Statistics & Probability

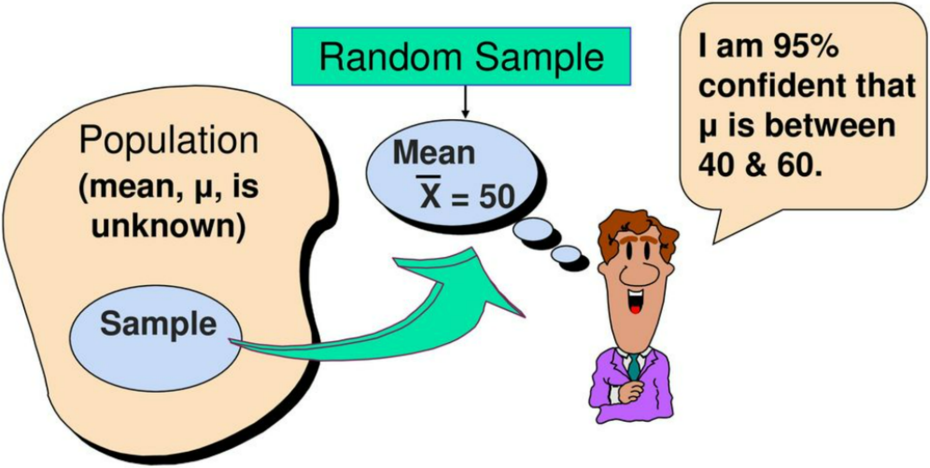
Chapter 8: STATISTICAL INTERVALS FOR A SINGLE SAMPLE

FPT University

Department of Mathematics

Quy Nhon, 2023

Table of Contents


Random Sample
Population
(mean, μ, is
unknown)
Sample
Mean
X̄ = 50
I am 95%
confident that
μ is between
40 & 60.

Confidence Interval For The Population Mean μ

Confidence Interval

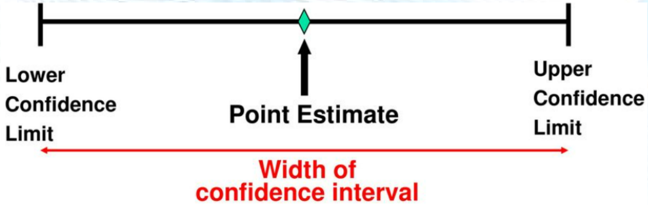
A **confidence interval** (CI) estimate for μ with $100(1-\alpha)\%$ confidence level is an interval of the form $l \leq \mu \leq u$ with l, u are computed from the sample data such that

$$P(l \leq \mu \leq u) = 1 - \alpha$$

where l, u are called lower-confidence limit and upper-confidence limit, respectively; and $1-\alpha$ is called the confidence coefficient.

Remarks

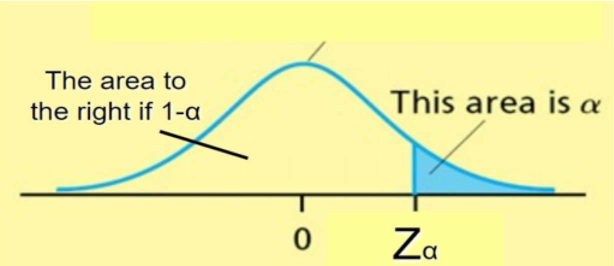
- A **confidence interval** provides additional information about variability.
- Stated in terms of level of confidence: can never be 100% confident.



Critical Value

Critical value z_α (or a **percentage point**) is the value such that

$$P(Z > z_\alpha) = \alpha.$$



Note

$$z_\alpha = \text{NORM.S.INV}(1-\alpha)$$

Example: $z_{0.025} = \text{NORM.S.INV}(0.975) = 1.96.$

Table of Contents

Recap: Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with unknown mean μ and known variance σ^2. The sample mean $\overline{X}$ is normally distributed with mean μ and variance $\dfrac{\sigma^2}{n}$. We **standardize** $\overline{X}$ as follows

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

which has a standard normal distribution by the CLT.

Our problem: Determine confidence interval for the population mean μ, known σ.

