Statistics for Engineers

Confidence intervals for a single sample

Confidence intervals for a single sample

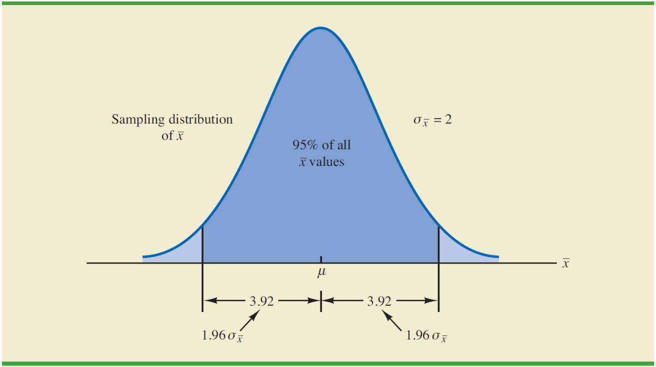
Contents

From point estimation to confidence interval estimation

- So far, we have consider the point estimation of an unknown population parameter which, assuming we had a SRS sample of n observations from X, would produce an educated guess about that unknown parameter.
- Point estimates however, do not take into account the variability of the estimation procedure due to, among other factors:
  - sample size - surely, larger samples should provide more accurate information about the population parameter
  - variability in the population - samples from populations with smaller variance should give more accurate estimates
  - whether other population parameters are known
  - etc

These drawbacks can be overcome by considering **confidence interval estimation**, that is, a method that gives a range of values (an interval) in which the parameter is likely to fall.

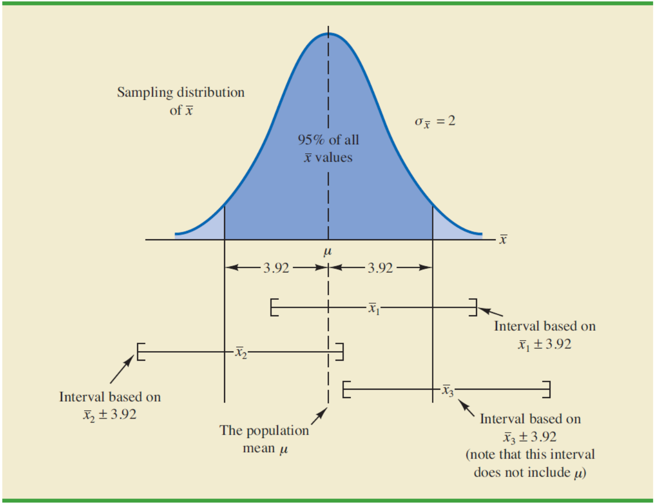
Confidence interval estimator and confidence interval



Typical levels of confidence

α	0.01	0.05	0.10
$100(1-\alpha)\%$	99%	95%	90%

Confidence interval estimator and confidence interval



Typical levels of confidence

α	0.01	0.05	0.10
$100(1-\alpha)\%$	99%	95%	90%

Finding confidence interval estimators: procedure

1. Use the upper $1-\alpha/2$ and lower $\alpha/2$ quantiles of that distribution and the definition of the confidence interval estimator to set up the equation

$$P(\overbrace{\alpha/2 \text{ quantile} < C(\underline{X}_n, \theta) < 1-\alpha/2 \text{ quantile}}^{\text{double inequality}}) = 1-\alpha$$

2. A $100(1-\alpha)\%$ confidence interval for θ is $(T_1(\underline{x}_n), T_2(\underline{x}_n))$

Confidence intervals formulae

Summary for one population

- Let $\underline{X}_n$ be a simple random sample from a population X with mean μ_X and variance σ_X^2

Parameter	Assumptions	Pivotal quantity	$(1-\alpha)$ Conf. Interval
Mean	Normal data Known variance	$\frac{\bar{X}-\mu_X}{\sigma_X/\sqrt{n}} \sim N(0,1)$	$\mu_X \in \left(\bar{x} - z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right)$
	Nonnormal data Large sample	$\frac{\bar{X}-\mu_X}{s_X/\sqrt{n}} \sim_{approx.} N(0,1)$	$\mu_X \in \left(\bar{x} - z_{1-\alpha/2}\frac{s_X}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{s_X}{\sqrt{n}}\right]$
	Bernoulli data Large sample	$\frac{\hat{p}_X - p_X}{\sqrt{\hat{p}_X(1-\hat{p}_X)/n}} \sim_{approx.} N(0,1)$	$p_X \in \left(\hat{p}_X \mp z_{1-\alpha/2}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}\right]$
	Normal data Unknown variance	$\frac{\bar{X}-\mu_X}{s_X/\sqrt{n}} \sim t_{n-1}$	$\mu_X \in \left(\bar{x} - t_{n-1,1-\alpha/2}\frac{s_X}{\sqrt{n}}, \bar{x} + t_{n-1,1-\alpha/2}\frac{s_X}{\sqrt{n}}\right)$
Variance	Normal data	$\frac{(n-1)s_X^2}{\sigma_X^2} \sim \chi^2_{n-1}$	$\sigma_X^2 \in \left(\frac{(n-1)s_X^2}{\chi^2_{n-1;1-\alpha/2}}, \frac{(n-1)s_X^2}{\chi^2_{n-1;\alpha/2}}\right)$
Standard dev.	Normal data	$\frac{(n-1)s_X^2}{\sigma_X^2} \sim \chi^2_{n-1}$	$\sigma_X \in \left(\sqrt{\frac{(n-1)s_X^2}{\chi^2_{n-1;1-\alpha/2}}}, \sqrt{\frac{(n-1)s_X^2}{\chi^2_{n-1;\alpha/2}}}\right)$

- **Target for this topic:** Students should be capable of constructing confidence intervals (right column) using the pivotal quantity column information.

