Tema 6. MOSTREIG I ESTIMACIÓ DE PARÀMETRES POBLACIONALS

Població i mostres

Una **població** és el conjunt total d'individus damunt el qual es vol fer un estudi estadístic.

Una **mostra** és el subconjunt de la població damunt el qual es fa l'estudi.

Mostreig: procés de selecció d'una mostra a partir d'una població

Per a cada element d'una mostra s'obté una dada estadística X_i.

El valor d'aquesta dada es pot considerar una **variable aleatòria**.

Una mostra de tamany n es diu **Mostra Aleatòriament Simple** (**M.A.S.**) si les variables aleatòries X_i es poden considerar independents i idènticament distribuïdes (i.i.d.).

Existeixen diferents **tècniques de mostreig** per a obtenir una M.A.S. representativa d'una població.

Paràmetres mostrals i poblacionals

Els paràmetres estadístics (mitjana, variància, etc.) obtinguts a partir d'una mostra es diuen **paràmetres mostrals**.

Els paràmetres obtinguts a partir d'una població es diuen **paràmetres poblacionals**.

Paràmetres mostrals habituals

Mitjana mostral
$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Variància mostral
$$\hat{s}_X^2 = \frac{n}{n-1}\left(\frac{\sum_{i=1}^{n} X_i^2}{n} - \bar{X}^2\right)$$

Desviació típica mostral
$$\hat{s}_X = \sqrt{\hat{s}_X^2}$$

Proporció mostral
$$\hat{p}_X = \frac{CF}{n}$$

On:

- n : nombre de valors mostrals
- X_i : ièsim valor mostral
- CF : casos favorables al succés que volem mesurar

Paràmetres mostrals habituals

Exemple: mostra de 10 individus, mesuram l'alçada de cada persona

Persona	Alçada (cm)
1	173
2	175
3	175
4	176
5	172
6	173
7	175
8	170
9	176
10	175

$$\bar{X}=\frac{173+175+175+176+172+173+175+170+176+175}{10}=174$$

$$\hat{s}_X^2=\frac{10}{9}\left(\frac{173^2+175^2+175^2+176^2+172^2+173^2+175^2+170^2+176^2+175^2}{10}-174^2\right)=3{,}78$$

$$\hat{s}_X=\sqrt{3{,}78}=1{,}94$$

Succés: més alt de 175 cm $\hat{p}_X=\frac{2}{10}=0{,}2$

Paràmetres mostrals i variables aleatòries

Els paràmetres mostrals són variables aleatòries

Paràmetre mostral (estadístic)	Esperança	Variància	Distribució de probabilitat	
$\bar{X}$	$E(\bar{X}) = \mu$	$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$	$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$	població normal, σ conegut
			$\frac{\bar{X}-\mu}{\hat{s}_X/\sqrt{n}} \sim t_{n-1}$	població normal, σ desconegut, $n \leq 30$
			$\bar{X} \sim N(\mu, \frac{\hat{s}^2_X}{n})$	σ desconegut, $n > 30$
$\hat{s}^2_X$	$E(\hat{s}^2_X) = \sigma^2$	$\mathrm{Var}(\hat{s}^2_X) = \frac{2\sigma^4}{n-1}$	$\frac{n-1}{\sigma^2}\hat{s}^2_X \sim \chi^2_{n-1}$	població normal
$\hat{p}_X$	$E(\hat{p}_X) = p$	$\mathrm{Var}(\hat{p}_X) = \frac{p(1-p)}{n}$	$\hat{p}_X \sim N(p, \frac{p(1-p)}{n})$	$n > 30$
			$\hat{p}_X \sim t_{n-1}$	població normal, $n \leq 30$

Altres distribucions mostrals:

