American University of Beirut

STAT 210: Elementary Statistics for Sciences

2022–2023 Fall

Siamak Taati

Chapter 8
Estimation of mean and proportion
Part 3

Estimation of population parameters

Estimation of population parameters

Interval estimate/estimator

A 95% confidence interval for a population parameter θ is an interval of the form

$$\text{point estimate} \pm \text{margin of error}$$

(calculated on the basis of a random sample) that is meant to estimate the true value of θ.

The procedure used to obtain a confidence interval is called an interval estimator.

Estimation of population proportion

Finding confidence intervals for the population proportion

Suppose we want to estimate the proportion p of the members of a population that have a certain characteristic $\circledast$ based on a *random sample* of size n.

Estimation of population proportion

Finding confidence intervals for the population proportion

Suppose we want to estimate the proportion p of the members of a population that have a certain characteristic $\circledast$ based on a *random sample* of size n.

We only consider the scenario in which all the following conditions are satisfied:

Estimation of population proportion

Finding confidence intervals for the population proportion

Suppose we want to estimate the proportion p of the members of a population that have a certain characteristic $\circledast$ based on a *random sample* of size n.

We only consider the scenario in which all the following conditions are satisfied:

I. The sample is large. [Rule of thumb: $np > 5$ and $n(1-p) > 5$]

Estimation of population proportion

Finding confidence intervals for the population proportion

Suppose we want to estimate the proportion p of the members of a population that have a certain characteristic $\circledast$ based on a *random sample* of size n.

We only consider the scenario in which all the following conditions are satisfied:

I. The sample is large. [Rule of thumb: $np > 5$ and $n(1-p) > 5$]

II. Either the sample is taken with replacement,
or the population is much larger than the sample.
[Rule of thumb: $\frac{n}{N} \leq 0.05$]

Estimation of population proportion

Finding confidence intervals for the population proportion

Suppose we want to estimate the proportion p of the members of a population that have a certain characteristic $\circledast$ based on a *random sample* of size n.

We only consider the scenario in which all the following conditions are satisfied:

I. The sample is large. [Rule of thumb: $np > 5$ and $n(1-p) > 5$]

II. Either the sample is taken with replacement,
or the population is much larger than the sample.
[Rule of thumb: $\frac{n}{N} \leq 0.05$]

The procedure is derived on the basis of the (approximate) distribution of the sample proportion.

Estimation of population proportion

I. The sample is large.
II. Either the sample is taken with replacement,
or the population is much larger than the sample.

Estimation of population proportion

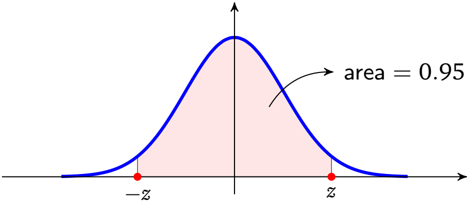
I. The sample is large.
II. Either the sample is taken with replacement,
or the population is much larger than the sample.

Confidence interval for p

Under conditions I and II, the 95% confidence interval for p is

$$\widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

where $\widehat{p}$ is the sample proportion and z is the value indicated below.



Standard normal distribution

Estimation of population proportion

I. The sample is large.

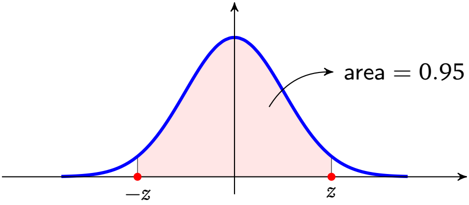
II. Either the sample is taken with replacement,
or the population is much larger than the sample.

Confidence interval for p

Under conditions I and II, the 95% confidence interval for p is

$$\widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

where $\widehat{p}$ is the sample proportion and z is the value indicated below.



Standard normal distribution

Estimation of population proportion

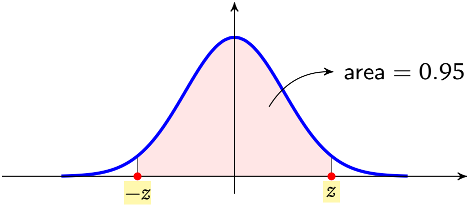
I. The sample is large.
II. Either the sample is taken with replacement,
or the population is much larger than the sample.

Confidence interval for p

Under conditions I and II, the 95% confidence interval for p is

$$\widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

where $\widehat{p}$ is the sample proportion and z is the value indicated below.