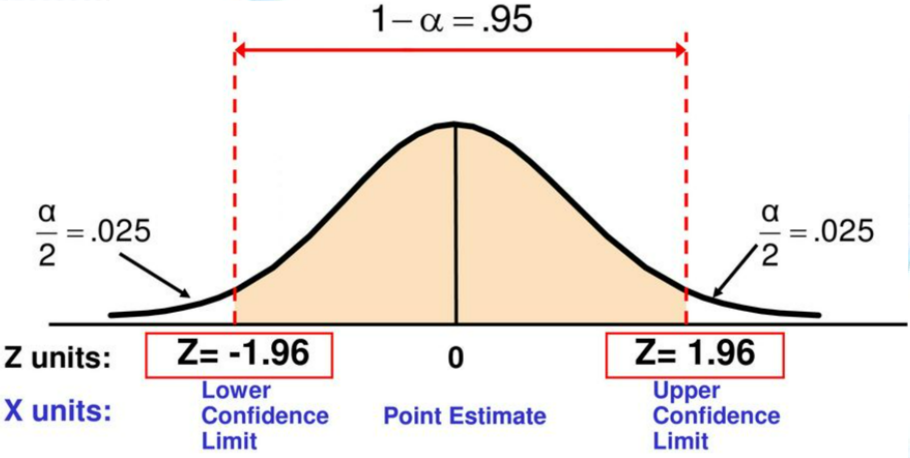
In our problem situation, we may write

$$P\left(-z_{\alpha/2} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1 - \alpha$$

which implies

$$P\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Consider a 95% confidence interval.



In this case, $Z = \pm 1.96$.

Two-Sided Confidence Interval on the Mean, Variance Known

If $\overline{x}$ is the sample mean of a random sample of size n from a normal population with known variance σ^2, a $100(1-\alpha)\%$ confidence interval on μ is given by

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

where $z_{\alpha/2}$ is the upper $100\alpha/2$ percentage point of the standard normal distribution.

Note

Commonly used confidence levels are 90%, 95% and 99%.

Confidence level	**Confidence coefficient** $1-\alpha$	$z_{\alpha/2}$
80%	0.80	1.28
90%	0.90	1.645
95%	0.95	1.96
98%	0.98	2.33
99%	0.99	2.575
99.8%	0.998	3.08
99.9%	0.999	3.27

Recall

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Example (Confidence Interval on the Mean, Variance Known)

A sample of 11 circuits from a large normal population has a mean resistance of 2.2 ohms. We know from past testing that the population standard deviation is 0.35 ohms. Determine a 95% confidence interval for the true mean resistance of the population.

Solution. Since $100(1-\alpha)\% = 95\%$, then $\alpha = 0.05$ and $z_{\alpha/2} = 1.96$. We have

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$2.2 - 1.96 \cdot \frac{0.35}{\sqrt{11}} \leq \mu \leq 2.2 + 1.96 \cdot \frac{0.35}{\sqrt{11}}$$

$$1.9932 \leq \mu \leq 2.4068.$$

Remarks

- We are 95% confident that the true mean resistance is between 1.9932 and 2.4068 ohms.
- Although the true mean may or may not be in this interval, 95% of intervals formed in this manner will contain the true mean.

Recall

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Quiz

In a sample of 36 randomly selected women, it was found that their mean height was 65.3 inches. From previous studies, it is assumed that the standard deviation of all women heights $\sigma = 2.5$ inches. Construct a 90% confidence interval for the mean height of all women.

