Business Statistics 41000

Lecture 6

Mladen Kolar

Outline of today's topics

Statistical Inference

- estimation and confidence intervals

Sampling distribution of $\bar{x}$, confidence intervals for μ

Sampling distribution of $\hat{p}$, confidence intervals for p

Statistical Inference

We have discussed possible models for data. Examples are:

- i.i.d. Bernoulli model
- i.i.d. Normal model
- random walk model
- regression model

Each of these models has unknown **parameters**.

How do we know the values for the parameters?

Example: Voting and the i.i.d. Bernoulli model

Suppose we have a large population of voters in Chicago and suppose each citizen will vote either Democrat or Republican.

We would like to know the fraction that will vote Democrat.
But the **population** is too large: **We can't ask them all!**

We **model** voters as i.i.d. Bernoulli. Let $X_i = 1$ if the i-th person votes Democrat and zero otherwise.

$$X_i \sim \mathrm{Bernoulli}(p) \quad \text{i.i.d.}$$

We do not know p!!!

Can we learn about p by taking a **sample** of voters?

Example: Auditing a pharma company

Suppose that you are auditing a pharmaceutical company that produces pills.
We want to make sure they put the correct amount of active ingredient is put into each pill.

How can we check that the claimed amount is actually put into each pill?
We cannot inspect every single pill.

We **model** the amount of active ingredient in each pill as i.i.d. normal.
Let X_i be the amount in milligrams in pill i.

$$X_i \sim N(\mu, \sigma^2) \quad \text{i.i.d.}$$

We do not know μ or σ^2!!!!

Can we learn about these parameters by looking at a sample of pills?

In both cases, we want to use observed data to learn something about the parameters of a model.

How do we do this?

Example 1:

We conduct a poll. We take a random sample of voters and we use
the proportion of Democrats in the sample as a guess or "estimate" of the
true proportion in the whole population.

Example 2:

We take a sample of pills and measure the amount of active ingredient
in each pill in our sample. We use the sample average and variance
as a guess for the expected amount of active ingredient and the variability of the process.

Example: Voting and the i.i.d. Bernoulli model

p is the fraction of Democrats in the whole population. This **parameter** is *unknown*.

$\hat{p}$: before we have taken the poll, we plan to use the proportion of Democrats in our sample of size n as our guess for the true value of p

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

We hope that $\hat{p} \approx p$.

Example: Voting and the i.i.d. Bernoulli model

p is the fraction of Democrats in the whole population. This **parameter** is *unknown*.

$\hat{p}$: before we have taken the poll, we plan to use the proportion of Democrats in our sample of size n as our guess for the true value of p

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

We hope that $\hat{p} \approx p$.

We treat our sample of voters from the poll $\{X_1, X_2, \ldots, X_n\}$ as the outcomes of i.i.d. Bernoulli r.v.'s from the population.

KEY IDEA: Recognize that there are many possible "samples" of voters that we could poll.

The value we get for $\hat{p}$ may be different for each sample.

Prior to conducting the poll, $\hat{p}$ is a random variable!! Why?

After conducting the poll, we observe an outcome for the random variable $\hat{p}$.

Example: Voting and the i.i.d. Bernoulli model

In real life, we (typically) only observe one sample.
For example, we only conduct one poll.

Given this sample, we get a specific value (a number) for p.

In general, the value we get for $\hat{p}$ from our sample will not be equal to the true, unknown value of p.

Today you will learn how to use the model to make a statement about how close $\hat{p}$ is to p.

Example: Auditing a pharma company

μ is the "true" expected amount of active ingredient. This **parameter** is *unknown*.

$\bar{x}$: before we get the data, we plan to use the sample average as a guess for μ

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

We hope that $\bar{x} \approx \mu$.

Example: Auditing a pharma company

μ is the "true" expected amount of active ingredient. This **parameter** is *unknown*.

$\bar{x}$: before we get the data, we plan to use the sample average as a guess for μ

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

We hope that $\bar{x} \approx \mu$.

We treat the amount of active ingredient measured in each pill, $X_1, X_2, \ldots, X_n$, as though it is an i.i.d. draw from a $N(\mu, \sigma^2)$ distribution.

KEY IDEA: Before we collect our data, $\bar{x}$ is a random variable.

This is because it will be different for each sample of pills we might observe.

Once we collect the data, we will observe an outcome for the random variable $\bar{x}$.

Example: Auditing a pharma company

In real life, we only observe one sample of pills.

Given this sample, we get a single value (that is, a specific number) for $\bar{x}$.

Question:

Can we use our model to make a statement about
how close $\bar{x}$ will be to μ?

Statistical terminology

An **estimator** is a function of the data that is used to infer the value of an unknown parameter in a statistical model. An estimator is a random variable because it is a function of the data, which are random variables.

- You can think of an estimator as a "rule" that determines how we will use our observed data to calculate our guess of the unknown parameter.

Example: $\hat{p}(X) = n^{-1} \sum_{i=1}^{n} X_i$ $\bar{x}(X) = n^{-1} \sum_{i=1}^{n} X_i$

An **estimate** is the observed outcome of the estimator for a specific data set. In other words, an estimate is a realization of the random variable (the estimator).

- Once you observe a specific data set $X_i = x_i$, you will be able to compute the estimate which is just a number.

Example: $\hat{p}(X) = 0.57$ $\bar{x}(X) = 50.12$

Statistical inference combines data analysis with probability.
We have a model in mind for our data but do not know what the parameters are.

1. Assume the data are i.i.d. draws from a model with unknown parameters.

2. Come up with an **estimator** of an unknown *parameter*.

   That is, a summary statistic that tells us about some aspect of the model.
   Before we see the data, our estimator is a random variable due to the fact that
   there are many samples we might observe.

3. Finally, after we have observed our sample, we get a numeric value that we call our **estimate**.
   Assuming our model is true, we use probability to say how close our estimator will be to the actual
   (unknown) value of the parameter.

The Sampling Distribution of an Estimator

How we use our model to make a statement about how close our estimator will be to the true value of the parameter?

We can ask questions like:

- On average, will our estimator tend to be too high or too low?
- Will the variance of the estimator be large or small?

The **sampling distribution** of an estimator is a probability distribution that describes all the possible outcomes of the estimator we **might see** if we could "repeat" our sample over and over again.

Key idea: imagine repeatedly drawing hypothetical samples of data from a large population defined by our model. For each sample, compute the value of the estimator. A histogram of the values is (approximately) the sampling distribution.

