Introduction to Point Estimation

One goal of statistics:

Draw insights/inference about certain aspects of the population using sample data.

Examples: might want to estimate

① Average gpa of students on campus.

② Average time spent on recreational activities by students at UMD.

③ Median age of everybody affiliated to UMD.

④ The median yearly income of people in the USA.

In each of these situations we need to identify

a) population of interest

b) a characteristic of the population that we are interested in.

Suppose, $\{X_1, X_2, X_3, \dots, X_n\} \longrightarrow$ random sample coming from a fixed population.

A "point estimator" for the population $\longrightarrow$ any statistic $\hat{\theta}$ for the random sample $\{X_1, X_2, \dots, X_n\}$

ie $\hat{\theta} : \mathcal{X}^n \longrightarrow \mathbb{R}$, $\mathcal{X}$:= values of X_i's.

We typically expect the values of $\hat{\theta}$ to be a "sensible value" of a certain popl'n characteristic $\longrightarrow$ "population parameter"

denoted by θ.

Note:

① The population parameter θ is fixed once we fix the population.

② If we have population data (ie a census)

can calculate the exact value of θ.

③ Usually the population is intractable.

need to resort to sample data to get an estimate for θ.

④ Suppose θ is a parameter of interest.

Given an estimator $\hat{\theta}$ for θ.

$\hat{\theta}$ is a statistic depending on the random sample $\{X_1, X_2, \dots, X_n\}$

the estimate $\hat{\theta}$ will change everytime sample data changes.

Q1) What do we expect the average value of $\hat{\theta}$ to be in relation to the true value θ?

ie the "Bias of $\hat{\theta}$"

If $\hat{\theta}$ is a point estimator for the popl'n param θ $\longrightarrow$ the "Bias of $\hat{\theta}$" is

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

"expected error" when we use $\hat{\theta}$ to estimate θ using a random sample of size n.

We say $\hat{\theta}$ is an "unbiased estimator" for θ $\xrightarrow{\text{if}}$ $\text{Bias}(\hat{\theta}) = 0$ for all possible choices of θ.

ie $E(\hat{\theta} - \theta) = 0$ for every possible θ.

⑤ We can have multiple estimators for a given parameter θ.

Question: Suppose we have estimators $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$, estimating θ.

Among all the $\hat{\theta}_i$'s, is there a notion of one estimator being "better" than the others?

Principle of unbiased Estimation

"Among several estimators for θ, choose one that is unbiased".

Examples:

① Population: $\{0, 1\}$

Population distribution $\longrightarrow$ Bernolli(p)

ie $P(X=1) = p$.

Let $\{X_1, X_2, \dots, X_n\}$ be a random sample

Recall:

$T_0 :=$ sample total

$= X_1 + X_2 + \dots + X_n$

$T_0 \sim \text{Bin}(n, p)$ $\longrightarrow$ sampling dist of the sample total is binomial with params: n, p.

Define:

$\hat{p} := \dfrac{T_0}{n}$ = the "sample proportion" of successes in a sample of size n.

$\hat{p}$ is an estimator for p (the true probability of getting a success in a single Bernolli trial).

Now,

$$E(\hat{p}) = E\left(\frac{T_0}{n}\right) = \frac{E(T_0)}{n} = \frac{n \cdot p}{n} = p$$

ie $E(\hat{p}) = p$ for any choice of p.

$\hat{p}$ is an "unbiased estimator" for p.

② Suppose $\{X_1, X_2, \dots, X_n\}$ is a random sample from a population with mean $= \mu$ and variance $= \sigma^2$.

a) Define $\hat{\theta}_1 = \overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$ $\longrightarrow$ sample mean

$\hat{\theta}_2 := \overline{X} + X_1 - X_n$ $\longrightarrow$ sample + 1st obs. − last obs.

Then,

$$E(\hat{\theta}_1) = E(\overline{X}) = \mu$$

$$E(\hat{\theta}_2) = E(\overline{X} + X_1 - X_n) = \mu + \mu - \mu$$

$$= \mu.$$

$\therefore$ $\hat{\theta}_1, \hat{\theta}_2$ are both unbiased estimators for the population mean μ.

unbiased estimation does not gaurantee a unique choice of estimator.

$\therefore$ We need additional criteria to rank $\hat{\theta}_1$ and $\hat{\theta}_2$.

"study the variance of an estimator"

Suppose $\hat{\theta}$ is an unbiased estimator for θ $\longrightarrow$ $\hat{\theta}$ is a statistic for a random sample $\{X_1, X_2, \dots, X_n\}$ such that

$E(\hat{\theta}) = \theta$ for every possible value of θ.

Then,

$\sigma^2_{\hat{\theta}} :=$ Variance of the sampling distribution of $\hat{\theta}$.

The principle of Minimum Variance Unbiased Estimation

"Among all unbiased estimators of θ, choose the one that has the minimum variance"

the resulting estimator is called the "Minimum Variance Unbiased Estimator (MVUE)" of θ.

Example 2) We know that $\hat{\theta}_1 = \overline{X}$ and $\hat{\theta}_2 = \overline{X} + X_1 - X_n$ are both unbiased estimators of μ.

