Tests of Significance

It is very common for a claim to be made about some characteristic of the population.

Tests of Significance

It is very common for a claim to be made about some characteristic of the population.

Possible Examples:

We know that 7% of the population is left handed.

Tests of Significance

It is very common for a claim to be made about some characteristic of the population.

Possible Examples:

We know that 7% of the population is left handed.

We know that the average height of college students is 6 feet 2 inches.

Tests of Significance

It is very common for a claim to be made about some characteristic of the population.

Possible Examples:

We know that 7% of the population is left handed.

We know that the average height of college students is 6 feet 2 inches.

We know that the average household uses 4 killowatt hours of electricity per month.

Tests of Significance

It is very common for a claim to be made about some characteristic of the population.

Possible Examples:

We know that 7% of the population is left handed.

We know that the average height of college students is 6 feet 2 inches.

We know that the average household uses 4 killowatt hours of electricity per month.

We know that the average age in Easton is 31.4 years.

Tests of Significance

It is very common for a claim to be made about some characteristic of the population.

Possible Examples:

We know that 7% of the population is left handed.

We know that the average height of college students is 6 feet 2 inches.

We know that the average household uses 4 killowatt hours of electricity per month.

We know that the average age in Easton is 31.4 years.

We know that the average speed on Route 495 is 81 mph.

Tests of Significance

It is very common for a claim to be made about some characteristic of the population.

Possible Examples:

We know that 7% of the population is left handed.

We know that the average height of college students is 6 feet 2 inches.

We know that the average household uses 4 killowatt hours of electricity per month.

We know that the average age in Easton is 31.4 years.

We know that the average speed on Route 495 is 81 mph.

We know that 2% of the population is homeless.

Tests of Significance

Naturally, we don't want to accept every claim as true, but there is usually no way of determining with 100% certainty that it is true (or false).

Tests of Significance

Naturally, we don't want to accept every claim as true, but there is usually no way of determining with 100% certainty that it is true (or false).

Suppose I have either a normal coin or a two-headed coin.

Tests of Significance

Naturally, we don't want to accept every claim as true, but there is usually no way of determining with 100% certainty that it is true (or false).

Suppose I have either a normal coin or a two-headed coin.

Let's say you want to know which coin I have.

You ask me to show you the coin and for five dollars I offer to toss the coin and tell you the outcome

Tests of Significance

Naturally, we don't want to accept every claim as true, but there is usually no way of determining with 100% certainty that it is true (or false).

Suppose I have either a normal coin or a two-headed coin.

Let's say you want to know which coin I have.

You ask me to show you the coin and for five dollars I offer to toss the coin and tell you the outcome

(truthfully)

Tests of Significance

Naturally, we don't want to accept every claim as true, but there is usually no way of determining with 100% certainty that it is true (or false).

Suppose I have either a normal coin or a two-headed coin.

Let's say you want to know which coin I have.

You ask me to show you the coin and for five dollars I offer to toss the coin and tell you the outcome

(truthfully)

I toss the coin, and tell you the result was tails.

You now know for sure that I have a normal coin.

Tests of Significance

Suppose instead the toss came up heads.

Tests of Significance

Suppose instead the toss came up heads.

If the coin is fair, the probability of heads is $1/2$.

Tests of Significance

Suppose instead the toss came up heads.

If the coin is fair, the probability of heads is $1/2$.

Of course if I have a two-headed coin, the probability of heads is 1.

Tests of Significance

Suppose instead the toss came up heads.

If the coin is fair, the probability of heads is $1/2$.

Of course if I have a two-headed coin, the probability of heads is 1.

You cannot rule out either possibility.

Tests of Significance

Suppose instead the toss came up heads.

If the coin is fair, the probability of heads is $1/2$.

Of course if I have a two-headed coin, the probability of heads is 1.

You cannot rule out either possibility.

In fact, unless tails comes up, you can *never* be certain that it is not a fair coin.

Tests of Significance

You pay the money, I toss the coin, and it comes up heads.

Tests of Significance

You pay the money, I toss the coin, and it comes up heads.

