

Table of Contents

Overview.....	4
Activity description.....	4
Grade levels / Target audience.....	4
Subjects.....	4
Concepts	5
Prerequisites.....	5
Process skills.....	5
Preparing for an Online Exploration.....	6
Preparation time	7
Execution time	7
Last Update	8
Education Standards	9
Computer Needs: Software and Hardware.....	15
Web browser	15
Hardware requirements	15
Operating systems.....	16
How to configure the browser	17
Troubleshooting.....	17
Lesson Plan.....	18
Purpose	18
Learning outcomes	18
Prerequisites.....	18
New vocabulary	19
General misconceptions	25

Preparation time	26
Execution time	26
Physical layout of room	26
Materials	27
Procedure / Directions	27
Evaluation / Assessment	32
Follow-up / Interdisciplinary activities	38
One-computer classrooms.....	40
Classrooms without computers or Internet access	41
Homeschoolers.....	42
Background Information	43
Math Background.....	43
1. What is population size?.....	43
2. What is sample size?.....	43
3. Why is sample size important?	44
4. What is simple random sampling?.....	44
5. How do you organize a simple random sampling?	44
6. What are the strengths and weaknesses of using simple random sampling?	45
7. What are some other types sampling?	45
8. What is sampling error?.....	46
9. What is random sampling error?.....	47
10. What is systematic error?	47
11. How does a statistic differ from a parameter?	47
12. How do we get from our sample statistic to an estimate of the population parameter?.....	47
13. How is the sampling error or standard error determined?.....	48
14. What is a normal distribution?.....	48
15. What is meant by a level of confidence or confidence level?.....	49
16. What is bias?	50
17. How can bias affect the accuracy of a sample?.....	50

18. Is a computer always unbiased?.....	51
19. What is a Box and Whisker Plot (or Boxplot) graph?.....	51
20. What is a frequency table?.....	52
21. Why do astronomers use statistics?.....	53
22. How do astronomers use sampling statistic techniques in their research?.....	53
Science Background.....	54
1. What is a galaxy?.....	54
2. Why do astronomers study galaxies?.....	54
3. Why do astronomers study distant galaxies?.....	55
4. When astronomers study distant galaxies what do they look at?.....	55
5. What is a “deep” field?.....	55
6. What are the Hubble Deep Fields?.....	56
7. How were the two Hubble Deep Field sites chosen?.....	57
8. If there are thousands of galaxies visible in the Hubble Deep Fields, why does the activity use just over 1000 as the populations of each HDF?.....	57
9. What is the importance of the HDF?.....	58
10. Why was a second Deep Field taken?.....	58
11. It seems hard to see the shape of some of the galaxies in the HDFs. How do astronomers classify them?.....	59
12. What is the most common type of galaxy in the nearby universe?.....	59
Words from a scientist	59
Grab Bag	61
Downloadable activity documents.....	61
Web resources	61
Books and printed material	64

Overview

Activity description

Galaxy Hunter is an Amazing Space online exploration that reinforces students' understanding of simple random sampling. A unique feature of this activity is its reliance on the statistical analysis of real astronomical data from the Hubble Deep Field-North (HDF-N) and the Hubble Deep Field-South (HDF-S). Students begin by selecting a sample of galaxies from one of the Hubble Deep Fields (HDFs). They determine if their sample data is valid by investigating bias in sampling techniques and the role of sample variability in determining an optimal sample size. A second sample, selected from the other Deep Field image, allows students to apply what they have learned about bias and use improved sampling techniques. Comparisons between the students' data and astronomers' results lead to a determination that the two HDFs are statistically similar. The activity concludes with a comparison between the distant and local universe and a determination that while the universe may look the same in all directions, it does not look different at different depths.

Grade levels / Target audience

Grades 10-12, Target Grade: 11

Subjects

- Math/Statistics
- Science

Concepts

- Statistics
 - Bias in selecting a sample invalidates the sample.
 - Optimal sample size can be established from the variability of the data as sample size increases.
- Science
 - The similarity of the two HDFs in terms of galaxy type and distribution indicates that the universe looks the same in both directions in which the HDF images were taken. This, along with other observations, supports the premise that the universe looks the same in all directions.
 - Comparison of the most common type of galaxy in the HDFs with that in the local/nearby universe illustrates that the universe **does not** look the same at different depths.

Prerequisites

Before completing this activity, students should:

- be able to construct and interpret frequency tables.
- have knowledge of simple random samples.
- have knowledge of min/max plots.
- be able to define range, mean, and median as they apply to statistics.

Process skills

- Analyzing data
- Interpreting graphs
- Comparing and contrasting
- Identifying patterns using scientific data
- Inferring relationships

Preparing for an Online Exploration

Before using the activity...

1. Preview the activity and decide if it meets your instructional needs and students' learning needs.
2. Read these Teaching Tips for a full understanding of the activity.
3. Work through the activity as your students would. As you go through the activity, pay attention to the following:
 - How will students navigate through the activity, and what difficulties may they encounter?
 - Can students complete the activity successfully?
 - Can the activity or selected parts of the activity be completed within your allotted instructional schedule or time-frame?
4. Check out your computers:
 - Review the **Computer Needs** section.
 - Reserve a time to use the computer lab if necessary.
 - Bookmark the activity on the Web browser of each student computer.
5. Determine a strategy for organizing your students. Options include:
 - A whole-group/educator-directed approach using one computer.
 - A cooperative team approach in a computer lab, with each student in a group having specific responsibilities while working on the activity (such as "mouse user," "note taker," and "oral reader").
 - A one-student-per-computer approach.

6. Think about how the activity aligns with other instructional materials that you already use, such as:
- Curriculum guides
 - Textbooks
 - Videos
 - Posters
 - Lab guides

Preparation time

- Allow time to preview the activity, read these Teaching Tips, and review the **Science Background** section. The Science Background will provide content information relevant to the activity and will help you answer questions posed by students.
- Allow time to download, print, and make copies of the Galaxy Hunter Travelogue for students. The travelogue contains the activity's questions and provides blank spaces for answers. Students fill in answers as they encounter questions in the activity.
- Allow time to explore the additional references and links provided in the **Grab Bag**, identify follow-up activities as appropriate, and gather needed supplies.

Execution time

Educators should allow approximately 20-30 minutes for students to complete the “Start Safari” section and 15-20 minutes to complete the “Bias” section of the activity. “Sample Size” will take approximately 30-45 minutes and “HDF-N vs. HDF-S” will take approximately 20-30 minutes. Allow 10-15 minutes for the “Last Stop” section of the activity. Depending on the length of a class, the needs of your students, and your instructional purpose, completion of the full activity could take three or four class periods. Be aware that the number of computers connected to the Internet and the

speed at which the computer systems can process images may significantly alter the time needed to complete this activity.

Last update

January 2019

[Back to Top](#)

Education Standards

Galaxy Hunter includes an introductory activity called “Start Safari” and the following sections: “Bias,” “Sample Size,” “HDF-N vs. HDF-S,” and “Last Stop.” It is appropriate for students in high school.

GRADE LEVEL: 10-12

TARGET GRADE: 11

Outcome 2:	
<p>In “Sample Size” students will select a simple random sample to draw conclusions from the data. Students also will use sample variability to determine optimal sample size.</p>	
<p>Core Standards for Mathematics</p>	<p>CCSS.Math.Content.HSS-IC.A.1: Understand statistics as a process for making inferences about population parameters based on a random sample from that population.</p>
<p>ITEA Standards for Technological Literacy</p>	<p>Standard 12: Students will develop the abilities to use and maintain technological products and systems.</p> <ul style="list-style-type: none"> • Grades 9-12, Benchmark P: Use computers and calculators to access, retrieve, organize, process, maintain, interpret, and evaluate data and information in order to communicate.

<p>NCTM Principles and Standards of School Mathematics</p>	<p>Data Analysis and Probability Standard, Grades 9-12</p> <p>Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them.</p> <ul style="list-style-type: none"> • Know the characteristics of well-designed studies, including the role of randomization in surveys and experiments. <p>Develop and evaluate inferences and predictions that are based on data.</p> <ul style="list-style-type: none"> • Use simulations to explore the variability of sample statistics from a known population and to construct sampling distributions.
<p>Common Core Standards for Mathematics</p>	<p>High School Statistics & Probability</p> <p>CCSS.Math.Content.HSS-IC.A.1: Understand statistics as a process for making inferences about population parameters based on a random sample from that population.</p>

Outcome 3:

In “HDF-N vs. HDF-S” students will compare sample data with the population parameter to determine accuracy of sampling techniques, and use statistical data to make conjectures about the universe.

