Tema 5. Teoría Elemental del Muestreo e Inferencia Paramétrica

Estadística

Ángel Serrano Sánchez de León

dice

- ntroducción: Población y Muestra
- ipos de muestreo
- istribuciones muestrales
  - De la media
  - Diferencia de medias
  - De una proporción
  - Diferencia de proporciones
- nferencia paramétrica: intervalos de confianza
  - De la media
  - Diferencia de medias
  - De una proporción
  - Diferencia de proporciones

2

troducción

a **Teoría Elemental del Muestreo** estudia la relación ntre una población y las muestras tomadas de ella.

oblación: Conjunto de elementos de referencia sobre los ue se realizan observaciones para extraer conclusiones.

- La población puede ser finita o infinita.
- En cualquier caso, suele ser muy costoso (o imposible) analizar todos los miembros de una población.

uestra: selección de elementos de la población.

- Debe ser **representativa** de la población.
- Para ello el muestreo debe ser **aleatorio**: todos los elementos de la población han de tener la misma probabilidad de ser seleccionados.
- Veremos que algunos muestreos no son aleatorios.

3

troducción

El muestreo puede ser de dos tipos:

- **Con reposición o reemplazo**: sacamos un elemento de la urna y lo volvemos a meter para la siguiente extracción.
  - Seleccionamos nombres al azar en un listín telefónico, admitiendo repeticiones.
  - Lanzamos 50 veces una moneda y contamos las caras.
- **Sin reposición o reemplazo**: los elementos extraídos de la urna no pueden volver a seleccionarse.
  - Sacamos 10 bolas sucesivamente de una urna que contiene 100 bolas, sin reponerlas.

troducción

La **Teoría Elemental del Muestreo** nos permite:
- Estimar magnitudes desconocidas de una población (**parámetros de la población**) a partir del conocimiento de magnitudes medidas en las muestras (**estadísticos de la muestra**).
- Determinar si las diferencias observadas entre dos muestras son debidas a causas fortuitas o son realmente significativas. Para ello se realizan **contrastes de hipótesis** y **tests de significación** (**Teoría de las Decisiones**).

El estudio de las inferencias hechas sobre una población a partir de sus muestras se llama **Inferencia Estadística**.

Tipos de Muestreo:
- No probabilístico.
- Probabilístico.

5

pos de muestreo

No probabilístico:

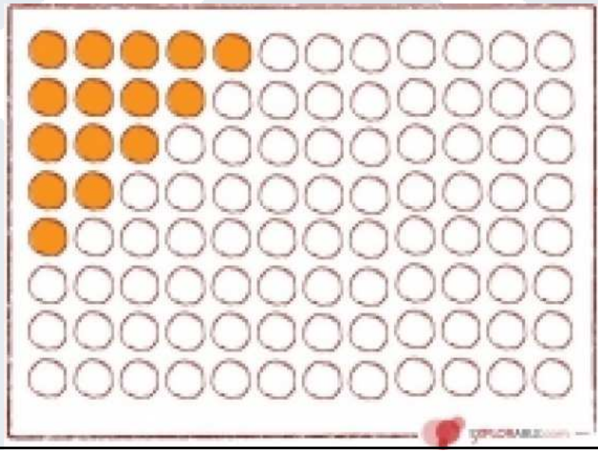
- No se usa el azar ni la probabilidad, sino el criterio del investigador, es decir, él decide si la muestra es o no representativa.
- Subtipos:
  - Por conveniencia o accidental.
  - Por bola de nieve.
  - Por cuotas.
  - Discrecional.

6

pos de muestreo no obabilístico

Muestreo por conveniencia o accidental:

- Los usuarios son fácilmente accesibles.
- Técnica fácil y barata.
- Ejemplo: cuando un profesor universitario hace una encuesta a sus alumnos.



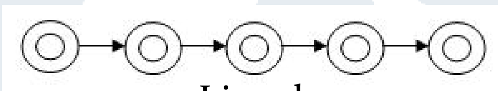
http://bit.ly/1yPnr1e

7

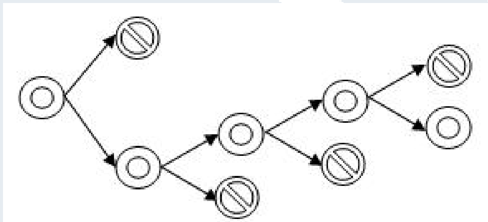
pos de muestreo no obabilístico

Muestreo por bola de nieve:

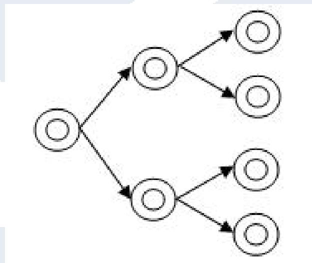
- Cuando los sujetos son difíciles de encontrar, a cada uno se le pide que dé el nombre de otro(s) sujeto(s) candidato(s) a participar en el estudio.
- Fácil, barata, sin planificación previa, pero difícil de controlar.



