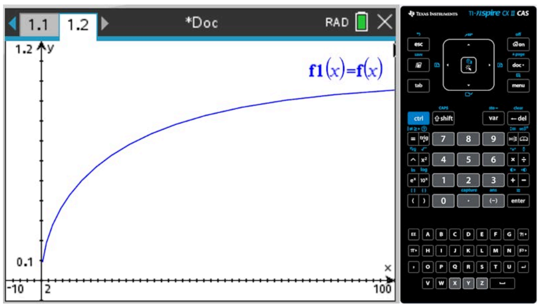
Student view

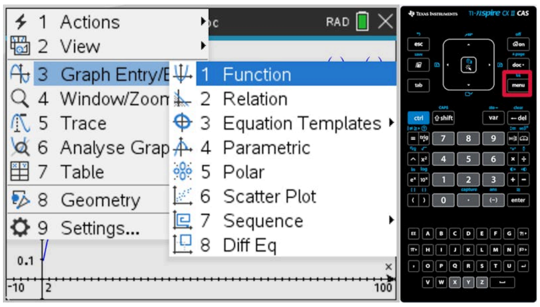
Overview
(/study/app/math-ai-hl/sid-132-cid-761618/ov

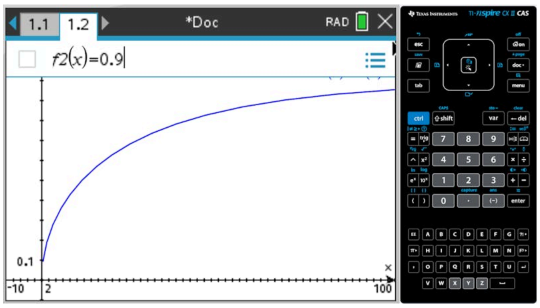
Steps	Explanation
At this point, the probability is defined as a function of the unknown sample size. To see the graph, add a new page to your document ...	
... choose to add a graph page.	
Tell the calculator, that the first function you want to graph is the one you just defined. Use the name you gave to the function.	

Student view

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

<table>
<tr><th>Steps</th><th>Explanation</th></tr>
<tr><td>You will need to adjust the window to see the relevant part of the graph.</td><td>


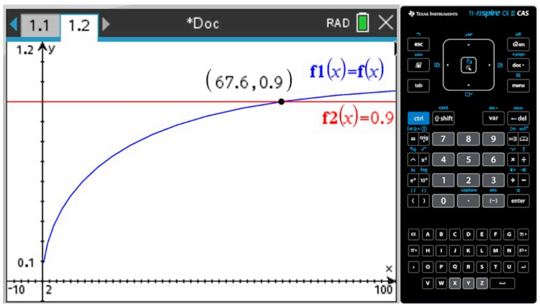
</td></tr>
<tr><td>To enter the target probability, open the menu and add a new function.</td><td>


</td></tr>
<tr><td>The question asked when the probability reaches 0.9.</td><td>


</td></tr>
</table>

Student
view

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

Steps	Explanation
You may need to readjust the window to see the intersection point. This is a good place to remind yourself on how to find the intersection point of two graphs.	

Be aware

In the definition above, it was stated that the sample needs to be large enough. This is a vague description because the necessary sample size depends on the original distribution of the random variable. For distributions that are already symmetrical, smaller sample sizes will work. Larger sample sizes are needed for the distributions that are not symmetrical.

Exam tip

In examinations, a sample size of $n > 30$ will be considered sufficient for any distribution of a random variable.

Example 3

★★★

The number of times that Ronnie goes to bed before midnight each week can be modelled with the random variable $X \sim \mathrm{B}\left(7, \frac{1}{3}\right)$. Find the probability that in a 52 week period (which is close to one year), Ronnie averages going to bed before midnight at least twice a week.

Steps	Explanation
$\mathrm{E}(X) = np = 7 \times \frac{1}{3} = \frac{7}{3}$ $\mathrm{Var}(X) = np(1-p) = 7 \times \frac{1}{3} \times \frac{2}{3} = \frac{14}{9}$	Recall from **section 4.8.1** how to find the mean and variance of a binomial distribution.