Confidence intervals formulae

Confidence interval for the population mean, normal population with known variance

Parameter	Assumptions	Pivotal quantity	$(1-\alpha)$ Conf. Interval
Mean	**Normal data** **Known variance**	$\boldsymbol{\frac{\bar{X}-\mu_X}{\sigma_X/\sqrt{n}} \sim N(0,1)}$	$\boldsymbol{\mu_X \in \left(\bar{x} - z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right)}$
	Nonnormal data Large sample	$\frac{\bar{X}-\mu_X}{s_X/\sqrt{n}} \sim_{approx.} N(0,1)$	$\mu_X \in \left(\bar{x} - z_{1-\alpha/2}\frac{s_X}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{s_X}{\sqrt{n}}\right]$
	Bernoulli data Large sample	$\frac{\hat{p}_X - p_X}{\sqrt{\hat{p}_X(1-\hat{p}_X)/n}} \sim_{approx.} N(0,1)$	$p_X \in \left(\hat{p}_X \mp z_{1-\alpha/2}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}\right]$
	Normal data Unknown variance	$\frac{\bar{X}-\mu_X}{s_X/\sqrt{n}} \sim t_{n-1}$	$\mu_X \in \left(\bar{x} - t_{n-1,1-\alpha/2}\frac{s_X}{\sqrt{n}}, \bar{x} + t_{n-1,1-\alpha/2}\frac{s_X}{\sqrt{n}}\right)$
Variance	Normal data	$\frac{(n-1)s_X^2}{\sigma_X^2} \sim \chi^2_{n-1}$	$\sigma_X^2 \in \left(\frac{(n-1)s_x^2}{\chi^2_{n-1;1-\alpha/2}}, \frac{(n-1)s_x^2}{\chi^2_{n-1;\alpha/2}}\right)$
Standard dev.	Normal data	$\frac{(n-1)s_X^2}{\sigma_X^2} \sim \chi^2_{n-1}$	$\sigma_X \in \left(\sqrt{\frac{(n-1)s_x^2}{\chi^2_{n-1;1-\alpha/2}}}, \sqrt{\frac{(n-1)s_x^2}{\chi^2_{n-1;\alpha/2}}}\right)$

Confidence interval for the population mean, normal population with known variance

1. Let $\underline{X}_n$ be a SRS of size n from X. Under the assumptions:
   - X follows a normal distribution with parameters μ_X and σ_X^2
   - σ_X^2 is known (rather unrealistic)
2. The mean sample distribution function is as follows:

$$\boxed{\frac{\bar{X} - \mu_X}{\sigma_X/\sqrt{n}} \sim N(0,1)}$$

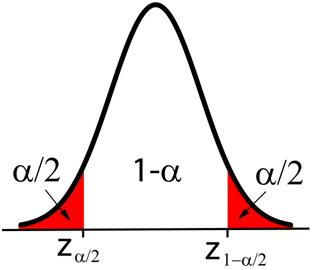
   - Note: the standard deviation of $\bar{X}$, $\sigma_X/\sqrt{n}$, (or any other stats) is called the **standard error**

Confidence interval for the population mean, normal population with known variance

3. Hence, if $z_{1-\alpha/2}$ and $z_{\alpha/2}$ are the $(1-\alpha/2)$ and $(\alpha/2)$ quantiles of the $N(0,1)$, we have

$$\boxed{P(z_{\alpha/2} < Z < z_{1-\alpha/2}) = 1-\alpha}$$

Standard normal density



Recall: If $Z \sim N(0,1)$ then $\mathrm{E}[Z] = 0$, $\mathrm{V}[Z] = 1$

4. Therefore $P(\overbrace{z_{\alpha/2}}^{-z_{1-\alpha/2}} < \overbrace{\frac{\bar{X}-\mu_X}{\sigma_X/\sqrt{n}}}^{Z} < z_{1-\alpha/2}) = 1-\alpha$

Confidence interval for the population mean, normal population with known variance

5. Solve the double inequality for μ_X:

$$-z_{1-\alpha/2} < \frac{\bar{X}-\mu_X}{\sigma_X/\sqrt{n}} < z_{1-\alpha/2}$$

$$-z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}} < \bar{X}-\mu_X < z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}$$

$$-z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}} - \bar{X} < -\mu_X < -\bar{X} + z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}$$

$$z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}} + \bar{X} > \mu_X > \bar{X} - z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}$$

   to obtain the confidence interval estimator

$$(\overbrace{\bar{X} - z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}}^{T_1(\underline{X}_n)}, \overbrace{\bar{X} + z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}}^{T_2(\underline{X}_n)})$$

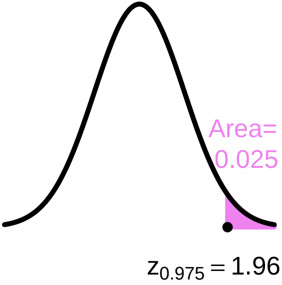
6. The confidence interval is:

$$\mathrm{CI}_{1-\alpha}(\mu_X) = \left(\bar{x} - z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right) = \left(\bar{x} \mp z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right)$$

Example: finding a confidence interval for μ_X

Example:8.2 (Newbold) A process produces bags of refined sugar. The weights of the contents of these bags are normally distributed with standard deviation 1.2 ounces. The contents of a random sample of twenty-five bags had mean weight 19.8 ounces. Find a 95% confidence interval for the true mean weight for all bags of sugar produced by the process.