Lineal



Discriminatorio exponencial



No discriminatorio exponencial

http://bit.ly/1JUdqCA

...pos de muestreo no ...obabilístico

...uestreo por cuotas:

- Se separan los sujetos por estratos (grupos homogéneos sin solapamiento).
- Después se eligen tantos sujetos de cada estrato según una cuota o proporción, donde la elección no es al azar sino según el criterio del investigador.
- Por ejemplo, para un estudio de mercado necesitamos conocer la opinión sobre un producto. Como sabemos que el 60% de los clientes potenciales son mujeres, entrevistamos a 60 mujeres y 40 hombres (esas son las cuotas). La elección de cada sujeto se hace sin seguir ningún criterio probabilístico adicional, buscando hombres o mujeres hasta completar el cupo.

...pos de muestreo no ...obabilístico

...Muestreo discrecional:

- Los sujetos elegidos para la muestra se escogen según lo que el investigador del estudio crea más conveniente.
- Ejemplo: si queremos analizar qué condiciones debemos cumplir para triunfar como empresario, podemos seleccionar a y entrevistar a personas que han tenido éxito con sus empresas.

10

pos de muestreo

Muestreo probabilístico:

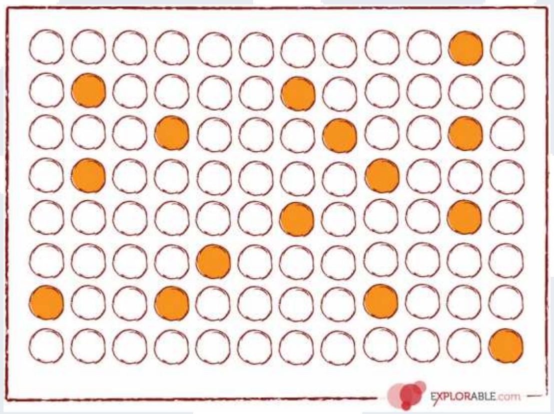
- En este caso los sujetos se eligen al azar de manera aleatoria, es decir, basándonos en la probabilidad.
- Es el que vamos a ver con más detalle.
- Subtipos:
  - Aleatorio simple.
  - Aleatorio sistemático.
  - Estratificado.
  - Por conglomerados.
  - Polietápico.

11

pos de muestreo probabilístico

Muestreo aleatorio simple:

- Se numera a cada uno de los sujetos y se van eligiendo al azar con igual probabilidad.
- Ejemplo: una lotería donde cada individuo tiene un número que le ha tocado al azar.



https://explorable.com/es/muestreo-aleatorio

12

...pos de muestreo probabilístico

...uestreo aleatorio sistemático:

- De manera aleatoria se elige un sujeto inicial *i*.
- A partir de él, se eligen los demás a saltos de tamaño *k* (*i*, *i*+*k*, *i*+2*k*, *i*+3*k*, etc.).
- Ejemplo: En las Escuelas Oficiales de Idiomas, para elegir la preferencia en el proceso de matriculación, se elige al azar una letra del abecedario. A partir de ahí, y de uno en uno (*k* = 1), tienen preferencia el resto de letras del abecedario.
- Ejemplo: de 100 individuos necesitamos una muestra de 12. Elegimos al azar el número 5. El intervalo es *k* = int(100/12) = 8, luego elegimos a los siguientes sujetos: 5, 13, 21, 29, 37, 45, 53, 61, 69, 77, 85, 93.

13

...pos de muestreo probabilístico

...uestreo aleatorio estratificado:

- Se separan los sujetos por estratos (grupos homogéneos sin solapamiento).
- Después se eligen los sujetos de cada estrato según su proporción aleatoriamente respecto de la población. Por lo tanto el muestreo se realiza en todos los estratos.
- Es la versión probabilística del muestreo por cuotas.
- Ejemplo: en una encuesta telefónica se eligen personas de cada provincia según la proporción que representa la población de dicha provincia respecto de todo el país.
- También se elegirán tantos hombres como mujeres, así como todos los rangos de edad, según la pirámide de población del lugar.

14

pos de muestreo probabilístico

Iuestreo por conglomerados:

Cuando los sujetos están agrupados en conglomerados, se eligen al azar los conglomerados o grupos.

Después se eligen aleatoriamente los sujetos de entre los conglomerados seleccionados. Por lo tanto, pueden quedar conglomerados sin muestrear.