Standard normal distribution

Estimation of population proportion

I. The sample is large.

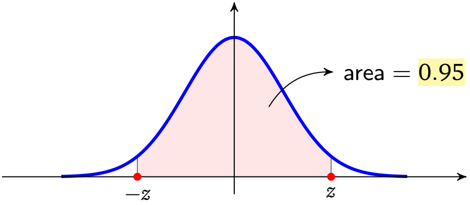
II. Either the sample is taken with replacement,
or the population is much larger than the sample.

Confidence interval for p

Under conditions I and II, the 95% confidence interval for p is

$$\widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

where $\widehat{p}$ is the sample proportion and z is the value indicated below.



Standard normal distribution

Estimation of population proportion

Example (Gene with two alleles)

A certain type of gene in humans has two alleles (versions) A and B. Researchers would like to know the proportion p of individuals with allele A.

Since they do not have access to the entire population, they decide to take a random sample of individuals, identify their alleles, and use this data as evidence to estimate the population proportion p.

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	···	50	# of As	# of Bs
B	B	A	A	···	B	31	19

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	···	50	# of As	# of Bs
B	B	A	A	···	B	31	19

Q1 What is a good estimate for the population proportion of individuals with allele A?

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	$\cdots$	50	# of As	# of Bs
B	B	A	A	$\cdots$	B	31	19

Q1 What is a good estimate for the population proportion of individuals with allele A?

A The sample proportion

$$\widehat{p} = \frac{}{}$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	⋯	50	# of As	# of Bs
B	B	A	A	⋯	B	31	19

Q1 What is a good estimate for the population proportion of individuals with allele A?

A The sample proportion

$$\widehat{p} = \frac{31}{}$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	···	50	# of As	# of Bs
B	B	A	A	···	B	31	19

Q1 What is a good estimate for the population proportion of individuals with allele A?

A The sample proportion

$$\widehat{p} = \frac{31}{50}$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	$\cdots$	50	# of As	# of Bs
B	B	A	A	$\cdots$	B	31	19

Q1 What is a good estimate for the population proportion of individuals with allele A?

A The sample proportion

$$\widehat{p} = \frac{31}{50} = \boxed{0.62} = 62\%$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	⋯	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	⋯	B	31	19	0.62

Q1 What is a good estimate for the population proportion of individuals with allele A?

A The sample proportion

$$\widehat{p} = \frac{31}{50} = \boxed{0.62} = 62\%$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	⋯	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	⋯	B	31	19	0.62

Q2 Find the 95% confidence interval for the population proportion of individuals with allele A?

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	···	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	···	B	31	19	0.62

Q2 Find the 95% confidence interval for the population proportion of individuals with allele A?

$\boxed{\text{A}}$ The sample is large, the population is much larger.

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	⋯	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	⋯	B	31	19	0.62

Q2 Find the 95% confidence interval for the population proportion of individuals with allele A?

A The sample is large, the population is much larger. Therefore,

$$95\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	⋯	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	⋯	B	31	19	0.62

Q2 Find the 95% confidence interval for the population proportion of individuals with allele A?

A The sample is large, the population is much larger. Therefore,

$$95\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	⋯	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	⋯	B	31	19	0.62

Q2 Find the 95% confidence interval for the population proportion of individuals with allele A?

A The sample is large, the population is much larger. Therefore,

$$95\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

$$\approx 0.62 \pm \qquad \cdot \sqrt{\frac{0.62 \cdot (1-0.62)}{}}$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	⋯	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	⋯	B	31	19	0.62

Q2 Find the 95% confidence interval for the population proportion of individuals with allele A?

A The sample is large, the population is much larger. Therefore,

$$95\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

$$\approx 0.62 \pm \qquad \cdot \sqrt{\frac{0.62 \cdot (1-0.62)}{}}$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	⋯	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	⋯	B	31	19	0.62

Q2 Find the 95% confidence interval for the population proportion of individuals with allele A?

A The sample is large, the population is much larger. Therefore,

$$95\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

$$\approx 0.62 \pm \qquad \cdot \sqrt{\frac{0.62 \cdot (1-0.62)}{50}}$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	···	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	···	B	31	19	0.62

Q2 Find the 95% confidence interval for the population proportion of individuals with allele A?

$\boxed{\text{A}}$ The sample is large, the population is much larger. Therefore,

$$95\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

$$\approx 0.62 \pm \qquad \cdot \sqrt{\frac{0.62 \cdot (1-0.62)}{50}}$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	⋯	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	⋯	B	31	19	0.62

Q2 Find the 95% confidence interval for the population proportion of individuals with allele A?

