Estadística

2º curso del Grado en Ciencias de la Actividad Física y el Deporte

---o0o---

Introducción a la Teoría de la Estimación

Bioestadística - Facultad de Medicina
Universidad de Granada (España)
http://www.ugr.es/~bioest

TEORÍA DE LA ESTIMACIÓN

Recordemos que la **Inferencia Estadística** aborda dos tipos de problemas distintos:

Inferencia Estadística:
- Teoría de la Estimación
  ¿Cúal es el tiempo medio que tarda en desparecer un fármaco analgésico de la sangre para no dar positivo en un control antidoping?
- Teoría de los Contrastes de Hipótesis
  ¿El entrenamiento tradicional (A) es igual de efectivo que el entrenamiento integrado (B)?

La **Teoría de la Estimación** es la parte de la Inferencia Estadística que sirve para determinar el valor de los parámetros poblacionales.

Distinguiremos dos formas de estimación:
1. Estimación puntual
2. Estimación por intervalo

No se trata de dos formas **alternativas** de estimación, sino **complementarias**. La estimación puntual representa el primer paso para obtener la estimación por intervalo, ¡que es la que siempre debemos dar!.

TEORÍA DE LA ESTIMACIÓN

ESTIMACIÓN PUNTUAL

Se trata de asignar al parámetro poblacional un (único) valor; que será un valor aproximado y que depende de la muestra

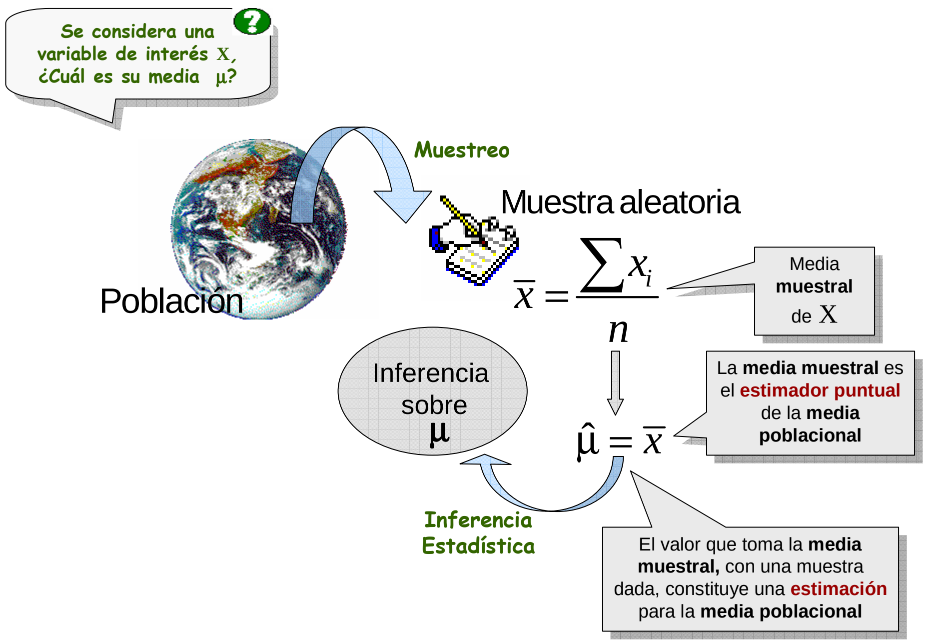
Concepto de Estimador

- Un **estimador (puntual)** para un parámetro es una función de los valores de la muestra cuyos valores son válidos como valores para ese parámetro.
- Notación: se dice, por ejemplo, que $\hat{\mu}$ es el estimador puntual de μ
- Cuando se haya obtenido la muestra y calculado $\hat{\mu}$ a partir de ella, se dice entonces que ese valor es una **estimación** de μ

En la práctica los estimadores habituales son (*método analógico*)**:**

Parámetros poblacionales de interés			Estimadores puntuales (insesgados)	
Media	μ	→	Media muestral	$\overline{x} = \hat{\mu}$
Desviación típica	σ	→	Desviación típica muestral	$s = \hat{\sigma}$
Proporción	p	→	Proporción muestral	$p = \hat{p}$

TEORÍA DE LA ESTIMACIÓN

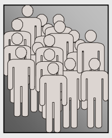


TEORÍA DE LA ESTIMACIÓN

Problema que presenta el uso de estimadores puntuales:

El problema de los estimadores puntuales es que **solo dan una idea de lo que puede valer el parámetro que estimamos**, **sin conocer como de buena es la aproximación**; es decir, simplemente proporcionan un valor (de los muchos posibles) que puede proponerse como valor del parámetro.

