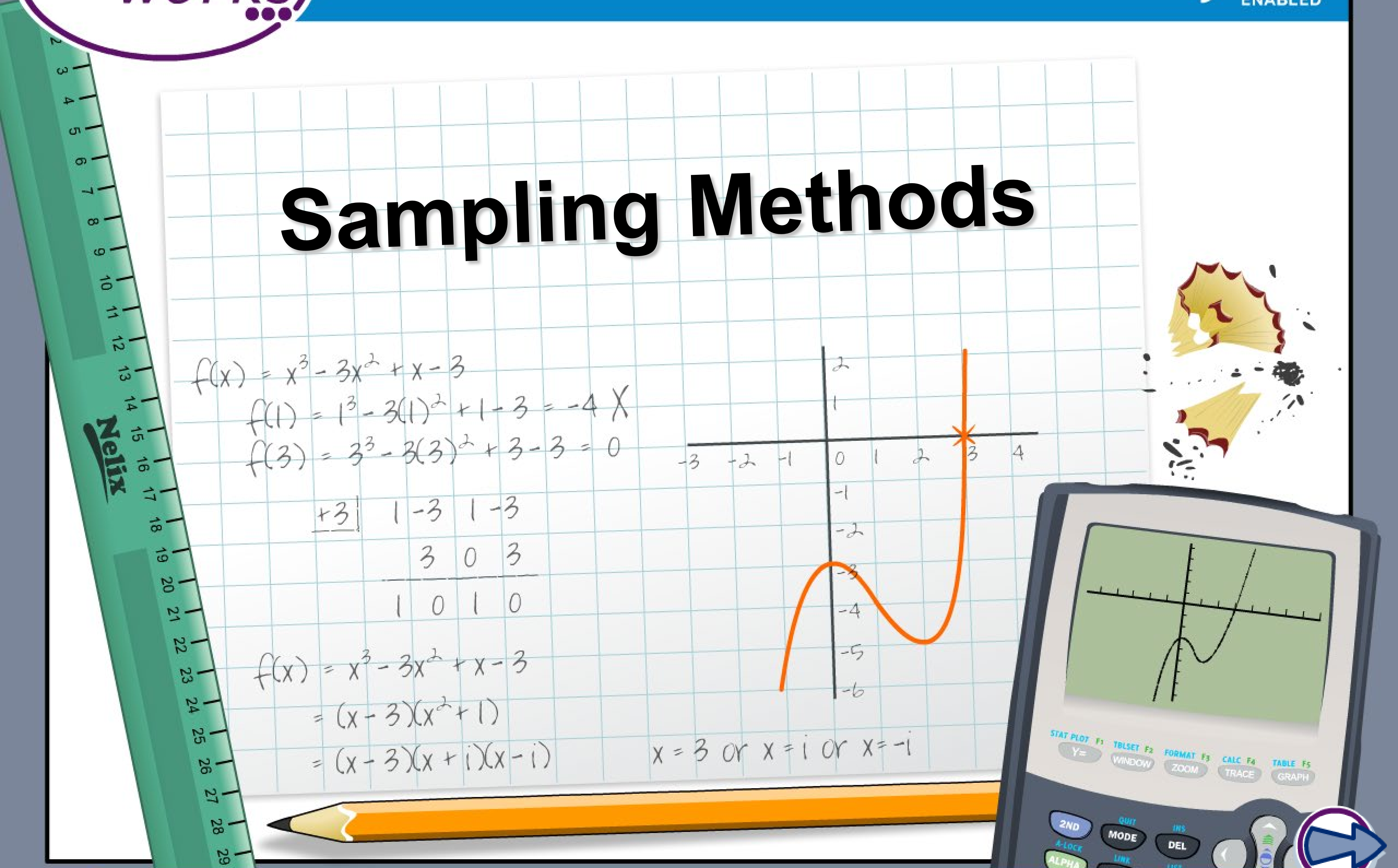
Sampling Methods



Information

Common core icons

This icon indicates a slide where the Standards for Mathematical Practice are being developed. Details of these are given in the Notes field.

MODELING

Slides containing examples of mathematical modeling are marked with this stamp.

This icon indicates an opportunity for discussion or group work.

The **Standards for Mathematical Practice** outlined in the Common Core State Standards for Mathematics describe varieties of expertise that mathematics educators at all levels should seek to develop in their students.

These are:

1) Make sense of problems and persevere in solving them.

2) Reason abstractly and quantitatively.

3) Construct viable arguments and critique the reasoning of others.

4) Model with mathematics.

5) Use appropriate tools strategically.

6) Attend to precision.

7) Look for and make use of structure.

8) Look for and express regularity in repeated reasoning.

This icon indicates that the slide contains activities created in Flash. These activities are not editable.

This icon indicates teacher's notes in the Notes field.

Populations

A **population** is the group of all the individuals fitting a description.

Give examples of hypotheses about the populations below.

