Parameter estimation

Parameter Estimates

Example (Agresti 2018): in the 2014 General Social Survey, a sample of n=2538 adults was considered (about one person out of 100,000 people in the population)

Amongst other items, the survey asked, “Do you believe there is a life after death?”

The (unknown) proportion of American believers is the **parameter** of interest

The **estimate** for the proportion of all Americans who would respond yes is obtained from the 2014GSS and equals 0.73

Point and Interval Estimation

A **point estimate** (or simply an estimate) is a single number that is the best guess for the parameter value.

An **interval estimate** is an interval of numbers around the point estimate that we believe contains the parameter value. This interval is also called a **confidence interval** (CI).

Example (Agresti 2018):

0.73 is the point **estimate** for the proportion of all American believers.

An interval estimate predicts that the population proportion responding yes falls between 0.71 and 0.75.

This CI tells us that the point estimate of 0.73 has a margin of error of 0.02. Thus, an interval estimate helps us gauge the precision of a point estimate

Stochastic model and model parameters

In statistical modelling we assume that the population of interest can be adequately describe by a random variable X.

For instance, we assume that the concentration of arsenic in water can be adequately described using a Gaussian random variable.
This means that, indicating by X the concentration of arsenic we suppose $X \sim N(\mu, \sigma^2)$ for some values of μ and σ^2.

If this approximation is adequate, we can use the model to predict values of X, calculate the probability that X is above some threshold endorsed by law or suggested by health recommendation, to calculate the average dose to which some human or animal population is exposed and so on.

Stochastic model and model parameters

In practice, we rarely know the value of μ and σ^2 and need to use data on X to estimate them.
μ and σ^2 are called **parameters** of the population of interest.

By **parameter** it is meant an unknown characteristic of the distribution that we assume to govern the population.

In statistical modelling it is assumed that the phenomenon of interest is described by a RV X whose distribution (i.e. the *population*) depends upon some parameter θ (i.e. the *characteristic* of the population) that is unknown i.e.

$$X \sim f(x, \theta)$$

The goal is *to estimate* this parameter using data taken form the considered distribution.

Estimates and Estimators

Let X be a random variable (RV) and θ an unknown characteristic (**parameter**) of X

Estimator: a particular type of statistic adopted for estimating a parameter θ

$$T = t(X_1,\ X_2,\ \dots,\ X_n)$$

Estimate is the value of T for a particular (observed) sample

$$t = t(x_1,\ x_2,\ \dots,\ x_n)$$

$X_1,\ X_2,\ \dots,\ X_n$ is the sample taken from the population of interest

Point Estimation

Let $f(x, \theta)$ the pdf or the pf of a RV X and θ an unknown parameter and $T = t(X_1,\ X_2,\ \ldots,\ X_n)$ and estimator of θ

Two issues

- There exists one (or more) general procedure to obtain an estimator of θ? ➔ Estimation methods
- How can we evaluate the performance of a given estimator? Since there might exist more than one estimator of θ, what estimator should be considered? ➔ Estimation properties

Naïve Point Estimates

Naïve point estimation is an intuitive and simple method to estimate an unknown parameter (that works in practice…).

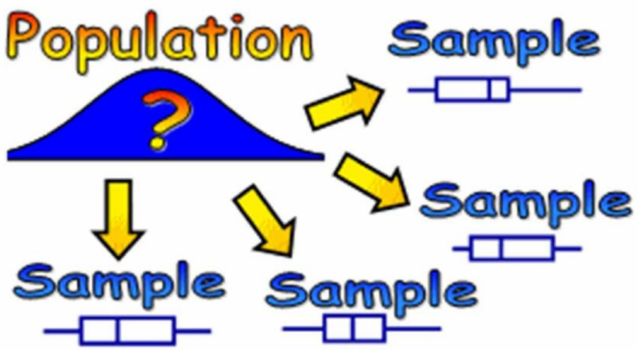
Just apply to the sample data the same “formula” that defines the parameter at the population level.