Subtipos:

- **Conglomerados en 1 etapa:** se eligen todos los sujetos de los conglomerados seleccionados.
- **Conglomerados en 2 etapas:** una vez elegidos los conglomerados, se seleccionan aleatoriamente algunos sujetos de los mismos.

15

Tipos de muestreo probabilístico

Muestreo mixto o polietápico:

- Es cuando el muestreo es tan complicado que se realiza por etapas o fases.
- En cada etapa se elige el tipo de muestreo más adecuado a cada caso.

Distribuciones muestrales

stribuciones muestrales

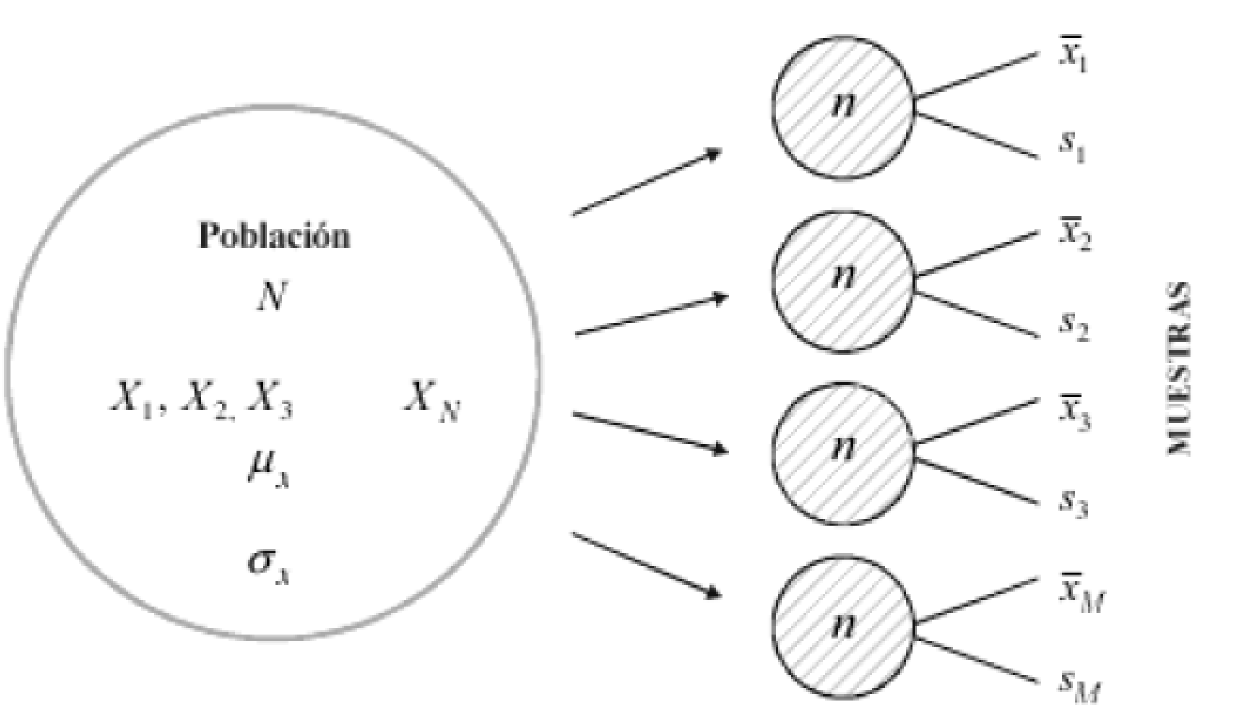
- Dada una **población de tamaño *N***, se pueden generar ***M*** **muestras** de ***n*** **elementos** (con o sin reposición).

$$= \binom{N}{n} = \frac{N!}{n!(N-n)!} \quad \text{(sin reposición)} \qquad M = N^n \quad \text{(con reposición)}$$

- Para cada muestra podemos calcular un **estadístico** (como la media, la desviación típica o una proporción), que variará de muestra a muestra.
  - El estadístico será una **variable aleatoria** que tendrá su propia **distribución muestral** o de muestreo.

18

stribuciones muestrales



19

stribuciones muestrales

ara una muestra:

- Media aritmética.
- Proporción.

ara dos muestras:

- Diferencia de medias.
- Diferencia de proporciones.

s habitual utilizar letras griegas para referirnos a los **arámetros de una población** y letras latinas para os **estadísticos de la muestra**.