The Sampling Distribution for $\hat{p}$

Example: Voting and the i.i.d. Bernoulli model

There are many potential voters in Chicago. A fraction p of them are Democrats. You would like to learn about p through a poll. You decide to call n voters. Let Y denote the number of Democrats in the polled sample.

$$Y \sim \mathrm{Binomial}(n, p)$$

Your best guess for the proportion of Democrats is $\hat{p} = \frac{Y}{n}$.

The sampling distribution of $\hat{p}$ is $\mathrm{Binomial}(n, p)$ divided by n.

The sampling distribution of the estimator depends on:

- The sample size n.
- The unknown parameter (in this case p).

See: https://mlakolar.shinyapps.io/CLT_prop/

The Sampling Distribution for $\hat{p}$

Even though we know the sampling distribution of $\hat{p}$, we will use the Central Limit Theorem to make our calculations easier.

For large n, the sampling distribution is (approximately):

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

since

$$\begin{aligned} E[\hat{p}] &= \frac{1}{n}\sum_{i=1}^{n} E[X_i] = p \qquad \text{(The estimator is unbiased.)} \\ \mathrm{Var}[\hat{p}] &= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}[X_i] = \frac{p(1-p)}{n} \end{aligned}$$

This gives us a simple way of thinking about what kind of estimate we might see.

Example: Voting and the i.i.d. Bernoulli model (recap)

We **model** voters in the U.S. as $\mathrm{Bernoulli}(p)$. We want to learn about the unknown parameter p based on a sample of voters.

1. Assume the data we are about to collect are i.i.d. $\mathrm{Bernoulli}(p)$.
2. We plan to use the sample proportion of Democrats, $\hat{p}$, as an estimator of the unknown parameter p. Before we see the data, $\hat{p}$ is (approximately) normally distributed as $N\left(p, \frac{p(1-p)}{n}\right)$.
3. We collect the data and compute $\hat{p}$ on the sample we collected. The number we get is an estimate of p.

Note: This is where people often get confused by two things!

- A lot of people think the true value of the parameter p is "random." NO! It's just a number, like 0.5 or 0.265. We just do not know what it is.
- People think "How can $\hat{p}$ be random?! I can compute it from the data!!" **Before we see the data, $\hat{p}$ is a random variable because it depends on which sample we observe.**

Confidence intervals for p

Our estimate $\hat{p}$ gives a specific value or "guess" as to what the unknown p is.

We now know what kind of sample proportion we can expect to get given the parameters.
What we really want to know is what we think the parameter is given the data.

The sampling distribution will allows us to construct the confidence interval
which tells us how uncertain we are about p.

Using the sampling distribution of $\hat{p}$,

$$P\left(p - 1.96\sqrt{\frac{p(1-p)}{n}} \leq \hat{p} \leq p + 1.96\sqrt{\frac{p(1-p)}{n}}\right) \approx 0.95.$$

Therefore, we can construct an interval

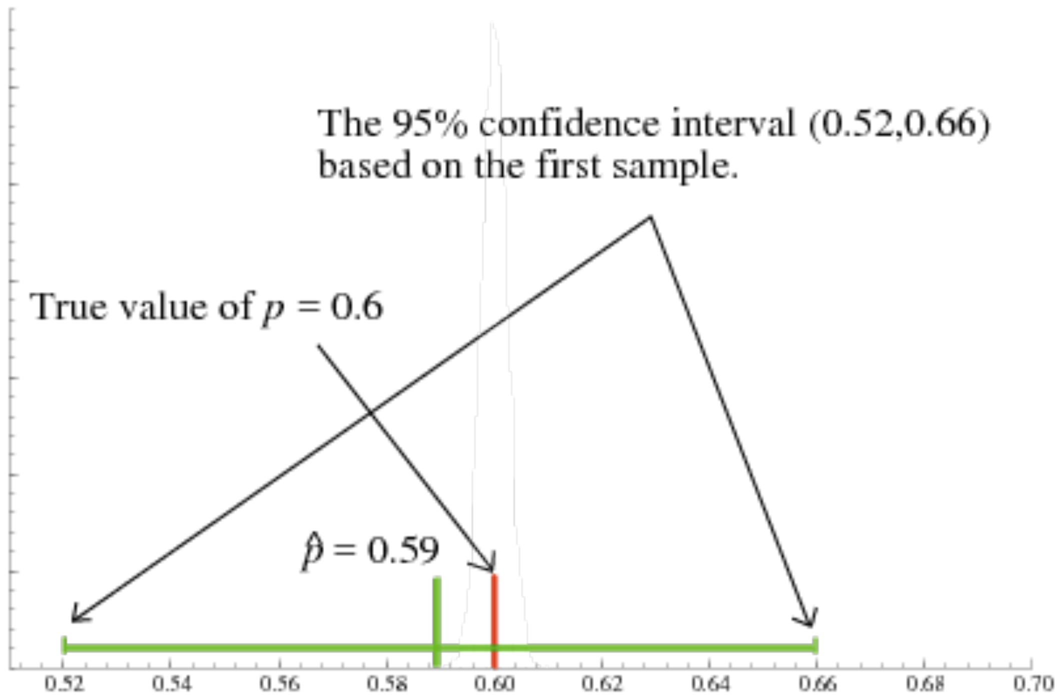
$$\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

that will include p in 95% of the samples we might observe.

95% Confidence intervals for p

Suppose the true value is $p = 0.6$ (remember, that you do not know this value)

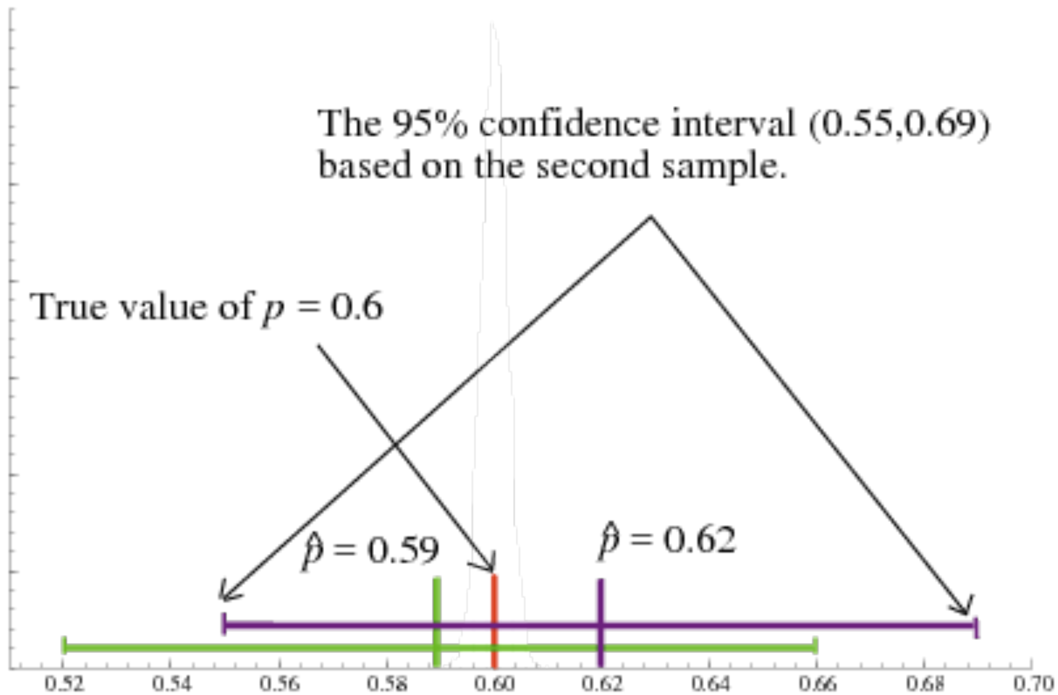
Consider a sample of size $n = 200$ from $\mathrm{Bernoulli}(0.6)$.