You still aren't sure which coin I have.

Tests of Significance

You pay the money, I toss the coin, and it comes up heads.

You still aren't sure which coin I have.

However, the probability of two heads in a row with a fair coin is $1/4$.

Tests of Significance

You pay the money, I toss the coin, and it comes up heads.

You still aren't sure which coin I have.

However, the probability of two heads in a row with a fair coin is $1/4$.

You pay another five dollars, and the coin comes up heads.

Tests of Significance

You pay the money, I toss the coin, and it comes up heads.

You still aren't sure which coin I have.

However, the probability of two heads in a row with a fair coin is $1/4$.

You pay another five dollars, and the coin comes up heads.

Now the probability of heads coming up three times in three tosses is $1/8$.

Tests of Significance

The coin comes up heads for the fourth time.

Tests of Significance

The coin comes up heads for the fourth time.

The probability of four heads in four tosses is $1/16$.

Tests of Significance

The coin comes up heads for the fourth time.

The probability of four heads in four tosses is $1/16$.

Each time we extend the streak of heads, the probability is reduced by a factor of $1/2$.

Tests of Significance

The coin comes up heads for the fourth time.

The probability of four heads in four tosses is $1/16$.

Each time we extend the streak of heads, the probability is reduced by a factor of $1/2$.

If we kept going for 10 tosses and they all came up heads, the probability of this outcome with a fair coin is $1/2^10$ or about 1 in 1000.

Tests of Significance

The coin comes up heads for the fourth time.

The probability of four heads in four tosses is $1/16$.

Each time we extend the streak of heads, the probability is reduced by a factor of $1/2$.

If we kept going for 10 tosses and they all came up heads, the probability of this outcome with a fair coin is $1/2^10$ or about 1 in 1000.

Of course it cost you 50 to be this sure.

Tests of Significance

The coin comes up heads for the fourth time.

The probability of four heads in four tosses is $1/16$.

Each time we extend the streak of heads, the probability is reduced by a factor of $1/2$.

If we kept going for 10 tosses and they all came up heads, the probability of this outcome with a fair coin is $1/2^10$ or about 1 in 1000.

Of course it cost you 50 to be this sure.

We can continue this indefinitely, but at some point you will run out of money.

Tests of Significance

At that point, you will have to decide which of the alternatives is the actual state of things:

Tests of Significance

At that point, you will have to decide which of the alternatives is the actual state of things:

1) I have a fair coin, and you just had a *very* unlikely event occur

Tests of Significance

At that point, you will have to decide which of the alternatives is the actual state of things:

1) I have a fair coin, and you just had a *very* unlikely event occur

2) The coin has both sides heads

Tests of Significance

At that point, you will have to decide which of the alternatives is the actual state of things:

1) I have a fair coin, and you just had a *very* unlikely event occur

2) The coin has both sides heads

The more tosses you paid for, the more convinced you will be that the coin is actually two-headed

Tests of Significance

At that point, you will have to decide which of the alternatives is the actual state of things:

1) I have a fair coin, and you just had a *very* unlikely event occur

2) The coin has both sides heads

The more tosses you paid for, the more convinced you will be that the coin is actually two-headed

This is because the more tosses, the smaller the probability that they all come up heads.

Tests of Significance

At that point, you will have to decide which of the alternatives is the actual state of things:

1) I have a fair coin, and you just had a *very* unlikely event occur

2) The coin has both sides heads

The more tosses you paid for, the more convinced you will be that the coin is actually two-headed

This is because the more tosses, the smaller the probability that they all come up heads.

While you can't be sure you are right if you say the coin has two heads, you can say that the probability you are wrong gets smaller with each toss.

Tests of Significance

This is the idea behind significance testing.

Tests of Significance

This is the idea behind significance testing.

You have some statement about the population, such as "the mean is 15.3"

Tests of Significance

This is the idea behind significance testing.

You have some statement about the population, such as "the mean is 15.3"

You draw a sample, and calculate the mean of the sample.

Tests of Significance

This is the idea behind significance testing.