"male high school students living in Minnesota"

"cars built between 1999 and 2005"

"pet cats belonging to adults with children"

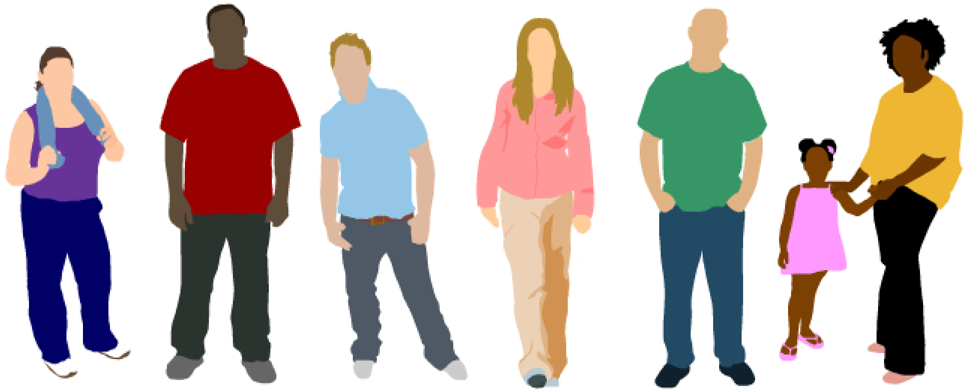
How would you gather data about these populations to investigate your hypotheses?

Would you be able to get information about every single member of the population?

Sampling

When investigating a hypothesis, it is usually more practical to gather data from a **sample** of people rather than surveying every member of the population.

A sample must be drawn carefully to ensure that it is representative of the whole population.



What could happen if the sample is not representative?

The results of the study could be **biased**, missing out the views or habits of a whole section of the population.

Any actions based on the conclusions of the study could unintentionally discriminate against people not included.

Representation

Kara wishes to find out what people in her town think about environmental issues.

Discuss the problems with each of these methods.

A) Asking her friends and relatives about their views on the issue.

B) Asking people using the local supermarket one Thursday morning.

C) Posting a questionnaire to a selection of houses in the town and asking people to return it to her.

D) Choosing a random page from the local telephone directory and phoning every 10th person listed.

How can Kara check whether a sample is representative?

Methods

A sample must be drawn carefully to ensure that it is representative of the whole population so that the data collected is not biased.

It is also important to have a sample large enough to ensure that results are reliable.

Press the buttons to find out more about different methods of choosing a sample.

random

systematic

quota

cluster

stratified

Stratified sampling practice (1)

A company is reviewing working conditions for all its employees. The human resources (HR) manager decides to conduct in-depth interviews with 60 employees.

The company employs both full- and part-time staff, with male and female employees.

	males	females
part-time	85	375
full-time	342	188

How many employees from each category should be interviewed to make a stratified sample?

Stratified sampling practice (2)

How many employees from each category should be interviewed to make a stratified sample?

	males	females	TOTAL
part-time	85	375	**460**
full-time	342	188	**530**
TOTAL	**427**	**563**	**990**

male part-time

$$\frac{85}{990} \times 60 = \mathbf{5.15}$$

female part-time

$$\frac{375}{990} \times 60 = \mathbf{22.73}$$

female full-time

$$\frac{188}{990} \times 60 = \mathbf{11.39}$$

male full-time

$$\frac{342}{990} \times 60 = \mathbf{20.73}$$

Interview 5 part-time male employees, 23 part-time female employees, 11 full-time female employees and 21 full-time male employees.

Check: 5 + 23 + 21 + 11 = 60

Multiple-choice quiz

How well do you know the differences between the sampling methods?

Press **start** to view examples of sampling and identify the method used.

How appropriate is each method for the context?

start

Investigating bias

Jason wants to investigate the cost of food by finding the average calories per dollar.

He treats the aisles at his local store as clusters, and randomly chooses an aisle and then selects an item from the middle shelf at 1 foot intervals along the aisle.

He gets an average of 500 calories per dollar.

Bianca repeats the study but systematically chooses items from every aisle. She gets an average of 250 calories per dollar.

Discuss possible causes of the difference in results.

Group work project

Use a sample to estimate the average number of hours students spend doing homework each week.

Work in groups of 4 or 5.

Each group should use a different sampling method to choose their sample.

Think about:

- how to find out how many hours each student spends on homework
- how many students should be included in the sample
- which categories might affect the results.

After every group has collected their results, share them with the class. How similar are they?

Statistics and parameters

Javier samples twenty students in his school and finds that the average time they spend doing homework each week is 2.3 hours.

2.3 hours is a **statistic**.

The actual value of the average in the whole population is a **parameter**.

Parameters are data about the whole population.

Statistics from the sample provide estimates for the value of the parameters.

The **larger** the sample, the **better** the estimate.

Sample size

You are investigating average working hours. Choose the size of your sample and press **select sample** to let the computer randomly choose individuals from the population.

Compare different samples of the same size. How consistent is the average at each size? What do you think the population average is?