Population:
X = **"weight of a sugar bag (in oz)"**
$X \sim N(\mu_X, \sigma_X^2 = 1.2^2)$
SRS: $n = 25$
Sample: $\bar{x} = 19.8$



Objective: $CI_{0.95}(\mu_X) = \left(\bar{x} \mp z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}}\right)$

$$\sigma_X = 1.2$$
$$n = 25\bar{x} = 19.8$$
$$1 - \alpha = 0.95 \Rightarrow \alpha/2 = 0.025$$
$$z_{1-\alpha/2} = z_{0.975} = 1.96$$
$$CI_{0.95}(\mu_X) = \left(19.8 \mp 1.96 \frac{1.2}{\sqrt{25}}\right)$$
$$= (19.8 \mp 0.47)$$
$$= (19.33, 20.27)$$

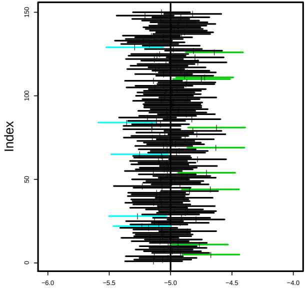
Interpretation: We can be 95% confident that μ_X is in $(19.33, 20.27)$

Frequency interpretation of the CI, conf. level effect

In this simulated example, 150 samples of the same size $n = 50$ were generated from $\boxed{X \sim N(\mu_X = -5, \sigma_X^2 = 1^2)}$ and 150 $CI_{1-\alpha}(\mu_X)$ were constructed with $\alpha = 0.1$ and $\alpha = 0.01$.

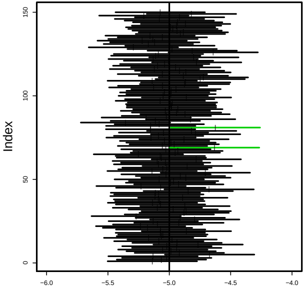
μ_X in approximately $150(0.9) = 135$ ints.
(but not in $150(0.1) = 15$)

$\boxed{(1-\alpha) = 0.9, n = 50}$



μ_X in approximately $150(0.99) = 148.5$ ints.
(but not in $150(0.01) = 1.5$)

$\boxed{(1-\alpha) = 0.99, n = 50}$



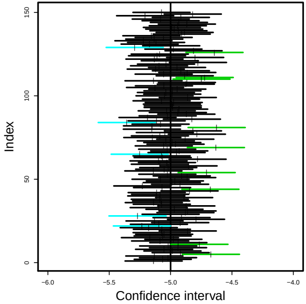
The **width** of the interval, $\boxed{w = \bar{x} + \frac{z_{1-\alpha/2}\sigma_X}{\sqrt{n}} - \left(\bar{x} - \frac{z_{1-\alpha/2}\sigma_X}{\sqrt{n}}\right) = 2\frac{z_{1-\alpha/2}\sigma_X}{\sqrt{n}}}$, increases with the increasing confidence level $(1-\alpha)$ (keeping everything else the same). Why?

Frequency interpretation of the CI, sample size effect

Here we collect 150 samples of size $n = 50$ and another 150 of size $n = 200$ from $X \sim N(\mu_X = -5, \sigma_X^2 = 1^2)$.

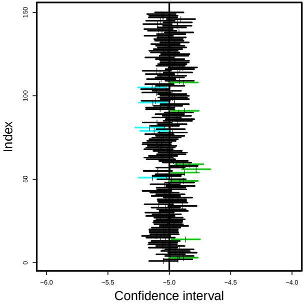
μ_X in approximately $150(0.9) = 135$ ints.
(but not in $150(0.1) = 15$)

$(1 - \alpha) = 0.9, n = 50$



μ_X in approximately $150(0.9) = 135$ ints.
(but not in $150(0.1) = 15$)

$(1 - \alpha) = 0.9, n = 200$



The width of the interval decreases with the increasing sample size (keeping everything else the same). Why?

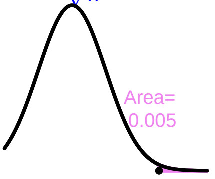
Question: What is the effect of σ on the width?

Example: estimating the sample size

Example: 8.14 (Newbold) The lengths of metal rods produced by an industrial process are normally distributed with standard deviation 1.8mm. Suppose that a production manager requires a 99% confidence interval extending no further than 0.5mm on each side of the sample mean. How large a sample is needed to achieve such an interval?

Population:
$X =$ **"length of a metal rod (in mm)"**
$X \sim N(\mu_X, \sigma_X^2 = 1.8^2)$
SRS: $n = ?$

$CI_{0.99}(\mu_X)$: $\overbrace{2\frac{z_{1-\alpha/2}\sigma_X}{\sqrt{n}}}^{\text{width}} \leq 2(0.5) = 1$



$z_{0.995} = 2.575$

Objective: n such that width ≤ 1

$$\begin{aligned} 2\frac{z_{1-\alpha/2}\sigma_X}{\sqrt{n}} &\leq 1 \\ 2z_{1-\alpha/2}\sigma_X &\leq \sqrt{n} \\ 85.93 = (2(2.575)(1.8))^2 &\leq n \end{aligned}$$

To satisfy the manager's requirement, a sample of at least 86 observations is needed.