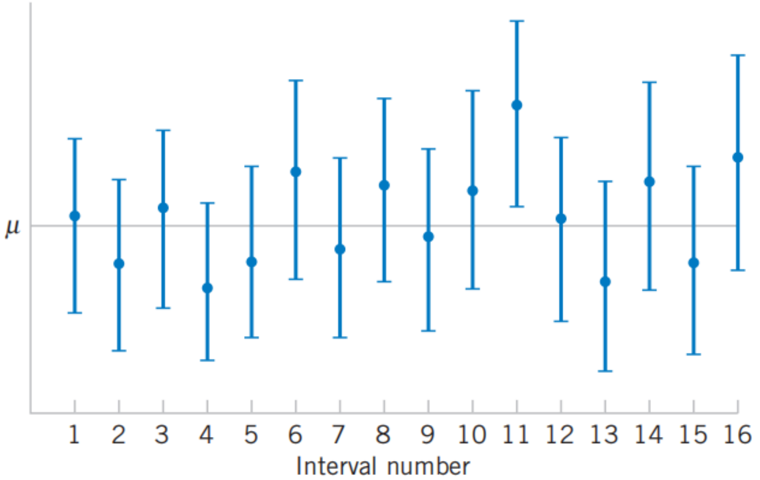


Figure. Repeated construction of a confidence interval for μ.

A $100(1-\alpha)\%$ confidence interval on μ is given by

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

- The **length** of the $100(1-\alpha)\%$ confidence interval is $2z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$.
- Using $\overline{x}$ to estimate μ, the error $|\overline{x}-\mu|$ is less than or equal to $z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ with confidence $100(1-\alpha)\%$.
- In situations the sample size can be controlled, we can choose n so that the error is less than a specified bound.

Sample Size for Specified Error on the Mean, Variance Known

If $\overline{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error $|\overline{x}-\mu|$ will not exceed a specified amount E when the sample size is

$$n = \left\lceil \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 \right\rceil.$$

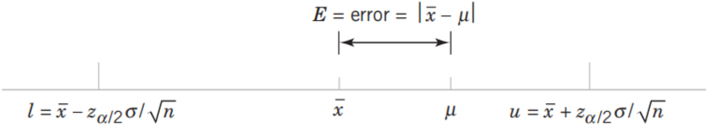


Figure. Error in estimate μ with $\overline{x}$.

FUQN

Recall

$$n = \left\lceil \left(\frac{z_{\alpha/2}\sigma}{E} \right)^2 \right\rceil.$$

Example

A doctor at a local hospital is interested in estimating the birth weight of infants. How large a sample must she select if she desires to be 95% confident that the true mean is within 3 ounces of the sample mean? The standard deviation of the birth weights is known to be 7 ounces.

Solution. Since the true mean is within 3 ounces of the sample mean, then $E = 3$. In addition, $\sigma = 7$ and $z_{\alpha/2} = z_{0.025} = 1.96$. Thus, the required sample size is

$$n = \left\lceil \left(\frac{z_{\alpha/2}\sigma}{E} \right)^2 \right\rceil = \left\lceil \left(\frac{1.96 \cdot 7}{3} \right)^2 \right\rceil = \lceil 20.9153 \rceil = 21.$$

Quiz

A confidence interval estimate is desired for the gain in a circuit on a semiconductor device. Assume that gain is normally distributed with standard deviation $\sigma = 20$. How large must n be if the length of the 95% confidence interval is to be 40?

Note

General relationship between sample size, desired length of the confidence interval $2E$, confidence level $100(1-\alpha)\%$, and standard deviation σ:

- As the desired length of the interval $2E$ decreases, the required sample size n increases for a fixed value of σ and specified confidence.
- As σ increases, the required sample size n increases for a fixed desired length $2E$ and specified confidence.
- As the level of confidence increases, the required sample size n increases for fixed desired length $2E$ and standard deviation σ.

One-Sided Confidence Bounds on the Mean, Variance Known

- A $100(1-\alpha)\%$ **upper-confidence bound** for μ is

$$\mu \leq u = \overline{x} + z_\alpha \frac{\sigma}{\sqrt{n}}.$$

- A $100(1-\alpha)\%$ **lower-confidence bound** for μ is

$$\overline{x} - z_\alpha \frac{\sigma}{\sqrt{n}} = l \leq \mu.$$

Quiz

The diameter of holes for a cable harness is known to have a normal distribution with $\sigma = 0.01$ inches. A random sample of size 15 yields an average diameter of 1.5 inches. Find a 98% lower-confidence bound for the population mean.

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with unknown mean μ and unknown variance σ^2. If the sample size n is large, by the CTL

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{and} \quad Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

However, σ is unknown, replacing σ by the sample standard deviation S when n is large.