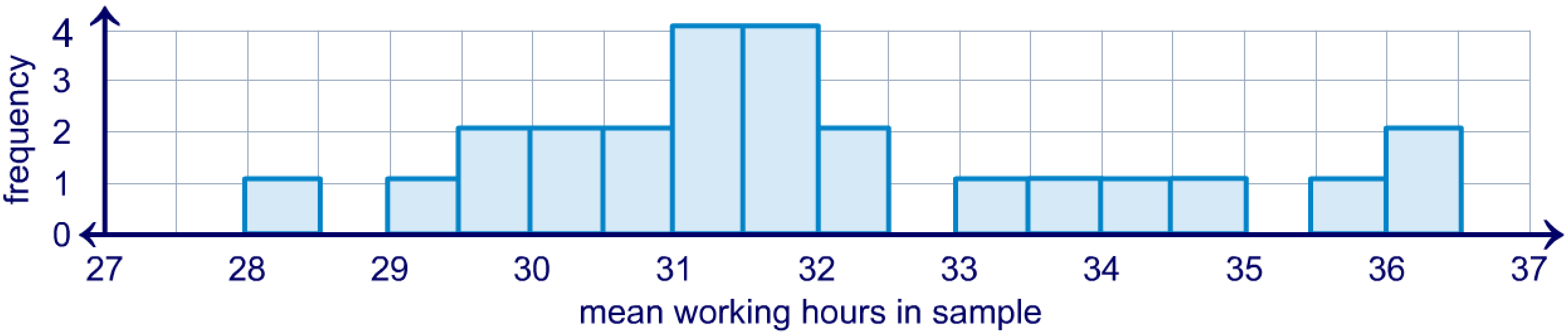
Press **start** to begin.

start

Comparing samples

Helena wants to compare sample means to a population mean. The population is 200 working adults. She calculates the mean number of hours worked per week: 31.365 hours.

She then takes 25 random samples of 15 from the population, and calculates the mean for each sample.



She makes a histogram of the results.

Is there a pattern in how the results are distributed?

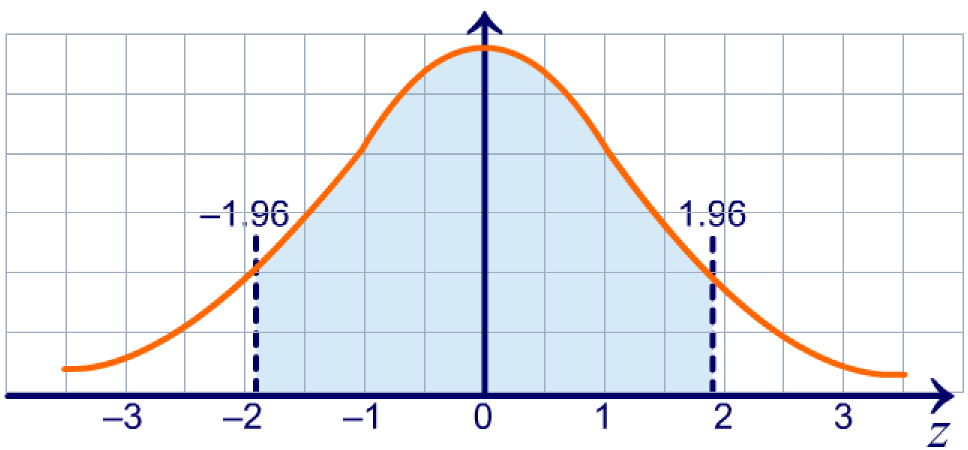
Sampling error

Sampling is subject to **random variation**. Sample means can vary and are unlikely to be the same as the population mean.

However, assume the population follows a normal distribution.

Sample statistics will tend to cluster around the population parameters. We can use the distribution to find the expected discrepancy between a statistic and the parameter.

We can construct a interval around a sample statistic with a given **confidence** of capturing the population parameter.



Margin of error formula

Suppose that, in a random sample of n individuals, the proportion or percent responding a certain way is p.

The **standard error** is:

$$\text{standard error} = \sqrt{\frac{p(1-p)}{n}}$$

The **margin of sampling error** (ME) is then defined for a confidence level, e.g. 95%, by multiplying the standard error by the z-score of the confidence level, e.g. 1.96 for 95%.

$$ME = 1.96\sqrt{\frac{p(1-p)}{n}}$$

The interval from $p - ME$ to $p + ME$ is called a **95% confidence interval**.

Margin of error example

A survey conducted on students' favorite sport in a school showed that 72% of the 450 students asked chose football as their favorite sport.

What is the 95% confidence interval?

Find the missing values to complete the calculations.

size of sample: $n =$

proportion in sample: $p =$

margin of error: $ME = 1.96\sqrt{\frac{p(1-p)}{n}}$

Margin of error practice

In a survey of 650 high school students selected at random, 65% of the students surveyed said they were involved in school-sponsored extracurricular activities.

What is the margin of sampling error to the nearest tenth of a percent?

$$ME = 1.96\sqrt{\frac{p(1-p)}{n}}$$

$$= 1.96\sqrt{\frac{0.65 \times 0.35}{650}}$$

$$= 0.03668 = \mathbf{3.7\%}$$

There is a 95% chance that the percentage of students in the whole population who are involved in extracurricular activities is between
65 – 3.7 = 61.3%
and 65 + 3.7 = 68.7%.