For instance use …

- the sample variance $\tilde{S}^2$ to estimate the variance of the population $\mathrm{Var}(X)$
- the sample mean $\overline{X}$ to estimate the mean of the population $E(X)$
- the sample median to estimate the median of the population
- the sample proportion to estimate the relative frequency of the population
- … … …

Sampling variability

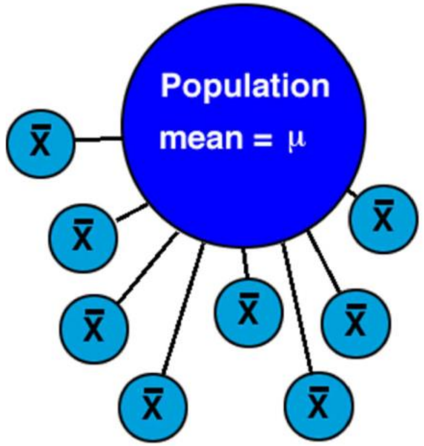
We typically do not know the sample in advance, and we can get different samples while sampling the same population in similar conditions.



The sample is then a set of random variables.

A **random sample** is a sample where $X_1,\ X_2,\ \dots,\ X_n$ are independent random variables all being distributed like X.

Being a function of random variables $T = t(X_1,\ X_2,\ \dots,\ X_n)$ is also a random variable. The properties of an estimator are defined in terms of the density distribution of T.



Estimation Properties

Let $(X_1, \dots, X_n)$ be a random sample taken from $\mathrm{f}(\mathrm{x}, \theta)$ and $T = T(X_1, \dots, X_n)$ an estimator of θ.

Since T is a statistic adopted for estimating a parameter θ, T *is a random variable* as well.

The performance of T can be evaluated by investigating the properties of the distribution of T.

Estimation Properties: Unbiased estimator

An estimator is **unbiased** if its sampling distribution centres around the parameter.

Def. T is unbiased for θ if $E(T-\theta)=0,\ \forall\theta$

If the estimator were calculated repeatedly with different samples,

in the long run the overestimates counterbalance the underestimates.

If $E(T-\theta)>0$ (<0), the estimator tends to overestimate (underestimate) the parameter.

If $E(T-\theta)\neq 0\ \forall\theta$ the estimator is **biased.**

In general, we tend to prefer an unbiased estimator.

Properties of the expected value and Bias

If T is unbiased for θ *then* $E(T) = \theta \;\; \forall\theta$

If T is biased

$$\text{Bias(T)} = E(T - \theta) = E(T) - \theta$$

is called the **estimator Bias**.

Thus overestimation (underestimation) means $E(T) > \theta$ ($< \theta$).

Example: the sample range

Sample range $\mathrm{R} = X_{(n)} - X_{(1)}$

- $X_{(1)} = \min(X_1, \ldots, X_n)$ is the minimum of the sample values of X
- $X_{(n)} = \mathrm{Max}(X_1, \ldots, X_n)$ is the maximum of the sample values of X

The sample range can never be larger than the population range

Thus, R is a biased estimator of the population range (being always narrower, i.e. the sample range underestimates the range of X)

Example

Let X be a random variable such that $E(X) = \mu$ *and* $Var(X) = \sigma^2$

$$\boxed{\mathrm{E}(\overline{X}) = \mu, \qquad \forall\ \mu}$$

$$\boxed{\mathrm{E}(S^2) = \sigma^2, \qquad \forall\ \sigma^2}$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Therefore, the sample mean, and S^2 are unbiased estimators of μ and σ^2 respectively.

Example

Consider the estimator of σ^2 given by $\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$

In this case $\mathrm{E}(\tilde{S}^2) = \mathrm{E}\left(\frac{n-1}{n}S^2\right) = \frac{n-1}{n}\mathrm{E}(S^2) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$

Therefore $\tilde{S}^2$ is biased for σ^2 and

$$\mathrm{Bias}(\tilde{S}^2) = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n} < 0$$

Bias and Variability

Which one would you prefer?

Unbiasedness alone is note completely satisfying.

The variability of T is also an issue!

Mean square Error (MSE)

Let $(X_1, \ldots, X_n)$ be a random sample taken from $f(x, \theta)$ and $T = T(X_1, \ldots, X_n)$ an estimator of θ.

The mean square error of T is defined by

$$MSE(T) = E(T - \theta)^2$$

$MSE(T)$ is a measure of closeness between the parameter and the estimator.

It is desirable that this measure is the smallest possible for any θ.

An **efficient estimator** has the minimum MSE possible.

It might be challenging to find it and the theory on this is a little bit advanced (a key result is the Theorem of Rao and Cramér) and won't be considered here.

Mean square Error (2)

Property $MSE(T) = Var(T) + (Bias(T))^2$

where $Var(T) = \mathrm{E}(T - E(T))^2$ is the variance of T

Therefore, if T is an unbiased estimator of θ we get

$$MSE(T) = Var(T)$$

Example. The sample mean is an unbiased estimator of $E(X) = \mu$.