95% Confidence intervals for p

Suppose the true value is $p = 0.6$ (remember, that you do not know this value)

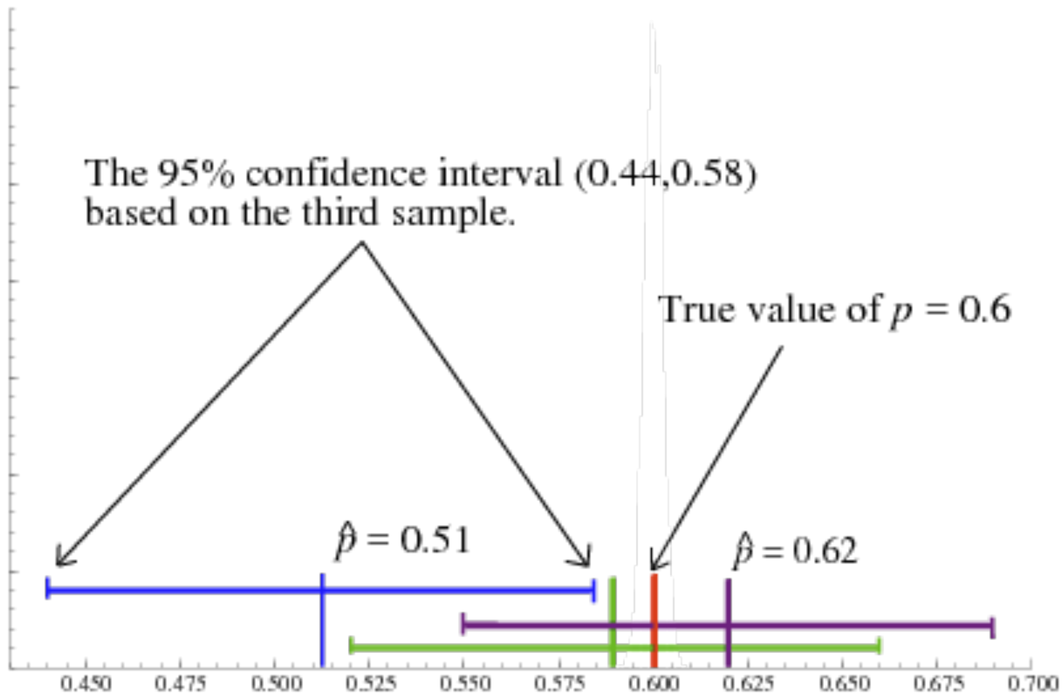
Consider a second sample of size $n = 200$ from $\mathrm{Bernoulli}(0.6)$.



95% Confidence intervals for p

Suppose the true value is $p = 0.6$ (remember, that you do not know this value)

Consider a third sample of size $n = 200$ from $\mathrm{Bernoulli}(0.6)$.



Confidence intervals for p (recap)

We have n i.i.d. observations $X_1, X_2, \ldots, X_n \sim \text{Bernoulli}(p)$

The (approximate) 95% confidence interval for p is given by

$$\hat{p} \pm 1.96 \times S.E.(\hat{p}) = \left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

The **standard error** of the sample proportion is **the standard deviation** of the sampling distribution:

$$S.E.(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- The standard error is the special name given to the standard deviation of the sampling distribution of an estimator.
- IMPORTANT: The formula for the standard error changes depending on the sampling distribution.

See: https://mlakolar.shinyapps.io/cinterval/

Example: Election 2009: New Jersey Governor (Rasmussen, 10/30/09)

"Republican Chris Christie continues to hold a three-point advantage… The latest Rasmussen Reports telephone survey in the state, conducted Thursday night, shows Christie with 46% $(\hat{p} = 0.46)$ of the vote and Corzine with 43%…"

"This statewide telephone survey of 1,000 Likely Voters in New Jersey was conducted by Rasmussen Reports October 29, 2009. The margin of sampling error for the survey is ± 3 percentage points with a 95% level of confidence."

$$2\sqrt{\frac{0.46 \cdot (1 - 0.46)}{1000}} = 0.0315$$

Here, p = the probability that a randomly chosen voter will vote for Christie

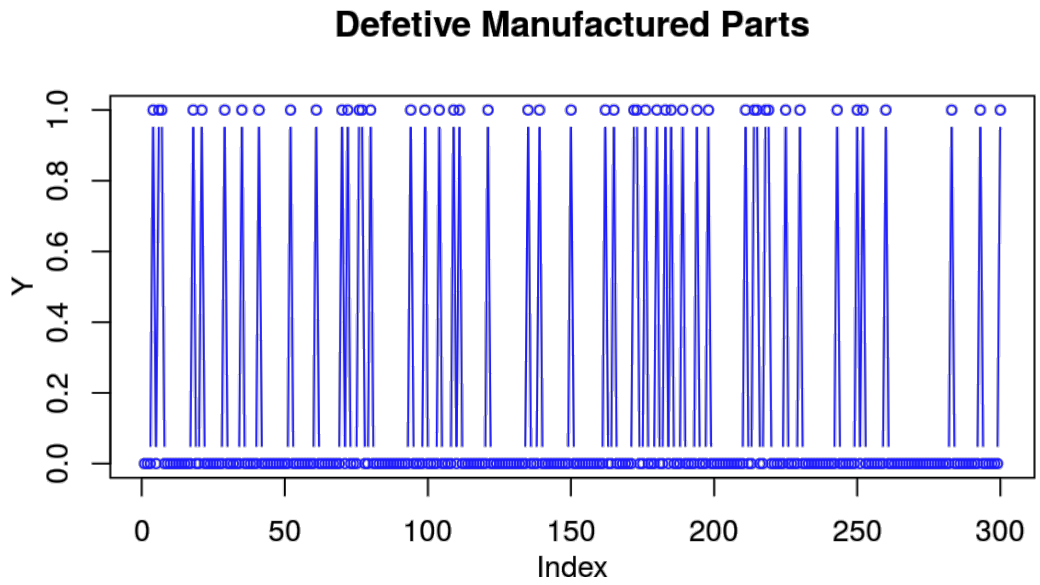
The 95% confidence interval for the true value of p is 0.46 ± 0.0315

This is equal to the "estimate ± 2 * standard error."

