Estimación puntual

Estadística inferencial

El problema usual de la estadística inferencial es:

- Queremos conocer el valor de una característica en una población
- No podemos medir esta característica en todos los individuos de la población
- Extraemos una muestra aleatoria de la población, medimos la característica en los individuos de esta muestra e inferimos el valor de la característica para la toda la población
  - ¿Cómo lo tenemos que hacer?
  - ¿Cómo tenemos que hacer la muestra?
  - ¿Qué información podemos inferir?

Ejemplos

Set de cada deu estudiants de la UIB practica el ciberplagi a l'hora de confeccionar els treballs acadèmics

<u>Fixa tècnica de la mostra de la UIB</u>

Univers: alumnat de primer i segon cicle de la UIB (N = 11.797 estudiants)
Punts de mostreig: 38 unitats/aules (una per cada estudi oficial)
Mostreig: mixt i polietàpic, estratificat per centres amb selecció de les unitats primàries (assignatures) de forma aleatòria amb afixació proporcional i de les unitats secundàries (alumnes) mitjançant mostreig incidental a l'aula.
Mostra: 727 unitats d'anàlisi (qüestionaris), amb un error per al conjunt de la mostra del 3,52 per cent estimat per a un nivell de confiança del 95 per cent i sota la condició més desfavorable de p = q = 0.05.

Ejemplos

Evaluación de la efectividad de una nueva vacuna contra la leptospirosis humana en grupos en riesgo

RESUMEN *Se realizó un estudio de cohorte prospectivo cuasi experimental que incluyó a los grupos en riesgo de enfermar de leptospirosis en la provincia de Holguín para evaluar la efectividad de la vacuna contra la leptospirosis humana. Se incluyeron 118 018 personas de 15 a 65 años que presentaban un riesgo permanente o temporal de contraer la enfermedad; de estas, 101 137 fueron inmunizadas con dos dosis de 0,5 mL por vía intramuscular profunda en el músculo deltoides del brazo no dominante, con un intervalo de 6 semanas, constituyendo la cohorte de vacunados, mientras que 16 881 personas no inmunizadas pasaron a integrar la cohorte de no vacunados. A los 21 días de aplicada la segunda dosis, el universo de estudio (previamente censado en un registro de modelo) fue seguido por el sistema local de vigilancia epidemiológica con el objetivo de detectar la enfermedad. El criterio de caso sospechoso y confirmado se conservó*

`http://www.scielosp.org/pdf/rpsp/v8n6/3956.pdf`

Definiciones básicas

Muestra aleatoria simple (m.a.s.) de tamaño n: De una población de N individuos, repetimos n veces el proceso consistente en escoger equiprobablemente un individuo de la población; *los individuos escogidos se pueden repetir*

Ejemplo: Escogemos al azar n estudiantes de la UIB (con reposición) para medirles la estatura

De esta manera, todas las muestras posibles de n individuos (posiblemente repetidos: multiconjuntos) tienen la misma probabilidad

Leeros la lección 2 de AprendeR2 sobre muestreo

Definiciones básicas

Estadístico (Estimador puntual): Una función que aplicada a una muestra nos permite estimar un valor que queramos conocer sobre toda la población

Ejemplo: La media de les estaturas de una muestra de estudiantes de la UIB nos permite estimar la media de les alturas de todos los estudiantes de la UIB

Formalmente

Una m.a.s. de tamaño n (de una v.a. X) es

- un conjunto de n copias independientes de X, o
- un conjunto de n variables aleatorias independientes $X_1, \ldots, X_n$, todas con la distribución de X

Ejemplo: Sea X la v.a. "escogemos un estudiante de la UIB y le medimos la altura". Una m.a.s. de X de tamaño n serán n copias independientes $X_1, \ldots, X_n$ de esta X.

Una realización de una m.a.s. son los n valores $x_1, \ldots, x_n$ que toman las v.a. $X_1, \ldots, X_n$

Formalmente

Un estadístico T es una función aplicada a la muestra $X_1, \ldots, X_n$:

$$T = f(X_1, \ldots, X_n)$$

Este estadístico se aplica a les realizaciones de la muestra

Ejemplo: La media muestral de una m.a.s. $X_1, \ldots, X_n$ de tamaño n es

$$\overline{X} := \frac{X_1 + \cdots + X_n}{n}$$

Estima $E(X)$

Ejemplo: La media muestral de las alturas de una realización de una m.a.s. de las alturas de estudiantes estima la altura media de un estudiante de la UIB.

Formalmente

Un estadístico T es una función aplicada a la muestra $X_1, \ldots, X_n$

$$T = f(X_1, \ldots, X_n)$$

Así pues, un estadístico es una (otra) variable aleatoria, con distribución, esperanza, etc.

La distribución muestral de T es la distribución de esta variable aleatoria.

Estudiando esta distribución muestral, podremos estimar propiedades de X a partir del comportamiento de una muestra

Error estándar de T: desviación típica de T

La vida real

En la vida real, las muestras aleatorias se toman, casi siempre, sin reposición (es decir sin repetición del mismo individuo de la población).
No son muestras aleatorias simples. pero:

- Si N es mucho más grande que n, los resultados para una m.a.s. son (aproximadamente) los mismos, ya que las repeticiones son improbables y las variables aleatorias que forman la muestra son prácticamente independientes.