20

stribución muestral de la media

ea una población normal de media μ y varianza σ^2, de la que xtraemos una muestra de tamaño n, formada por las bservaciones X_i, $i = 1, \ldots, n$. La media muestral es $\overline{X}$:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

odas las X_i siguen la misma distribución normal. Por eso, la **media muestral** sigue también una distribución normal con arámetros:

$$\mathrm{E}(\overline{X}) \equiv \mu_{\overline{X}} = \frac{\mathrm{E}(X_1) + \mathrm{E}(X_2) + \ldots + \mathrm{E}(X_n)}{n} = \frac{\overbrace{\mu + \ldots + \mu}^{n}}{n} = \mu$$

$$\mathrm{Var}(\overline{X}) \equiv \sigma^2_{\overline{X}} = \frac{\sigma^2_{X_1} + \sigma^2_{X_2} + \ldots + \sigma^2_{X_n}}{n^2} = \frac{\overbrace{\sigma^2 + \ldots + \sigma^2}^{n}}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

21

stribución muestral de la media

Recordando que la desviación típica mide el **grado de variabilidad** de los datos en torno a la media, podemos considerarla como el **error** en la estimación de la media de la población.

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Podemos reducir el error aumentando el tamaño de la muestra n, pero el error se reduce lentamente (es inversamente proporcional a la raíz de n).

- Para reducir el error a la mitad, debemos aumentar la muestra en un factor 4.

eorema del Límite Central

Para valores grandes del tamaño de muestra ($n \geq 30$ y $N > 2n$), la distribución muestral de medias es aproximadamente Normal con media $\mu_{\overline{X}}$ y desviación típica $\sigma_{\overline{X}}$, independientemente de la población.

- Dicho de otra forma: si n tiende a infinito (muestras muy grandes), la variable tipificada siguiente tiende a la normal estándar N(0,1).

$$Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

- Si la población está normalmente distribuida, entonces la distribución de medias también lo está, incluso para $n < 30$.

23

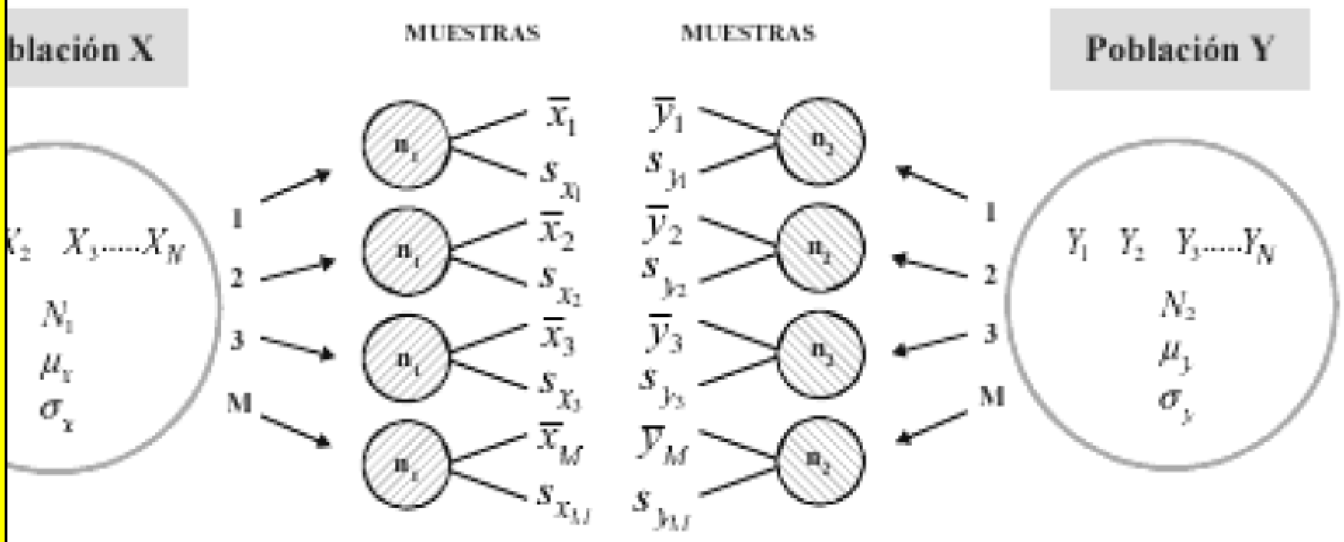
stribución muestral de diferencia e medias

- ean dos poblaciones, la primera con media μ_1 y arianza σ_1^2, y la segunda con media μ_2 y varianza σ_2^2.
- omamos dos muestras aleatorias **independientes** de ada población, de tamaño n_1 y n_2, respectivamente.
- ea $\overline{X}_1$ la media muestral de la primera población, y $\overline{X}_2$ a de la segunda.
- amos a estudiar un nuevo estadístico: la **diferencia le las medias**:

$$\overline{X}_1 - \overline{X}_2$$

24

...stribución muestral de diferencia ...e medias



25

stribución muestral de diferencia e medias

- La **distribución muestral de la diferencia de las medias** tiene por valor esperado la diferencia de las medias poblacionales.