Example: The IID defect model

Recall our defects example from Lecture 3.

You collect data on 300 parts that were made successively.
You assume that $X_i \sim \mathrm{Bernoull}(p)$.
49 turn out to be defective.



Example: The IID defect model

We estimate p as

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{49}{300} = 0.163$$

Given our model for the process, how wrong could we be about p?

$$2 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 2 \cdot \sqrt{\frac{0.163(0.837)}{300}} = 0.021$$

The 95 confidence interval is $(0.163 \pm 2 \cdot 0.021) = (0.121, 0.205)$

Interpreting the size of a confidence interval

How much do I know about the parameter?

- If the confidence interval is small: I know a lot.
- If the confidence interval is large: I know little.

Example:

Suppose $\hat{p} = 0.2$ and $n = 100$. Then $S.E.(\hat{p}) = 0.04$ and 95% CI is: $(0.12, 0.28)$.

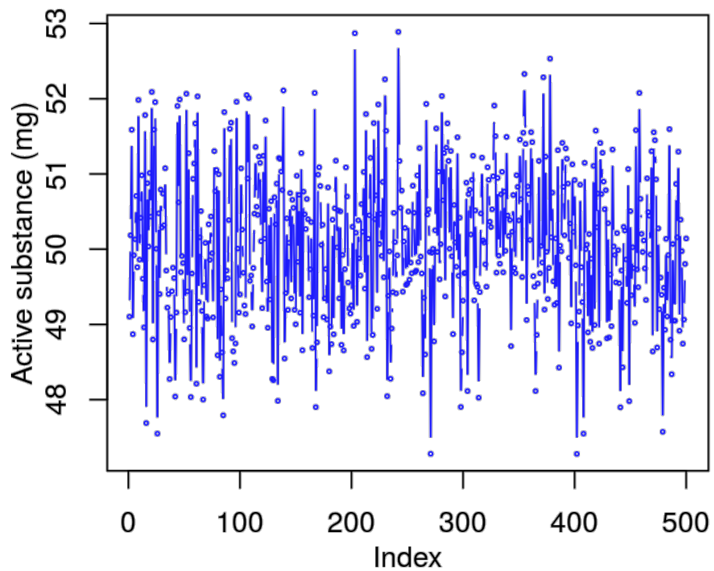
Instead, suppose $\hat{p} = 0.2$ and $n = 10000$. Then $S.E.(\hat{p}) = 0.004$ and 95% CI is: $(0.192, 0.208)$.

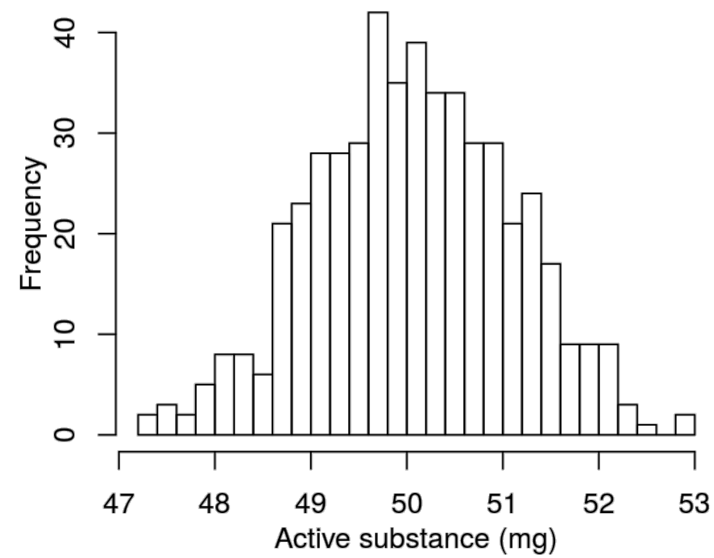
We increased n by a factor of 100 and decreased the standard error by 1/10.

The Sampling Distribution for $\bar{x}$

Consider a pharmaceutical company which makes pills. The company claims to put 50 milligrams of active substance in each pill. The amount that goes in each pill will not be the same for every pill so we can think of the amount a random outcome.

You are responsible for quality control and need to measure the amount of active substance actually put in each pill. You take a sample of $n = 500$ pills.





It looks like the data could be i.i.d. normal.
The sample mean turns out to be 50.06 – pretty close to 50!

We could model the process as: $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$ i.i.d.

Given our model, a parameter of interest is μ.
A reasonable estimator of μ is the sample mean, $\bar{x}$.

When we finally observe our data, we get a value for $\bar{x}$;
that is our estimate of the unknown parameter μ.

What are the properties of our estimator?
How bad might the sample average be as an estimate of μ?

See: https://mlakolar.shinyapps.io/CLT_mean/

We are about to take the next 500 observations.
We will estimate μ with the sample average.

What will we get for our sample average? We don't know.
Because it depends on the random outcomes we will get for the next 100 observations.

The reason there is variation in the sample average is because the average depends on the random outcomes we happen to get for this particular sample.

Our estimator is a linear combination of random variables

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Our estimator is a linear combination of random variables

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

The mean is $E[\bar{x}] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \mu$. The estimator is **unbiased**.

The variance is $\mathrm{Var}[\bar{x}] = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}[X_i] = \frac{\sigma^2}{n}$.

- The variance of the sample average depends on two things:
  - the variance of the population from which we are sampling (σ^2), and
  - the sample size n.

The sampling distribution is

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

See: https://mlakolar.shinyapps.io/CLT_mean/

Confidence intervals for μ

We now know what kind of sample average we can expect to get given the parameters. What we really want to know is what we think the parameter is given the data. We can use what we have learnt to develop a confidence interval for μ.

Using the sampling distribution of $\bar{x}$,

$$P\left(\mu - 1.96\frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95.$$

Therefore, we can construct an interval

$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

that will include μ in 95% of the samples we might observe.

