18. CONFIDENCE INTERVALS, INTRODUCTION

“Statistics is never having to say you're certain”.
(Tee shirt, American Statistical Association).

The confidence interval is one way of conveying our uncertainty about a parameter.

It’s misleading (and maybe dangerous) to pretend we’re certain.

It is not enough to provide a guess (*point estimate*) for the parameter.

We also have to say something about how far such an estimator is likely to be from the true parameter value.
With a confidence interval, we report a range of numbers, in which we hope the true parameter will lie. The interval is centered at the estimated value, and the width ("margin of error") is an appropriate multiple of the standard error.

We can think of the margin of error as "fuzz", introduced to account for sampling variability.

Example: “A recent poll indicates that 74 percent of the people are in favor of President Bush as president. Margin of error is ± 3 percent.” (Gallup)

We have “confidence” that the method will work, since we can control the probability that such an interval will fail to contain the true proportion of people favoring Bush.
Why We Need to Provide Standard Errors

**Example:** The Presidential election in 2000 pitted Gore against Bush. A poll taken just before the election showed that 51% of the sample intended to vote for Bush. The sample consisted of 2,500 registered voters. The polling organization said the election was too close to call. Why?

First, remember that people can still change their minds, there may have been biases in the polling methods, etc.

Besides such problems, there will always be inexactness in the sample proportion, due to sampling variation.
The sample proportion for Bush ($\hat{p} = 0.51$) is a point estimate (a single guess) of the true proportion ($p$) of all voters who are for Bush.

But $\hat{p}$ is very unlikely to be exactly equal to $p$.

A different sample would have yielded a different guess.

By just reporting the value you happened to get (0.51), you give people the false impression that your estimate is actually the true value.

This gives you a bogus air of certainty.
You’re pretending you know that Bush is winning.

You’re ignoring the “margin of error” inherent in the sampling process.

It’s only human to deny the uncertainty and simply say that Bush is winning, but you must fight this urge.

A way around the problem is to report standard error information along with the estimator.

The standard error measures the uncertainty in the sample proportion. The smaller the standard error, the more reliable (accurate, exact) the estimate is.
If we only polled 20 people, we would not expect a sample proportion (such as 0.51) to estimate the true proportion with much accuracy, since the standard error is high.

With 2,500 people polled, the sample proportion has a lower standard error, and is therefore more accurate than above.

But we could still easily find 51% of those polled in favor of Bush even if Bush were actually losing \( (p < 50\%) \). (Detailed calculation below).

Even in large sample sizes, \( \hat{p} \) is a fuzzy measurement of \( p \). Fortunately, we can quantify this fuzziness by using standard errors, together with the Empirical Rule.
The Margin of Error: Typical Distance Between \( \hat{p} \) and \( p \)

If we happen to get a spectacularly non-representative sample, we could get any value at all for \( \hat{p} \). (The people polled could all be for Bush; A bag of M&Ms could contain all orange.) So we can’t absolutely guarantee that \( \hat{p} \) will be close to \( p \). What we can do, however, is to use \( \hat{p} \) to make bets on \( p \).

We could ask, for example, how likely it is that \( \hat{p} \) and \( p \) will differ by more than one percent.

Alternatively, we can ask how far \( \hat{p} \) will be from \( p \) most of the time. (Here, we’ll define “most of the time” to be 95%, or 19 times out of 20.)
So let’s try to find a range in which the difference between \( \hat{p} \) and \( p \) will lie 95% of the time. This tells us something about how “fuzzy” \( \hat{p} \) is, and allows us to make bets about \( p \).

Suppose for a moment that the true value of \( p \) is 0.51. Then the standard error of the sample proportion is

\[
\sqrt{\frac{0.51 \times 0.49}{2500}} = 0.01
\]

(This standard error won't change much if the true \( p \) is slightly different from 0.51).
According to the normal approximation to the binomial distribution, we can assume that the sample proportion is normally distributed with a mean of $p$ and a standard error of 0.01.

Therefore, by the Empirical Rule, 95% of the time the sample proportion would be within 0.02 (that is, two standard errors) of $p$.

Thus, it’s not so clear that Bush was really winning. Instead, it seems plausible that the true proportion of voters for Bush could be anywhere within two standard errors of the sample proportion, or in this case, between 0.49 and 0.53.

So it would be dangerous to call the results of an election based solely on the given sample proportion!

The interval (0.49, 0.53) is called a 95% confidence interval for $p$. 
Let’s turn now to the similar problem of estimating the mean of a population. (We’ll come back to proportions later).

Suppose we are trying to make inferences about a population mean $\mu$ based on a sample of size $n$.

The sample mean $\bar{x}$ is a point estimator of the parameter $\mu$. Used by itself, $\bar{x}$ is of limited usefulness because it contains no information about its own reliability.

Furthermore, the reporting of $\bar{x}$ alone may leave the false impression that $\bar{x}$ estimates $\mu$ with complete accuracy.

Remember the sampling lab. The $\bar{x}$ values you got were all different. So how can you seriously believe that YOUR $\bar{x}$ is actually equal to $\mu$?
• **Confidence Interval**: An interval with random endpoints which contains the parameter of interest (in this case, $\mu$) with a prespecified probability, denoted by $1 - \alpha$.

The confidence interval automatically provides a margin of error to account for the sampling variability of $\overline{X}$.

**Example**: A machine is supposed to fill “2-Liter” bottles of Pepsi. To see if the machine is working properly, we randomly select 100 bottles recently filled by the machine, and find that the average amount of Pepsi is 1.985 liters. Can we conclude that the machine is not working properly?
No! By simply reporting that $\bar{x} = 1.985$ liters, we are neglecting the fact that the amount of Pepsi varies from bottle to bottle and that the value of the sample mean depends on the luck of the draw. It is possible that a value as low as 1.985 is within the range of natural variability for $\bar{X}$, even if the average amount for all bottles is in fact $\mu = 2$ liters.

Suppose we know from past experience that the amounts of Pepsi in bottles filled by the machine have a standard deviation of $\sigma = 0.05$ liters.

Since $n = 100$, we can assume (using the Central Limit Theorem) that $\bar{X}$ is normally distributed with mean $\mu$ (unknown) and standard error

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 0.005$$
From the Empirical Rule, the probability is about 95% that $\bar{X}$ will be within two standard errors of its mean.
So the probability is about 0.95 that \( \mu \) will be within 0.01 ounces of \( \bar{X} \).

Thus, the interval \( \bar{X} \pm 0.01 \) will contain \( \mu \) with probability about 0.95.

In general, the interval \( \bar{X} \pm 2 \frac{\sigma}{\sqrt{n}} \) will contain \( \mu \) with probability about 0.95.

Therefore, the interval provides (approximately) a 95% confidence interval for \( \mu \).