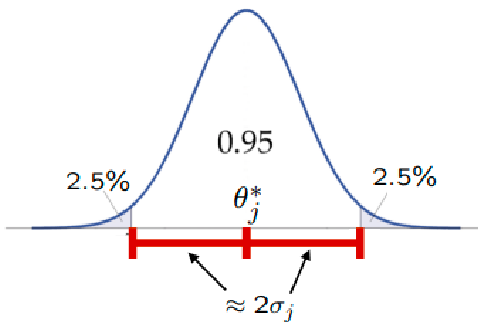
Confidence Intervals and Hypothesis Testing

Now we will see how the standard errors can be used to construct **Confidence Intervals.**

Confidence Interval (CI):



One key point to remember is that there's a true value of theta ($\theta^* j$) and we're looking at the jth component of θ .
Our estimate is going to be somewhere near that true value and is described by a normal distribution.

$$\widehat{\Theta}_j \sim \mathcal{N}(\theta_j^*, \sigma_j^2)$$

That normal distribution has a certain standard deviation. So our estimate is going to be within two standard deviations from the true value with a probability of 95%.

And with a probability of 95%, the error is going to be less than two standard deviations or less than two standard errors.

- With probability 95%: $|\text{error}| = |\widehat{\Theta}_j - \theta_j^*| \leq 2\sigma_j$

Another way of writing this statement is that with a probability of 95%, the true value of the $\theta^* j$ is going to be inside this interval.

- 95%-CI: $\left[\widehat{\Theta}_j - 2\hat{\sigma}_j,\ \ \widehat{\Theta}_j + 2\hat{\sigma}_j\right]$

$$\mathbb{P}(\theta_j^* \in \text{CI}) \approx 0.95$$

So, we know that with a probability of 95% the error is at most two standard deviations, which means θ^*_j will be within two standard deviations from the estimates with this probability of 95%. So, this interval is called a **Confidence Interval**.

The confidence interval is calculated by adding or subtracting two standard deviations from the estimates.

The defining property of a confidence interval is that with a probability of 95%, the true value of the parameter is inside the confidence interval.

There's nothing special about 95%. We could also use a 99% confidence interval. In that case, the confidence interval would be wider and instead of having a standard deviation range of 2, we would have some larger value.

However, one important caveat, we need to interpret this below statement, what exactly does it mean?

$$P(\theta^*_j \in CI) \simeq 0.95 ,$$

It's a statement about probability. It's a statement about something being random. The random thing happening here is not θ^*_j . θ^*_j is assumed to be a constant. The random thing here is the confidence interval.

So the way to think about the confidence interval is the following.

There's a true value which is θ^*_j. We do our estimation that produces an estimate and the confidence interval. And we go and report to the confidence interval.

Now if we carry out a similar estimation again using a new data set then we get another confidence interval. Similarly if we do the estimation again using another data set we get a different confidence interval.

And so what the probability statement above is saying is that, 95% of the time when we use this estimation procedure, the confidence is going to capture the true value.

So think of a person who keeps using regression day after day many times. So each time he uses it, he gets a different confidence interval. And, 95% of those confidence intervals will capture the correct value. So, 95% of the time our confidence intervals will be good in the sense that the correct value will happen to be in there.

However, there's nothing random about $\theta^{*}j$.
If we get a confidence interval for example, CI = [2, 4]

Then we are not allowed to say that $\theta^{*}j$ is in there with probability 95%. That's a wrong statement. Because $\theta^{*}j$ is not random. We cannot talk about the probability of $\theta^{*}j$ falling in there or not falling in there.
Rather, the correct statement is the so-called "frequentist" interpretation that tells us we're running our regression many times and 95% of those times we're capturing the correct value.

Now, let us use this idea of confidence intervals to actually test the hypothesis.