We do not know σ, so replace it by the sample standard deviation to get

$$\left(\bar{x} - 1.96\frac{s_x}{\sqrt{n}}, \bar{x} + 1.96\frac{s_x}{\sqrt{n}}\right)$$

Confidence intervals for μ

Based on n i.i.d. observations $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$

The (approximate) 95% confidence interval for μ is given by

$$\bar{x} \pm 1.96 \times S.E.(\bar{x}) = \left(\bar{x} - 1.96 \frac{s_x}{\sqrt{n}}, \bar{x} + 1.96 \frac{s_x}{\sqrt{n}}\right)$$

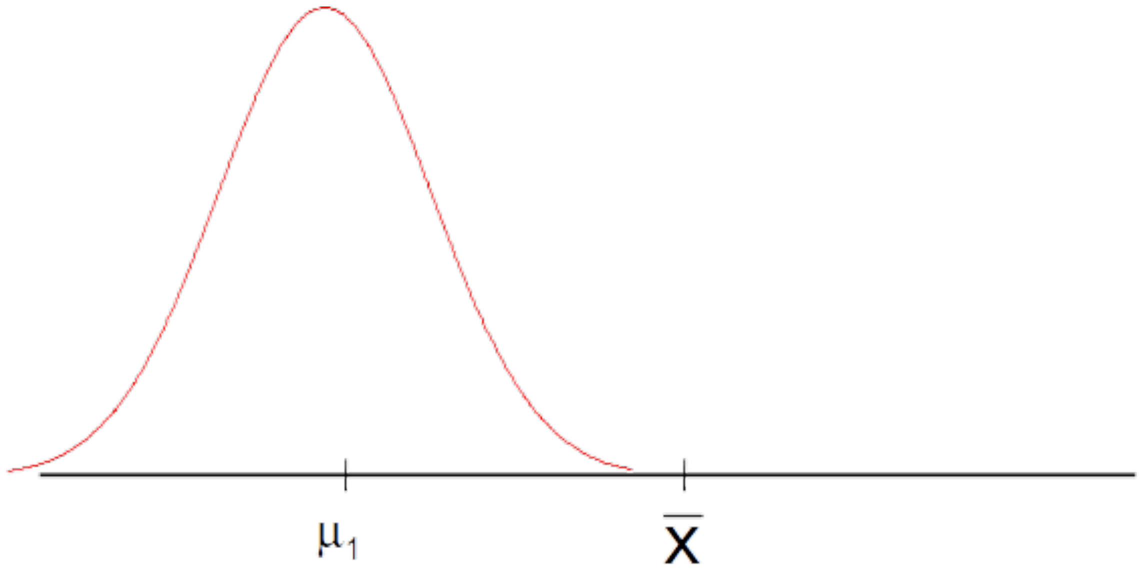
The **standard error** of the sample mean is **the standard deviation** of the sampling distribution:

$$S.E.(\bar{x}) = \frac{s_x}{\sqrt{n}}$$

See: https://mlakolar.shinyapps.io/cinterval/

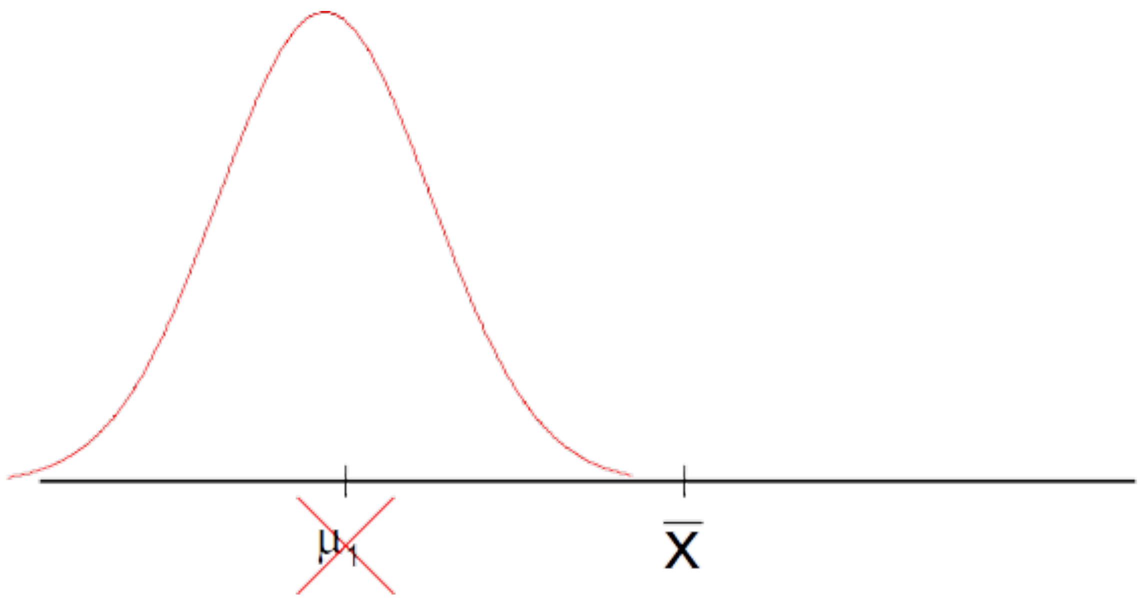
Our sample gives us a value for $\bar{x}$. We want to ask what values for μ are *reasonable*.

Our sample gives us a value for $\bar{x}$. We want to ask what values for μ are *reasonable*.



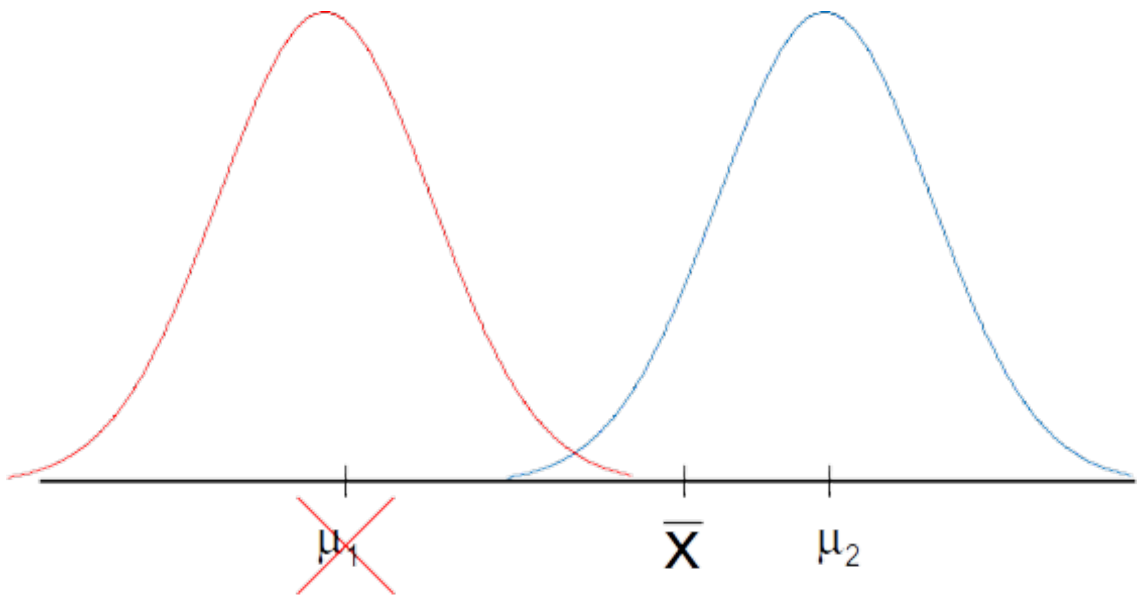
Consider a possible value μ_1. The red curve is the sampling distribution of $\bar{x}$ under this model. Is this "reasonable"?