You have some statement about the population, such as "the mean is 15.3"

You draw a sample, and calculate the mean of the sample.

The closer the sample mean is to 15.3, the more inclined you are to accept the statement as true.

Tests of Significance

This is the idea behind significance testing.

You have some statement about the population, such as "the mean is 15.3"

You draw a sample, and calculate the mean of the sample.

The closer the sample mean is to 15.3, the more inclined you are to accept the statement as true.

If the mean is nowhere near 15.3, you would tend to reject the statement as false.

Tests of Significance

This is the idea behind significance testing.

You have some statement about the population, such as "the mean is 15.3"

You draw a sample, and calculate the mean of the sample.

The closer the sample mean is to 15.3, the more inclined you are to accept the statement as true.

If the mean is nowhere near 15.3, you would tend to reject the statement as false.

This is actually very similar to the confidence interval problem.

Tests of Significance

In fact, you can use confidence intervals to test the likelihood that statements of this kind are true.

Tests of Significance

In fact, you can use confidence intervals to test the likelihood that statements of this kind are true.

The procedure works this way. A claim is made about some measurable quantity.

Tests of Significance

In fact, you can use confidence intervals to test the likelihood that statements of this kind are true.

The procedure works this way. A claim is made about some measurable quantity.

The claim is usually referred to as the *null hypothesis*

Tests of Significance

In fact, you can use confidence intervals to test the likelihood that statements of this kind are true.

The procedure works this way. A claim is made about some measurable quantity.

The claim is usually referred to as the *null hypothesis*

Another way to describe the null hypothesis is that it is the status quo, if there is no effect.

Tests of Significance

In fact, you can use confidence intervals to test the likelihood that statements of this kind are true.

The procedure works this way. A claim is made about some measurable quantity.

The claim is usually referred to as the *null hypothesis*

Another way to describe the null hypothesis is that it is the status quo, if there is no effect.

The opposite position on the issue is called the *alternative hypothesis*

Tests of Significance

For example, in clinical trials, the null hypothesis is usually the statement that the drug is not effective.

Tests of Significance

For example, in clinical trials, the null hypothesis is usually the statement that the drug is not effective.

The alternate hypothesis is that the drug is effective.

Tests of Significance

For example, in clinical trials, the null hypothesis is usually the statement that the drug is not effective.

The alternate hypothesis is that the drug is effective.

In evaluation of school departments, the null hypothesis would probably be that the school is within the range you would expect for similar schools in the state.

Tests of Significance

For example, in clinical trials, the null hypothesis is usually the statement that the drug is not effective.

The alternate hypothesis is that the drug is effective.

In evaluation of school departments, the null hypothesis would probably be that the school is within the range you would expect for similar schools in the state.

The alternate hypothesis might be that they are outside the range expected.

Tests of Significance

For example, in clinical trials, the null hypothesis is usually the statement that the drug is not effective.

The alternate hypothesis is that the drug is effective.

In evaluation of school departments, the null hypothesis would probably be that the school is within the range you would expect for similar schools in the state.

The alternate hypothesis might be that they are outside the range expected.

It could also be that they are worse.

Tests of Significance

The procedure is to take a sample of the measure in question.

Tests of Significance

The procedure is to take a sample of the measure in question.

Next compute a confidence interval for the mean with the desired level of confidence.

Tests of Significance

The procedure is to take a sample of the measure in question.

Next compute a confidence interval for the mean with the desired level of confidence.

If the confidence interval includes the value of the measure when the null hypothesis is true, we *accept* the null hypothesis.

Tests of Significance

The procedure is to take a sample of the measure in question.

Next compute a confidence interval for the mean with the desired level of confidence.

If the confidence interval includes the value of the measure when the null hypothesis is true, we *accept* the null hypothesis.

If it does not, we *reject* the null hypothesis.

Tests of Significance

The procedure is to take a sample of the measure in question.

Next compute a confidence interval for the mean with the desired level of confidence.

If the confidence interval includes the value of the measure when the null hypothesis is true, we *accept* the null hypothesis.

If it does not, we *reject* the null hypothesis.