One-sided confidence bounds for the population mean, normal population with known variance

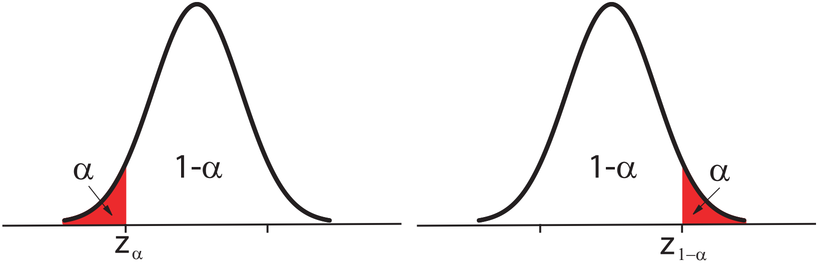
- The confidence intervals and resulting confidence bounds discussed thus far are *two-sided*.
- However, there are many applications in which only one bound is sought.
- One-sided confidence bounds are developed in the same fashion as two-sided intervals. For instace, from $P\left(\frac{\overline{X}-\mu_X}{\sigma_X\sqrt{n}} < z_{1-\alpha}\right) = 1-\alpha$ we can derive: $P\left(\mu_X < \overline{X} + z_{1-\alpha}\sigma_X/\sqrt{n}\right) = 1-\alpha$
- In general, If $\overline{X}$ is the mean of a random sample of size n from a population with variance σ_X^2, the one-sided $100(1-\alpha)$ % confidence bounds for μ_X are given by:

$$\text{upper one-sided bound:} \quad \mu_X \leq \overline{X} + z_{1-\alpha}\sigma_X/\sqrt{n}$$

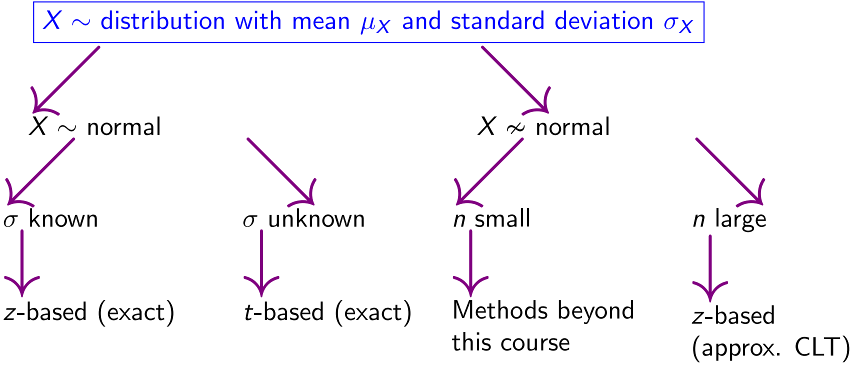
$$\text{lower one-sided bound:} \quad \mu_X \geq \overline{X} - z_{1-\alpha}\sigma_X/\sqrt{n}$$

- Similar one-sided bounds can be derived for all the CI that will be introduced in the following sections.

One-sided confidence bounds for the population mean, normal population with known variance



Confidence intervals for the population mean: when to use what?



Confidence intervals formulae

Confidence interval for the population mean in large samples

Parameter	Assumptions	Pivotal quantity	$(1-\alpha)$ Conf. Interval
Mean	Normal data Known variance	$\frac{\bar{X}-\mu_X}{\sigma_X/\sqrt{n}} \sim N(0,1)$	$\mu_X \in \left(\bar{x} - z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right)$
	Nonnormal data Large sample	$\boldsymbol{\frac{\bar{X}-\mu_X}{s_X/\sqrt{n}} \sim_{approx.} N(0,1)}$	$\boldsymbol{\mu_X \in \left(\bar{x} - z_{1-\alpha/2}\frac{s_X}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{s_X}{\sqrt{n}}\right]}$
	Bernoulli data Large sample	$\frac{\hat{p}_X - p_X}{\sqrt{\hat{p}_X(1-\hat{p}_X)/n}} \sim_{approx.} N(0,1)$	$p_X \in \left(\hat{p}_X \mp z_{1-\alpha/2}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}\right]$
	Normal data Unknown variance	$\frac{\bar{X}-\mu_X}{s_X/\sqrt{n}} \sim t_{n-1}$	$\mu_X \in \left(\bar{x} - t_{n-1,1-\alpha/2}\frac{s_X}{\sqrt{n}}, \bar{x} + t_{n-1,1-\alpha/2}\frac{s_X}{\sqrt{n}}\right)$
Variance	Normal data	$\frac{(n-1)s_X^2}{\sigma_X^2} \sim \chi^2_{n-1}$	$\sigma_X^2 \in \left(\frac{(n-1)s_X^2}{\chi^2_{n-1;1-\alpha/2}}, \frac{(n-1)s_X^2}{\chi^2_{n-1;\alpha/2}}\right)$
Standard dev.	Normal data	$\frac{(n-1)s_X^2}{\sigma_X^2} \sim \chi^2_{n-1}$	$\sigma_X \in \left(\sqrt{\frac{(n-1)s_X^2}{\chi^2_{n-1;1-\alpha/2}}}, \sqrt{\frac{(n-1)s_X^2}{\chi^2_{n-1;\alpha/2}}}\right)$