  En estos casos cometeremos el abuso de lenguaje de decir que es una m.a.s.

- Si n es relativamente grande, se suelen dar versiones corregidas de los estadísticos

Convenio

LOS ESTADÍSTICOS, EN MAYÚSCULAS; las realizaciones, en minúsculas

Ejemplo:

- $X_1, \ldots, X_n$ una m.a.s. y

  $$\overline{X} := \frac{X_1 + \cdots + X_n}{n}$$

  la media muestral

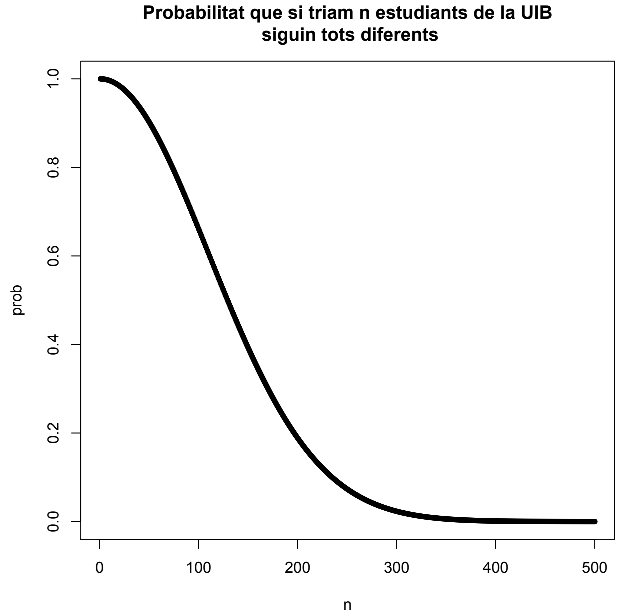
- $x_1, \ldots, x_n$ una realización de esta m.a.s. y

  $$\overline{x} := \frac{x_1 + \cdots + x_n}{n}$$

  la media (muestral) de esta realización

La vida real

Ejemplo: La UIB tiene unos 15000 estudiantes



media muestral

Sea $X_1, \ldots, X_n$ una m.a.s. de tamaño n de una v.a. X de esperanza μ_X y desviación típica σ_X

La media muestral es

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n}$$

Teorema

En estas condiciones

$$E(\overline{X}) = \mu_X, \quad \sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}$$

$\sigma_{\overline{X}}$ es el error estándar de $\overline{X}$

Media muestral

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n}$$

- es un estimador puntual de μ_X
- $E(\overline{X}) = \mu_X$
  - El valor esperado de $\overline{X}$ es μ_X
  - Si tomamos muchas veces una m.a.s. y calculamos la media muestral, el valor medio de estas medias tiende con mucha probabilidad a ser μ_X
- $\sigma_{\overline{X}} = \sigma_X/\sqrt{n}$: la variabilidad de los resultados de $\overline{X}$ tiende a 0 a medida que tomamos muestras más grandes

Media muestral

```r
# tests.txt=notas de un examen  sobre estadística descriptiva
notas=read.table("http://bioinfo.uib.es/~recerca/matIII_GIN/notas.txt", header=TRUE)
str(notas)


## 'data.frame': 185 obs. of  1 variable:
##  $ x: num  54.5 60.1 53 57.3 54.4 ...


mean(notas$x)


## [1] 55.32254


set.seed(100)
medias=replicate(10^4,mean(sample(notas$x,40,rep=TRUE)))
head(medias)


## [1] 55.86225 55.90700 54.74575 56.39950 56.02750 54.99475


mean(medias)


## [1] 55.33279


#sd, por ir  deprisa
c(sd(notas$x)/sqrt(40),sd(medias))


## [1] 0.5209303 0.5156658
```

Ejemplo

Se toma una m.a.s. de 10 estudiantes de los estudios del grado de informática (GIN), y se miden sus estaturas. Se obtuvieron estos resultados:

$$1.62, 1.75, 1.64, 1.69, 1.83, 1.85, 1.72, 1.61, 1.93, 1.62$$

Podemos estimar la estatura media de los estudiantes del GIN:

$$\overline{x} = \frac{1.62 + 1.75 + 1.64 + \cdots + 1.62}{10} = 1.726$$

¿Cuál es la precisión de esta estimación?

“.....The solution is comming!!!!"¡No os perdáis las próximas lecciones!