Paràmetre mostral	Distribució de probabilitat	
Diferència de mitjanes	$\bar{X}-\bar{Y} \sim N(\mu_X-\mu_Y, \frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m})$	si poblacions normals i variàncies conegudes
	$\frac{\frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{\sqrt{\frac{1}{n}+\frac{1}{m}}}}{\sqrt{\frac{n\hat{s}_X^2+m\hat{s}_Y^2}{n+m-2}}} \sim t_{n+m-2}$	si poblacions normals i variàncies desconegudes però iguals i $n \leq 30$
	$\bar{X}-\bar{Y} \sim N(\mu_X-\mu_Y, (\frac{n\hat{s}_X^2+m\hat{s}_Y^2}{n+m-2})(\frac{1}{n}+\frac{1}{m}))$	si poblacions normals i variàncies desconegudes però iguals i $n > 30$
	$\bar{X}-\bar{Y} \sim N(\mu_X-\mu_Y, \frac{\hat{s}_X^2}{n}+\frac{\hat{s}_Y^2}{m})$	si poblacions normals i variàncies desconegudes i diferents
Diferència de proporcions	$\hat{p}_X-\hat{p}_Y \sim N(p_X-p_Y, \frac{\hat{p}_X(1-\hat{p}_X)}{n}+\frac{\hat{p}_Y(1-\hat{p}_Y)}{m})$	n gran

Estimadors

Els paràmetres mostrals s'anomenen **estimadors** quan permeten *estimar* els valors poblacionals.

Notació: θ (valor que es vol estimar)

$\hat{\theta}$ (valor de l'estimador)

Per a un estimador ideal es verifica:

(i) la mitjana dels valors de l'estimador és igual al valor que es vol estimar (**estimador insesgat**)

$$E(\hat{\theta}) = \theta$$

(ii) la variància dels valors de l'estimador és petita

Estimadors

Mostra		**Població**
Mitjana mostral ($\bar{X}$)	estimador insesgat de	Esperança (μ)
Variància mostral ($\hat{s}_X^2$)	estimador insesgat de	Variància (σ^2)
Proporció mostral ($\hat{p}_X$)	estimador insesgat de	probabilitat (p)

Intervals de confiança

Donen el rang de valors més probable d'un paràmetre poblacional

Es donen sempre en funció de:

(i) el **nivell de confiança** del paràmetre (probabilitat que el valor poblacional estigui en l'interval)

o bé

(ii) el **nivell de significació** (α) o error permès (probabilitat que el valor poblacional estigui *fora* de l'interval)

Intervals de confiança més habituals

Paràmetre mostral	Interval de confiança	
Mitjana	$\bar{X} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$	si la població segueix una llei normal i σ és conegut
	$\bar{X} \pm t_{n-1,\alpha/2}\dfrac{\hat{s}_X}{\sqrt{n}}$	si la població segueix una llei normal, σ no és conegut i $n \leq 30$
	$\bar{X} \pm z_{\alpha/2}\dfrac{\hat{s}_X}{\sqrt{n}}$	si $n > 30$
Variància	$\left[\dfrac{n-1}{\chi^2_{n-1,1-\alpha/2}}\hat{s}^2_X, \dfrac{n-1}{\chi^2_{n-1,\alpha/2}}\hat{s}^2_X\right]$	si la població segueix una llei normal
Proporció	$\hat{p}_X \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}_X(1-\hat{p}_X)}{n}}$	si $n > 30$

Altres intervals de confiança:

Paràmetre mostral	Interval de confiança	
Diferència de mitjanes	$\bar{X}-\bar{Y}\pm z_{\alpha/2}\sqrt{\dfrac{\sigma_X^2}{n}+\dfrac{\sigma_Y^2}{m}}$	si poblacions normals i variàncies conegudes
	$\bar{X}-\bar{Y}\pm t_{n+m-2,\alpha/2}\sqrt{\dfrac{n\hat{s}_X^2+m\hat{s}_Y^2}{n+m-2}}\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}$	si poblacions normals i variàncies desconegudes però iguals i $n \leq 30$
	$\bar{X}-\bar{Y}\pm z_{\alpha/2}\sqrt{\dfrac{n\hat{s}_X^2+m\hat{s}_Y^2}{n+m-2}}\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}$	si poblacions normals i variàncies desconegudes però iguals i $n > 30$
	$\bar{X}-\bar{Y}\pm z_{\alpha/2}\sqrt{\dfrac{\hat{s}_X^2}{n}+\dfrac{\hat{s}_Y^2}{m}}$	si poblacions normals i variàncies desconegudes i diferents
Diferència de proporcions	$\hat{p}_X-\hat{p}_Y\pm z_{\alpha/2}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}+\frac{\hat{p}_Y(1-\hat{p}_Y)}{m}}$	n gran