$$E(\overline{X}_1 - \overline{X}_2) \equiv \mu_{\overline{X}_1 - \overline{X}_2} = \mu_{\overline{X}_1} - \mu_{\overline{X}_2} = \mu_1 - \mu_2$$

- Por otro lado, la **varianza de la diferencia de las medias** cumple:

$$\mathrm{Var}(\overline{X}_1 - \overline{X}_2) \equiv \sigma^2_{\overline{X}_1 - \overline{X}_2} = \sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

26

stribución muestral de diferencia e medias

- uando el tamaño de ambas muestras, n_1 y n_2, tiende al nfinito, la variable tipificada siguiente sigue proximadamente una distribución normal estándar N(0,1):

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- sto ya se cumple típicamente si $n_1 + n_2 > 30$ y $n_1 \approx n_2$.
- i las poblaciones son normales, la distribución muestras e las diferencias es normal sin importar los tamaños de las nuestras.

27

...so de varianza poblacional no ...nocida

- ...i no conocemos la varianza poblacional σ^2, se ...tilizada la varianza muestral como aproximación.
- ...e define **varianza muestral** S^2 de n variables ...leatorias $X_1, \ldots, X_n$, como:

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

- ...sta es la **varianza insesgada**, llamada así porque su ...alor esperado coincide con el de la varianza ...oblacional.

$$\mathrm{E}(S^2) \equiv \mu_{S^2} = \sigma^2$$

28

stribución muestral de una oporción

- Analicemos ahora una distribución binomial.
- ea el caso de una población infinita (o muy grande) le la que extraemos una muestra de tamaño *n* (=*n* nsayos).
- ea $\hat{P} = X/n$ el cociente que representa la tasa o proporción de éxitos *X* obtenidos en la muestra con *n* nsayos, lo cual será un buen estimador de la tasa de xitos *p* de la población.

29

stribución muestral de una oporción

Como la muestra es grande ($n \geq 30$), se puede aplicar el Teorema del Límite Central.

Por tanto esta variable $\hat{P}$ sigue aproximadamente una **distribución normal** con **media**:

$$\mathrm{E}(\hat{P}) \equiv \mu_{\hat{P}} = \mathrm{E}\left(\frac{X}{n}\right) = \frac{\mathrm{E}(X)}{n} = \frac{np}{n} = p$$

Por otro lado, su **varianza** cumple:

$$\mathrm{Var}(\hat{P}) \equiv \sigma^2_{\hat{P}} = \sigma^2_{X/n} = \frac{\sigma^2_X}{n^2} = \frac{npq}{n^2} = \frac{pq}{n} = \frac{p(1-p)}{n}$$

30

emplo

Un jugador de baloncesto lanza 100 tiros libres y acierta el
o%. Calcular la distribución muestral de la proporción.

n este caso el tamaño de la muestra es $n = 100$.

$$\mu_{\hat{P}} = \frac{X}{n} = \frac{80}{100} = 0{,}80 = p$$

una desviación típica:

$$\sigma_{\hat{P}} = \sqrt{\frac{0{,}8 \times 0{,}2}{100}} = 0{,}04 = 4\%$$

omo $n = 100 \geq 30$, la aproximación normal es válida.

cierta entre 76 y 84 % de los tiros libres un 68,3 % de las
eces.

31

stribución muestral de la ferencia de proporciones

- ambién podemos calcular la **distribución muestral le la diferencia de las proporciones** de dos oblaciones distribuidas binomialmente con arámetros (n_1, p_1) y (n_2, p_2).
- i las muestras son grandes, $n_1 + n_2 > 30$, la listribución muestral de la diferencia de proporciones e puede aproximar por la normal con parámetros:

$$\mathrm{E}(\hat{P}_1 - \hat{P}_2) \equiv \mu_{\hat{P}_1 - \hat{P}_2} = \mu_{\hat{P}_1} - \mu_{\hat{P}_2} = p_1 - p_2$$

$$\mathrm{Var}(\hat{P}_1 - \hat{P}_2) \equiv \sigma^2_{\hat{P}_1 - \hat{P}_2} = \sigma^2_{\hat{P}_1} + \sigma^2_{\hat{P}_2} = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

32

Inferencia paramétrica

ferencia paramétrica

a **inferencia paramétrica** es la estimación de los parámetros de una oblación (como media y varianza poblacionales) a partir de los stadísticos de la muestra (como media y varianza muestrales).

stimación puntual: Se calcula el valor del parámetro como un único alor numérico.

- Ejemplo: La longitud media de un lote de cables creados en una fábrica es de 10,326 m.
- Ejemplo: La intención de voto del partido A es 45%.

- Existe un método de estimación puntual denominado **Método de Máxima Verosimilitud** (no lo vamos a estudiar).