**National
Science
Education
Standards**

Content Standard A: Science as Inquiry

As a result of activities in grades 9-12, all students should develop understandings about scientific inquiry.

- Mathematics is essential in scientific inquiry. Mathematical tools and models guide and improve the posing of questions, gathering data, constructing explanations and communicating results.

AAAS Project 2061 Standards	The Nature of Mathematics: B. Mathematics, Science, and Technology By the end of the 12th grade, students should know that: <ul style="list-style-type: none">• Mathematics provides a precise language for science and technology-to describe objects and events, to characterize relationships between variables, and to argue logically. 2B/H3
NCTM Principles and Standards of School Mathematics	Data Analysis and Probability Standard, Grades 9-12 Develop and evaluate inferences and predictions that are based on data. <ul style="list-style-type: none">• Understand how sample statistics reflect the values of population parameters and use sampling distributions as the basis for informal inference.

Common Core Standards for Mathematics	High School Statistics & Probability CCSS.Math.Content.HSS-IC.A.1: Understand statistics as a process for making inferences about population parameters based on a random sample from that population. CCSS.Math.Content.HSS-IC.A.2: Decide if a specified model is consistent with results from a given data-generating process, e.g., using simulation. For example, a model says a spinning coin falls heads up with probability 0.5. Would a result of 5 tails in a row cause you to question the model?
--	---

Outcome 4:

In “Last Stop” students make a prediction based on data collection and analysis.

NCTM Principles and Standards of School Mathematics

Data Analysis and Probability Standard, Grades 9-12

Develop and evaluate inferences and predictions that are based on data.

- Understand how sample statistics reflect the values of population parameters and use sampling distributions as the basis for informal inference.

Common Core Standards for Mathematics

High School Statistics & Probability

CCSS.Math.Content.HSS-IC.A.1: Understand statistics as a process for making inferences about population parameters based on a random sample from that population.

[Back to Top](#)

Computer Needs: Software and Hardware

Web browser

This activity was designed for use with standard compliant browsers. This includes:

- [Internet Explorer](#) 8 and later, on Windows
- [Safari](#) on Macintosh
- [Chrome](#)
- [Firefox](#)

Although this activity may not format properly on non-compliant browsers (including Internet Explorer 5.01 and earlier versions, Internet Explorer for the Macintosh, and Netscape 4.0 and earlier), all content should be available. This activity has not been tested with any other browsers.

Hardware requirements

You will need an Internet-connected computer capable of running one of the following Web browsers: [Firefox](#), [Chrome](#), [Safari](#), or [Internet Explorer](#) 8, or later.

PC Hardware Requirements:	
CPU	Any
Main Memory	Operating system basic requirements (typically 4GB)
Operating System	Windows 7, 8, or later

Display	Color required: Minimum 256 colors Color recommended: 16K colors or better Resolution: Configured for 1024x768, or larger
Printer	None required
Disk Space	Enough space for installation of a browser

Mac Hardware Requirements:

CPU	Any
Main Memory	Operating system basic requirements (typically 4GB)
Operating System	Mac OS X, or later
Display	Color required: Minimum 256 colors Color recommended: 16K colors or better Resolution: Configured for 1024x768, or larger
Printer	None required
Disk Space	Enough space for installation of a browser

Operating systems

- Any operating system that can support [Firefox](#), [Chrome](#), [Safari](#), or [Internet Explorer](#) 8, or later. This includes Windows 7 and 8 and Macintosh OS X, or later.
- Most Unix systems

How to configure the browser

- Set your display to 1024x768 pixels, or better.
- Maximize the browser window to fill the screen.
- Configure the browser to hide the Location Toolbar and Personal Toolbar. This can be done from the “View” menu item.
- Make sure JavaScript is enabled. Because browsers update frequently, a handy website to visit is enable-javascript.com, which compiles up-to-date methods for enabling JavaScript in different browsers. Alternatively, a [Google](#) or [Bing](#) search for the following search phrases will lead you to directions on how to enable JavaScript in the browser you are using:
 - "firefox enable javascript"
 - "safari enable javascript"
 - "IE enable javascript"

Troubleshooting

For many problems, try clearing the browser's cache. To clear the cache in Firefox:

1. Go to the *History* menu item.
2. Select *Clear Recent History*. Select the *Cache* checkbox.
3. In *Time Range to Clear*, select *Everything*.
4. Click on *Clear Now*.
5. Try the activity again. Some pages may need to be reloaded.

Note: Successful completion of the “Bias” and “Sample Size” sections requires selecting a sample of galaxies. If you attempt to complete these sections before completing “Start Safari” you will get a message that says, “**Oops! In order to do this you must first pick some galaxies...**” By clicking the “close” button on the message, you will be taken to a screen for selecting galaxies from either the HDF-N or HDF-S. After selecting some galaxies, the activity will return you to the section you started.

Lesson Plan

Purpose

The purpose of **Galaxy Hunter** is to reinforce students' understanding of simple random sampling through the analysis of real astronomical data. The activity uses real data from the Hubble Deep Fields to strengthen students' understanding of simple random samples. Students also investigate bias in sampling techniques and use min/max plots to determine optimal sample size based on the variability of different-sized samples. Students then use statistical data and make comparisons in order to make conjectures about the universe.

Learning outcomes

Students will...

- analyze and identify sampling methods that reduce bias.
- select a simple random sample to draw conclusions from the data.
- use sample variability to determine optimal sample size.
- compare sample data with the population parameter to determine accuracy of sampling techniques.
- use statistical data to make conjectures about the universe.
- make a prediction based on data collection and analysis.

Prerequisites

Before completing this activity, students should:

- be able to construct and interpret frequency tables.
- have knowledge of simple random samples.
- have knowledge of min/max plots.
- be able to define range, mean, and median as they apply to statistics.

New vocabulary

Math Terms:

BIAS

A systemic error in sample statistics that can occur from the use of poor sampling methods.

DATA

Information (such as age, color, or shape) gathered concerning a sample or population.

MEAN

The average value found by adding the values for all the members of the sample and dividing the sum by the number of members. The mean can be skewed to one side or the other by an extreme score - one that is out of line with the others.

MEASURES OF CENTRAL TENDENCY

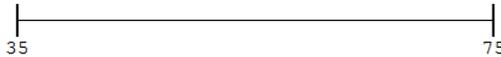
Numerical values that are located, in some sense, in the middle of a sample or population. These include the mean, median, and mode.

MEDIAN

The value that falls in the middle of a sample; there are as many members above the median as below it. Since the median falls in the middle of the sample, it is not skewed by extreme scores like the mean.

MIN/MAX PLOT

A visual representation of the range. It shows the dispersion of the data by using a line to connect the lowest-valued (smallest) piece of data to the highest-valued (largest) one. For example, the min/max plot for the data: 35, 47, 58, 70, 71, 75 would look like this:



MODE

The number, or range of numbers, in a set that occurs most frequently.

POPULATION

All members of some set that, as a whole, is generally too large to study. Sometimes the term population refers to people or objects; other times it refers to measurements of those people or objects. Both meanings are used and the context usually tells which one is meant.

POPULATION PARAMETER

The theoretical descriptive measure on the scores of the population. Generally a population is too large to record all the scores but theoretically, the parameters such as the mean, median and standard deviation for the population exist. The goal of sample statistics is to be able to infer something meaningful about a population parameter based on a sample statistic.

RANDOM NUMBER TABLE

A list of random numbers from which single digit (0-9), double digit (00 - 99) or multi-digit numbers can be selected (usually by randomly pointing to a spot on the table). These numbers can then be used to identify the members of a population to be included in a sample or to simulate an experiment. Most random number tables include a suggestion for how to use them.

RANGE

The simplest measure of dispersion. It is the interval between the highest-valued (largest) and lowest-valued (smallest) pieces of data, expressed as the difference between them.

SAMPLE

A subset of the population obtained by collecting information only about some members of the population. Sometimes the term sample refers to people or objects; other times it refers to measurements of those people or objects. Both meanings are used and the context usually tells which one is meant.

SAMPLE STATISTIC

Any descriptive measure on the scores of a sample, such as the mean, median, or standard deviation.