Testing the hypothesis $\theta^{*}j = 0$:

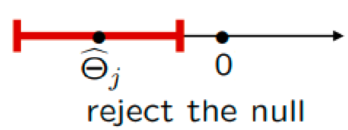
A hypothesis might be something like, "Perhaps a specific coefficient is zero."
For example a hypothesis might be that, the coefficient associated with **newspaper advertisement** is zero.
Now, the question is does the data support this hypothesis?

- Is the data compatible with the **null hypothesis** $\theta^{*}j = 0$?

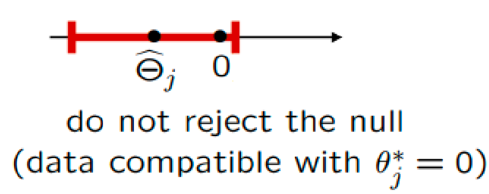
We call this the null hypothesis. It's called 'null' in the sense that nothing is happening under the null hypothesis and this is the default hypothesis.
And we have a very simple way of deciding about that which is called Wald test and it goes as follows.

Wald Test:

We're asking whether theta ($\theta^{*}j$) might be 0. We form an estimate and also a confidence interval around that estimate. If that confidence interval does not capture 0, as shown in the picture, then we have evidence that the true value of theta ($\theta^{*}j$) is not 0. And in this case, we reject the null hypothesis, which is like making an assertion that the true theta ($\theta^{*}j$) is non-zero.



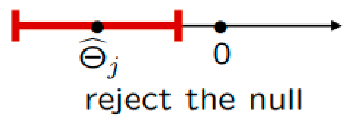
If on the other hand, the picture is like the one shown here i.e here the confidence interval does contain 0. We consider that to be evidence that 0 is compatible with the data that we have seen. And we do not reject the null hypothesis.



So this is a pretty simple way of doing these tests. When we have confidence intervals in our hands, we just check whether 0 is inside or outside and we use that to make a decision.

Now, when does this test get things wrong?

Suppose that the true hypothesis is $\theta^{*}j = 0$.
Now, how likely is it that we're going to make a mistake and reject it?
That's going to happen if the confidence interval happens to be like this picture below.



Now, how likely is it that the true value is away from the confidence interval or how likely is it that the confidence interval misses the true value of zero?
By definition of a confidence interval, this value is just 5%. The way that we build the confidence interval is that 95% of the time it captures the true thing, so it misses it only 5% of the time.

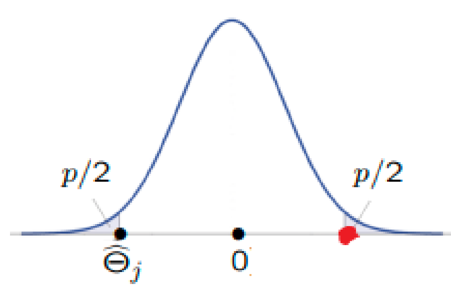
$$P\ (\text{reject} \mid \theta^{*}j = 0)$$
$$= P\ (\text{the CI “misses” } 0 \mid \theta^{*}j = 0)$$
$$\simeq 5\ \%$$

Besides running the hypothesis, another indirect way of checking various hypotheses is by reporting **p-values**. And p-values are very popular in reporting such results.

p-value:

Here, we're taking the hypothesis where the true value is zero i.e $\theta^{*}j = 0$.

We have the probability distribution of $\widehat{\theta}$ under that hypothesis.
And then we observe the value of the estimate and we want to quantify how far is the value.
Well, say if the value of the estimate is the red point in the picture, then we could look at the probability on it's side. And similarly the corresponding probability on the other side. That is a certain number p which is the so-called p-value.



So **the p-value is the probability of getting something that extreme as the observed $\hat{\theta}_j$, under $\theta^* j = 0$.**

If $\theta^* j = 0$, then how likely is it that we will get something that extreme?
That's the probability under the tails of that normal that we have shown in the above picture.

If the p-value is less than **0.05**, then the probability of getting something so extreme is only 5%. So the data seem to be incompatible with the hypothesis. So **we will reject the hypothesis** and consider that $\theta^* j$ is non-zero.

Note that p-values can also be misinterpreted or misleading.