Our sample gives us a value for $\bar{x}$. We want to ask what values for μ are *reasonable*.



Consider a possible value μ_1. The red curve is the sampling distribution of $\bar{x}$ under this model. Is this "reasonable"? NO. If that were the true value of μ, it's extremely unlikely we would see a $\bar{x}$ like the one we got in the data.

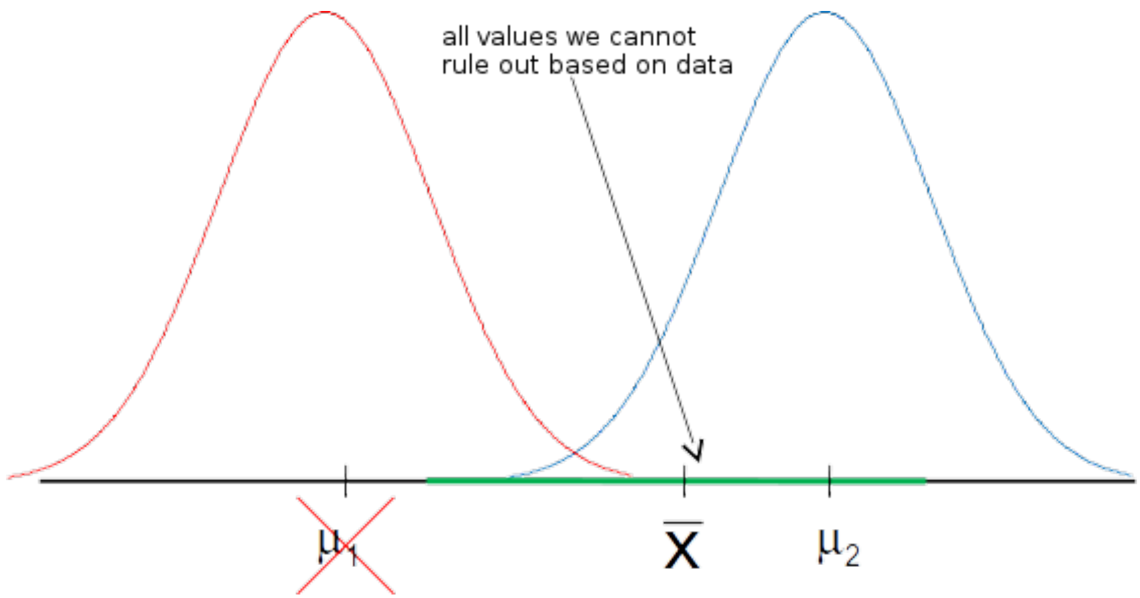
Our sample gives us a value for $\bar{x}$. We want to ask what values for μ are *reasonable*.



Consider a possible value μ_1. The red curve is the sampling distribution of $\bar{x}$ under this model. Is this "reasonable"? NO. If that were the true value of μ, it's extremely unlikely we would see a $\bar{x}$ like the one we got in the data.

Consider μ_2. If this were the right value of μ, it is perfectly possible we would see $\bar{x}$ like the one we saw in the data.

Our sample gives us a value for $\bar{x}$. We want to ask what values for μ are *reasonable*.

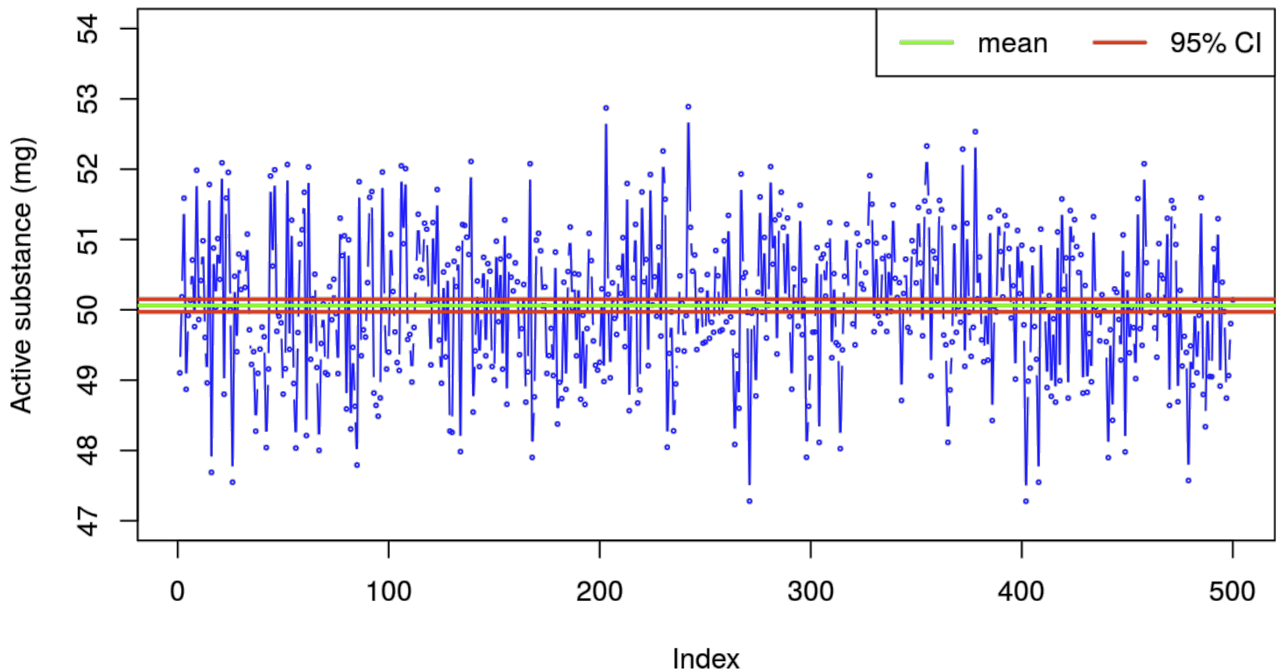


Consider a possible value μ_1. The red curve is the sampling distribution of $\bar{x}$ under this model. Is this "reasonable"? NO. If that were the true value of μ, it's extremely unlikely we would see a $\bar{x}$ like the one we got in the data.