$\mathrm{E}(\overline{X} - \mu) = 0$ ➔ $MSE(\overline{X}) = Var(\overline{X}) = \frac{\sigma^2}{n}$ where $Var(X) = \sigma^2$ and n is the sample size

$\sqrt{Var(T)}$ is known as the **standard error** of T, $se(T)$

Example $se(\overline{X}) = \frac{\sigma}{\sqrt{n}}$ is the standard error of the sample mean

Estimation methods

Estimation methods are mathematical criteria for deriving an estimator of a given parameter

Several methods exist

- Maximum Likelihood (ML)
- Methods of Moments (MoM)
- Least Squares (LS)
- Bayes estimators
- …

We do not consider estimation methods in this course.

Interval Estimation

An **interval estimate** is an interval of numbers around the point estimate that we believe contains the parameter value. The interval estimator is also called a **confidence interval** (CI).

A CI is built assuming that the interval includes the true and unknown parameter with a given probability.

Example (Agresti 2018)

0.73 is the point **estimate** for the proportion of all American believers.

An interval estimate predicts that the population proportion responding yes falls between 0.71 and 0.75.

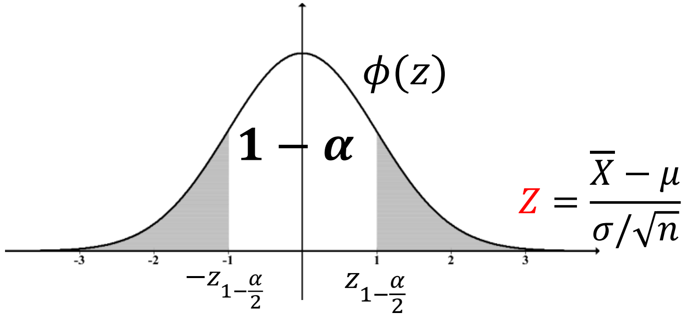
This CI tells us that the point estimate of 0.73 has a margin of error of 0.02. Thus, an interval estimate helps us gauge the precision of a point estimate

Confidence Interval (CI) for the mean of a Normal Population - known variance

$(X_1,\ X_2,\ \ldots,\ X_n)$ random sample from $\mathrm{N}(\mu,\sigma^2)$, σ^2 known.

The CI is constructed with a confidence level of $1-\alpha$, hence we select first a probability value α (typically small, say 0.1, 0.05, 0.01).

It can be proved that $Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is a standard normal therefore we get



$$1-\alpha=P\left(-z_{1-\frac{\alpha}{2}}<\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}<z_{1-\frac{\alpha}{2}}\right)\quad \text{and}$$

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

Hence

$$\left(\bar{X}-z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\ \bar{X}+z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

is the desired CI where $z_{1-\alpha/2}$ is the quantile of the standard normal distribution

Interval estimate for the mean of a Normal Population – known variance

If $(x_1,\ x_2,\ \ldots,\ x_n)$ is an observed sample (lowercase letters) the observed CI of μ is

$$\left(\bar{x} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\qquad \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

- $z_{1-\frac{\alpha}{2}}$ is the quantile of the standard normal distribution
- $\bar{x}$ is the sample mean obtained on the actual sample $(x_1,\ x_2,\ \ldots,\ x_n)$

Example

X: concentration of arsenic in drinking water in μg per liter

It is assumed that $X \sim \mathrm{N}(\mu, 0.03^2)$ $(\sigma = 0.03)$

A random sample of $n = 120$ observations of X is selected where $\bar{x} = 19.4996\ \mu g$

If $\alpha = 0.05$ ➔ $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$. The interval estimate is $19.4996 \mp 1.95\frac{0.03}{\sqrt{120}} = (19.49, 19.51)$. This interval is extremely likely to include the true and unknown parameter $\mu = E(X)$

Confidence Interval (CI) definition

$(X_1,\ X_2,\ \dots,\ X_n)$ random sample from $f(x, \theta)$,

θ unknown parameter of interest

Let $L_1 = l_1(X_1,\ X_2,\ \dots,\ X_n)$ and $L_2 = l_2(X_1,\ X_2,\ \dots,\ X_n)$ be two sample statistics such that

$$L_1 < L_2 \quad \text{and} \quad P(L_1 < \theta < L_2) = 1 - \alpha$$

where α is a small probability not depending upon θ.

(L_1, L_2) is called the **confidence interval** of θ.

