What Is **Statistics***?*

What Is **Statistics***?*

The word **statistics** entered the English language in the 1790s as a term to describe the measurement of characteristics of nations or states (hence the term **stat**istics).

$\approx$

What Is **Statistics***?*

In modern usage, statistics is, quite simply, the science of data.

$\approx$

OK, So What Are Data?

OK, So What Are Data?

Data are facts that convey information.

OK, So What Are Data?

Data are facts that convey information. Here are some examples:

OK, So What Are Data?

Data are facts that convey information. Here are some examples:

- Yearly average global surface air temperatures.

OK, So What Are Data?

Data are facts that convey information. Here are some examples:

- Yearly average global surface air temperatures.
- Lifetimes of electrical transformers.

OK, So What Are Data?

Data are facts that convey information. Here are some examples:

- Yearly average global surface air temperatures.
- Lifetimes of electrical transformers.
- Breaking strengths of metal pins.

OK, So What Are Data?

Data are facts that convey information. Here are some examples:

- Yearly average global surface air temperatures.
- Lifetimes of electrical transformers.
- Breaking strengths of metal pins.
- Percentage of lymphoma patients treated at various medical centers who survive more than 5 years from diagnosis.

OK, So What Are Data?

Data are facts that convey information. Here are some examples:

- Yearly average global surface air temperatures.
- Lifetimes of electrical transformers.
- Breaking strengths of metal pins.
- Percentage of lymphoma patients treated at various medical centers who survive more than 5 years from diagnosis.

Here are some numbers that convey information, but just don't add up:

$$\approx$$


NEW CUYAMA
Population 562
Ft. above sea level 2150
Established 1951
TOTAL 4663

What Is this Course About?

What Is this Course About?

In this course, we will focus on **statistical inference**, which is the use of a subset of a population to draw conclusions about the entire population.

What Is this Course About?

In this course, we will focus on **statistical inference**, which is the use of a subset of a population to draw conclusions about the entire population.

The subset is called a **sample**. To enable us to do statistical inference in a scientific way, we will need to select the sample in a statistically valid way. This will mean choosing it according to a known chance mechanism (think of tossing a coin).

$\approx$

What Is this Course About?

You might think the results would be random:

What Is this Course About?

You might think the results would be random: and they are!

What Is this Course About?

You might think the results would be random: and they are!

However, they are not completely random. There is just enough regularity to the resulting randomness to enable us to draw useful scientific conclusions.

$\approx$

What Is this Course About?

The course will be based on four main themes:

What Is this Course About?

The course will be based on four main themes:

- **Producing Data:** How to obtain data in a statistically valid way; types of studies done with the resulting data.

What Is this Course About?

The course will be based on four main themes:

- **Producing Data:** How to obtain data in a statistically valid way; types of studies done with the resulting data.
- **Summarizing Data:** Appropriate graphical and numerical summaries of data.

What Is this Course About?

The course will be based on four main themes:

- **Producing Data:** How to obtain data in a statistically valid way; types of studies done with the resulting data.
- **Summarizing Data:** Appropriate graphical and numerical summaries of data.
- **Estimation:** Using data to estimate population quantities of interest.

What Is this Course About?

The course will be based on four main themes:

- **Producing Data:** How to obtain data in a statistically valid way; types of studies done with the resulting data.
- **Summarizing Data:** Appropriate graphical and numerical summaries of data.
- **Estimation:** Using data to estimate population quantities of interest.
- **Hypothesis Tests:** Using data to test hypotheses about population quantities of interest.

$\approx$

So let's get started on the first of these: Producing Data.

$\approx$