La combinación lineal de normales es normal

Teorema

Si $Y_1, \dots, Y_n$ son v.a. normales independientes, cada $Y_i \sim N(\mu_i, \sigma_i)$, y $a_1, \dots, a_n, b \in \mathbb{R}$ entonces

$$Y = a_1 Y_1 + \cdots + a_n Y_n + b$$

es una v.a. $N(\mu, \sigma)$ con μ y σ las que correspondan:

- $E(Y) = a_1 \cdot \mu_1 + \cdots + a_n \cdot \mu_n + b$
- $\sigma(Y)^2 = a_1^2 \cdot \sigma_1^2 + \cdots + a_n^2 \cdot \sigma_n^2$

Caso en el que X tiene distribución normal

Teorema

Sea $X_1, \dots, X_n$ una m.a.s. de una v.a. X de esperanza μ_X y desviación típica σ_X. Si X es $N(\mu_X, \sigma_X)$, entonces

$$\overline{X} \text{ es } N\Big(\mu_X, \frac{\sigma_X}{\sqrt{n}}\Big)$$

y por lo tanto

$$Z = \frac{\overline{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} \text{ es } N(0,1)$$

Z es la expresión tipificada de la media muestral

Teorema Central del Límite

Teorema

Sea $X_1, \dots, X_n$ una m.a.s. de una v.a. X cualquier de esperanza μ_X y desviación típica σ_X. Cuando $n \to \infty$,

$$\overline{X} \to N\Big(\mu_X, \frac{\sigma_X}{\sqrt{n}}\Big)$$

y por lo tanto

$$Z = \frac{\overline{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} \to N(0,1)$$

(estas convergencias se refieren a las distribuciones.)

Teorema Central del Límite

"Teorema"

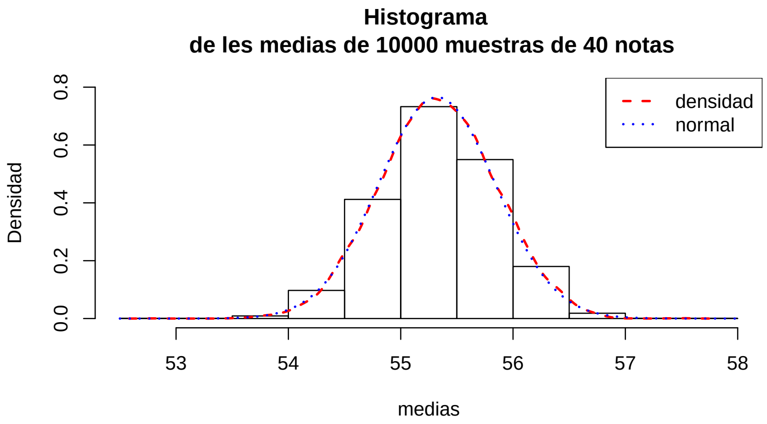
Si n es grande ($n \geqslant 30$ o **40**), $\overline{X}$ es aproximadamente normal, con esperanza μ_X y desviación típica $\dfrac{\sigma_X}{\sqrt{n}}$

Ejemplo: Tenemos una v.a. X de media $\mu_X = 3$ y desv. típ. $\sigma_X = 0.2$. Tomamos muestras aleatorias simples de tamaño 50. La distribución de la media muestral $\overline{X}$ es aproximadamente

$$N\left(3, \frac{0.2}{\sqrt{50}}\right) = N(3, 0.0283)$$

Teorema Central del Límite

```r
# con los datos de las  10000 medias de muestras
#de tamaño 40  las notas anteriores
hist(medias,freq=FALSE, main="Histograma
 de les medias de 10000 muestras de 40 notas",ylab="Densidad",ylim=c(0,0.8))
lines(density(medias),lty=2,lwd=2,col="red")
curve(dnorm(x,mean(notas$x),sd(notas$x)/sqrt(40)),
 lty=3,lwd=2,col="blue",add=TRUE)
legend("topright",legend=c("densidad","normal"),
 lwd=c(2,2),lty=c(2,3),col=c("red","blue"))
```



Ejemplo

El tamaño en megabyte (MB) de un tipo de imágenes comprimidas tiene un valor medio de 115 MB, con una desviación típica de 25. Tomamos una m.a.s. de 100 imágenes de este tipo.

¿Cuál es la probabilidad de que la media muestral del tamaño de los ficheros sea $\leqslant 110$ MB?

$$Z = \frac{\overline{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} = \frac{\overline{X} - 115}{2.5} \text{ es (aproximadamente) } N(0,1)$$

$$\begin{aligned} P(\overline{X} \leqslant 110) &= P\left(Z \leqslant \frac{110-115}{2.5}\right) = P(Z \leqslant -2) \\ &= 0.0228 \end{aligned}$$

Ejemplo

El tamaño en megabyte (MB) de un tipo de imágenes comprimidas tiene un valor medio de 115 MB, con una desviación típica de 25. Tomamos una m.a.s. de 100 imágenes de este tipo.

¿Cuál es la probabilidad de que la media muestral del tamaño de las imágenes esté entre 113 MB y 117 MB?

Media muestral en muestras sin reposición

Sea $X_1, \ldots, X_n$ una m.a. sin reposición de tamaño n de una v.a. X de esperanza μ_X y desviación típica σ_X.

Si n es pequeño en relación al tamaño N de la población, todo lo que hemos contado funciona (aproximadamente)

Si n es grande en relación a N, entonces

$$E(\overline{X}) = \mu_X, \quad \sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

(factor de población finita)

El T.C.L. ya no funciona exactamente en este último caso

Proporción muestral

Sea X una v.a. Bernoulli de parámetro p_X (1 éxito, 0 fracaso). Sea $X_1, \ldots, X_n$ una m.a.s. de tamaño n de X.