Being $(x_1, \dots, x_n)$ an observed sample an **interval estimate** of θ is

$$(l_1 = l_1(x_1, \dots, x_n), \qquad l_2 = l_2(x_1, \dots, x_n))$$

Confidence Interval (CI) for the mean of a Normal Population - unknown variance

$(X_1,\ X_2,\ \dots,\ X_n)$ random sample from $\mathrm{N}(\mu,\sigma^2)$

The CI is constructed with a confidence level of $1-\alpha$ hence we select first this probability.

In this case $Var(X)=\sigma^2$ is unknown and must be estimated by $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$

If σ is replaced by S then $\boxed{\mathrm{T}=\frac{\overline{X}-\mu}{S/\sqrt{n}}}$ is a Student's t with $n-1$ we get

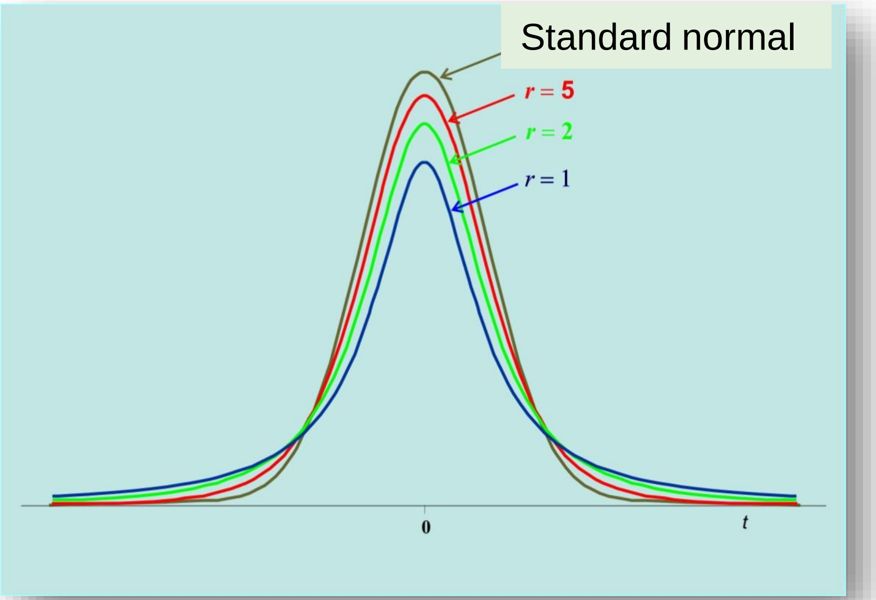
$$1-\alpha=P\left(-t_{n-1,1-\frac{\alpha}{2}}<\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}<t_{n-1,1-\frac{\alpha}{2}}\right)=P\left(-t_{n-1,1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}<\overline{X}-\mu<t_{n-1,1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right)$$
$$=P\left(\overline{X}-t_{n-1,1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}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

$t_{n-1,1-\frac{\alpha}{2}}$ is the quantile of the Student's t distribution with $n-1$

Hence $\left(\bar{X}-t_{n-1,1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}},\bar{X}+t_{n-1,1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right)$ is the desired CI

Student’s T random variable

$T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$ is a Student’s T random variable with $r = n - 1$ df



Note that

- $\mathrm{E}(T) = 0$
- T is symmetric about 0
- The density function equation is a bit awkward ….
- if $r >> 0$ then T is approximated by the standard normal distribution

Approximated Confidence Interval (CI) of the mean for general populations

$(X_1, X_2, \ldots, X_n)$ random sample from $f(x)$ with $E(X) = \mu$

The CI is built at the $1 - \alpha$ level, hence the probability α is first selected

Since $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is **approximately** normally distribute if $n \gg 0$ (Central Limit Theorem)

$$1 - \alpha \approx P\left(-z_{1-\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{1-\frac{\alpha}{2}}\right) = P\left(\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

Hence $\left(L_1 = \bar{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, L_2 = \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$ is the desired CI

In practice σ is unknown, hence the CI can be calculated using S instead

$$\left(L_1 = \bar{X} - z_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, L_2 = \bar{X} + z_{1-\alpha/2}\frac{S}{\sqrt{n}}\right)$$

Final considerations

For parameters of the population distribution other than the mean even large sample (i.e. $n \gg 0$) approximations are difficult to find.

Some classes of estimators (for instance Maximum Likelihood estimators) permit one to obtain such approximated CI.

For this reason, these methodologies are very often adopted in practice for estimation.

Other methods exist, however, to build good approximated CI even in case of samples of moderate or small size.