Por ejemplo:
Considere la población de estudiantes de la Universidad de Granada. Estamos interesados en investigar el valor del parámetro **μ=‘***peso medio de los estudiantes de la U. de Gr.*’. Para ello seleccionamos aleatoriamente dos muestras de 20 estudiantes cada una y obtenemos el valor del peso medio en cada caso. ¿coincidirán las dos medias muestrales? ¿cuál es mejor como estimador? ¿qué error se comete al asumir como valor de **μ** el ofrecido por una de estas medias?



$$\left.\begin{array}{l}\text{Muestra } M_1: \ \bar{x}_1 = 67Kg \\ \text{Muestra } M_2: \ \bar{x}_2 = 71Kg\end{array}\right\} \Rightarrow \mu?$$

TEORÍA DE LA ESTIMACIÓN

ESTIMACIÓN POR INTERVALOS DE CONFIANZA

Se trata de asignar al parámetro poblacional desconocido, por ejemplo **μ**, un intervalo de valores, digamos (a, b) entre los cuales está **μ** con una cierta **confianza** (1- **α). ** Es decir, si se cumple que

$$P(a \le \mu \le b) = 1 - \alpha$$

diremos entonces que (a, b) es un **intervalo de confianza** para el parámetro **μ** construido al **(1-α)% de confianza** o, lo que es lo mismo, al **α% de error.**

La **confianza de un intervalo** debe interpretarse en el sentido siguiente: Por cada 100 intervalos que construyamos para estimar un mismo parámetro (a partir de otras tantas muestras aleatorias y para un valor α prefijado) el (1-α)% de los intervalos obtenidos recogerán en su interior al verdadero valor del parámetro, mientras que el α% restante, por cosas del azar, pueden resultar ‘equivocados’. Si elegimos el nivel de confianza (1-α)% grande, por ejemplo del 95%, entonces nuestra esperanza al elaborar un intervalo a partir de una muestra es que sea uno de los 95 de cada 100 intervalos ‘acertados’, el ejemplo siguiente ilustra esto.

TEORÍA DE LA ESTIMACIÓN

Ejemplo:

Supongamos que se estudia el nivel de glucosa en sangre en la población cuyos valores aparecen en el recuadro sombreado (esta población es pequeñísima, pero nos vale como ejemplo). Seleccionamos **de forma aleatoria** 5 muestras de tamaño n=5 y elaboramos, en cada caso, el intervalo de confianza para el nivel medio de glucemia (consideraremos un nivel de confianza del 95%). Observemos los resultados:

Población

108	118
112	120
123	133
109	127
121	125
136	115
124	117
113	125
118	117
129	110

$\mu = 120$

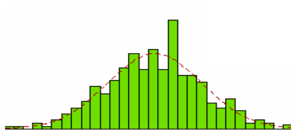
Muestra	Datos	$\bar{x}$	s	Intervalo (95% conf.)
M1	123 125 118 125 113	120.80	5.215	**(114.325 ; 127.275)**
M2	124 110 115 133 112	118.80	9.576	**(106.912 ; 130.688)**
M3	125 113 117 123 124	120.40	5.177	**(113.973 ; 126.827)**
M4	133 110 136 125 110	122.80	12.357	**(107.459 ; 138.141)**
M5	118 113 117 110 112	114.00	3.391	**(109.790 ; 118.210)** !

Observemos que:

- Las estimaciones puntuales varían de muestra a muestra, como ya sabemos
- Los 5 intervalos tienen diferente amplitud ¿por qué?
- Los cuatro primeros intervalos contienen al verdadero valor de la media (que, excepcionalmente, por conocer a toda la población sabemos que vale 120), sin embargo, en la 5ª muestra los valores obtenidos son, por azar, mas bajos de la cuenta y dan lugar a un intervalo que ¡no contiene a **μ**!