$S = \sum_{i=1}^{n} X_i$ es el nombre de éxitos observados es $B(n, p)$.

La proporción muestral es

$$\widehat{p}_X = \frac{S}{n}$$

y es un estimador de p_X

Notemos que $\widehat{p}_X$ es un caso particular de $\overline{X}$, por lo que todo lo que hemos dicho para medias muestrales es cierto para proporciones muestrales.

Proporción muestral

$$\widehat{p}_X = \frac{S}{n}$$

- $E(\widehat{p}_X) = p_X$
- $\sigma_{\widehat{p}_X} = \sqrt{\frac{p_X(1-p_X)}{n}}$, l'error estándar de la proporción muestral
- Si la muestra es sin reposición y n es relativamente grande, $\sigma_{\widehat{p}_X} = \sqrt{\frac{p_X(1-p_X)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

Proporción muestral

Por el T.C.L.:

"Teorema"

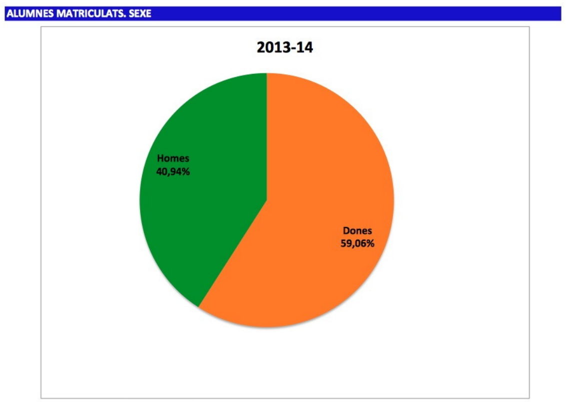
Si n es grande ($n \geqslant 30$ o **40**) y la muestra es aleatoria simple,

$$\frac{\widehat{p}_X - p_X}{\sqrt{\frac{p_X(1-p_X)}{n}}} \approx N(0,1)$$

Ejemplo

En una muestra aleatoria de 60 estudiantes de la UIB del curso 2013-14, 37 son mujeres. Estimar la fracción de mujeres entre los estudiantes de la UIB

$$\frac{37}{60} = 0.6167$$



Ejemplo

Un 59% de los estudiantes de la UIB son mujeres. Si tomamos una m.a.s. de 60 estudiantes, ¿cuál es la probabilidad que la proporción muestral de mujeres sea superior al 61%?

Varianza muestral

Sea $X_1, \ldots, X_n$ una m.a.s. de tamaño n de una v.a. X de esperanza μ_X y desviación típica σ_X

La varianza muestral es

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

La desviación típica muestral es

$$\widetilde{S}_X = +\sqrt{\widetilde{S}_X^2}$$

A mes, escriurem

$$S_X^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n} = \frac{(n-1)}{n}\widetilde{S}_X^2 \quad \text{y} \quad S_X = +\sqrt{S_X^2}$$

Varianza muestral: Propiedades

- $S_X^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n} = \left(\frac{\sum_{i=1}^n X_i^2}{n} - \overline{X}^2\right)$
- $\widetilde{S}_X^2 = \frac{n}{n-1}\left(\frac{\sum_{i=1}^n X_i^2}{n} - \overline{X}^2\right)$

Teorema

Si la v.a. X es normal, entonces $E(\widetilde{S}_X^2) = \sigma_X^2$ y la v.a.

$$\frac{(n-1)\widetilde{S}_X^2}{\sigma_X^2}$$

té distribución χ_{n-1}^2

La distribución χ_n^2

La distribución χ_n^2 (χ: en catalán, khi; en castellano, ji; en inglés, chi), on n es un parámetro llamado grados de libertad:

- es la de
$$X = Z_1^2 + Z_2^2 + \cdots + Z_n^2$$
on $Z_1, Z_2, \ldots, Z_n$ son v.a. independientes $N(0,1)$
- Su función de densidad es:
$$f_{\chi_n^2}(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{(n/2)-1} e^{-x/2} \quad \text{si } x \geqslant 0$$
on $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$ si $x > 0$

La distribución χ^2_n

- La distribución está tabulada (Tenéis unas tablas en Aula Digital), y con R es `chisq`
- Si $X_{\chi^2_n}$ es una v.a. con distribución χ^2_n,

$$E(X_{\chi^2_n}) = n, \quad Var(X_{\chi^2_n}) = 2n$$

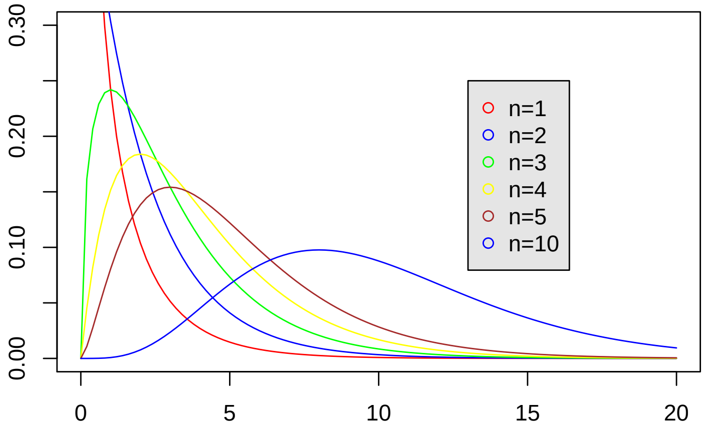
- χ^2_n se aproxima a una distribución normal $N(n, \sqrt{2n})$ para n grande ($n > 40$ o 50)