SIMPLE RANDOM SAMPLE (SRS)

A simple random sample is gathered by selecting each member without replacement, which means that a member is not available to be selected more than once. The members are selected by chance so that each member has an equal opportunity of being selected. The sample, of size n , is representative of the population as a whole.

SMALLEST REASONABLE SAMPLE

Statisticians look for the smallest sample that is reasonable. The smallest sample minimizes time and costs. A reasonable sample is one that is unbiased and large enough to adequately represent the population it is derived from.

Science Terms:

BULGE

The bulge is a spherical structure at the center of spiral galaxies composed primarily of old stars, with a little gas and dust. The bulge of the Milky Way is about 30,000 light-years across.

DISK

The disk is a pancake-shaped structure in spiral galaxies and is composed primarily of young and middle-aged stars, with abundant gas and dust. Some old stars are also present. The disk contains the spiral arms and slices through the bulge and halo of a spiral galaxy. The disk in the Milky Way is about 100,000 light-years across and 2,000 light-years thick.

ELLIPTICAL GALAXY

A galaxy having an oval or nearly spherical shape. Some are more elongated than others. Resembling a bulge and halo, it is composed mostly of old stars and contains very little gas and dust. The smallest elliptical galaxies (called dwarf ellipticals) are probably the most common type of galaxy in the nearby universe.

GALAXY

A galaxy is an enormous collection of a few million to trillions of stars, gas, and dust held together by gravity. They can be several thousand to hundreds of thousands of light-years across.

HALO

The halo extends outward from the bulge and disk, and contains clusters of old stars (globular clusters), individual stars, and a small amount of gas and dust. In the Milky Way, the halo measures about 130,000 light-years across.

HUBBLE DEEP FIELDS (HDFs)

There are two HDFs - one in the northern sky and one in the southern sky. Hubble Deep Field North (HDF-N) is a tiny region of the northern sky near the Big Dipper toward which the Hubble Space Telescope was pointed for ten consecutive days in 1995. Because this observation was designed to detect very faint light from the most distant galaxies Hubble can observe, the field contains few bright celestial objects. Seemingly devoid of light, this small area provided a “keyhole” view of the universe’s past, reaching across space and time to see infant galaxies. Hubble Deep Field South (HDF-S) is a tiny region of the southern sky near the Southern Cross toward which the Hubble Space Telescope was pointed for ten consecutive days in 1998. By probing these remote regions of space, astronomers are gaining more information on galaxy development.

HUBBLE SPACE TELESCOPE

An automated reflecting telescope which orbits the Earth, built by the National Aeronautics and Space Administration and the European Space Agency. It contains instruments capable of receiving many types of radiation.

IRREGULAR GALAXY

A galaxy whose shape is neither elliptical nor spiral. It contains both young and old stars and is often rich in gas and dust. These galaxies often have active regions of star formation. Sometimes the irregular shape of these galaxies results from interactions or collisions between galaxies. Observations such as the Hubble Deep Fields show that irregular galaxies were more common in the distant (early) universe.

LIGHT-YEAR

The distance traveled by light in a full year - some 10 trillion kilometers (about 6 trillion miles). To calculate the distance light travels in one year, multiply the speed of light, 300,000 km/s, by the length of a year expressed in seconds: $1\text{LY} = (300,000 \text{ km/s})(31,536,000 \text{ sec}) = 10 \text{ trillion kilometers}$.

MILKY WAY

The specific galaxy to which our solar system belongs, so named because most of its visible stars appear overhead on a clear, dark night as a milky band of light extending across the sky. The Milky Way is a spiral galaxy.

REDSHIFT

When an object is moving away from an observer, the light it emits appears to be redder than it would be if the object were at rest. This effect is similar to the apparent change in pitch associated with moving objects that emit sound waves (the Doppler Effect). For example, the pitch of a fire truck's siren sounds higher (having a shorter wavelength) as it approaches a car stopped along the side of the road. The pitch drops significantly as the truck passes the car and sounds much lower (having a longer wavelength) as the truck continues to move away. When dealing with light, the shift to longer wavelengths is referred to as redshift because red light has the longest wavelength within the optical spectrum. (The wavelength shift occurring when objects are moving toward an observer is called a blueshift.) The cosmological redshift tells astronomers how fast the universe is expanding.

SPIRAL ARMS

Pinwheel shaped features located in the disk of a spiral galaxy. The new bright blue stars that are born there make the spiral pattern visible.

SPIRAL GALAXY

A galaxy made up of a disk with spiral (pinwheel-shaped) arms, a bulge near its center, and a halo. The sizes of the disk and bulge vary. The galaxy is composed of a mixture of old and young stars as well as gas and dust. The spiral arms are sites of active star formation. The majority of large galaxies in the nearby universe are spirals.

WHOLE POPULATION OF HDFs

Astronomers counted and classified all the galaxies that they could reliably count and classify in each of the Deep Fields.

General misconceptions

Students may think that:

- any way they choose data will give results representative of a population. Humans are not good at selecting data in a truly random fashion, since they tend to introduce unconscious bias, whereas computers do not.
- closing one's eyes and picking a sample by pointing a finger produces an unbiased sample. Bias is introduced when the student decides which item the finger is covering. Also, the student will have a mental image of the area from which the sample is taken and may consciously move the finger to different locations in order to avoid being biased which introduces bias.
- galaxies are static and do not change with time. In fact, galaxies are dynamic and change over millions of years.

Note: For a list of additional misconceptions and the facts to dispel them, see [Myths vs. realities: Galaxies](#) on Amazing Space.

Preparation time

- Allow time to preview the activity, read these Teaching Tips, and review the **Science Background** section. The **Science Background** will provide content information relevant to the activity and will help you answer questions posed by students.
- Allow time to download, print, and make copies of the Galaxy Hunter Travelogue for students. The travelogue contains the activity's questions and provides blank spaces for answers. Students fill in answers as they encounter questions in the activity.
- Allow time to explore the additional references and links provided in the **Grab Bag**, identify follow-up activities as appropriate, and gather needed supplies.

Execution time

Educators should allow approximately 20-30 minutes for students to complete the “Start Safari” section and 15-20 minutes to complete the “Bias” section of the activity. “Sample Size” will take approximately 30-45 minutes and “HDF-N vs. HDF-S” will take approximately 20-30 minutes. Allow 10-15 minutes for the “Last Stop” section of the activity. Depending on the length of a class, the needs of your students, and your instructional purpose, completion of the full activity could take three or four class periods. Be aware that the number of computers connected to the Internet and the speed at which the computer systems can process images may significantly alter the time needed to complete this activity.

Physical layout of room

You may decide to have students work individually or in small groups of two or three. To maximize learning, no more than three students should share a computer at a time. Adaptations can be made to accommodate a classroom that has one computer

with Internet access (See **One-computer classrooms** below). Suggestions include leading a whole-class activity and discussion by connecting the computer to a projector or monitor.

Materials

- Computers with Internet access for as many students as possible. See the **Computer Needs** section.
- A printout of the four-page Galaxy Hunter Travelogue for each student. The travelogue contains activity questions and blank spaces for student responses. Students answer the travelogue questions as they complete the Galaxy Hunter activity.

Procedure / Directions

Galaxy Hunter consists of the following sections or levels:

START SAFARI

After choosing one of the Deep Fields to work with (HDF-N or HDF-S) and selecting a sample of galaxies, students view a summary of the galaxy types present in their sample. Students then compare the number of irregular galaxies in their sample to the number of irregular galaxies in the whole HDF field. This leads students to consider bias and sample size, both of which affect the validity of results.

BIAS

In this section, students investigate several different sampling methods and decide whether they feel each one is unbiased or biased. By comparing results for various sampling methods, students learn that computers are the best tools to use for selecting an unbiased sample.

SAMPLE SIZE

In this section, optimal sample size is determined by exploring sample variability. The concept of variability is introduced through a min/max plot of five computer-generated samples that are equal in size to the student sample. The concept is then extended to the whole population, using increasingly larger sample sizes. The mean and median are added to the smaller sample sizes in order to pinpoint the spot where variability settles down and the measures of central tendency approach a constant value. The point where this first occurs is the smallest reasonable sample size. A check for understanding is provided when students are asked to determine a small, reasonable sample of galaxies based on their selected HDF image.