Confidence interval for the population mean in large samples

1. Let $\underline{X}_n$ be a SRS of size n from X. Under the assumptions:
   - X follows a nonnormal distribution with parameters μ_X and σ_X^2
   - the sample size n is large ($n \geq 30$)
2. The pivotal quantity for μ_X based on the **Central Limit Theorem** is

$$\boxed{\frac{\bar{X} - \mu_X}{s_X/\sqrt{n}} \sim_{\text{approx.}} N(0,1)}$$

3. The confidence interval is:

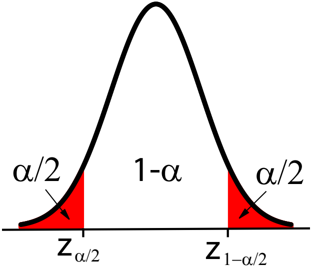
$$\text{CI}_{1-\alpha}(\mu_X) = (\bar{x} - z_{1-\alpha/2}\frac{s_x}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{s_x}{\sqrt{n}})$$

Confidence interval for the population mean in large samples

3. Hence, if $z_{1-\alpha/2}$ and $z_{\alpha/2}$ are the $(1-\alpha/2)$ and $(\alpha/2)$ quantiles of the $N(0,1)$, we have

$$\boxed{P(z_{\alpha/2} < Z < z_{1-\alpha/2}) = 1 - \alpha}$$

Standard normal density 



4. Therefore $P(\overbrace{z_{\alpha/2}}^{-z_{1-\alpha/2}} < \overbrace{\frac{\bar{X} - \mu_X}{s_X/\sqrt{n}}}^{Z} < z_{1-\alpha/2}) = 1 - \alpha$

Confidence intervals formulae

Confidence interval for the population proportion in large samples

Parameter	Assumptions	Pivotal quantity	$(1-\alpha)$ Conf. Interval
Mean	Normal data Known variance	$\frac{\bar{X}-\mu_X}{\sigma_X/\sqrt{n}} \sim N(0,1)$	$\mu_X \in \left(\bar{x} - z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right)$
	Nonnormal data Large sample	$\frac{\bar{X}-\mu_X}{s_X/\sqrt{n}} \sim_{approx.} N(0,1)$	$\mu_X \in \left(\bar{x} - z_{1-\alpha/2}\frac{s_X}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{s_X}{\sqrt{n}}\right]$
	Bernoulli data Large sample	$\boldsymbol{\frac{\hat{p}_X-p_X}{\sqrt{\hat{p}_X(1-\hat{p}_X)/n}} \sim_{approx.} N(0,1)}$	$\boldsymbol{p_X \in \left(\hat{p}_x \mp z_{1-\alpha/2}\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n}}\right]}$
	Normal data Unknown variance	$\frac{\bar{X}-\mu_X}{s_X/\sqrt{n}} \sim t_{n-1}$	$\mu_X \in \left(\bar{x} - t_{n-1,1-\alpha/2}\frac{s_X}{\sqrt{n}}, \bar{x} + t_{n-1,1-\alpha/2}\frac{s_X}{\sqrt{n}}\right)$
Variance	Normal data	$\frac{(n-1)s_X^2}{\sigma_X^2} \sim \chi^2_{n-1}$	$\sigma_X^2 \in \left(\frac{(n-1)s_x^2}{\chi^2_{n-1;1-\alpha/2}}, \frac{(n-1)s_x^2}{\chi^2_{n-1;\alpha/2}}\right)$
Standard dev.	Normal data	$\frac{(n-1)s_X^2}{\sigma_X^2} \sim \chi^2_{n-1}$	$\sigma_X \in \left(\sqrt{\frac{(n-1)s_x^2}{\chi^2_{n-1;1-\alpha/2}}}, \sqrt{\frac{(n-1)s_x^2}{\chi^2_{n-1;\alpha/2}}}\right)$

Confidence interval for the population proportion in large samples

Application of CIs for the population mean in large samples

Let $\underline{X}_n$, $n \geq 30$ be a SRS from a Bernoulli distr. with parameter p_X ($\mu_X = \mathrm{E}[X] = p_X$ and $\sigma_X = \sqrt{p_X(1-p_X)}$). The sample proportion $\hat{p}_X$ is a special case of the sample mean of zero-one observations, $\hat{p}_X = \bar{X}$.

Thus, from the CLT

$$\underbrace{\frac{\hat{p}_X - p_X}{\sqrt{p_X(1-p_X)/n}}}_{\sigma_X/\sqrt{n}} \sim^{\text{approx.}} N(0,1)$$

This result remains true if we use an estimate for the population standard deviation

$$\underbrace{\frac{\hat{p}_X - p_X}{\sqrt{\hat{p}_X(1-\hat{p}_X)}/\sqrt{n}}}_{\hat{\sigma}_X/\sqrt{n}} \sim^{\text{approx.}} N(0,1)$$

Thus, in large samples, the confidence interval for p_X is:

$$\mathrm{CI}_{1-\alpha}(p_X) = \left(\hat{p}_x - z_{1-\alpha/2}\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n}}, \hat{p}_x + z_{1-\alpha/2}\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n}}\right)$$

One-sided confidence bounds for the population proportion in large samples

- Let $\underline{X}_n$, $n \geq 30$ be a SRS from a Bernoulli distribution with parameter p_X ($n \geq 30$), the one-sided $100(1-\alpha)$ % confidence bounds for p_X are given by:

$$\text{upper one-sided bound:} \quad p_X \leq \hat{p}_X + z_{1-\alpha}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}$$

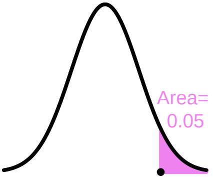
$$\text{lower one-sided bound:} \quad p_X \geq \hat{p}_X - z_{1-\alpha}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}$$

Example: finding a confidence interval for p_X

Example: 8.6 (Newbold) A random sample of 344 industrial buyers were asked: "What is your firm's policy for purchasing personnel to follow on accepting gifts from vendors?". For 83 of these buyers, the policy of the firm was for the buyer to make his/her own decision. Find a 90% confidence interval for the population proportion of all buyers who are allowed to make their own decisions.