TEORÍA DE LA ESTIMACIÓN

Intervalos de confianza para la media de una variable normal



$X \rightarrow N(\mu ; \sigma)$

1. Si $x \rightarrow N\{\mu;\sigma\} \Rightarrow x \in \mu \pm 1{,}96\sigma$ el 95% de las veces (por cap. II)
2. Si $\bar{x} \rightarrow N\left\{\mu;\frac{\sigma}{\sqrt{n}}\right\} \Rightarrow \bar{x} \in \mu \pm 1{,}96\frac{\sigma}{\sqrt{n}}$ el 95% de las veces (por cap. II)
3. Operando: $\mu \in \bar{x} \pm 1{,}96\frac{\sigma}{\sqrt{n}}$ el 95% de las veces (resultado)

El problema es que σ suele ser desconocido

TEORÍA DE LA ESTIMACIÓN

Intervalos de confianza para la media de una variable normal

Dada una muestra con
- tamaño: n
- media: $\bar{x}$
- desviación típica: s

y fijado α de antemano. El intervalo de confianza para μ se calcula como:

$$\mu \in \bar{x} \pm t_{\alpha;n-1} \frac{s}{\sqrt{n}}$$

Este conjunto de valores cumple con la condición

$$P\left(\bar{x} - t_{\alpha;n-1} \frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha;n-1} \frac{s}{\sqrt{n}} \right) = 1 - \alpha$$

En donde $t_{\alpha;n-1}$ es el valor de la ***t de Student*** con n-1 grados de libertad (tabla 3).

TEORÍA DE LA ESTIMACIÓN

Tabla de la t de Student
Tabla 6

Grados de libertad

Error α

g.l \ α	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.001
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.929
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
35	0.682	0.852	1.052	1.306	1.690	2.030	2.438	2.724	3.592
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.705	3.551
45	0.680	0.850	1.049	1.301	1.679	2.014	2.412	2.690	3.521
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.497
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660	3.461
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.417
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.391
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

¿Cómo buscar un valor $t_{\alpha;n-1}$?

Supongamos una muestra de 10 observaciones:

$$n = 10 \quad \Rightarrow \quad n-1 = 9 = g.l.$$

Para una confianza del 95% :

$$(1-\alpha) = 0.95$$

Entonces $\alpha = 0.05$

Así que tenemos que

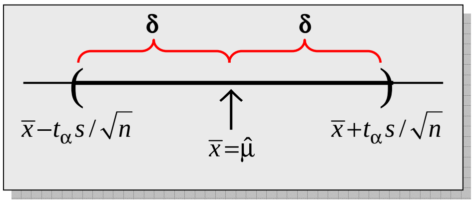
$$t_{0.05;9} = 2.262$$

TEORÍA DE LA ESTIMACIÓN

Fijar α implica, para un tamaño de muestra dado fijar $t_{\alpha;n-1}$

Si $\bar{x} \pm t_{\alpha;n-1} \dfrac{s}{\sqrt{n}}$ es el intervalo para μ, entonces se dice que

$\delta = t_{\alpha;n-1} \dfrac{s}{\sqrt{n}}$ es su **precisión**



¿De qué depende la precisión de un intervalo?

Pero s depende de la muestra, ¡después habrá que comprobar que esto se cumple!

Tamaño mínimo de muestra para estimar μ: $n \geq \left(\dfrac{t_{\alpha;n-1}\, s}{\delta} \right)^2$

TEORÍA DE LA ESTIMACIÓN

Ejemplo 4.1:
Para determinar la estatura media de los varones adultos españoles, se tomó una muestra al azar de 10 de ellos, en la que se obtuvieron los valores siguientes (en cm):

162, 176, 169, 165, 171, 169, 172, 168, 167, 175

Estimar el valor de la estatura media

1) Valores muestrales: $n = 10;\quad \bar{x} = 169.40;\quad s = 4.30$

2) Fijemos una confianza del 95% $\Rightarrow \alpha = 0.05$ y $t_{0.05;9} = 2.262$

3) Entonces: $\mu \in \bar{x} \pm t_{\alpha;n-1} \dfrac{s}{\sqrt{n}} \Rightarrow \mu \in 169.40 \pm 2.262 \dfrac{4.30}{\sqrt{10}}$

con una confianza del 95%, es decir:

$$\mu \in 169.4 \pm 3.08 \Rightarrow \mu \in (166.32; 172.48)$$

Lo que **se interpreta** diciendo que la estatura media de los españoles es, con un 95% de confianza, un valor comprendido entre 166.32 y 172.48 cm.

Observe que la precisión de nuestra estimación es de δ=3.08 cm.
¿Podríamos aumentarla? ¿cómo?