HDF-N VS. HDF-S

In this section, students apply what they have learned in previous parts of the activity to generate a small, reasonable sample of galaxies from the other HDF image (the one they did not initially select). Students then compare the number of galaxy types in both HDF images to determine if they are statistically similar. This provides students with a real-life example of how statistics can be used by scientists. Since the images are statistically similar, this also provides evidence to support the premise that the universe looks the same in all directions.

THE LAST STOP

One of the fundamental questions in astronomy asks how galaxies formed and evolved. The Hubble Deep Fields help to answer to this question by providing a glimpse into the history of our universe and showing what galaxies looked like 2 to 10 billion years ago. In this section, students are asked whether or not they think the distant universe looks the same as the nearby universe. The activity concludes by revealing that irregular galaxies are much more common in the distant universe, whereas in the nearby universe, elliptical galaxies are more common.

BEFORE USING THE ACTIVITY...

1. Preview the activity and decide if it meets your instructional needs and students' learning needs.
2. Read these Teaching Tips for a full understanding of the activity.
3. Work through the activity as your students would. As you go through the activity, pay attention to the following:
 - How will students navigate through the activity and what difficulties may they encounter?
 - Can students complete the activity successfully?
 - Can the activity or selected parts of the activity be completed within your allotted instructional schedule or timeframe?
4. Check out your computers:
 - Review the **Computer Needs** section.
 - Reserve a time to use the computer lab, if necessary.
 - Bookmark the activity on the Web browser of each student computer.
5. Determine a strategy for organizing your students. Options include:
 - A whole-group/educator-directed approach using one computer.
 - A cooperative team approach in a computer lab, with each student in a group having specific responsibilities while working on the activity (such as “mouse user,” “note taker,” and “oral reader”).
 - A one-student-per-computer approach.

6. Think about how the activity aligns with other instructional materials that you already use, such as:
 - Curriculum guides
 - Textbooks
 - Videos
 - Posters
 - Lab guides

PREPARING STUDENTS FOR THE ACTIVITY...

1. Give students a pre-assessment to determine their computer experience and/or their background knowledge about galaxies and/or statistics.
2. Use the [Myths vs. realities: Galaxies](#) to identify student misconceptions about galaxies.
3. Introduce the exploration with an engagement activity. Engagement activities include, but are not limited to, the following:
 - Have students write down as many things as they know about galaxies. This offers an opportunity to elicit and identify student misconceptions.
 - Display the Hubble Deep Field image and ask students what kinds of statistical treatments might be possible. Hubble Deep Field posters can be obtained from your local [NASA Educator Resource Center](#). Hubble Deep Field lithographs can be downloaded from Amazing Space here:
http://amazingspace.org/resource_page/43/galaxies/type#resource_tab.
 - Download and display side-by-side images of the HDF-N and HDF-S. Links to these images can be found in the Grab Bag. Have students compare the two images and decide if they look similar and/or different, and in what ways. Lead a discussion in which students identify the types of statistical evidence that can be collected from the images.

4. Share the activity's outcomes and vocabulary with students.
5. Use a large monitor, projector, or transparencies to provide an overview of the activity for students and to demonstrate how to navigate it.
6. Organize students in such a way that more experienced computer users are paired with less experienced ones.
7. Consider using one of the suggested activities in the **Assessment / Evaluation** section to measure student learning.
8. Consider using one of the suggested activities in the **Follow-up / Interdisciplinary activities** section to provide students with additional exploration or research opportunities.

WHILE STUDENTS ARE DOING THE ACTIVITY...

1. Help students navigate through the activity as necessary.
2. Provide options for students who finish the activity early:
 - Have students review the activity to define key vocabulary words.
 - Have students visit related websites to conduct additional research.
 - Have students complete some type of assessment activity. Suggested assessment activities can be found in the **Evaluation / Assessment** section.

[Back to Top](#)

Evaluation / Assessment

Several formative assessment questions are embedded within the Galaxy Hunter activity. The Galaxy Hunter Travelogue can be used as an assessment activity to measure student learning. **The answers to the Travelogue questions are:**

1. What method, if any, did you use to select your sample of galaxies?

Students may have selected the galaxies by size, shape, color, or location on the map. They may have tried to make the selection random by closing their eyes and pointing to different locations.

2. Record your data from the HDF- N or S (circle one) in the table at left, below.

Students' answers will vary depending on the size of their sample and how they selected their sample.

3. Record the actual result for percentage of irregulars in the table above, right. Describe the variation between your result for percentage of irregular galaxies and the actual result. What do you think accounts for this difference?

The actual percentage of irregulars for the whole HDF-N is 70.7%; for the whole HDF-S, it is 67.2%. Variations between the accepted result and the students' results could be caused by too few galaxies in their samples and/or bias in their sampling methods. For example, the brightest galaxies in either HDF tend to be spirals and ellipticals. If the student chose bright galaxies, their percentage of irregulars would be lower than the value for the whole HDF. If they chose only dim galaxies, they might come up with a value very close to or even higher than that for the whole HDF.

4. Define bias.

Bias is a systematic error in sample statistics that can occur from the use of poor sampling methods. The students might include a statement about humans unconsciously basing choices on a trait or quality of the sample member.

5. Based on your current knowledge, was your method of selecting the sample of galaxies described in question 1 biased? What effect might this have had on your results?

Students will probably say that their method was biased and that the results are not reliable due to this bias. Student results will probably be too low but they could be very close to the actual results or even slightly higher depending on how they selected their galaxies.

6. If a computer randomly chooses a fixed number of galaxies, 5 different times, would you expect it to keep getting the same result for percentage of irregulars each time, since the sample sizes were the same each time? Why or why not?

The results will vary each time a sample is taken because, with small sample sizes, the likelihood of selecting the same proportion of irregulars each time is low. With sample sizes of 5 or 10, the variability of results will be great, with percentages of irregulars ranging from 30% or less up to 90% or more. The smaller the sample size, the greater the variability of results. If the sample size is larger than 35, the students will not see as great a variability as they would if the sample size were lower, but there will still be some variability. At a sample size of 35, percentages of irregulars will generally range from 60-85%.

7. Compare the positions of the mean and median on any single min/max plot. How would those positions change if the lowest value on the min/max plot were zero?

The positions of the mean and median may be close to each other if the distribution of results is close to a bell curve. For example, if the five values are: 40%, 50%, 60%, 70% and 80%, then the median is 60% and the mean is also 60%. If the data is skewed, then the median and mean will be farther apart. For example, if the five values are 60%, 60%, 60%, 80% and 90%, then the median would be close to 60% and the mean would be 70%. If the lowest value on the min/max plot were zero, the median would probably remain the same but the mean would be lower.

8. Explain why the variability of sample results approaches zero as the sample size approaches the population size.

When the whole population is sampled, there is no variability. As the sample size approaches the population size, the likelihood of obtaining a different sample population decreases because only a few of the members are "left out" of the sample. Since so many of the members of the population are included, missing a few won't greatly affect the proportions of the sample, so the variability is slight.

9. Predict a range of acceptable percentages for irregulars based on the length of the min/max bar for your best sample size. The range of percentages for the whole chart is 40% to 100%.

Estimates will vary based on student sample size. Generally speaking, the min/max bar for the best sample size is 10 units long, so the range is about 10%.

10. Copy the table, making sure you fill in your sample size and the full name of the HDF you used. Compare your value for the percentage of irregulars with your predicted range of acceptable percentages from above.

Student values will vary for their sample. Students' values should fall within 10% of the astronomers' results for the whole HDF. The astronomers' results for the whole HDFs are as follows:

Actual Percent of Galaxies for HDF-N and HDF-S		
Galaxy Type	HDF-N	HDF-S
Ellipticals	4.6%	5.1%
Spirals	24.7%	27.7%
Irregulars	70.7%	67.2%

11. Now that you know about sample size and bias, do you think the percentage of irregulars in your original sample is within acceptable range of the actual results? Explain. (Your results and the actual value for percentage of irregulars can be found in question 2 above.)

Students will probably decide that their original sample size was too small and subject to bias so their results were not within the acceptable range.

12. State the sample size you will use and explain why you chose this size.

Students should choose the same-sized sample for the other HDF as they choose for the first one, since the populations of the two fields are about the same.

13. Do you think your results are close to the astronomers'? Explain.

If the students selected the correct sample size, their results will fall close to those of the astronomers'. We are considering their results close if they are within a few percentage points of the astronomers' results.