Population:
$X = 1$ **if a buyer makes their own decision and 0 otherwise**
$X \sim Bernoulli(p_X)$
SRS: $n = 344$ large

Sample: $\hat{p}_x = \frac{83}{344} = 0.241$



Objective: $CI_{0.9}(p_X) = \left(\hat{p}_X \mp z_{1-\alpha/2}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}\right)$

$$\hat{p}_X = 0.241 \qquad n = 344$$
$$1 - \alpha = 0.9 \Rightarrow \alpha/2 = 0.05$$
$$z_{1-\alpha/2} = z_{0.95} = 1.645$$
$$CI_{0.9}(p_X) = \left(0.241 \mp 1.645\sqrt{\frac{0.241(1-0.241)}{344}}\right) = (0.241 \mp 0.038) = (0.203, 0.279)$$

Interpretation: We can be 90% confident that the proportion of buyers who make their own decision, p_X, falls in $(0.203, 0.279)$

Confidence intervals formulae

Confidence interval for the population mean, normal population with unknown variance

Parameter	Assumptions	Pivotal quantity	$(1-\alpha)$ Conf. Interval
Mean	Normal data Known variance	$\frac{\bar{X}-\mu_X}{\sigma_X/\sqrt{n}} \sim N(0,1)$	$\mu_X \in \left(\bar{x} - z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right)$
	Nonnormal data Large sample	$\frac{\bar{X}-\mu_X}{s_X/\sqrt{n}} \sim_{approx.} N(0,1)$	$\mu_X \in \left(\bar{x} - z_{1-\alpha/2}\frac{s_X}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{s_X}{\sqrt{n}}\right]$
	Bernoulli data Large sample	$\frac{\hat{p}_X - p_X}{\sqrt{\hat{p}_X(1-\hat{p}_X)/n}} \sim_{approx.} N(0,1)$	$p_X \in \left(\hat{p}_X \mp z_{1-\alpha/2}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}\right]$
	Normal data Unknown variance	$\boldsymbol{\frac{\bar{X}-\mu_X}{s_X/\sqrt{n}} \sim t_{n-1}}$	$\boldsymbol{\mu_X \in \left(\bar{x} - t_{n-1,1-\alpha/2}\frac{s_X}{\sqrt{n}}, \bar{x} + t_{n-1,1-\alpha/2}\frac{s_X}{\sqrt{n}}\right)}$
Variance	Normal data	$\frac{(n-1)s_X^2}{\sigma_X^2} \sim \chi^2_{n-1}$	$\sigma_X^2 \in \left(\frac{(n-1)s_X^2}{\chi^2_{n-1;1-\alpha/2}}, \frac{(n-1)s_X^2}{\chi^2_{n-1;\alpha/2}}\right)$
Standard dev.	Normal data	$\frac{(n-1)s_X^2}{\sigma_X^2} \sim \chi^2_{n-1}$	$\sigma_X \in \left(\sqrt{\frac{(n-1)s_X^2}{\chi^2_{n-1;1-\alpha/2}}}, \sqrt{\frac{(n-1)s_X^2}{\chi^2_{n-1;\alpha/2}}}\right)$

Confidence interval for the population mean, normal population with unknown variance

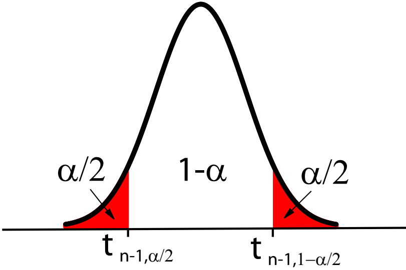
1. Let $\underline{X}_n$ be a SRS of size n from X. Under the assumptions:
   - X follows a normal distribution with parameters μ_X and σ_X^2
   - σ_X^2 is unknown (quite realistic)
2. The pivotal quantity for μ_X is

$$\boxed{\frac{\bar{X} - \mu_X}{s_X/\sqrt{n}} \sim t_{n-1}}$$

Confidence interval for the population mean, normal population with unknown variance

3. Hence, if $t_{n-1;1-\alpha/2}$ and $t_{n-1;\alpha/2}$ are the $(1-\alpha/2)$ and $(\alpha/2)$ quantiles of the t distribution with $n-1$ degrees of freedom (df), we have

$$P(t_{n-1;\alpha/2} < \overbrace{T}^{\sim\, t_{n-1}} < t_{n-1;1-\alpha/2}) = 1-\alpha$$



t (Student) density

Recall: if $T \sim t_n$, $\mathrm{E}[T] = 0$, $\mathrm{V}[T] = \frac{n}{n-2}$

4. Therefore $P(\overbrace{t_{n-1;\alpha/2}}^{-t_{n-1;1-\alpha/2}} < \overbrace{\frac{\bar{X}-\mu_X}{s_X/\sqrt{n}}}^{T\,\sim\, t_{n-1}} < t_{n-1;1-\alpha/2}) = 1-\alpha$