Large-Sample Confidence Interval on the Mean

When n is large, the quantity

$$\frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has an approximate standard normal distribution. Consequently,

$$\overline{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}$$

is a **large sample confidence interval** for μ, with confidence level of approximately $100(1-\alpha)\%$.

Do You Ever Truly Know σ^2?

1. Probably not!
2. If there is a situation where σ^2 is known, then μ is also known since to calculate σ^2 you need to know μ.
3. If you truly know μ there would be no need to gather a sample to estimate it.

To determine confidence interval on the mean of normal distribution in the case we do not know the population variance σ^2,

- we need to replace the population standard deviation σ by a sample standard deviation s.
- using t-distribution instead of normal distribution.

Question. What is t-distribution?

Table of Contents

Confidence Interval on the Mean of a Normal Distribution, Variance Unknown

t-Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with unknown mean μ and unknown variance σ^2. Then, the random variable

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has a t-distribution with $n-1$ degrees of freedom, denoted as d$f = n-1$.

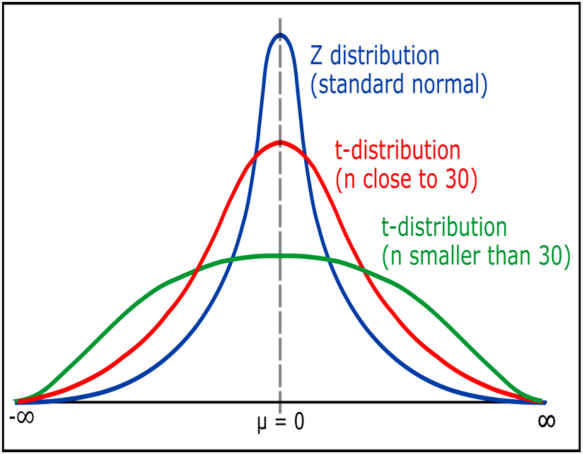
The t probability density function is

$$f(x) = \frac{\Gamma[(k+1)/2]}{\sqrt{\pi k}\Gamma(k/2)} \cdot \frac{1}{[(x^2/k)+1]^{(k+1)/2}}, \quad -\infty < x < +\infty,$$

where k is the number of degrees of freedom. The mean and variance of the t-distribution are

$$\mu = 0 \quad \text{and} \quad \sigma^2 = \frac{k}{k-2},\ k > 2.$$

Note. The t is a family of distributions and the t value depends on degrees of freedom (the number of observations that are free to vary after sample mean has been calculated).



Note

1. The t-distribution is similar to the standard normal distribution in that both distribution are symmetric, unimodal, and the maximum ordinate value is reached when the mean $\mu = 0$.
2. The t-distribution has heavier tails than the normal, that is, it has more probability in the tails than the normal distribution.

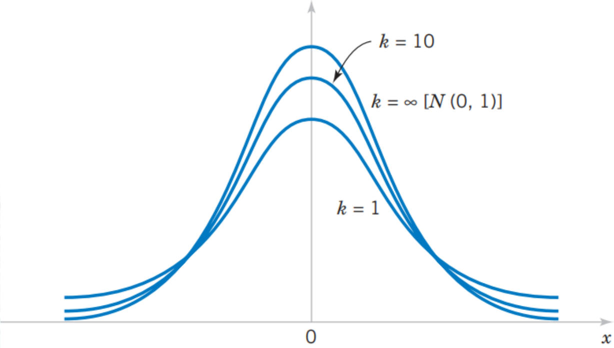


Figure. Probability density functions of several t-distribution.

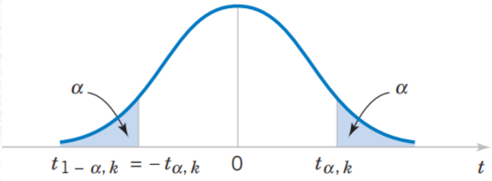


Figure. Percentage points of the t-distribution.

Note

Critical value $t_{\alpha,\mathrm{d}f} = \mathrm{T.INV.2T}(2\alpha, \mathrm{d}f)$. In addition, $t_{1-\alpha,n} = -t_{\alpha,n}$.