14. Comparing the two HDFs, could you say that the universe probably looks the same in these two directions? Explain.

Since the proportions of galaxies for the two fields are similar, students should conclude that the universe probably looks the same in these two directions.

15. Using what you've learned about sample size and bias, could you use the HDFs to make a general statement concerning the uniformity of the universe? Explain.

No. Each HDF represents only a tiny fraction of the sky, and could easily be subject to bias. In order to make a general statement about the uniformity of the sky, many more HDFs would need to be randomly selected from the whole sky. The fact that the two fields are similar to each other is consistent with astronomers' supposition that the universe looks the same in all directions.

16. What is the most common type of galaxy in (a) the local universe and (b) the faraway universe pictured in the HDFs?

(a) When one counts both large and small galaxies, dwarf ellipticals (small ellipticals) are probably the most common type of galaxy in the nearby universe. Since these galaxies are small and faint, the exact number of these galaxies is not well known. The majority of large, bright galaxies in the nearby universe are spirals. Large bright elliptical galaxies are relatively rare.

(b) In the faraway universe pictured in the HDFs, the most common type of galaxy is irregular. Astronomers hope that this information will help them understand the fundamental question of how galaxies form and evolve.

Other suggestions include the following:

- Have students revisit their engagement activity responses/findings and compare them to what they now know about galaxies and statistical sampling.
- Students can identify other instances of when statistics and sampling are important – for example, when interviewing other students about their favorite foods or sports. Students can apply what they have learned to determine a reasonable yet representative sample for their class, grade, school, or neighborhood.
- Students can devise a plan for conducting a project, as noted above, and implement it. Students can formulate questions that can be addressed with data and determine a strategy for selecting a random, unbiased sample. Students can then collect and analyze data, and synthesize and report their findings.

[Back to Top](#)

Follow-up / Interdisciplinary activities

The following suggested activities can provide students with additional opportunities to explore statistical analyses and/or galaxies. They also can help reinforce skills across multiple content areas, such as science, math, and English/language arts.

SCIENCE:

- Students can use the Internet to investigate how astronomers assign coordinates to different points in the sky.

MATH:

- Students can use the Internet to investigate what mathematics is necessary for a career in astronomy.
- For students who have previous experience with confidence intervals...

Have students consider the challenge of scientists taking a random sample of the entire sky. Have students determine how many galaxies would be needed in order for the sample to be representative of the whole population of galaxies in the universe. More advanced students can calculate the minimum sample size by using the following formula where the sample proportion value is unknown:

$$n = \frac{(z_{\alpha/2})^2 pq}{E^2}$$

Where n = sample size; $z_{\alpha/2}$ = critical value of z ; E = maximum error of estimates; p = estimated population proportion; and $q = 1 - p$. If no approximate value of p is known, then $p = .5$ and $q = .5$ ($1 - .5$). Note: There are websites listed in the Grab Bag that will calculate sample sizes. These require entering a confidence level and interval.

The answer to the equation above, a value of about 350, allows one to conclude that the population of galaxies within the HDFs is large enough to be representative of the universe as a whole, since the population for HDF-N is 1067 and the population for HDF-S is 1275. However, the galaxies in the HDF images were not selected randomly so the sample could be subject to bias.

Note: In the **Galaxy Hunter** activity, students determine the minimum sample size that is representative of the HDFs, not the universe.

ENGLISH LANGUAGE ARTS:

- Students can use a graphic organizer to compare how science fiction stories, movies, and TV shows depict deep space in contrast to reality, as in the example below:

Fiction	Reality
All stars look white.	Stars come in a variety of colors: blue, white, yellow, orange, and red.

- Students can submit/research questions and obtain answers through the following websites:
 - [NASA's Ask the Space Scientist](#)
 - [Curious About Astronomy? Ask an Astronomer!](#)
 - [Enchanted Learning](#)

One-computer classrooms

It is recommended that educators use a projector or large monitor connected to the computer in order to display the activity. To facilitate a more organized and predictable large-group presentation, and avoid last-minute glitches, bookmark the activity or pages you wish to use and/or save selected activity screens for offline viewing. This will eliminate the inconvenience of unexpectedly losing Internet connectivity.

Take advantage of the pace of a one-computer presentation or demonstration by leading discussions at key points throughout the activity. Some discussion points include...

- For "Start Safari," after data has been generated from the clickable map of either the HDF-N or HDF-S, lead a class discussion on sampling methods. In the case of a one-computer classroom, use the "Back" button and allow several students to generate samples and record the data. A box plot of the data for one type of galaxy can be constructed to show variability in the data. ***Note: Because this data is biased, values should not be averaged at this point.***
- For "Bias," before leaving the page titled "Banishing Bias" and moving on to the page titled "Comparing Sampling Methods," have students predict what the percentage of irregulars might be for each sampling method. Record these values and then move on to "Comparing Sampling Methods" to see how close the students' predictions were.

- For "Bias," record data from the page titled "Comparing Sampling Methods," which shows the results for each sampling method. Have students generate a box plot of the different percentages for each galaxy type. These box plots should show how widely spread the data is for the different sampling techniques. ***Note: The box plot is used to graphically display the great variation between the sampling methods. Because this data is biased, values should not be averaged and there should be no discussion of median at this point.***

Classrooms without computers or Internet access

Several strategies are available to educators working in environments without access to computers or the Internet. If your classroom is not equipped with a computer and/or an Internet connection, try one of the following:

- If you have a computer with Internet access and a printer in the school library or at home, you can...
 - print all or selected parts of the activity as paper copies or transparencies.
 - print the information provided in the **Science Background**, which can be used as content reading for students.
 - check the **Grab Bag** for a variety of other educational resources and links that can be used to create your own activity, as reading material for students, or as information sources for student research.
- If your school has a computer lab, students may be able to complete the activity in the computer lab, either individually or in small groups.
- If you have a classroom computer but no Internet connection, some software programs and Web browsers provide offline access to Web pages. These programs allow you to save Web pages to a local hard drive. You can then open the downloaded Web pages locally and view them as if you are online.
- Students may have computers with Internet access at home. If so, consider assigning the **Galaxy Hunter** as a homework or extra-credit assignment.

Homeschoolers

This activity is easily followed without additional support if the prerequisites are met. Parents can preview the activity and examine these Teaching Tips ahead of time. More information for homeschooling parents can be found at:

- [American Homeschool Association](#)
- [Yahoo Homeschooling Groups](#)
- [The Home School Learning Network](#)
- Griffith, Mary. (March 1999). *The Homeschooling Handbook, 2nd edition*. Three Rivers Press: New York.

[Back to Top](#)

Background Information

The following information provides additional background about statistics and galaxies. It also can be used as an information source for student research projects, or to support review activities and group discussions.

Math Background:

1. What is population size?

The population size includes all the individuals in the identified group to be studied. This may be the number of people in a city, or the number of people who buy new cars. Often you may not know the exact population size, which is not a problem. The mathematics of probability proves that the size of the population is irrelevant, unless the size of the sample exceeds a few percent of the total population you are examining. This means that a sample of 500 people is equally useful in examining the opinions of a state of 15,000,000 as it would a city of 100,000. For this reason, The Survey System ignores the population size when it is "large" or unknown. A large population is referred to as infinite, while a small population is considered finite. Population size is only likely to be a factor when you work with a relatively small, known, finite group of people (e.g., the members of an association).

2. What is sample size?

The sample size is the number of individuals included in a study and represents only a subset of the population. This subset is selected in a way that gives every member of the population an equal chance of being chosen. The larger your sample, the more sure you can be that it truly reflects the population.

3. Why is sample size important?

It is essential to use the correct sample size to accurately represent the population. Choosing a sample size that is too small may not give an accurate representation of the population distribution. Too large a sample size is wasteful and sometimes impossible to complete. For example, you want to change something in a school with a population of 500 students, and decide to survey only ten people. Is this truly representative of the entire school population? No! Ten people are not enough to accurately represent the school. Suppose you tried to ask every person in the school. Sometimes this is not easily accomplished and can be unnecessary. In this case, a sample of 23 should be enough to represent the population. Reasonable sample size is dependent on population size and how much sampling error is tolerated.