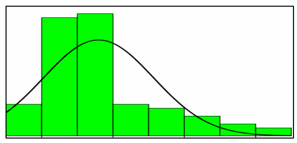
TEORÍA DE LA ESTIMACIÓN

T. Estimación . - **13**

Resúmenes: 4.3a *ii)*

Intervalos de confianza para la media de una variable NO NORMAL

Estudiamos una variable aleatoria X, que es continua pero su distribución no es la Normal



$x \not\rightarrow N(\mu\,;\sigma)$

Dada una muestra con $\begin{cases}\text{tamaño: n}\\ \text{media: } \overline{x}\\ \text{desviación típica: s}\end{cases}$

Si n≥60 entonces, para a fijado de antemano, el intervalo de confianza para μ se calcula como:

$$\mu \in \overline{x} \pm z_{\alpha}\frac{s}{\sqrt{n}}$$

Con Z_{α} en la tabla 1. De forma que tenemos

$$P\left(\overline{x} - z_{\alpha}\frac{s}{\sqrt{n}} \le \mu \le \overline{x} + z_{\alpha}\frac{s}{\sqrt{n}}\right) \cong 1-\alpha$$

La aproximación es tanto mejor cuanto mayor sea n.

TEORÍA DE LA ESTIMACIÓN

T. Estimación . - **14**

Resúmenes: 4

Tabla de la normal estándar
Tabla 2

α	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	∞	2.576	2.326	2.170	2.054	1.960	1.881	1.812	1.751	1.695
0.1	1.645	1.598	1.555	1.514	1.476	1.440	1.405	1.372	1.341	1.311
0.2	1.282	1.254	1.227	1.200	1.175	1.150	1.126	1.103	1.080	1.058
0.3	1.036	1.015	0.994	0.974	0.954	0.935	0.915	0.896	0.878	0.860
0.4	0.842	0.824	0.806	0.789	0.772	0.755	0.739	0.722	0.706	0.690
0.5	0.674	0.659	0.643	0.628	0.613	0.598	0.583	0.568	0.553	0.539
0.6	0.524	0.510	0.496	0.482	0.468	0.454	0.440	0.426	0.412	0.399
0.7	0.385	0.372	0.358	0.345	0.332	0.319	0.305	0.292	0.279	0.266
0.8	0.253	0.240	0.228	0.215	0.202	0.189	0.176	0.164	0.151	0.138
0.9	0.126	0.113	0.100	0.088	0.075	0.063	0.050	0.038	0.025	0.013

Tabla para los pequeños valores de α

α	0.002	0.001	0.000 1	0.000 01	0.000 001	0.000 000 1
z_{α}	3.090	3.291	3.891	4.417	4.892	5.327

¿Cómo buscar un valor Z_{α}?

Para una confianza del 95% :

$$(1-\alpha) = 0.95$$

Entonces $\alpha = 0.05$

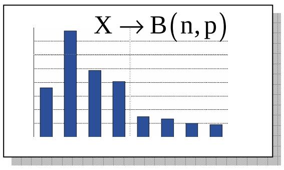
Así que tenemos que

$$z_{0.05} = 1.96$$

$$\alpha = 0.05 = 0.0 + 0.05 \Rightarrow Z_{0.05} = 1.96$$

TEORÍA DE LA ESTIMACIÓN

Intervalos de confianza para la proporción (X binomial aprox. a la normal)



Dada una muestra con

$$\begin{cases} \text{tamaño: n} \\ x \text{ casos que cumplen con la característica de interés} \\ n - x \text{ casos que no cumplen con la característica de interés} \end{cases}$$

el estimador puntual de la proporción de casos que verifican la característica de interés es

$$\hat{p} = x / n$$

El uso de la distribución binomial puede resultar complicado, pero si ***x* y *n-x* son valores mayores que 5**, entonces se puede elaborar un intervalo de confianza para **p** como sigue (aproximación a la normal)

TEORÍA DE LA ESTIMACIÓN

Intervalo de confianza para la proporción

Fijado α de antemano, si ***x*** y ***n-x*** son ambos mayores que 5, entonces:

$$p \in \frac{1}{n + z_\alpha^2}\left\{(x \pm 0.5) + \frac{z_\alpha^2}{2} \pm z_\alpha \sqrt{\frac{z_\alpha^2}{4} + (x \pm 0.5)\left(1 - \frac{x \pm 0.5}{n}\right)}\right\}$$

Cuando ***x* y *n-x* son mayores que 20** la cosa se simplifica y podemos utilizar la siguiente expresión:

$$p \in \hat{p} \pm \left(z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} + \frac{1}{2n}\right)$$

¡Atención, las **condiciones de validez** indicadas hay que comprobarlas **siempre**!

En estas expresiones z_α es el valor de la distribución normal (tabla 2) correspondiente al (1-α)% de confianza (observe que ahora no hablamos de grados de libertad).