$\boxed{\text{A}}$ The sample is large, the population is much larger. Therefore,

$$95\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

$$\approx 0.62 \pm \qquad \cdot \sqrt{\frac{0.62 \cdot (1-0.62)}{50}}$$

Using the applet for normal distribution, $z \approx 1.95996$.

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	···	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	···	B	31	19	0.62

Q2 Find the 95% confidence interval for the population proportion of individuals with allele A?

A The sample is large, the population is much larger. Therefore,

$$95\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

$$\approx 0.62 \pm 1.95996 \cdot \sqrt{\frac{0.62 \cdot (1-0.62)}{50}}$$

Using the applet for normal distribution, $z \approx 1.95996$.

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	⋯	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	⋯	B	31	19	0.62

Q2 Find the 95% confidence interval for the population proportion of individuals with allele A?

A The sample is large, the population is much larger. Therefore,

$$
\begin{aligned}
95\%\text{-CI} &= \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \\
&\approx 0.62 \pm 1.95996 \cdot \sqrt{\frac{0.62 \cdot (1-0.62)}{50}} \\
&\approx \boxed{0.62 \pm 0.135} = (0.485, 0.755)
\end{aligned}
$$

Using the applet for normal distribution, $z \approx 1.95996$.

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	···	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	···	B	31	19	0.62

Q3 What would be the level of confidence if we require the margin of error for estimating the population proportion to be 0.05?

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	···	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	···	B	31	19	0.62

Q3 What would be the level of confidence if we require the margin of error for estimating the population proportion to be 0.05?

A The sample is large, the population is much larger.

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	···	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	···	B	31	19	0.62

Q3 What would be the level of confidence if we require the margin of error for estimating the population proportion to be 0.05?

$\boxed{\text{A}}$ The sample is large, the population is much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

We require

$$0.05 = \text{margin of error}$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	⋯	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	⋯	B	31	19	0.62

Q3 What would be the level of confidence if we require the margin of error for estimating the population proportion to be 0.05?

$\boxed{\text{A}}$ The sample is large, the population is much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm \underbrace{z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}}_{\text{margin of error}}$$

We require

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	⋯	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	⋯	B	31	19	0.62

Q3 What would be the level of confidence if we require the margin of error for estimating the population proportion to be 0.05?

A The sample is large, the population is much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm \underbrace{z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}}_{\text{margin of error}}$$

We require

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = \quad \cdot \sqrt{\quad}$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	$\cdots$	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	$\cdots$	B	31	19	0.62

Q3 What would be the level of confidence if we require the margin of error for estimating the population proportion to be 0.05?

$\boxed{\text{A}}$ The sample is large, the population is much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

margin of error

We require

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = \quad \cdot \sqrt{\frac{0.62(1-0.62)}{}}$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	···	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	···	B	31	19	0.62

Q3 What would be the level of confidence if we require the margin of error for estimating the population proportion to be 0.05?

A The sample is large, the population is much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

margin of error

We require

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = \quad \cdot \sqrt{\frac{0.62(1-0.62)}{50}}$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	···	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	···	B	31	19	0.62

Q3 What would be the level of confidence if we require the margin of error for estimating the population proportion to be 0.05?

$\boxed{A}$ The sample is large, the population is much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

margin of error

We require

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = z \cdot \sqrt{\frac{0.62(1-0.62)}{50}}$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	⋯	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	⋯	B	31	19	0.62

Q3 What would be the level of confidence if we require the margin of error for estimating the population proportion to be 0.05?

A The sample is large, the population is much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

margin of error

We require

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = z \cdot \sqrt{\frac{0.62(1-0.62)}{50}}$$

Solving for z, we get $z \approx$.

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	⋯	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	⋯	B	31	19	0.62

Q3 What would be the level of confidence if we require the margin of error for estimating the population proportion to be 0.05?

$\boxed{\text{A}}$ The sample is large, the population is much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

margin of error

We require

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = z \cdot \sqrt{\frac{0.62(1-0.62)}{50}}$$

Solving for z, we get $z \approx 0.7283957$.

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	$\cdots$	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	$\cdots$	B	31	19	0.62

Q3 What would be the level of confidence if we require the margin of error for estimating the population proportion to be 0.05?

A The sample is large, the population is much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm \underbrace{z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}}_{\text{margin of error}}$$

We require

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = z \cdot \sqrt{\frac{0.62(1-0.62)}{50}}$$

Solving for z, we get $z \approx 0.7283957$.