4. What is simple random sampling?

Simple random sampling is the basic sampling technique in which a group of subjects, or a sample, is selected for study from a larger group or population. Each individual is chosen entirely by chance and each member of the population has an equal chance of being included in the sample. Every possible sample of a given size has the same chance of selection; i.e. each member of the population is equally likely to be chosen at any stage in the sampling process.

5. How do you organize a simple random sampling?

A simple random sample is formed by assigning each member of the population a number and then indiscriminately selecting from these numbers. One way to make the selection random is to use a random number table or let a computer generate a series of random numbers. Each member of the population is assigned a unique number, or perhaps a number is already assigned to each member, such as a social security number or telephone number. The members of the population chosen for the sample

will be those whose numbers are identical to the ones extracted from the random number table (or computer), in succession, until the desired sample size is reached. For example, suppose a committee is to be formed whose members are randomly selected from a group of 25 people. To obtain a simple random sample, each person is assigned a number, the numbers are placed in a hat and mixed, and then blindly drawn to form the committee.

6. What are the strengths and weaknesses of using simple random sampling?

The simple random sample requires less knowledge about the population than other techniques, but it does have two major drawbacks. One is if the population is large, a great deal of time must be spent listing and numbering the members. The other is that a simple random sample will not adequately represent many population attributes (characteristics) unless the sample is relatively large. That is, if you are interested in choosing a sample to be representative of a population on the basis of characteristics such as gender, age, and economic status, a simple random sample will need to be very large to ensure the distribution of these characteristics is equivalent to (or representative of) the population at-large.

7. What are some other types of sampling?

- **Systematic sampling** - Similar to simple random sampling, but instead of selecting random numbers from tables, you move through a list (sample frame) picking every n th name. For example, pick every 10th name from an alphabetical list of students enrolled in a school.
- **Random route sampling** - Used in market research surveys and mainly used for sampling households, shops, garages and other premises in urban areas. A starting address is randomly selected and, taking alternate left- and right-hand turns at road junctions, every n th address is selected.

- **Stratified sampling** - All people in a sampling frame are divided into "strata" (groups or categories). Within each stratum, a simple random sample or systematic sample is selected. For example, a politician wishes to poll his/her constituents regarding taxation. The constituents are broken into income brackets and then each bracket is polled.
- **Cluster or Area random sampling** - In cluster sampling, the population is divided into clusters (usually along geographic boundaries). The clusters are randomly sampled and all units within the sampled cluster are measured. For example, a survey of town governments that requires going to the towns personally could be done by using county boundaries as the clusters and randomly selecting five counties. All of the town governments in these selected counties would then be surveyed.
- **Multi-stage cluster sampling** - As the name implies, this involves drawing several different samples. The first stage would be a cluster sample as described above but then another sample is taken from these samples. For example, a face-to-face survey of a state's could be done by first selecting a sample of counties within the state, and then selecting another sample, such as a systemic sample, of the residents in the selected counties. Thus the cost of interviewing is minimized.

There are many other methods of sampling that are more advanced. Visit the sampling method websites listed in the **Grab Bag** to learn more.

8. What is sampling error?

Every survey contains some form of error. Even a complete census of all known members of a population is subject to random error or potential measurement error. There are two major forms of sampling error that might be encountered in a survey: random error and systematic error.

9. What is random sampling error?

Random error occurs when a particular sample is not representative of the population of interest due to random variation. It can be expressed as the difference between the sample results and the true results. Even if all aspects of a sampling method are executed properly, the results are still subject to a certain amount of error because of random, chance variation.

10. What is a systematic error?

A systematic error occurs when something is wrong with the technique being used or when an instrument is not calibrated correctly. This results in an error throughout the sample.

11. How does a statistic differ from a parameter?

A statistic is a generalization concerning an entire sample, such as the mean, mode or median. A parameter is a generalization for an entire population, such as the mean, mode, or median. In order to get a parameter, the entire population is involved whereas a statistic is derived from a sample of that population.

12. How do we get from our sample statistic to an estimate of the population parameter?

There are an infinite number of samples that can be taken from a large population. One sample from a population might yield a slightly different statistic than another sample taken from the same population but the statistics should be similar to each other. If more and more samples of the same size were taken from the population, the sampling distribution of the statistic would resemble a bell curve or normal distribution. The average of the sampling distribution is essentially equivalent to the parameter. The standard deviation of the sampling distribution, called sampling error, tells us something about how different samples would be distributed. This, in turn,

tells us how far the statistic is from the parameter. A low sampling error means that there is relatively less variability or range in the sampling distribution and we are closer to the parameter.

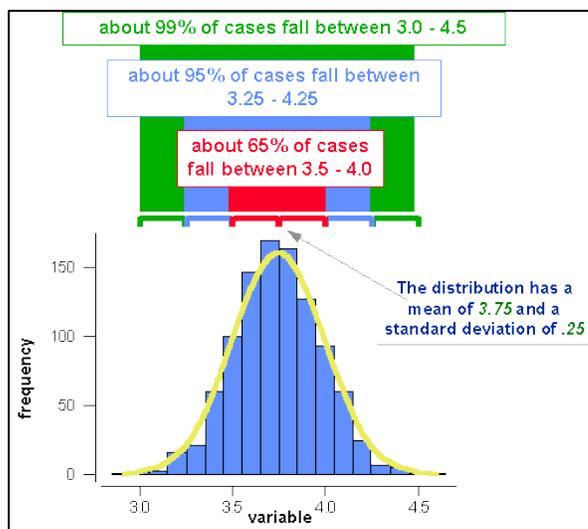
13. How is the sampling error or standard error determined?

Calculation of sampling error (also called standard error) is based on the standard deviation of the sample – the greater the standard deviation, the greater the sampling error. The sampling error is also related to the sample size. The greater your sample size, the smaller the sampling error. This error cannot be avoided. It can only be reduced by increasing the sample size. It is possible to estimate the range of random error at a particular level of confidence. Suppose we surveyed 500 people and found that 65% of them said that vanilla is their favorite ice cream. For a sample of 500, the sampling error is 4 percent. This means that we can expect our sample results to be within 4 percentage points of the actual figure for the population. In other words, as high as 69% or as low as 61%. As sample size increases, sampling error decreases. Sampling error is 10% for a sample of 100 and 3% for a sample of 1000.

14. What is a normal distribution and how does standard error relate to this distribution?

A normal distribution is a bell curve that extends to infinity in both directions. The high point represents the mean. If the area under the curve is defined to be 1 and you multiply that by 100 then there is a 100% chance that any value you name will be somewhere in the distribution. Because half the area of the curve is below the mean and half is above the mean, there is a 50% chance that a randomly chosen value will be above the mean and the same chance that it will be below it. The area under the normal curve is equivalent to the probability of randomly drawing a value in that range. The area is greatest in the middle where the "hump" is and thins out toward the tails.

When the area of the standard normal curve is divided into sections by standard error above and below the mean, the area in each section is a known quantity. The areas above and below the mean can be added together to get the probability of obtaining a value within (plus or minus) a given number of standard errors. There is a 65% chance of a value falling within one standard error of the mean, a 95% chance within two standard errors and a 99% chance that it will be within three. Suppose a normal distribution has a mean of 3.75 (highest point on the graph below) and a standard deviation of .25. Then 65% of the values will fall between 3.5 and 4.0 as shown below.



Taken from <http://trochim.human.cornell.edu/kb/sampstat.htm>

15. What is meant by a level of confidence or confidence level?

Confidence levels are used when two sets of data are being compared. Confidence level, also called significance level, is the likelihood of obtaining a particular result by chance as opposed to there being a truly significant difference in the sets of data. The smaller the significance level, the more stringent the test, and the greater the likelihood the conclusion is correct. Common confidence levels are 0.05 (1 in 20 chance), 0.01 (1 in 100 chance), and 0.001 (1 in 1000 chance).

16. What is bias?

Bias is a systematic error in sample statistics that can occur from the use of poor sampling methods. Samples may be biased for a number of reasons: frame error, population specification error, or selection error. The sampling frame is the list of population elements or members from which the sample is selected.

- **Frame error:** Frame error results when the sampling frame does not represent a true cross-section of the target population. For example, suppose you survey your neighborhood and only talk to the people on the street. Any data collected in this manner is heavily biased because not everybody in the neighborhood had a chance to respond. Any conclusions drawn about your neighborhood using this method of sampling will not be representative of the population, the entire neighborhood.
- **Selection error:** Selection error results from a systematic bias in the manner in which respondents are selected for participation in a survey. Even if the sampling frame is properly defined to include the appropriate population members, selection error can still occur. Incomplete or improper procedures for selecting participants will lead to selection error. For example, if a sample list was sorted by zip code and interviewers selected survey participants by contacting names in order from the beginning of the list, selection error would occur because members of the population appearing at the end of the list (larger zip codes) would never be contacted.