Let $t_{\alpha/2,n-1}$ be the upper $100\alpha/2$ percentage point of the t-distribution with $n-1$ degrees of freedom, we may write

$$P(-t_{\alpha/2,n-1} \leq T \leq t_{\alpha/2,n-1}) = 1-\alpha$$

or

$$P\left(-t_{\alpha/2,n-1} \leq \frac{\overline{X}-\mu}{S/\sqrt{n}} \leq t_{\alpha/2,n-1}\right) = 1-\alpha.$$

Therefore,

$$P\left(\overline{X} - t_{\alpha/s,n-1}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right) = 1-\alpha.$$

Two-Sided Confidence Interval on the Mean, Variance Unknown

Let $\overline{x}$ and s are the mean and standard deviation of a random sample from a normal distribution with unknown variance σ^2. Then, a $100(1-\alpha)\%$ **confidence interval** on μ is given by

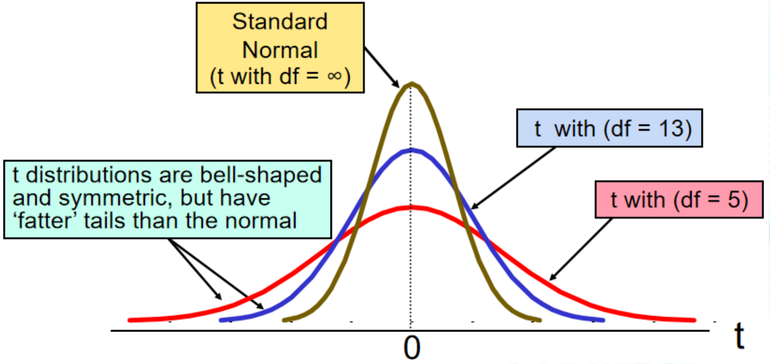
$$\overline{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

where $t_{\alpha/2,n-1}$ is the upper $100\alpha/2$ percentage point of the t-distribution with $n-1$ degrees of freedom.

t-**distribution values with comparison to the** Z **value**.

Confidence level	$t_{\alpha/2,10}$	$t_{\alpha/2,20}$	$t_{\alpha/2,30}$	$z_{\alpha/2}$
80%	1.372	1.325	1.310	1.28
90%	1.812	1.725	1.697	1.645
95%	2.228	2.086	2.042	1.96
98%	2.764	2.528	2.457	2.33
99%	3.169	2.845	2.750	2.575

Note. $t \to Z$ as n increases.



$$\overline{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \le \mu \le \overline{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

Example

A random sample of $n = 25$ has $\overline{X} = 50$ and $S = 8$. Form a 95% CI for μ.

Solution. We have

$$\mathrm{d}f = n - 1 = 24 \text{ and } \alpha = 0.05, \text{ so } t_{\alpha/2,n-1} = t_{0.025,24} = 2.0639.$$

The confidence interval is

$$\overline{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \le \mu \le \overline{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

$$50 - 2.0639 \cdot \frac{8}{\sqrt{25}} \le \mu \le 50 + 2.0639 \cdot \frac{8}{\sqrt{25}}$$

$$46.698 \le \mu \le 53.302.$$

Quiz

The distribution of weights of all products of a company has a normal distribution. A random sample of products has the following weights (in kg):

$$1.9,\ 2.0,\ 2.0,\ 2.1,\ 1.8,\ 2.2,\ 1.8.$$

Construct a 95% confidence interval for the true average weight of all products.

One-Sided Confidence Interval on the Mean, Variance Unknown

1. A $100(1-\alpha\%)$ upper-confidence bound for μ is

$$\mu \leq \overline{x} + t_{\alpha,n-1}\frac{s}{\sqrt{n}}.$$

2. A $100(1-\alpha)\%$ lower-confidence bound for μ is

$$\mu \geq \overline{x} - t_{\alpha,n-1}\frac{s}{\sqrt{n}}.$$

Quiz

The distribution of weights of all products of a company has a normal distribution. A random sample of products has the following weights (in kg):

$$1.9,\ 2.0,\ 2.0,\ 2.1,\ 1.8,\ 2.2,\ 1.8.$$

Construct a 95% lower-confidence bound for the average weight of all products.