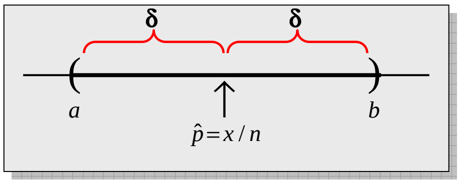
T. Estimación . - **17**
Resúmenes: 4.4b

TEORÍA DE LA ESTIMACIÓN

Obtenido el intervalo, que tendrá la forma (a, b), la precisión de la estimación viene dada por

$$\delta = \frac{(b-a)}{2}$$

que será, claro está, un valor expresable porcentualmente.



El tamaño mínimo de muestra para obtener una precisión deseada de δ (x100%) viene dado, para α prefijado, por la expresión:

$$n \geq \frac{z_{\alpha}^{2} \times 0.25}{\delta^{2}}$$

Observe que **no depende de valores muestrales**. Esta expresión **garantiza** la precisión deseada

T. Estimación . - **18**
Resúmenes: 4.4b

TEORÍA DE LA ESTIMACIÓN

Si se tiene información a partir de una muestra piloto, el tamaño mínimo de muestra para obtener una precisión deseada de δ (x100%) viene dado, para α prefijado, por la expresión:

Esta expresión **no garantiza** la precisión deseada

$$n \geq \frac{z_{\alpha}^{2} \times p \times q}{\delta^{2}}$$

donde p es el valor del intervalo de confianza $(p_1 \, , p_2)$ más próximo a 0.5 y $q = 1 - p$. En el caso de que el intervalo de confianza contenga al valor 0.5, entonces

$$n \geq \frac{z_{\alpha}^{2} \times 0.25}{\delta^{2}}$$

TEORÍA DE LA ESTIMACIÓN

Ejemplo:
Se va a elegir un cargo directivo mediante votación de todos los socios de un club. Si de 100 socios encuestados 30 se manifiestan a favor de un determinado candidato, ¿qué porcentaje de votos obtendría dicho candidato de celebrarse en ese momento la votación?

1) Valores muestrales: $n = 100;\quad x = 30;\quad n - x = 100 - 30 = 70$
El estimador puntual de **p** será entonces $\hat{p} = x/n = 30/100 = 0.3 \quad (30\%)$
Se suele indicar $\hat{q} = 1 - \hat{p} \quad (= 70\%)$

2) Fijemos una confianza del 95% $\Rightarrow$ $\alpha = 0.05$ y $Z_{0.05} = 1.96$

3) Como *x* y *n*-*x* son mayores que 20 podemos utilizar la versión mas simple del intervalo de confianza:

$$p \in \hat{p} \pm \left(z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} + \frac{1}{2n} \right) \Rightarrow p \in 0.30 \pm \left(1.96 \sqrt{\frac{(0.30)(0.70)}{100}} + \frac{1}{(2)(100)} \right)$$

4) Entonces: $p \in (0.2052; 0.3948)$ con una confianza del 95%.
Lo que **se interpreta** diciendo que la proporción de votos obtenidos por dicho partido sería, con un 95% de confianza, un valor comprendido entre el 20.52% y el 39.48%.

TEORÍA DE LA ESTIMACIÓN

Ficha técnica de una encuesta

1). Es legalmente obligatorio que las encuestas publicadas vengan acompañadas de sus garantías. La ficha adjunta se corresponde con una encuesta de satisfacción de los soldados españoles:

Universo: Individuos mayores de 18 años residentes en la Península y Baleares. *Muestra*: 1.294 entrevistados, seleccionados de forma polietápica por cuotas de sexo, edad, nivel de actividad y profesión del cabeza de familia. *Margen de error*: de ± 2,8% con probabilidad del 95,5% *Campo*: Del 24 al 29 de octubre de 1985. *Realización*: Instituto ECO S.A.

2). *n* = 1.294 (por un muestreo diferente del aleatorio simple), *d*=2,8%, *α*=4,5% $\Rightarrow$ cualquier proporción $\hat{p}$ de la encuesta indica que la verdad es $\hat{p} \pm 2{,}8\%$.

TEORÍA DE LA ESTIMACIÓN

Un posible esquema de trabajo

Paso 1. Identifique claramente la **información muestral** y el **estimador puntual** correspondiente

Paso 2. Establezca el **nivel de confianza (1-α)**

Paso 3. Compruebe si se cumplen las **condiciones de validez** que pudiera haber

Paso 4. Escriba la expresión del intervalo de confianza y realice los **cálculos** necesarios para obtenerlo

Paso 5. **Interprete los resultados obtenidos**, indicando claramente cuál es el intervalo y **con qué confianza** se ha obtenido. Conviene comentar qué representan estos valores, si la precisión es aceptable, ...