17. How can bias affect the accuracy of a sample?

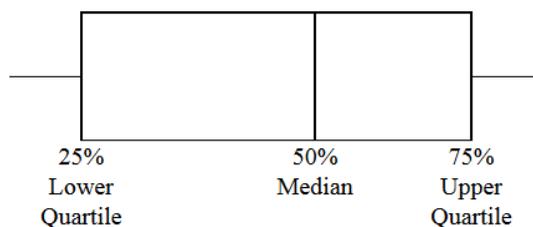
When bias occurs, results are skewed from the normal distribution. A negatively skewed curve has a thicker tail on the side below the mean while a positively skewed distribution has a larger tail on the side above the mean. In either case, the accuracy of the results will be compromised. Note: a skewed sample does not necessarily mean it is biased.

18. Is a computer always unbiased? Or, do computers always produce random samples?

The answer is no. A computer's random number generator could be programmed in such a manner as to yield a biased sample. However, for the purposes of this activity, computers are considered unbiased.

19. What is a Box and Whisker Plot (or Boxplot) graph?

A box and whisker plot is a way of summarizing a set of data measured on an interval scale. It is often used in exploratory data analysis to show the shape of the distribution, its central value, and variability. The picture below consists of the most extreme values in the data set (maximum and minimum values at the ends of the line), the lower and upper quartiles (edges of the box), and the median (vertical line through the box). A box plot, as it is often called, is especially helpful for indicating whether a distribution is skewed and whether there are any unusual observations (outliers) in the data set. Box and whisker plots are also very useful when large numbers of observations are involved and when two or more data sets are being compared. **Note:** *The lines extending from the box may be adjusted to represent a certain fraction of the data. They could be set at 5% and 95% or they could represent the minimum and maximum values.*



20. What is a frequency table?

A frequency table is a way of summarizing a set of data. It is a record of how often each value (or set of values) of the variable in question occurs. It may be enhanced by the addition of percentages that fall into each category. A frequency table is used to summarize categorical, nominal, and ordinal data. It may also be used to summarize continuous data once the data set has been divided up into sensible groups.

Example: Suppose that in thirty shots at a target, a marksman makes the following scores:

5 2 2 3 4
 4 3 2 0 3
 0 3 2 1 5
 1 3 1 5 5
 2 4 0 0 4
 5 4 4 5 5

The frequencies of the different scores can be summarized as:

Score	Frequency	Frequency (%)
0	4	13%
1	3	10%
2	5	17%
3	5	17%
4	6	20%
5	7	23%

21. Why do astronomers use statistics?

Astronomers use statistics because they can't manipulate the universe in a laboratory the way a chemist can manipulate a compound or a biologist can manipulate a specimen. Since it is impossible to perturb some part of the population in order to see its effect, astronomers rely on standard sampling design and estimation methods in order to make conclusions regarding the universe. Also, processes in the universe take place over a very large time scale so noticeable changes are rare and tend to be studied in detail. As an example, consider stellar evolution. No one has ever observed a star go through its life cycle since the shortest cycles are about 10 million years long. However, astronomers can observe many stars at different stages in their life cycles and make predictions.

22. How do astronomers use sampling statistic techniques in their research?

Astronomers use two different sampling designs depending on the population being studied. If the population is finite in size, such as a cluster of stars or the HDFs, simple random sampling is chosen. If the population is very large and considered infinite, then more complex designs are used, depending on the characteristics of the population and the property being studied. Active galactic nuclei and halo stars are two populations that are considered infinite.

[Back to Top](#)

Science Background:

1. What is a galaxy?

A galaxy is an enormous collection of a few million to trillions of stars, gas, and dust held together by gravity. They can be several thousand to hundreds of thousands of light-years across. Galaxies can be placed into three main classes:

- **Elliptical galaxy:** A galaxy having an oval or nearly spherical shape. Some are more elongated than others. Elliptical galaxies are mostly composed of old stars and contain very little gas and dust. The smallest elliptical galaxies (called "dwarf ellipticals") are probably the most common type of galaxy in the nearby universe.
- **Spiral galaxy:** A galaxy made up of a disk with spiral (pinwheel-shaped) arms, a bulge near its center, and a halo. The sizes of the disk and bulge vary. The galaxy is composed of a mixture of old and young stars as well as gas and dust. The spiral arms are sites of active star formation. The majority of large galaxies in the nearby universe are spirals.
- **Irregular galaxy:** A galaxy whose shape is neither elliptical nor spiral. It contains both young and old stars and is often rich in gas and dust. These galaxies often have active regions of star formation. Sometimes the irregular shape of these galaxies results from interactions or collisions between galaxies. Observations such as the Hubble Deep Fields show that irregular galaxies were more common in the distant (early) universe.

2. Why do astronomers study galaxies?

Galaxies are visible to vast distances, and trace the structure of the visible universe with their collections of billions of stars, gas, and dust. By studying other galaxies, astronomers learn more about the Milky Way galaxy, the galaxy that contains our solar system. Answers to such questions as: "Do all galaxies have the same shape?,"

"Are all galaxies the same size?," "Do they all have the same number of stars?," and "How and when did galaxies form?" help astronomers learn about the history of the universe.

3. Why do astronomers study distant galaxies, if they are faint and hard to observe?

When we study astronomical objects, we are actually looking back in time. Light from the Sun takes 8 minutes to reach Earth. The light we see today from the next nearest star was emitted about four years ago. Light from the nearest galaxy that is like our own, Andromeda, takes over 2 million years to reach us. That is, we see Andromeda as it appeared more than 2 million years ago! Observations of distant galaxies show us what the universe looked like at an earlier time in the history of the universe. By studying the properties of galaxies at different epochs, we can map the evolution of the universe.

4. When astronomers study distant galaxies what do they look at?

They observe many properties of each galaxy including size, shape, brightness, color, amount of star formation, and distance from us. This information helps astronomers to determine how these structures may have formed and evolved.

5. What is a "deep" field?

In astronomical terms, a deep field is a long exposure observation taken to view very faint objects. Light from these objects is collected over a large period of time, so the detectors have a chance to gather as much light as possible. Objects can be very far away and appear faint to us due to the vast distances over which the light must travel. However, objects can lie close to us and be faint because they don't give off much light. So "deep" doesn't necessarily mean far. However, in the case of the Hubble Deep Fields, deep does mean far away since the images were taken in locally empty

areas.

6. What are the Hubble Deep Fields?

The Hubble Deep Field project was inspired by some of the first deep images to return from the Hubble Space Telescope after the 1993 servicing mission. These images showed that the early universe contained galaxies in a bewildering variety of shapes and sizes. Some had the familiar elliptical and spiral shapes seen among normal galaxies, but there were many peculiar shapes as well. Such images of the early universe are likely to be one of the enduring legacies of the Hubble Space Telescope.

Impressed by the results of earlier observations such as the Hubble Medium Deep Survey, a special advisory committee convened by Robert Williams, then Director of the Space Telescope Science Institute (STScI), recommended that he use a significant fraction of his annual director's discretionary time to take the deepest optical picture of the universe, by aiming Hubble for 150 consecutive orbits on a single piece of sky. The research was done by pointing the telescope at one spot in the northern sky for 10 days in December of 1995.

Thousands of never-before-seen galaxies were visible in the image, called the "Hubble Deep Field" (later named the HDF-North). Besides the classical spiral galaxies seen in the image, the variety of other galaxy shapes and colors serve as important clues to understanding the evolution of the universe. Some of the galaxies may have formed less than one billion years after the Big Bang. Hubble took a second deep look in the southern hemisphere in October of 1998, the HDF-South, to see if a similar result would be obtained.

Images from the Hubble Deep Field project were made available to astronomers around the world shortly after completion.