Using the applet for normal distribution, $? \approx$.

Estimation of population proportion

Example (Gene with two alleles)

The researchers take a sample of 50 individuals and identify their alleles of this gene:

1	2	3	4	⋯	50	# of As	# of Bs	$\widehat{p}$
B	B	A	A	⋯	B	31	19	0.62

Q3 What would be the level of confidence if we require the margin of error for estimating the population proportion to be 0.05?

A The sample is large, the population is much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

margin of error

We require

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = z \cdot \sqrt{\frac{0.62(1-0.62)}{50}}$$

Solving for z, we get $z \approx 0.7283957$.

Using the applet for normal distribution, $? \approx \boxed{0.5336} \approx 53.4\%$.

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

A The sample will be larger than 50, and the population still much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

A The sample will be larger than 50, and the population still much larger. Therefore,

$$? \%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Ideally, we would like

$$0.95 = \text{confidence level} = ?$$

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

A The sample will be larger than 50, and the population still much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Ideally, we would like

$$0.95 = \text{confidence level} = ?$$

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Given the confidence level $? = 0.95$, we can use the applet for normal distribution to find $z \approx$.

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

A The sample will be larger than 50, and the population still much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Ideally, we would like

$$0.95 = \text{confidence level} = ?$$

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Given the confidence level $? = 0.95$, we can use the applet for normal distribution to find $z \approx 1.95996$.

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

A The sample will be larger than 50, and the population still much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Ideally, we would like

$$0.95 = \text{confidence level} = ?$$

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.95996 \cdot \sqrt{\frac{\phantom{\widehat{p}(1-\widehat{p})}}{n}}$$

Given the confidence level $? = 0.95$, we can use the applet for normal distribution to find $z \approx 1.95996$.

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

A The sample will be larger than 50, and the population still much larger. Therefore,

$$? \%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Ideally, we would like

$$0.95 = \text{confidence level} = ?$$

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.95996 \cdot \sqrt{\frac{\qquad\qquad}{n}}$$

Q What shall we do with $\widehat{p}$?

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

A The sample will be larger than 50, and the population still much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Ideally, we would like

$$0.95 = \text{confidence level} = ?$$

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.95996 \cdot \sqrt{\frac{\qquad}{n}}$$

Q What shall we do with $\widehat{p}$?

A1 Conservative choice: substitute $\widehat{p}$ with 0.5. This is the worst case scenario!

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

A The sample will be larger than 50, and the population still much larger. Therefore,

$$? \%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Ideally, we would like

$$0.95 = \text{confidence level} = ?$$

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.95996 \cdot \sqrt{\frac{0.5(1-0.5)}{n}}$$

Q What shall we do with $\widehat{p}$?

A1 Conservative choice: substitute $\widehat{p}$ with 0.5. This is the worst case scenario!

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

$\boxed{\text{A}}$ The sample will be larger than 50, and the population still much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Ideally, we would like

$$0.95 = \text{confidence level} = ?$$

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.95996 \cdot \sqrt{\frac{0.5(1-0.5)}{n}}$$

Q What shall we do with $\widehat{p}$?

$\boxed{\text{A1}}$ Conservative choice: substitute $\widehat{p}$ with 0.5. This is the worst case scenario!

Solving for n, we get $n \approx$.

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

A The sample will be larger than 50, and the population still much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Ideally, we would like

$$0.95 = \text{confidence level} = ?$$

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.95996 \cdot \sqrt{\frac{0.5(1-0.5)}{n}}$$

Q What shall we do with $\widehat{p}$?

A1 Conservative choice: substitute $\widehat{p}$ with 0.5. This is the worst case scenario!

Solving for n, we get $n \approx 384.15$.

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

$\boxed{\text{A}}$ The sample will be larger than 50, and the population still much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Ideally, we would like

$$0.95 = \text{confidence level} = ?$$

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.95996 \cdot \sqrt{\frac{0.5(1-0.5)}{n}}$$

Q What shall we do with $\widehat{p}$?

$\boxed{\text{A1}}$ Conservative choice: substitute $\widehat{p}$ with 0.5. This is the worst case scenario!

Solving for n, we get $n \approx 384.15$. [Warning: n must be a whole number!]

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

$\boxed{\text{A}}$ The sample will be larger than 50, and the population still much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Ideally, we would like

$$0.95 = \text{confidence level} = ?$$

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.95996 \cdot \sqrt{\frac{0.5(1-0.5)}{n}}$$

Q What shall we do with $\widehat{p}$?