Consider μ_2. If this were the right value of μ, it is perfectly possible we would see $\bar{x}$ like the one we saw in the data.

The values of μ that are consistent with our $\bar{x}$ are the values with $\pm 1.96 \cdot S.E.\bar{x}$.

Example: Audit of a pharma company



The sample mean is 50.06 and the sample standard deviation is 1.03.
The standard error is $S.E.(\bar{x}) = s_x/\sqrt{n} = 1.03/\sqrt{500} = 0.046$

The 95% confidence interval is $(50.06 \pm 1.96 \cdot 0.046)$

Confidence Interval for μ (recap)

Confidence intervals answer the basic questions,
what do you think the parameter is and how sure are you.

In particular, a 95% CI means that if we took 100 samples and created 100 different confidence intervals, we would expect 95 of them to contain the true (but unknown) value μ.

- Small interval: good, you know a lot
- Big interval: bad, you don't know much.

We do not really need to assume our data is normal. The Central Limit theorem tells us that when n is large,

$$\bar{x} \sim N\left(\mu, \sigma^2/n\right)$$

This justifies the confidence interval

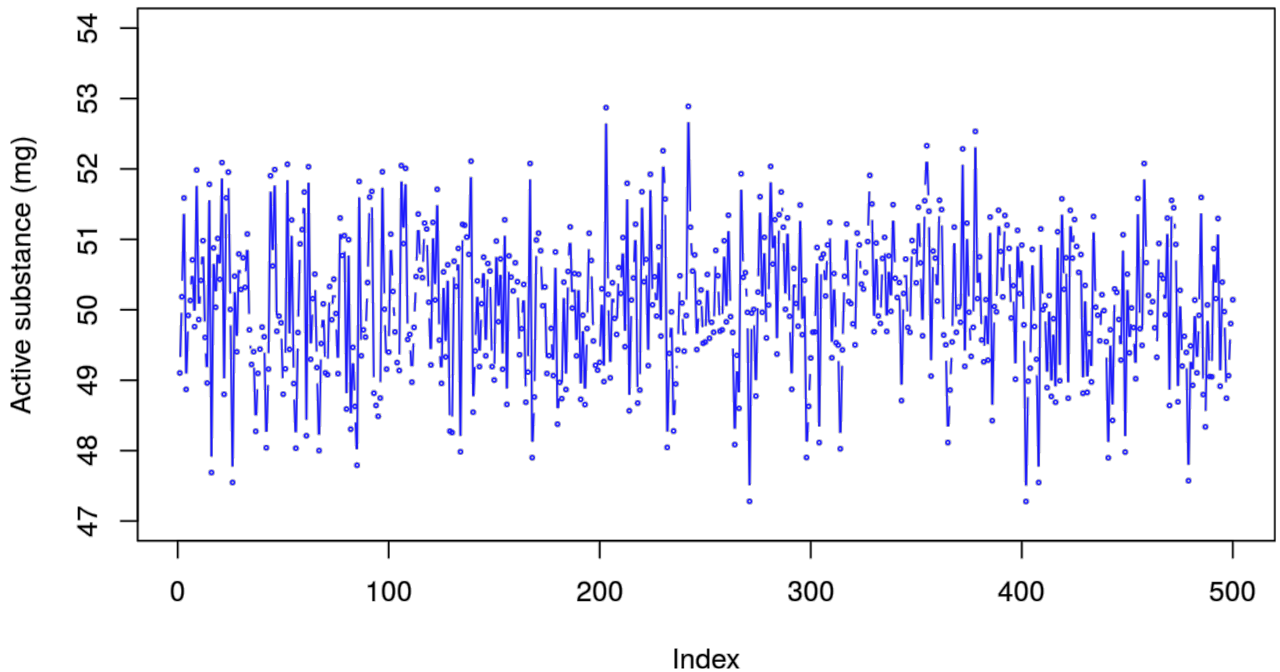
$$\bar{x} \pm 1.96 \cdot s_x/\sqrt{n}$$

even when the data is not normally distributed if the sample size n is large and the data are i.i.d.

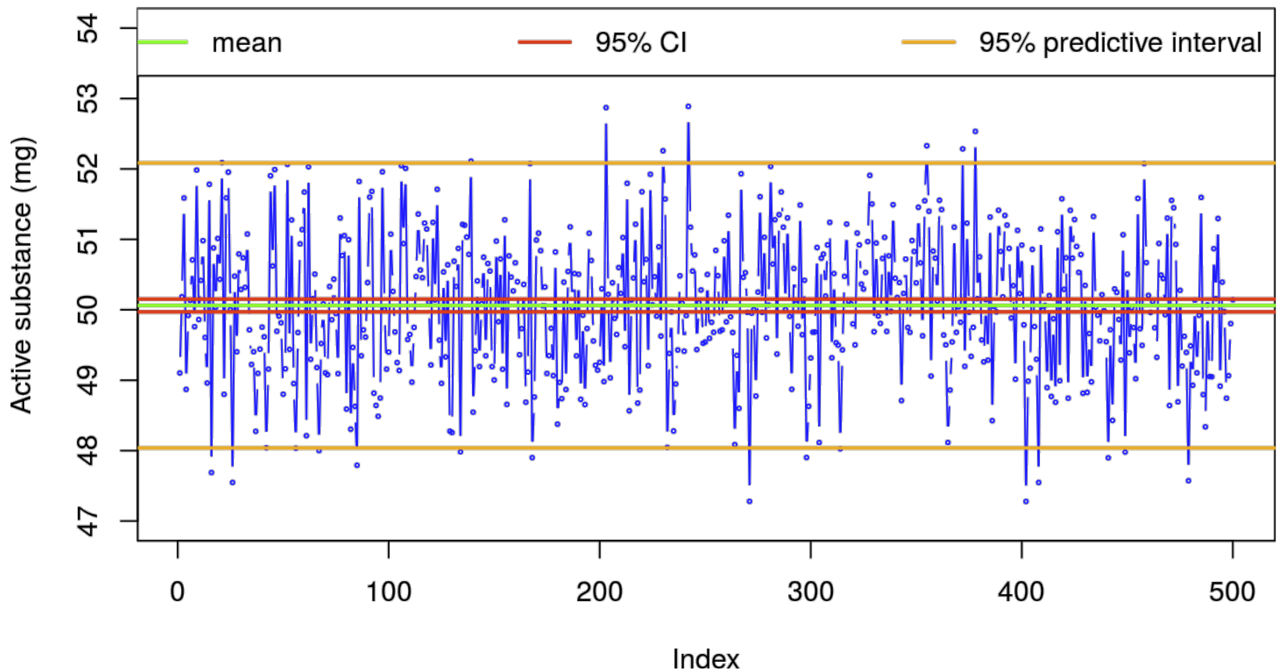
The Plug-in Predictive Interval for the IID Normal Model

Suppose a machine is about to create a new pill. How much active ingredient will go in the pill?