7. How were the two Hubble Deep Field sites chosen?

Each of the Hubble Deep Fields represents a "carefully selected random spot on the sky." To allow the Hubble Space Telescope to peer deeply into the sky, astronomers selected a special region of Hubble's orbit where Hubble can view the sky without being blocked by the Earth or experiencing interference from the Sun or Moon. The field also had to be far away from the plane of our own galaxy, to avoid being cluttered with objects in our galaxy. Finally, the field needed to have nearby guide stars, used to keep Hubble pointed at the field. These criteria led to the selection of a spot of sky near the handle of the Big Dipper in the northern hemisphere, and a spot of the sky in the constellation Tucana in the southern hemisphere.

8. If there are thousands of galaxies visible in the Hubble Deep Fields, why does the activity use just over 1,000 as the populations of each HDF?

Because of the way astronomers' instruments work, they can be reasonably sure that they have detected all galaxies with a certain range of brightnesses in the Hubble Deep Fields. Astronomers may be able to identify fainter objects, but they cannot be sure that they have detected all of the fainter objects that exist. When studying populations of objects, astronomers need to make sure that the sample they choose is representative. The very faintest objects do not form a representative sample, since astronomers do not know if they have detected all of the faintest objects. Therefore, astronomers limit their sample to objects in a certain brightness range. The sample is then said to be "statistically complete" to that brightness level. For the HDF-N, the statistically complete sample consists of 1067 galaxies.

9. What is the importance of the HDF?

The HDF contains some of the faintest galaxies we've ever been able to see over a large range of distances. Since seemingly "empty" spots were chosen, most of the galaxies in the Deep Fields lay billions of light-years away. The images show that the early universe contained galaxies in a bewildering variety of shapes and sizes. Some had the familiar elliptical and spiral shapes seen among galaxies today, but there were many peculiar shapes as well. Some of the galaxies may have formed less than one billion years after the Big Bang. The HDFs are important because they can help answer questions such as: How many galaxies are there in the universe? How did large-scale structure evolve in the universe? How were galaxies assembled? Is the universe open or closed? What is the age of the Universe? The HDF has been compared to a geological core sample of the earth, done on the sky.

10. Why was a second Deep Field taken?

The HDF-N covers a very small fraction of the sky. It would take 27 million fields and well over 500,000 years to use Hubble to survey the entire sky to the depths of the HDF. So, astronomers must rely on a thin "looking-through-a-soda-straw" view across the cosmos to infer the history of star and galaxy formation. Taking a second Deep Field helps astronomers to confirm that the HDF-N is representative, and that it is not unusual in some way. The two HDFs are, in fact, consistent with the common assumption that the universe should look largely the same in any direction we look.

11. It seems hard to see the shape of some of the galaxies in the HDFs. How do astronomers classify them?

They use the colors of the galaxies. Different types of galaxies tend to be different colors. For example, elliptical galaxies have reddish colors because they are mostly composed of old red stars. Astronomers study the colors of nearby elliptical, spiral, and irregular galaxies and compare these colors to those of the galaxies in the Hubble Deep Fields. Comparing the colors allows them to classify the galaxies.

12. What is the most common type of galaxy in the nearby universe?

When one counts both large and small galaxies, dwarf ellipticals (small ellipticals) are probably the most common type of galaxy in the nearby universe. Since these galaxies are small and faint, the exact number of these galaxies is not well known. The majority of large, bright galaxies in the nearby universe are spirals. Large bright elliptical galaxies are relatively rare.

Words from a scientist

by Denise Smith, Ph.D., Deputy Head-Office of Public Outreach, Space Telescope Science Institute

Our Milky Way galaxy is just one of the many billions of galaxies that inhabit the universe. These galaxies come in different shapes, sizes, and colors. You will notice this variety in morphology as you capture galaxies in "Galaxy Hunter."

Studying the appearance of galaxies in the Hubble Deep Fields can help us understand how galaxies formed and how their properties have changed over time. For example, the presence of many irregularly shaped, blue objects in the Hubble Deep Fields may indicate that collisions between galaxies and episodes of rapid star formation, or

"starbursts," were more common in the past. This possibility is very exciting to me, since I study starbursts in nearby galaxies.

Understanding results from the Hubble Deep Fields ultimately rests in mathematics, however. Our studies are based upon samples of galaxies drawn from the Hubble Deep Fields, and thus rely heavily on our understanding of sampling techniques and statistics. Come join us in "Galaxy Hunter" to see how important mathematics can be!

[Back to Top](#)

Grab Bag

Downloadable activity documents

- Images of [HDF-N](#) and [HDF-S](#)
- [Student Travelogue](#)

Web resources

One of the unique features of using the Internet is the ability to move quickly and easily to other links of related topics. If you find an appropriate related link, don't forget to bookmark it for future use. These related links can play an important role in enhancing your lesson. They can provide the latest information on astronomy or more information for a research topic. They also can be used to introduce, reinforce, or extend a concept or topic. Sample Web resources from the [Space Telescope Science Institute](#) and others are listed below.

WEB RESOURCES FROM THE SPACE TELESCOPE SCIENCE INSTITUTE (STSCI):

Amazing Space

An award-winning website that includes a comprehensive collection of standards-based science education materials. Materials and activities are organized by topic and type.

HubbleSite

Contains news releases of Hubble's science and remarkable discoveries, illustrated facts about the telescope and its instruments, and gorgeous pictures of stars, planets, galaxies, nebulae, and more.

James Webb Space Telescope

Contains news releases and information about the James Webb Space Telescope, which will use infrared sensing technology to look into the farthest reaches of the universe.

HubbleSite News Release Archive: Galaxies

OTHER WEB RESOURCES:

Below you will find links to additional information that comes from sources other than STScI. Some of the resources are quite in-depth, and it is easy to get lost. It is suggested that you bookmark this page before you examine each source. That way, should you get lost, you can use the bookmark to return to this page. Since other parties run the sites listed below, there is no way of guaranteeing their stability. You may occasionally find a site "under construction," or you may find that a site is completely gone. There is no way to control such situations because the sites are out of our control. This page is only a fraction of the information available on the Web.

If you do not find what you are looking for here, [GO FIND IT!](#)

- [Sample Size Calculator](#) This site calculates sample size given a desired confidence level and interval. It will also calculate confidence intervals.
- [Rice Virtual Lab in Statistics](#) This site has simulations and demonstrations, case studies, an analysis lab and a link to an online statistics book.
- [Research Randomizer](#) This site provides an online random number generator.

- [Astro-physical Calculator](#) A site that provides a chart of astronomical measurements and an online calculator.
- [Hawaii Active Catalog of the Hubble Deep Field](#) This site features a clickable map with accompanying data for the HDF-N.
- [Hubble's Messier Catalog](#) A guide to 110 objects in the Messier catalog of celestial objects.
- [Windows to the Universe](#) This site contains information on galaxy types including information on Edwin Hubble's classification system.
- [Bad Astronomy](#) A site maintained by an astronomer that details the “bad astronomy” reflected or portrayed in movies, print, and TV media.
- [From the Big Bang to the End of the Universe: The Mysteries of Deep Space Timeline](#) A timeline for the universe and its evolution. The site also contains links to other sites at PBS, including some lesson plans.
- [The National Historical Geographic Information System](#) A resource that provides easy access to summary tables and time series of population, housing, agriculture, and economic data, along with GIS-compatible boundary files, for years from 1790 through the present and for all levels of U.S. census geography, including states, counties, tracts, and blocks.
- [USNO](#) A database site for creating tables for sunrise/sunset, moonrise/moonset, or the beginning and end of twilight, for one year for any location.
- [Project SkyMath: Making Mathematical Connections](#) A middle school level curriculum integrating math skills using temperature as the context. Some simple statistics and sampling activities are presented.
- [AstroCappella](#) AstroCappella is a marriage of astronomy and music, developed by astronomers and professionally recorded by the a cappella group, The Chromatics. There is a statistics activity listed with the song "HST-Bop."

Books and printed material

- Moore, David S. and McCabe, George P. (1997). *Introduction to the Practice of Statistics, 2nd edition*. W. H. Freeman and Company.
- Triola, Mario F. (1994). *Elementary Statistics, 6th edition*. Addison Wesley.
- Senk, Sharon L., et. al. (1998). *The UCSMP Functions, Statistics, and Trigonometry, Integrated Mathematics, 2nd edition*. Scott Foresman.
- Bohan, James F. (2000). *AMSCO's AP Statistics, Preparing for the Advanced Placement Examination*. AMSCO School Publication Inc.

Contact us: Send your comments about this activity or these Teaching Tips to amazing-space@stsci.edu

[Back to Top](#)