To rank them, we calculate $V(\hat{\theta}_1)$ and $V(\hat{\theta}_2)$.

$$V(\hat{\theta}_1) = V\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right)$$

$$= V\left(\frac{X_1}{n} + \frac{X_2}{n} + \dots + \frac{X_n}{n}\right)$$

$$= \frac{1}{n^2}\left(\sigma^2 + \sigma^2 + \dots + \sigma^2\right) = \frac{n\sigma^2}{n^2}$$

$$\boxed{V(\hat{\theta}_1) = \frac{\sigma^2}{n}}$$

Similarly, can show that

$$V(\hat{\theta}_2) = V(\overline{X} + X_1 - X_n) = \frac{\sigma^2}{n} + 2\sigma^2$$

$$= \boxed{\frac{(2n+1)}{n}\sigma^2}$$

$\because \dfrac{(2n+1)}{n} > \dfrac{1}{n} \;\; \forall n \implies V(\hat{\theta}_2) > V(\hat{\theta}_1)$

If we were to choose only betwn $\hat{\theta}_1$ and $\hat{\theta}_2$, we would go with $\hat{\theta}_1 = \overline{X}$.

Note: This does not prove that $\overline{X}$ is the MVUE!

Thm: If $\{X_1, X_2, \dots, X_n\}$ is a random sample from a normally dist popl'n, ie $N(\mu, \sigma^2)$.

Then the estimator $\overline{X}$ is the MVUE for μ.

The "standard error" of a point estimator $\hat{\theta}$ $\longrightarrow$ $\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$

provides a "measure of precision" of the point estimator $\hat{\theta}$.

the standard deviation of the sampling distribution of $\hat{\theta}$.

Note: If $\hat{\theta}$ has a continuous distribution,

$P(\hat{\theta} = \theta) = P$(the estimator $\hat{\theta}$ takes the value θ, the true parameter value)

$= 0.$

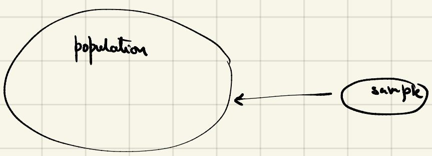
Even so the point-estimator provides an "exact" estimate for θ, we have zero confidence that the calculated point estimate will equal θ!!!

Point Estimates

$\{1, 2, 1, 1, 6, 3, 2, 7, 1, 1, 2, 2, 5, 1, 1, 2\}$

need statistics to summarize the data.

max, min, mean, median, quartiles.



Population: Students in this class. (~ 270).

characteristic of interest: height. $\rightsquigarrow X$

To get a picture of the heights that involves much less info than 270 numbers $\rightsquigarrow$ $P(5.5 \leq X \leq 6.0) = ?$

Make a guess about how the hieghts might be distributed.

Can use a normal distribution to approximate the true dist.

Need to specify: mean : μ

variance : σ^2.

We will assume $X \sim N(\mu, \sigma^2)$

to actually calculate P(height of a random selected student is between 5'5 and 6')

$=$

$P(5'5'' \leq X \leq 6'')$?

to be able to calculate this we need to know μ, σ^2 $\rightsquigarrow$ typically unknown.

finding "reasonable" estimates for μ and σ^2

Note: μ and σ^2 $\rightsquigarrow$ "population parameters"

fixed once we fix population.

A "point estimator" $\hat{\theta}$ $\rightsquigarrow$ statistic that depends on a random sample.

typically we want the estimator to estimate some population parameter of interest.

Example: Heights example $\rightsquigarrow$ want to estimate μ and σ^2.

Want to discuss "principles" that identify if an estimator is "good"

Suppose we want to estimate μ,

important considerations (i) expectation that your estimate will take the value of the param

(ii) want a measure of accuracy for each prediction that you make.

Bias of an estimator

Suppose $\hat{\mu}$ is an estimator for μ.

given sample data $\{x_1, x_2, \ldots, x_n\}$

$\hat{\mu}(x_1, \ldots, x_n) \longrightarrow$ point estimate for μ.

$\mu - \hat{\mu}(x_1, x_2, \ldots, x_n) =$ deviation/error in your estimate.

Want: Avg error in our estimates to be zero.

expected value.

"Expected error in our estimates must be zero!"

Suppose $\hat{\theta}$ is an estimator for θ.

$$\text{Bias}(\hat{\theta}) = \underbrace{E(\hat{\theta}-\theta)}_{\substack{\text{expected deviation/error of the} \\ \text{estimator from the true value } \theta.}}$$

We say that $\hat{\theta}$ is a unbiased estimator for θ $\rightsquigarrow$ if $E(\hat{\theta}-\theta)=0$ for all possible choices of θ.

"Principle of unbiased Estimation"

$\Downarrow$

"If choosing from several estimators for θ select an estimator that is unbiased"

Suppose population : $\{0,1\}$

$\text{Ber}(p)$ $\quad p =$ prob of a success.

random sample from $\{0,1\}$ $= \{X_1, X_2, X_3, \ldots, X_n\}$

$X_i \sim \text{Ber}(p)$

T_0 = sample total

$= X_1 + X_2 + \cdots + X_n$

= # of success in a sample of size n

Goal: get an estimate for p.

if we set $\hat{p} = \dfrac{\text{\# success in the sample}}{\text{sample}}$

$= \dfrac{T_0}{n}$.

$\hat{p} = \dfrac{T_0}{n}$ $\rightsquigarrow$ estimator for p. (also the sample mean)

is $\hat{p}$ unbiased ?

$$E(\hat{p}) = E\left(\frac{T_0}{n}\right) = \frac{1}{n}E(T_0)$$

$$= \frac{1}{n}E(X_1 + X_2 + \cdots + X_n)$$

$$= \frac{1}{n}\cdot(p + p + \cdots + p)$$

$$= \frac{1}{n}\cdot np = p$$

$E(\hat{p}) = p$ $\rightsquigarrow$ true for all possible choices of p.

$\Rightarrow$ $\hat{p}$ is an unbiased estimator

p = true prob of getting a success.

Suppose $\{(x_1,y_1), (x_2,y_2), \ldots, (x_n,y_n)\}$