Student view

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

Steps	Explanation
$\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$ $\sim \mathrm{N}\left(\frac{7}{3}, \frac{\frac{14}{9}}{52}\right)$ $\sim \mathrm{N}\left(\frac{7}{3}, \frac{14}{468}\right)$	Since there are 52 weeks in one year, the average week can be modelled by.
$\mathrm{P}\left(\bar{X} > 2\right) = 0.9730257\ldots$ $\approx 97.3\%$	Use your calculator to find the probability that Ronnie averages going to bed early more than 2 nights each week.

International Mindedness

Sleep is a requirement for every human being. However, the way you sleep may very well depend on the culture in which you were raised. The time you go to bed, whether you nap during the day, whether you expect to sleep right through the night, where you sleep and what you sleep on can vary between different countries.

The big idea for this subtopic encouraged you to think about the idea of 'normal'. Continuing with that idea, it may be that the sleeping habits you consider to be normal could be strange to someone in a different culture!

3 section questions

4. Probability and statistics / 4.15 The central limit theorem

Checklist

Section Student... (0/0) Feedback Print (/study/app/math-ai-hl/sid-132-cid-761618/book/checklist-id-27549/print/) Assign

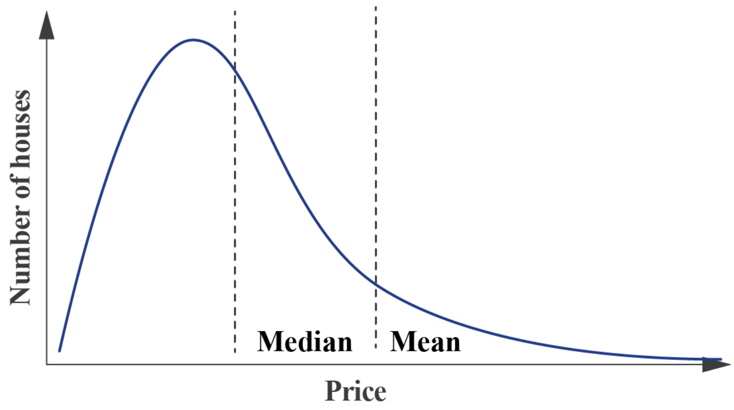
What you should know

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

- state that a linear combination of n independent normal random variables is normally distributed
- find the distribution of a combination of normal random variables by using the rule $aX_1 \pm bX_2 \sim \mathrm{N}\left(a\mu \pm b\mu,\ a^2{\sigma_1}^2 + b^2{\sigma_2}^2\right)$
- calculate the expectation and variance of an average random sample of a normally distributed random variable using the formulae $\mathrm{E}\left(X\right) = \mu$ and $\mathrm{Var}\left(X\right) = \frac{\sigma^2}{n}$
- use the central limit theorem to calculate probabilities that random samples will fall within specified ranges
- state that the necessary sample size depends on the distribution of the population. However, for most cases, a sample size of $n > 30$ is sufficient.

Student view

Investigation



More information

The graph illustrates the distribution of house prices and is skewed to the right. It features the number of houses on the Y-axis and the price on the X-axis. The Y-axis does not display specific numerical values or units, while the X-axis represents price with annotations for 'Median' and 'Mean'. The graph line begins at the origin, rises sharply indicating a higher number of houses at lower prices, peaks, and then gradually declines, with the line extending further as prices increase. The median is shown at a lower price point than the mean, highlighting the skewness where a small number of very expensive homes pull the mean price higher, distorting the average house price calculations.

[Generated by AI]

The graph shown above is a distribution of house prices that is typical for most places in the world. The distribution is skewed as the price of a small number of very expensive homes pulls the mean price away from the median price. This skewness means that the median home price of an area often gives a clearer picture of its typical housing cost than the mean price.

Activity

In this investigation, you will use the central limit theorem to explore the effects of sample size on the distribution of random samples of houses within a particular area of the United Kingdom.

1. Go to the website HM Land Registry (https://landregistry.data.gov.uk/app/ppd) and search for houses in a certain area of the United Kingdom by entering either the name of the town/ctiy or a postcode.
2. Download your data and then design a sampling method that allows you to randomly choose a certain number of houses from your area.
3. Consider factors such as type of house, the date/year of transaction and sample size when designing your method.
4. Find the median and mean house prices for you method and compare.
5. Repeat your experiment several times. Comment on your findings.

6. Change the sample size and repeat the experiment again several times. Comment on your findings.