La distribución χ^2_n

$n\backslash\alpha$	0.995	0.99	0.975	0.95	0.9	0.85	0.8	0.75	0.7	0.65
1	7.879	6.635	5.024	3.841	2.706	2.072	1.642	1.323	1.074	0.873
2	10.597	9.21	7.378	5.991	4.605	3.794	3.219	2.773	2.408	2.1
3	12.838	11.345	9.348	7.815	6.251	5.317	4.642	4.108	3.665	3.283
4	14.86	13.277	11.143	9.488	7.779	6.745	5.989	5.385	4.878	4.438
5	16.75	15.086	12.833	11.07	9.236	8.115	7.289	6.626	6.064	5.573
6	18.548	16.812	14.449	12.592	10.645	9.446	8.558	7.841	7.231	6.695
7	20.278	18.475	16.013	14.067	12.017	10.748	9.803	9.037	8.383	7.806
8	21.955	20.09	17.535	15.507	13.362	12.027	11.03	10.219	9.524	8.909
9	23.589	21.666	19.023	16.919	14.684	13.288	12.242	11.389	10.656	10.006
10	25.188	23.209	20.483	18.307	15.987	14.534	13.442	12.549	11.781	11.097
11	26.757	24.725	21.92	19.675	17.275	15.767	14.631	13.701	12.899	12.184
12	28.3	26.217	23.337	21.026	18.549	16.989	15.812	14.845	14.011	13.266
13	29.819	27.688	24.736	22.362	19.812	18.202	16.985	15.984	15.119	14.345
14	31.319	29.141	26.119	23.685	21.064	19.406	18.151	17.117	16.222	15.421
15	32.801	30.578	27.488	24.996	22.307	20.603	19.311	18.245	17.322	16.494
16	34.267	32	28.845	26.296	23.542	21.793	20.465	19.369	18.418	17.565
17	35.718	33.409	30.191	27.587	24.769	22.977	21.615	20.489	19.511	18.633
18	37.156	34.805	31.526	28.869	25.989	24.155	22.76	21.605	20.601	19.699
19	38.582	36.191	32.852	30.144	27.204	25.329	23.9	22.718	21.689	20.764
20	39.997	37.566	34.17	31.41	28.412	26.498	25.038	23.828	22.775	21.826
21	41.401	38.932	35.479	32.671	29.615	27.662	26.171	24.935	23.858	22.888
22	42.796	40.289	36.781	33.924	30.813	28.822	27.301	26.039	24.939	23.947
23	44.181	41.638	38.076	35.172	32.007	29.979	28.429	27.141	26.018	25.006

$F_{\chi^2_{10}}(25.188) = 0.995$, $F_{\chi^2_{20}}(26.5) \approx 0.85$

La distribución χ^2_n



Función de densidad de χ^2_n para algún n

Ejemplo

Supongamos que el aumento diario del la ocupación de una granja de discos duros medido en Gigas sigue distribución normal con desviación típica 1.7. Se toma una muestras de 12 discos. Supongamos que esta muestra es pequeña respecto del total de la población de la granja de discos.

¿Cual es la probabilidad de que la desviación típica muestral sea $\leqslant 2.5$?

Sea X= aumento diario en Gigas de un disco duro elegido al azar. Sabemos que $\sigma_X^2 = (1.7)^2 = 2.89$. Como que X es normal y $n = 12$, tenemos que

$$\frac{11 \cdot \widetilde{S}_X^2}{2.89} = \frac{(n-1)\widetilde{S}_X^2}{\sigma_X^2} \sim \chi^2_{11}$$

Ejemplo

Supongamos que el aumento diario del la ocupación de una granja de discos duros medido en Gigas sigue distribución normal con desviación típica 1.7. Se toma una muestras de 12 discos. Supongamos que esta muestra es pequeña respecto del total de la población de la granja de discos.

¿Probabilidad de que la desviación típica muestral sea $\leqslant 2.5$?

$$\frac{11\widetilde{S}_X^2}{2.89} \sim \chi_{11}^2$$

$$P(\widetilde{S}_X < 2.5) = P\left(\widetilde{S}_X^2 < 2.5^2\right)$$
$$= P\left(\frac{11\cdot\widetilde{S}_X^2}{2.89} < \frac{11\cdot 2.5^2}{2.89}\right) = P(\chi_{11}^2 < 23.79)$$

Ejemplo

Supongamos que el aumento diario del la ocupación de una granja de discos duros medido en Gigas sigue distribución normal con desviación típica 1.7. Se toma una muestras de 12 discos. Supongamos que esta muestra es pequeña respecto del total de la población de la granja de discos.