Table of Contents

Large-Sample Confidence Interval For A Population Proportion

Normal Approximation for a Binomial Proportion

Let X be a binomial random variable with parameters n and p. Then, n is large, the distribution of

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\hat{P} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$$

is approximately standard normal distribution.

To construct the confidence interval on p, note that $P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) \cong 1 - \alpha$, i.e.,

$$P\left(-z_{\alpha/2} \leq \frac{\hat{P} - p}{\sqrt{\dfrac{p(1-p)}{n}}} \leq z_{\alpha/2}\right) \cong 1 - \alpha.$$

It follows that

$$P\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{P} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) \cong 1 - \alpha.$$

The quantity $\sqrt{p(1-p)/n}$ is called standard error of the point estimator $\hat{P}$.

Approximate Two-Sided Confidence Interval on a Binomial Proportion

Let $\hat{p}$ is the proportion of observations in a random sample of size n that belongs to a class of interest. Then, an approximate $100(1-\alpha)\%$ confidence interval on the proportion p of the population that belongs to this class is

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution.

Exampel (Left-Handers)

A random sample of 100 people shows that 25 are left-handed. Form a 95% confidence interval for the true proportion of left-handers.

Solution. We have $n = 100$, $\hat{p} = 0.25$ and $\alpha/2 = 0.025$,

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\frac{25}{100} - 1.96\sqrt{\frac{0.25(1-0.25)}{100}} \le p \le \frac{25}{100} + 1.96\sqrt{\frac{0.25(1-0.25)}{100}}$$

$$0.1651 \le p \le 0.3349.$$

Remarks (Left-Handers)

- We are 95% confident that the true percentage of left-handers in the population is between 16.51% and 33.49%.
- Although this range may or may not contain the true proportion, 95% of intervals formed from samples of size 100 in this manner will contain the true proportion.

Recall

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Quiz 1

In a random sample of 85 automobile engine crankshaft bearing, 10 have a surface finish that is rougher than the specifications allow. Determine a 95% two-sided confidence intervals for p.

Quiz 2

Of 1000 randomly selected cases of lung cancer, 750 resulted in death within 10 years. Calculate a 95% two-sided confidence interval of the death rate from lung cancer.

Sample Size for a Specified Error on a Binomial Proportion

1. The required sample size that the error estimating $|\hat{P} - p|$ not exceed E is
$$n = \left\lceil \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p) \right\rceil.$$
2. If an estimate $\hat{p}$ from a previous sample is available, change $p(1-p)$ by $\hat{p}(1-\hat{p})$,i.e.,
$$n = \left\lceil \left(\frac{z_{\alpha/2}}{E}\right)^2 \hat{p}(1-\hat{p}) \right\rceil.$$
3. Else, use
$$n = \left\lceil \left(\frac{z_{\alpha/2}}{E}\right)^2 \cdot 0.25 \right\rceil.$$

Example

How large a sample would be necessary to estimate the true proportion defective in a large population within $\pm 3\%$, with 95% confidence? Assume a pilot sample yields $\hat{p} = 0.12$.

Solution. For 95% confidence, use $z_{\alpha/2} = z_{0.025} = 1.96$ and $E = 0.03$, we have
$$n = \left\lceil \left(\frac{z_{\alpha/2}}{E}\right)^2 \hat{p}(1-\hat{p}) \right\rceil = \left\lceil \left(\frac{1.96}{0.03}\right)^2 0.12(1-0.12) \right\rceil = \lceil 450.74 \rceil = 451.$$

Quiz

Of 1000 randomly selected cases of lung cancer, 750 resulted in death within 10 years. What sample size is needed to be 95% confident that the error in estimating the true value of p is less than 4%?

Approximate One-Sided Confidence Bounds on a Binomial Proportion

The approximate $100(1-\alpha)\%$ lower and upper confidence bounds are

$$\hat{p} - z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \quad \text{and} \quad p \le \hat{p} + z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$

respectively.

Quiz

A survey of 250 homeless persons showed that 47 were veterans. Construct a 95% upper confidence bound for the proportion of homeless persons who are veterans.

Thank you!