$\boxed{\text{A1}}$ Conservative choice: substitute $\widehat{p}$ with 0.5. This is the worst case scenario!

Solving for n, we get $n \approx 384.15$. [Warning: n must be a whole number!]

Thus, a sample of size $n \geq \boxed{385}$ would be sufficient.

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

$\boxed{\text{A}}$ The sample will be larger than 50, and the population still much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Ideally, we would like

$$0.95 = \text{confidence level} = ?$$

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.95996 \cdot \sqrt{\frac{\qquad}{n}}$$

Q What shall we do with $\widehat{p}$?

$\boxed{\text{A2}}$ Use the estimate provided by the earlier (smaller scale) sample, that is, 0.62.

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

A The sample will be larger than 50, and the population still much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Ideally, we would like

$$0.95 = \text{confidence level} = ?$$

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.95996 \cdot \sqrt{\frac{0.62(1-0.62)}{n}}$$

Q What shall we do with $\widehat{p}$?

A2 Use the estimate provided by the earlier (smaller scale) sample, that is, 0.62.

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

$\boxed{\text{A}}$ The sample will be larger than 50, and the population still much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Ideally, we would like

$0.95 = \text{confidence level} = ?$

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.95996 \cdot \sqrt{\frac{0.62(1-0.62)}{n}}$$

Q What shall we do with $\widehat{p}$?

$\boxed{\text{A2}}$ Use the estimate provided by the earlier (smaller scale) sample, that is, 0.62.

Solving for n, we get $n \approx$.

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

A The sample will be larger than 50, and the population still much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Ideally, we would like

$$0.95 = \text{confidence level} = ?$$

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.95996 \cdot \sqrt{\frac{0.62(1-0.62)}{n}}$$

Q What shall we do with $\widehat{p}$?

A2 Use the estimate provided by the earlier (smaller scale) sample, that is, 0.62.

Solving for n, we get $n \approx 362.02$.

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

A The sample will be larger than 50, and the population still much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Ideally, we would like

$0.95 = \text{confidence level} = ?$

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.95996 \cdot \sqrt{\frac{0.62(1-0.62)}{n}}$$

Q What shall we do with $\widehat{p}$?

A2 Use the estimate provided by the earlier (smaller scale) sample, that is, 0.62.

Solving for n, we get $n \approx 362.02$. [Warning: n must be a whole number!]

Estimation of population proportion

Example (Gene with two alleles)

The researchers decide to take a larger sample that allows them to estimate the population proportion of allele A with

margin of error ≤ 0.05 confidence level $\geq 95\%$

Q4 How large must the sample be?

A The sample will be larger than 50, and the population still much larger. Therefore,

$$?\%\text{-CI} = \widehat{p} \pm z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Ideally, we would like

$$0.95 = \text{confidence level} = ?$$

$$0.05 = \text{margin of error} = z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 1.95996 \cdot \sqrt{\frac{0.62(1-0.62)}{n}}$$

Q What shall we do with $\widehat{p}$?

A2 Use the estimate provided by the earlier (smaller scale) sample, that is, 0.62.

Solving for n, we get $n \approx 362.02$. [Warning: n must be a whole number!]

Thus, a sample of size $n \geq \boxed{363}$ would be sufficient.

Estimation of population parameters

Interval estimate/estimator

A 95% confidence interval for a population parameter θ is an interval of the form

$$\text{point estimate} \pm \text{margin of error}$$

(calculated on the basis of a random sample) that is meant to estimate the true value of θ.

The procedure used to obtain a confidence interval is called an interval estimator.

Estimation of population parameters

Interpretation of confidence intervals

The confidence level of 95% is a property of the interval estimator (i.e., the procedure).

Estimation of population parameters

Interpretation of confidence intervals

The confidence level of 95% is a property of the interval estimator (i.e., the procedure). It means the following:

Estimation of population parameters

Interpretation of confidence intervals

The confidence level of 95% is a property of the interval estimator (i.e., the procedure). It means the following:

Consider the following experiment:

> Take a random sample of size n and use the interval estimator to form a confidence interval for θ.

Estimation of population parameters

Interpretation of confidence intervals

The confidence level of 95% is a property of the interval estimator (i.e., the procedure). It means the following:

Consider the following experiment:

> Take a random sample of size n and use the interval estimator to form a confidence interval for θ.

If we repeat this experiment many many times (each time with a fresh sample, independently of the previous ones), then

- 95% of the times, the formed confidence interval will contain the true value of θ, and
- the remaining 5% of the times it misfires.