¿Probabilidad de que la desviación típica muestral sea $\leqslant 2.5$?

$$\frac{11\widetilde{S}_X^2}{2.89} \sim \chi_{11}^2$$

$$P(\widetilde{S}_X < 2.5) = P\left(\widetilde{S}_X^2 < 2.5^2\right)$$
$$= P\left(\frac{11\cdot\widetilde{S}_X^2}{2.89} < \frac{11\cdot 2.5^2}{2.89}\right) = P(\chi_{11}^2 < 23.79)$$
$$\approx P(\chi_{11}^2 < 24.725) = 0.99$$

Ejemplo

Supongamos que el aumento diario del la ocupación de una granja de discos duros medido en Gigas sigue distribución normal con desviación típica 1.7. Se toma una muestras de 12 discos. Supongamos que esta muestra es pequeña respecto del total de la población de la granja de discos.

¿Probabilidad de que la desviación típica muestral sea $\leqslant 2.5$?

$$\frac{11\widetilde{S}_X^2}{2.89} \sim \chi_{11}^2$$

$$P(\widetilde{S}_X < 2.5) = P\left(\widetilde{S}_X^2 < 2.5^2\right)$$
$$= P\left(\frac{11\cdot\widetilde{S}_X^2}{2.89} < \frac{11\cdot 2.5^2}{2.89}\right) = P(\chi_{11}^2 < 23.79)$$
$$= \texttt{pchisq(23.7889,11)} = 0.986$$

Estimadores insesgado

¿Cuándo un estimador es bueno?

Un estimador puntual $\widehat{\theta}$ de un parámetro poblacional θ es insesgado, no sesgado o sin sesgo cuando su valor esperado es precisamente el valor del parámetro:

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

Entonces se dice que el estimador puntual es no sesgado.

El sesgo de $\widehat{\theta}$ es $E(\widehat{\theta}) - \theta$

Estimadores no segados

Ejemplos

- $E(\overline{X}) = \mu_X$: $\overline{X}$ es estimador no sesgado de μ_X
- $E(\widehat{p}_X) = p_X$: $\widehat{p}_X$ es estimador no sesgado de p_X
- $E(\widetilde{S}_X^2) = \sigma_X^2$ si X es normal: $\widetilde{S}_X^2$ es estimador no sesgado de σ_X^2 si X es normal
- $E(S_X^2) = \dfrac{n-1}{n}\sigma_X^2$ si X es normal; por lo tanto S_X^2 es sesgado, con sesgo

$$E(S_X^2) - \sigma_X^2 = \frac{n-1}{n}\sigma_X^2 - \sigma_X^2 = -\frac{\sigma_X^2}{n} \underset{n\to\infty}{\longrightarrow} 0$$

Estimadores eficientes

¿Cuando un estimador un estimador es bueno?

Cuando es no segado y tiene poca variabilidad (así es más probable que aplicado a una m.a.s. de un valor más cercano al valor esperado)

Error estándar de un estimador $\widehat{\theta}$: es su desviación típica

$$\sigma_{\widehat{\theta}} = \sqrt{Var(\widehat{\theta})}$$

Estimadores eficientes

Dados dos estimadores $\widehat{\theta}_1$, $\widehat{\theta}_2$ no sesgados (o con sesgo que $\underset{n\to\infty}{\longrightarrow} 0$) del mismo parámetro θ, diremos que

$$\widehat{\theta}_1 \text{ es más eficiente que } \widehat{\theta}_2$$

cuando

$$\sigma_{\widehat{\theta}_1} < \sigma_{\widehat{\theta}_2},$$

es decir, cuando

$$Var(\widehat{\theta}_1) < Var(\widehat{\theta}_2)$$

Estimadores eficientes

Ejemplo: Sea X una v.a. con media μ_X y desviación típica σ_X

Consideremos la mediana $Me = Q_{0.5}$ de la realización de una m.a.s. de X como estimador puntual de μ_X

Si X es normal,

$$E(Me) = \mu_X,$$
$$Var(Me) \approx \frac{\pi}{2}\frac{\sigma_X^2}{n} \approx \frac{1.57\sigma_X^2}{n} = 1.57 \cdot Var(\overline{X})$$

Por lo tanto, Me es un estimador no sesgado de μ_X, pero menos eficiente que $\overline{X}$

Estimadores eficientes

- Si la población es normal, la media muestral es el estimador no sesgado más eficiente de la media poblacional
- Si la población es Bernoulli, la proporción muestral es el estimador no sesgado más eficiente de la proporción poblacional
- Si la población es normal, la varianza muestral es el estimador no sesgado más eficiente de la varianza poblacional

Estimadores eficientes

Como hemos visto si la población es normal, la varianza muestral es el estimador no sesgado más eficiente de la varianza poblacional