34

timaciones por intervalo de nfianza

stimación por intervalo de confianza: Se da un ntervalo en el cual tenemos una cierta probabilidad o nivel e confianza de que se encuentra el valor del parámetro oblacional que estamos estimando.

- Ejemplo: La longitud media de un lote de cables creados en una fábrica es de 10,326 ± 0,016 m.
- Ejemplo: La intención de voto del partido A es 45 ± 10 %.

ea β el parámetro poblacional (desconocido) y L_1 y L_2 los ímites del intervalo. Se define el **nivel de confianza** $1-\alpha$ omo la probabilidad de que β se encuentre en el intervalo L_1, L_2]:

$$P(L_1 \leq \beta \leq L_2) = 1 - \alpha$$

timaciones por intervalo de nfianza

l intervalo $[L_1, L_2]$ se llama **intervalo de confianza** lel $(1 - \alpha)\cdot 100\%$.

- Ejemplo: Si α vale 0,05, el intervalo de confianza es del 95%, es decir, la probabilidad de que el parámetro β esté dentro del intervalo de confianza dado es del 95%.

upongamos que la distribución muestral de un stadístico B es **aproximadamente normal** (lo cual curre si la población de partida es normal, o bien, si el amaño de la muestra es grande $n > 30$).

n este caso el estadístico B sigue una distribución ormal $N(\mu_B, \sigma_B)$.

36

...timaciones por intervalo de ...nfianza

...or las propiedades de la distribución normal:

$$P(\mu_B - \sigma_B \leq B \leq \mu_B + \sigma_B) = 0{,}68267$$

...sta ecuación corresponde a la variable tipificada $z = \pm 1$.

...eordenando los términos:

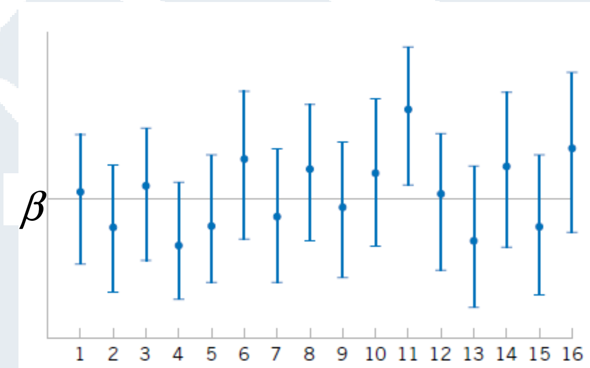
$$P(B - \sigma_B \leq \mu_B \leq B + \sigma_B) = 0{,}68267$$

...i el estadístico B no tiene sesgo, entonces $\mu_B = \beta$.

$$P(B - \sigma_B \leq \beta \leq B + \sigma_B) = 0{,}68267$$

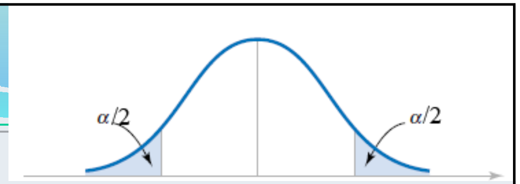
timaciones por intervalo de nfianza

- s decir, existe una probabilidad del 68,3% de que el alor del parámetro poblacional que buscamos, β, esté ontenido en el intervalo $[B - \sigma_B, B + \sigma_B]$.
- es conocido a partir de la muestra, y σ_B o bien es onocido, o bien se estima con la desviación típica nuestral S (la insesgada, con $n - 1$ en el denominador).



Nº de muestra

38



oeficientes de confianza

e llama **coeficiente de confianza $z_{\alpha/2}$** al factor que nultiplica a σ_B y que determina la longitud del ntervalo de confianza $[B - z_{\alpha/2}\, \sigma_B,\, B + z_{\alpha/2}\, \sigma_B]$.

ongitud del intervalo = $2\, z_{\alpha/2}\, \sigma_B \equiv 2\, \varepsilon$

α	Nivel de confianza $(1 - \alpha)\cdot 100\%$	$z_{\alpha/2}$
0,50	50%	$z_{0,25} = 0{,}6745$
0,3173	68,27%	$z_{0,1587} = 1{,}00$
0,05	95%	$z_{0,025} = 1{,}96$
0,0455	95,45%	$z_{0,02275} = 2{,}00$
0,01	99%	$z_{0,005} = 2{,}58$
0,0027	99,73%	$z_{0,00135} = 3{,}00$

tervalo de confianza para la edia

uposición: población normal N(μ, σ) con σ^2 onocida. Además, población infinita o finita con eposición. Sea una muestra.

n este caso:

$$\mu_{\overline{X}} = \mu \approx \overline{X}, \quad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

ntonces:

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

l intervalo de confianza de nivel (1 – α) para la media n el caso de una población infinita o con reposición le varianza conocida es:

$$\mu = \overline{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

40

tervalo de confianza para la edia

- **uposición:** cualquier tipo de población con σ^2 lesconocida y muestra grande ($n \geq 30$).
- Como no conocemos σ, la estimamos con la desviación ípica muestral S ("$n - 1$" en el denominador).
- Al ser muestra grande, por el Teorema del Límite Central la distribución muestral de la media es normal.
- Entonces el intervalo de confianza para la media es:

$$\mu = \overline{X} \pm z_{\alpha/2} \frac{S}{\sqrt{n}}$$

41

tervalo de confianza para la ferencia de medias

uposición: poblaciones normales N(μ_1, σ_1) y N(μ_2, σ_2) con σ_1^2 y σ_2^2 conocidas. Además, poblaciones nfinitas o finitas con reposición. Sean 2 muestras.

a sabemos que:

$$\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2 \approx \bar{X}_1 - \bar{X}_2, \qquad \sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

ntonces el intervalo de confianza de nivel (1 – α) para a diferencia de medias es:

$$\mu_1 - \mu_2 = \left(\bar{X}_1 - \bar{X}_2\right) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

42

tervalo de confianza para la ferencia de medias

uposición: cualquier tipo de población con σ_1^2 y σ_2^2 lesconocidas, y muestras grandes ($n_1 + n_2 \geq 30$) y imilares ($n_1 \approx n_2$). Además, poblaciones infinitas o initas con reposición.

lo conocemos σ_1 ni σ_2, así que las estimamos a partir le las desviaciones típicas muestrales S_1 y S_2.

ntonces:

$$\mu = \left(\overline{X}_1 - \overline{X}_2\right) \pm z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

43

tervalo de confianza para una oporción

uposición: población binomial de parámetro p lesconocido y una muestra grande ($n \geq 30$). Además, oblación infinita o finita con reposición.

l ser muestra grande, podemos aplicar la proximación normal a la distribución binomial.

n este caso:

$$\mu_{\hat{P}} = p \approx \hat{P}, \quad \sigma_{\bar{P}} = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

uego el intervalo de confianza de nivel ($1 - \alpha$) para la roporción de éxito de la población es:

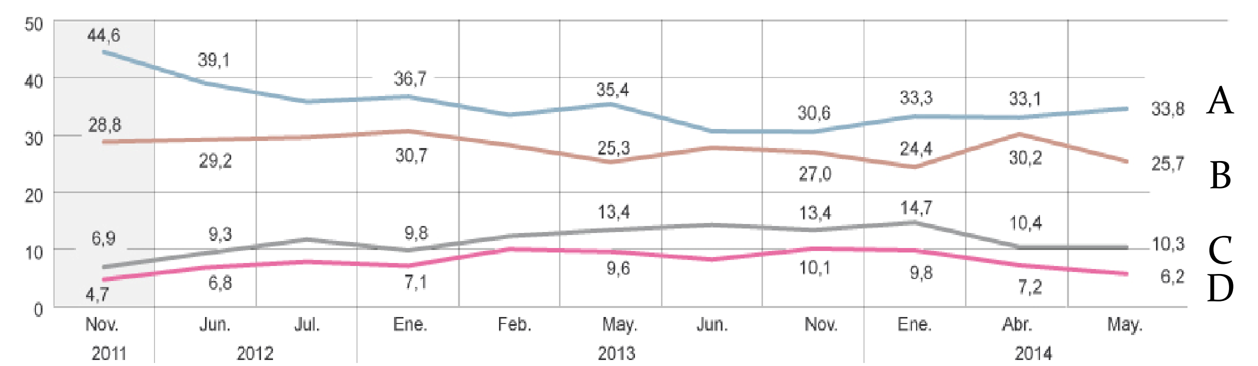
$$p = \hat{P} \pm z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

44

tervalo de confianza para una oporción: Ejemplo

ondeo electoral (El Mundo, 19/05/2014)

Evolución de la intención de voto desde los últimos comicios generales



CHA TÉCNICA. Universo: Mayores de 18 años. Ámbito: Nacional. Muestra: 1.111 entrevistas con un margen de error ±3% para los datos globales, con un nivel de confianza del ,5% (dos sigma) y un p/q=50/50. Selección: Estratificada, aleatoria. Entrevista: Telefónica. Fecha del trabajo de campo: Del 13 al 15 de mayo de 2014. Realización: SIGMA DOS. rección: José Miguel de Elías. Nota: para el cálculo de la estimación de voto y su proyección en la pregunta de intención de voto a nivel global, hay un NS/NC del 36,7%, que una z aplicada la pregunta de simpatía se reduce al 15,7%.