According to our model, $X_i \sim N(\mu, \sigma^2)$ i.i.d., the amount of cereal in the next box will be an independent draw from the same normal distribution the observed data came from.



The sample mean is 50.06 and the sample standard deviation is 1.03.
If the data is $N(\mu, \sigma^2)$, there is a 95% chance the next observation will be in the interval: $\mu \pm 1.96 \cdot \sigma$
Plugging in our estimates for the true values we obtain $(50.06 \pm 1.96 \cdot 1.03)$.



Predictive interval (orange) answers: "what will the next observation be?"

Confidence interval (red) answers: "what do we think μ is?"

IMPORTANT: Inference versus Prediction

In the confidence interval, we are using all the data to learn about an unknown parameter μ.

In the predictive interval, we are trying to make a prediction for a single random outcome.

The formulas for these two intervals look similar. But, they are doing VERY different things.

The confidence interval shrinks as n grows because with more data we gather more information about the unknown parameter μ.

The prediction interval does not shrink. Having more data does not make us any more certain about what happens for any single outcome.

Confidence Intervals other than 95%

The confidence intervals we have considered look like:

$$P(\text{estimate} - 2 \times S.E. < \text{par} < \text{estimate} + 2 \times S.E.) = 0.95$$

Here, "par" is the unknown parameter of interest, for example, p or μ.

There is nothing special about 95% confidence intervals (other than convention).

Confidence Intervals other than 95%

We can construct an $(1-\alpha)\%$ confidence interval for the unknown parameter as

$$\text{par} \pm z_{1-\alpha/2} \cdot S.E.(\text{par})$$

where z_α is a number such that

$$P(Z \leq z_\alpha) = \alpha$$

and $Z \sim N(0,1)$ is the standard normal random variable.

In R you can obtain z_α as

$$z_\alpha = \texttt{qnorm}(\alpha)$$

Where does the formula come from?

The main observation is that

$$\frac{\widehat{\text{par}} - \text{par}}{S.E.(\widehat{\text{par}})} \approx N(0,1)$$

Therefore

$$P\left(-z_{1-\alpha/2} < \frac{\widehat{\text{par}} - \text{par}}{S.E.(\widehat{\text{par}})} < z_{1-\alpha/2}\right) = 1 - \alpha$$

$$P\left(\widehat{\text{par}} - z_{1-\alpha/2} \cdot S.E.(\widehat{\text{par}}) < \text{par} < \widehat{\text{par}} + z_{1-\alpha/2} \cdot S.E.(\widehat{\text{par}})\right) = 1 - \alpha$$

See https://mlakolar.shinyapps.io/distributionCalculator/

Example: Confidence for difference of two means

We observe that 34 patients receiving treatment have a mean reduction in weight of 5 pounds with standard deviation of 4 pounds. The 60 patients in the placebo group show a mean reduction in weight of 3 pounds with a standard deviation of 6 pounds.

Example: Confidence for difference of two means

We observe that 34 patients receiving treatment have a mean reduction in weight of 5 pounds with standard deviation of 4 pounds. The 60 patients in the placebo group show a mean reduction in weight of 3 pounds with a standard deviation of 6 pounds.

Let $X_1, \ldots, X_n \sim N(\mu_x, \sigma_x^2)$ i.i.d. denote weight reduction for patients receiving the treatment.
Let $Y_1, \ldots, Y_m \sim N(\mu_y, \sigma_y^2)$ i.i.d. denote weight reduction for patients receiving placebo pill.

Using the central limit theorem $\bar{x} \approx N(\mu_x, \sigma_x^2/n) \text{ and } \bar{y} \approx N(\mu_y, \sigma_y^2/m)$

Example: Confidence for difference of two means

We observe that 34 patients receiving treatment have a mean reduction in weight of 5 pounds with standard deviation of 4 pounds. The 60 patients in the placebo group show a mean reduction in weight of 3 pounds with a standard deviation of 6 pounds.

Let $X_1, \ldots, X_n \sim N(\mu_x, \sigma_x^2)$ i.i.d. denote weight reduction for patients receiving the treatment.
Let $Y_1, \ldots, Y_m \sim N(\mu_y, \sigma_y^2)$ i.i.d. denote weight reduction for patients receiving placebo pill.

Using the central limit theorem $\bar{x} \approx N(\mu_x, \sigma_x^2/n) \text{ and } \bar{y} \approx N(\mu_y, \sigma_y^2/m)$

Therefore

$$\bar{x} - \bar{y} \approx N(\mu_x - \mu_y, \sigma_x^2/n + \sigma_y^2/m) \qquad \text{and} \qquad \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{\sqrt{s_x^2/n + s_y^2/m}} \approx N(0, 1)$$

Example: Confidence for difference of two means

We observe that 34 patients receiving treatment have a mean reduction in weight of 5 pounds with standard deviation of 4 pounds. The 60 patients in the placebo group show a mean reduction in weight of 3 pounds with a standard deviation of 6 pounds.

Let $X_1, \ldots, X_n \sim N(\mu_x, \sigma_x^2)$ i.i.d. denote weight reduction for patients receiving the treatment.
Let $Y_1, \ldots, Y_m \sim N(\mu_y, \sigma_y^2)$ i.i.d. denote weight reduction for patients receiving placebo pill.

Using the central limit theorem $\bar{x} \approx N(\mu_x, \sigma_x^2/n) \text{ and } \bar{y} \approx N(\mu_y, \sigma_y^2/m)$

Therefore

$$\bar{x} - \bar{y} \approx N(\mu_x - \mu_y, \sigma_x^2/n + \sigma_y^2/m) \qquad \text{and} \qquad \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{\sqrt{s_x^2/n + s_y^2/m}} \approx N(0,1),$$

which gives us the following $(1-\alpha)\%$ confidence interval for $\mu_x - \mu_y$

$$(\bar{x} - \bar{y}) \pm z_{1-\alpha/2} \cdot \sqrt{s_x^2/n + s_y^2/m}$$