El estimador "varianza"

$$S_X^2 = \frac{(n-1)}{n}\widetilde{S}_X^2$$

aunque sea más eficiente, tiene sesgo $\underset{n\to\infty}{\longrightarrow} 0$

Si n es pequeño ($\leqslant 30$ o 40), es mejor utilizar la varianza muestral $\widetilde{S}_X^2$ para estimar la varianza, ya que el sesgo influye, pero si n es grande, el sesgo ya no es tan importante y se puede utilizar S_X^2

Ejemplo: Estimación de poblaciones

Tenemos una población numerada $1, 2, \ldots, N$

Tomamos una m.a.s. $x_1, \ldots, x_n$; sea $m = \max(x_1, \ldots, x_n)$

Teorema

El estimador no segado más eficiente de N es

$$\widehat{N} = m + \frac{m-n}{n}$$

Un problema de relevancia histórica:

http://en.wikipedia.org/wiki/German_tank_problem

Ejemplo: Estimación de poblaciones

Ejemplo: Sentados en una terraza de un bar del Paseo Marítimo de Palma hemos anotado el número de licencia de los 40 primeros taxis que hemos visto pasar:

```
taxis=c(1217,600,883,1026,150,715,297,137,508,134,38,961,538,1154,
        314,1121,823,158,940,99,977,286,1006,1207,264,1183,1120,
        498,606,566,1239,860,114,701,381,836,561,494,858,187)
```

Supondremos que estas observaciones son una m.a.s. de los taxis de Palma. Entonces, estimamos que el número de taxis de Palma es

```
N=max(taxis)+(max(taxis)-length(taxis))/length(taxis)
N

## [1] 1268.975
```

En realidad, hay 1246 http://www.caib.es/eboibfront/es/2014/10195/551436/departamento-de-movilidad-seccion-de-transportes-r

Estimadores máximo verosímiles

¿Cómo encontramos buenos estimadores?

Sea X una v.a. discreta con función de probabilidad

$$f_X(x;\theta)$$

que depende de un parámetro desconocido θ

Sea $X_1,\ldots X_n$ una m.a.s. de X, y sea $x_1,x_2,\ldots,x_n$ una realización de esta muestra

La función de verosimilitud de la muestra es la probabilidad condicionada siguiente:

$$\begin{aligned} L(\theta|x_1,x_2,\ldots,x_n) &:= P(x_1,x_2,\ldots,x_n|\theta) \\ &= P(X_1=x_1)\cdots P(X_n=x_n) \\ &= f_X(x_1;\theta)\cdots f_X(x_n;\theta) \end{aligned}$$

Estimadores máximo verosímiles

Dada la función de verosimilitud $L(\theta|x_1,\ldots,x_n)$ de la muestra, indicaremos por

$$\hat{\theta}(x_1,\ldots,x_n)$$

el valor del parámetro θ en el que se alcanza el máximo de $L(\theta|x_1,\ldots,x_n)$. Será una función de $x_1,\ldots,x_n$.

Definición

Un estimador $\hat{\theta}$ de un parámetro θ es máximo verosímil (MV, en inglés EM) cuando,para cada m.a.s, la probabilidad de observarlo es máxima cuando el parámetro toma el valor del estimador aplicado a la muestra, es decir, si la función de verosimilitud

$$L(\theta|x_1,x_2,\ldots,x_n)=P(x_1,x_2,\ldots,x_n|\theta)$$

alcanza su máximo.

Estimadores máximo verosímiles

Ejemplo: Supongamos que tenemos una v.a. Bernoulli X de probabilidad de éxito p desconocida

Para cada m.a.s. $x_1,\ldots,x_n$ de X, sean $\widehat{p}_x$ su proporció muestral y $P(x_1,\ldots,x_n\mid p)$ la probabilidad de obtenerla cuando el verdadero valor del parámetro es p

Teorema

El valor de p para el que $P(x_1,\ldots,x_n\mid p)$ es máximo es $\widehat{p}_x$.

La proporción muestral es un estimador MV de p. Veámoslo.

Estimadores máximo verosímiles

Observación

En general, al ser ln una función decreciente, en lugar de maximizar $L(\theta|x_1,\ldots,x_n)$, maximizamos

$$\ln(L(\theta|x_1,\ldots,x_n))$$

que suele ser más simple (productos $\to$ sumas).

Estimadores máximo verosímiles

Sea $X_1, \ldots X_n$ una m.a.s. de una v.a. Bernoulli X de parámetro p (desconocido). Denotemos $q = 1 - p$

$$f_X(1; p) = P(X = 1) = p, \quad f_X(0; p) = P(X = 0) = q$$

es a decir, para $x \in \{0, 1\}$, resulta que

$$f_X(x; p) = P(X = x) = p^x q^{1-x}.$$

La funció de verosimilitud es:

$$\begin{aligned} L(p|x_1, \ldots, x_n) &= f_X(x_1; p) \cdots f_X(x_n; p) \\ &= p^{x_1} q^{1-x_1} \cdots p^{x_n} q^{1-x_n} \\ &= p^{\sum_{i=1}^n x_i} q^{\sum_{i=1}^n (1-x_i)} = p^{\sum_{i=1}^n x_i} q^{n-\sum_{i=1}^n x_i} \\ &= p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} \end{aligned}$$