ENTE: SIGMA DOS — EL MUNDO

http://mun.do/PionqZ

tervalo de confianza para una oporción: Ejemplo

l caso crítico es el de un partido que obtiene una ntención de voto $\hat{P}$ = 50% = 0,5. El intervalo de onfianza para ese caso extremo es:

$$\left.\begin{array}{l} = 50\% = 0{,}500 \\ = 1111 \\ ,5\% \Rightarrow z_{\alpha/2} = 2{,}0 \end{array}\right\} \Rightarrow \; p = 0{,}500 \pm 2{,}0\sqrt{\frac{0{,}5\times 0{,}5}{1111}} = 0{,}500 \pm 0{,}030 = (50{,}0 \pm 3{,}0)\%$$

l valor de $\hat{P}$ = 0,5 es aquel para el cual la longitud del ntervalo de confianza es máxima (mayor error).

$$f(p) = 2 z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = c\sqrt{p-p^2}$$

$$\frac{df(p)}{dp} = c\frac{1-2p}{2\sqrt{p-p^2}} = 0 \;\Rightarrow\; 1-2p = 0 \;\Rightarrow\; p = 1/2 = 0{,}5$$

46

tervalo de confianza para una oporción: Ejemplo

Veamos los intervalos de confianza para cada uno de stos partidos:

Partido A: $p_A = 0{,}338 \pm 2{,}0\sqrt{\dfrac{0{,}338 \times 0{,}662}{1111}} = 0{,}338 \pm 0{,}028 = (33{,}8 \pm 2{,}8)\%$

Partido B: $p_B = 0{,}257 \pm 2{,}0\sqrt{\dfrac{0{,}257 \times 0{,}743}{1111}} = 0{,}257 \pm 0{,}026 = (25{,}7 \pm 2{,}6)\%$

Partido C: $p_C = 0{,}103 \pm 2{,}0\sqrt{\dfrac{0{,}103 \times 0{,}897}{1111}} = 0{,}103 \pm 0{,}018 = (10{,}3 \pm 1{,}8)\%$

Partido D: $p_D = 0{,}062 \pm 2{,}0\sqrt{\dfrac{0{,}062 \times 0{,}938}{1111}} = 0{,}062 \pm 0{,}014 = (6{,}2 \pm 1{,}4)\%$

deo predice que ninguno de los partidos obtendrá mayoría absoluta con el 95,5% de nza, ya que ningún intervalo incluye el 50% de los votos.

tervalo de confianza para la ferencia de proporciones

uposición: poblaciones binomiales de parámetros p_1 p_2 desconocidos y 2 muestras grandes ($n_1 + n_2 \geq 30$). demás, población infinita o finita con reposición.

l ser muestra grande, podemos aplicar la proximación normal a la distribución binomial.

n este caso:

$$= p_1 - p_2 \approx \hat{P}_1 - \hat{P}_2, \quad \sigma_{\hat{P}_1 - \hat{P}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \approx \sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}$$

ntonces:

$$p_1 - p_2 \approx \left(\hat{P}_1 - \hat{P}_2\right) \pm z_{\alpha/2} \sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}$$

48

tervalo de confianza para la ferencia de proporciones

Ejemplo: Calculemos el intervalo de confianza de la liferencia entre los dos primeros partidos, según el ondeo anterior.

$$\left.\begin{array}{l} n_2 = 1111 \\ 33{,}8\% = 0{,}338 \\ 25{,}7\% = 0{,}257 \\ \% \Rightarrow z_{\alpha/2} = 2{,}0 \end{array}\right\} \Rightarrow p_1 - p_2 = (0{,}338 - 0{,}257) \pm 2{,}0\sqrt{\frac{0{,}338 \times 0{,}662}{1111} + \frac{0{,}257 \times 0{,}743}{1111}} =$$

$$= 0{,}081 \pm 0{,}039 = (8{,}1 \pm 3{,}9)\%$$

egún este sondeo, el Partido A sacará entre 4,2 y 12,0 puntos más que l Partido B, con una confianza del 95,5 %. Luego el sondeo predice que l Partido A ganará las elecciones.

asos especiales

Estadístico	Si el intervalo de confianza incluye el valor	A ese nivel de significación el resultado es compatible con	En caso contrario
Media	0	Media positiva, cero o negativa	Media estrictamente positiva o negativa
Diferencia de medias	0	Medias iguales	Una media estrictamente mayor que la otra
Proporción	0,5	Mayoría absoluta no garantizada	Consigue la mayoría absoluta (>0,5) o no la consigue (<0,5)
Diferencia de proporciones	0	Empate técnico entre ambos candidatos	Un candidato gana al otro

50