Confidence interval for the population mean, normal population with unknown variance

5. Solve the double inequality for μ_X:

$$-t_{n-1;1-\alpha/2} \quad < \frac{\bar{X}-\mu_X}{s_X/\sqrt{n}} < \quad t_{n-1;1-\alpha/2}$$

   to obtain the confidence interval estimator

$$(\overbrace{\bar{X} - t_{n-1;1-\alpha/2}\frac{s_X}{\sqrt{n}}}^{T_1(\underline{X}_n)}, \overbrace{\bar{X} + t_{n-1;1-\alpha/2}\frac{s_X}{\sqrt{n}}}^{T_2(\underline{X}_n)})$$

6. The confidence interval is:

$$\mathrm{CI}_{1-\alpha}(\mu_X) = (\bar{x} - t_{n-1;1-\alpha/2}\frac{s_x}{\sqrt{n}}, \bar{x} + t_{n-1;1-\alpha/2}\frac{s_x}{\sqrt{n}})$$

One-sided confidence bounds for the population mean, normal population with unknown variance

- If $\overline{X}$ is the mean of a random sample of size n from a normal population with unknow variance σ_X^2, the one-sided $100(1-\alpha)$ % confidence bounds for μ_X are given by:

$$\text{upper one-sided bound:} \quad \mu_X \leq \overline{X} + t_{n-1;1-\alpha} s_X/\sqrt{n}$$
$$\text{lower one-sided bound:} \quad \mu_X \geq \overline{X} - t_{n-1;1-\alpha} s_X/\sqrt{n}$$

Example: finding a confidence interval for μ_X

Example: 8.4 (Newbold) A random sample of six cars from a particular model year had the following fuel consumption figures, in mpg: 18.6, 18.4, 19.2, 20.8, 19.4, 20.5. Find a 90% confidence interval for the population mean fuel consumption, assuming that the population distribution is normal.

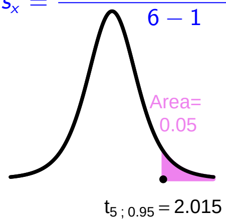
Population:
X = **"mpg of a car from the model year"** $X \sim N(\mu_X, \sigma_X^2)$

σ_X^2 **unknown**

SRS: $n = 6$ small

Sample: $\bar{x} = \frac{116.9}{6} = 19.4833$

$$s_x^2 = \frac{2282.41 - 6(19.4833)^2}{6-1} = 0.96$$



Objective: $CI_{0.9}(\mu_X) = \left(\bar{x} \mp t_{n-1;1-\alpha/2} \frac{s_x}{\sqrt{n}}\right)$

$$\begin{aligned}
s_x &= \sqrt{0.96} = 0.98 \\
n = 6 & \qquad \bar{x} = 19.48 \\
1 - \alpha = 0.9 &\Rightarrow \alpha/2 = 0.05 \\
t_{n-1;1-\alpha/2} &= t_{5;0.95} = 2.015 \\
CI_{0.9}(\mu_X) &= \left(19.48 \mp 2.105 \frac{0.98}{\sqrt{6}}\right) \\
&= (19.48 \mp 0.81) \\
&= (18.67, 20.29)
\end{aligned}$$

Interpretation: We can be 90% confident that the population mean fuel consumption for these cars, μ_X, is between 18.67 and 20.29

Confidence intervals formulae

Confidence interval for the population variance, normal population

Parameter	Assumptions	Pivotal quantity	$(1-\alpha)$ Conf. Interval
Mean	Normal data Known variance	$\frac{\bar{X}-\mu_X}{\sigma_X/\sqrt{n}} \sim N(0,1)$	$\mu_X \in \left(\bar{x} - z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right)$
	Nonnormal data Large sample	$\frac{\bar{X}-\mu_X}{s_X/\sqrt{n}} \sim_{approx.} N(0,1)$	$\mu_X \in \left(\bar{x} - z_{1-\alpha/2}\frac{s_X}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{s_X}{\sqrt{n}}\right]$
	Bernoulli data Large sample	$\frac{\hat{p}_X - p_X}{\sqrt{\hat{p}_X(1-\hat{p}_X)/n}} \sim_{approx.} N(0,1)$	$p_X \in \left(\hat{p}_X \mp z_{1-\alpha/2}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}\right]$
	Normal data Unknown variance	$\frac{\bar{X}-\mu_X}{s_X/\sqrt{n}} \sim t_{n-1}$	$\mu_X \in \left(\bar{x} - t_{n-1,1-\alpha/2}\frac{s_X}{\sqrt{n}}, \bar{x} + t_{n-1,1-\alpha/2}\frac{s_X}{\sqrt{n}}\right)$
Variance	**Normal data**	$\boldsymbol{\frac{(n-1)s_X^2}{\sigma_X^2} \sim \chi_{n-1}^2}$	$\boldsymbol{\sigma_X^2 \in \left(\frac{(n-1)s_X^2}{\chi_{n-1;1-\alpha/2}^2}, \frac{(n-1)s_X^2}{\chi_{n-1;\alpha/2}^2}\right)}$
Standard dev.	**Normal data**	$\boldsymbol{\frac{(n-1)s_X^2}{\sigma_X^2} \sim \chi_{n-1}^2}$	$\boldsymbol{\sigma_X \in \left(\sqrt{\frac{(n-1)s_X^2}{\chi_{n-1;1-\alpha/2}^2}}, \sqrt{\frac{(n-1)s_X^2}{\chi_{n-1;\alpha/2}^2}}\right)}$