Ejemplo

La función de verosimilitud es

$$\begin{aligned} L(p|x_1, \ldots, x_n) &= p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} \\ &= p^{n\overline{x}} (1-p)^{n-n\overline{x}} \end{aligned}$$

donde $\overline{x} = \dfrac{\sum_{i=1}^n x_i}{n}$

Queremos encontrar el valor de p en el que se alcanza el máximo de esta función (donde $\overline{x}$ es un parámetro: la variable es p)

Maximizaremos su logaritmo:

$$\ln(L(p|x_1, \ldots, x_n)) = n\overline{x} \ln(p) + n(1-\overline{x}) \ln(1-p)$$

Ejemplo

Derivamos respecto de p:

$$\begin{aligned} &\ln(L(p|x_1, \ldots, x_n))' \\ &\quad = n\overline{x}\frac{1}{p} - n(1-\overline{x})\frac{1}{1-p} \\ &\quad = \frac{1}{p(1-p)}\Big((1-p)n\overline{x} - pn(1-\overline{x})\Big) \\ &\quad = \frac{1}{p(1-p)}(n\overline{x} - pn) = \frac{n}{p(1-p)}(\overline{x} - p) \end{aligned}$$

Estudiamos el signo:

$$\begin{aligned} \ln(L(p|x_1, \ldots, x_n))' \geqslant 0 \quad &\Leftrightarrow \overline{x} - p \geqslant 0 \\ &\Leftrightarrow p \leqslant \overline{x} \end{aligned}$$

Ejemplo

Por lo tanto

$$\ln(L(p|x_1, \ldots, x_n)) \begin{cases} \text{creciente hasta } \overline{x} \\ \text{decreciente a partir de } \overline{x} \\ \color{red}{\text{tiene un máximo en } \overline{x}} \end{cases}$$

El resultado queda demostrado.
$L(\widehat{p}_X|x_1, \ldots, x_n) \geqslant L(p|x_1, \ldots, x_n)$ para cualquier p

Algunos estimadores MV

- $\widehat{p}_x$ es el estimador MV del parámetro p de una v.a. Bernoulli
- $\overline{X}$ es el estimador MV del parámetro θ de una v.a. Poisson
- $\overline{X}$ es el estimador MV del parámetro μ de una v.a. normal
- S_X^2 (no $\widetilde{S}_X^2$) es el estimador MV del parámetro σ^2 de una v.a. normal
- El máximo (no $\widehat{N}$) es el estimador MV de la N en el problema de los taxis

Ejemplo: θ para una Poisson

Ejercicio

Sea X una característica de una población que sigue una ley de distribución $Po(\theta)$, con $\theta > 0$ desconocido. Tomamos una muestra aleatoria simple $X_1, \ldots, X_n$ de aquesta población y obtenemos los resultados $x_1, \ldots, x_n$.

Encontremos el estimador máximo verosímil de θ para $x_1, \ldots, x_n$.

Ejemplo: Marca-recaptura

En una población hay N individuos, capturamos K, los marcamos y los volvemos a soltar. Ahora volvemos a capturar n, de los que k están marcados. A partir de estos datos, queremos estimar N.

Supongamos que N y K no han cambiado de la primera a la segunda captura

$X =$"Un individuo esté marcado" es $Be(p)$ con $p = \dfrac{K}{N}$

$X_1, \ldots, X_n$ la muestra capturada la segunda vez: $\widehat{p}_X = \frac{k}{n}$

Ejemplo: Marca-recaptura

$\widehat{p}_X$ es un estimador máximo verosímil p: estimamos que

$$\frac{K}{N} = \frac{k}{n} \Rightarrow N = \frac{n \cdot K}{k}$$

Por lo tanto, el estimador

$$\widehat{N} = \frac{n \cdot K}{k}$$

maximiza la probabilidad de la observación "k marcados de n capturados".

Por lo que $\hat{N}$ es el estimador máximo verosímil de N.

Ejemplo: Marca-recaptura

Supongamos que hemos marcado 15 peces del lago, y que en una captura de 10 peces, hay 4 marcados. ¿Cuántos peces estimamos que contiene el lago?

$$\widehat{N} = \frac{15 \cdot 10}{4} = 37.5$$

Estimamos que habrá entre 37 y 38 peces en el lago

Ejemplo: Marca-recaptura

El estimador

$$\widehat{N} = \frac{n \cdot K}{k}$$

es sesgado, con sesgo $\underset{n\to\infty}{\longrightarrow} 0$

El estimador de Chapman

$$\widehat{N} = \frac{(n+1)\cdot(K+1)}{k+1} - 1$$

es menos segado para muestras pequeñas, y no sesgado si $K + n \geqslant N$ (pero no máximo verosímil)

Ejemplo: Marca-recaptura

$$P(k \text{ marcados de } n \text{ capturados}) = \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}}$$

```
N=15:100
p=choose(15,4)*choose(N-15,6)/choose(N,10)
plot(N,p,type="h",xaxp=c(15,100,17))
```