Confidence interval for the population variance, normal population

1. Let $\underline{X}_n$ be a SRS of size n from X. Under the assumptions:
   - X follows a normal distribution with parameter σ_X^2
2. The pivotal quantity for σ_X^2 is

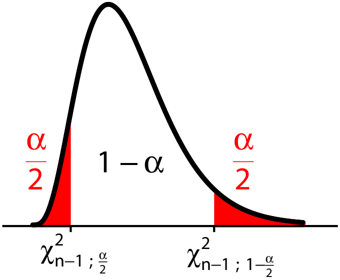
$$\boxed{\frac{(n-1)s_X^2}{\sigma_X^2} \sim \chi^2_{n-1}}$$

Confidence interval for the population variance, normal population

3. Hence, if $\chi^2_{n-1;1-\alpha/2}$ and $\chi^2_{n-1;\alpha/2}$ are the $(1-\alpha/2)$ and $(\alpha/2)$ quantiles of the chi-square distribution with $n-1$ degrees of freedom, we have

$$\boxed{P(\chi^2_{n-1;\alpha/2} < \chi^2_{n-1} < \chi^2_{n-1;1-\alpha/2}) = 1-\alpha}$$

Chi-square density



Recall: $\mathrm{E}[\chi^2_n] = n$, $\mathrm{V}[\chi^2_n] = 2n$

4. Therefore $P(\chi^2_{n-1;\alpha/2} < \overbrace{\frac{(n-1)s^2_X}{\sigma^2_X}}^{\chi^2_{n-1}} < \chi^2_{n-1;1-\alpha/2}) = 1-\alpha$

Confidence interval for the population variance, normal population

5. Solve the double inequality for σ_X^2:

$$\begin{aligned}
\chi^2_{n-1;\alpha/2} &< \frac{(n-1)s_X^2}{\sigma_X^2} < \chi^2_{n-1;1-\alpha/2} \\
\frac{1}{\chi^2_{n-1;\alpha/2}} &> \frac{\sigma_X^2}{(n-1)s_X^2} > \frac{1}{\chi^2_{n-1;1-\alpha/2}} \\
\frac{(n-1)s_X^2}{\chi^2_{n-1;\alpha/2}} &> \sigma_X^2 > \frac{(n-1)s_X^2}{\chi^2_{n-1;1-\alpha/2}}
\end{aligned}$$

   to obtain the confidence interval estimator

$$\left(\frac{(n-1)s_X^2}{\chi^2_{n-1;1-\alpha/2}}, \frac{(n-1)s_X^2}{\chi^2_{n-1;\alpha/2}} \right)$$

6. The confidence interval is:

$$\mathrm{CI}_\alpha(\sigma_X^2) = \left(\frac{(n-1)s_x^2}{\chi^2_{n-1;1-\alpha/2}}, \frac{(n-1)s_x^2}{\chi^2_{n-1;\alpha/2}} \right)$$

One-sided confidence bounds for the population variance, normal population

- Let $\underline{X}_n$ be a SRS of size n from X. Under the assumptions:
  - X follows a normal distribution with parameter σ_X^2

  the one-sided $100(1-\alpha)$ % confidence bounds for σ_X^2 are given by:

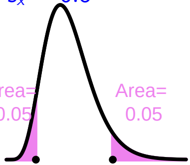
$$\text{upper one-sided bound:} \quad \sigma_X^2 \leq \frac{(n-1)s_X^2}{\chi^2_{n-1;\alpha}}$$

$$\text{lower one-sided bound:} \quad \sigma_X^2 \geq \frac{(n-1)s_X^2}{\chi^2_{n-1;1-\alpha}}$$

Example: finding a confidence interval for σ_X^2 and σ_X

Example: 8.8 (Newbold) A SRS of 15 pills for headache relief showed a quasi standard deviation of 0.8% in the concentration of the active ingredient, which follows a normal distribution. Find a 90% confidence interval for the population variance. How would you obtain a CI for the population standard deviation?

Population:
X = "concentration of an active ingredient in a pill (in %)" $X \sim N(\mu_X, \sigma_X^2)$
SRS: $n = 15$
Sample: $s_x = 0.8$



Objective: $CI_{0.9}(\sigma_X^2) = \left(\frac{(n-1)s_x^2}{\chi^2_{n-1;1-\alpha/2}}, \frac{(n-1)s_x^2}{\chi^2_{n-1;\alpha/2}} \right)$

$s_x^2 = 0.8^2 = 0.64 \qquad n = 15$

$$1 - \alpha = 0.9 \Rightarrow \alpha/2 = 0.05$$

$$\chi^2_{n-1;1-\alpha/2} = \chi^2_{14;0.95} = 23.68$$

$$\chi^2_{n-1;\alpha/2} = \chi^2_{14;0.05} = 6.57$$

$$CI_{0.9}(\sigma_X^2) = \left(\frac{14(0.64)}{23.68}, \frac{14(0.64)}{6.57} \right) = (0.378, 1.364) \Rightarrow$$

$$CI_{0.9}(\sigma_X) = (\sqrt{0.378}, \sqrt{1.364}) = (0.61, 1.17)$$

To obtain $CI(\sigma_X)$ we apply $\surd$ to the end-points of $CI(\sigma_X^2)$