Estimación

proporciones

Ejemplo

Un artículo relacionado con la salud, reporta los siguientes datos sobre la incidencia de disfunciones importantes entre recién nacidos con madres fumadoras de marihuana y de madres que no la fumaban:

	Usuaria	No Usuaria
Tamaño Muestral	1246	11178
Número de disfunciones	42	294

Encuentra el intervalo de confianza del 99% para la diferencia de
proporciones.

Solución

Datos

$x_1=42$
$x_2=294$
$n_1=1246$
$n_2=11178$
$IC=90\ \%$

Fórmula

$$(P_1-P_2)=(p_1-p_2)-z\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}$$

$$IC=0.5-(0.99/2)=0.005$$

z=2.575

Un artículo relacionado con la salud, reporta los siguientes datos sobre la incidencia de disfunciones importantes entre recién nacidos con madres fumadoras de marihuana y de madres que no la fumaban:

	Usuaria	No Usuaria
Tamaño Muestral	1246	11178
Número de disfunciones	42	294

p_1 = 42/1246= 0.0337; p_2=294/11178= 0.0263

q_1=1- 0.033 =0.967; q_2=1- 0.0263 =0 .9737

$$\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}=\sqrt{\frac{0.033\times0.967}{1246}+\frac{0.0263\times0.9737}{11178}}=$$

$$\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}=\sqrt{2.561\times10^{-5}+2.291\times10^{-6}}=5.2821\times10^{-3}$$

$$z\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}=(2.575)(5.2821\times10^{-3})=0.01360$$

$$(P_1-P_2)=(0.00337-0.0263)\pm0.01360$$

$$(P_1-P_2)=6.7\times10^{-3}\pm0.01360$$

$-0.0069 \leq P_1 - P_2 \leq 0.0203$

Cálculo del tamaño de la muestra para estimar una media.

Pregunta

¿Qué tan grande debe ser una muestra si la media muestral se va a usar para estimar la media poblacional?

Respuesta

Depende del error estándar de la media, si éste fuera cero, entonces se necesitaría una sola media que será igual necesariamente a la media poblacional desconocida μ, porque s = 0.

Este caso extremo no se encuentra en la práctica, pero refuerza el hecho de que mientras menor sea el error estándar de la media, menor es el tamaño de muestra necesario para lograr un cierto grado de precisión.

Se estableció con el teorema de limite central, que una forma de disminuir el error de estimación es aumentar el tamaño de la muestra ($\sigma_x = \sigma/\sqrt{n}$), si éste incluye el total de la población, entonces $\mu - x$ sería igual a cero.

Con esto en mente, parece razonable que para un nivel de confianza *fijo*, sea posible determinar un tamaño de la muestra tal que el error de estimación sea tan pequeño como queramos, para ser mas preciso, dado un nivel de confianza y un error de estimación ε fijos, se puede escoger un tamaño de muestra n tal que P(|x -μ | < ε) = Nivel de confianza. Con el propósito de determinar n. El error máximo de estimación esta dado por

$$\varepsilon = z\frac{\sigma}{\sqrt{n}}$$

Que al elevarla al cuadrado y despejar n se obtiene :

$$n = z^2\frac{\sigma^2}{\varepsilon^2}$$

Como *n* debe de ser un número entero, *redondeamos* hacia arriba todos los resultados fraccionarios. En el caso que se tenga una población finita y/o un muestreo sin reemplazo, el error de estimación se convierte en:

$$\varepsilon = z\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Que al despejar n se obtiene:

$$n = \frac{\left(z\frac{\sigma}{\varepsilon}\right)^2\frac{N}{N-1}}{1+\left(z\frac{\sigma}{\varepsilon}\right)^2\frac{1}{N-1}} = \frac{(z\sigma)^2 N}{\varepsilon^2(N-1)+(z\sigma)^2}$$

Ejemplo

En un negocio de importaciones de artesanías se requiere estimar el peso promedio de las cunas fabricadas por artesanos Totziles. Un estudio anterior de 10 cunas mostró que la desviación estándar de sus pesos es de 7.32 Kg ¿Qué tan grande debe ser una muestra para que el negociante tenga el 95% de confianza de que el error de estimación es a lo más de 2.4 Kg?

Solución

Datos

$\varepsilon=2.4$

$n=10$

$\sigma=7.32$

$IC=95\ \%$

Fórmula

$$n = z^2 \frac{\sigma^2}{\varepsilon^2}$$

$$IC = 0.5-(0.95/2)=0.025$$

z=1.96

En un negocio de importaciones de artesanías se requiere estimar el peso promedio de las cunas fabricadas por artesanos Totziles. Un estudio anterior de 10 cunas mostró que la desviación estándar de sus pesos es de 7.32 Kg ¿Qué tan grande debe ser una muestra para que el negociante tenga el 95% de confianza de que el error de estimación es a lo más de 2.4 Kg?

$$n = z^2 \frac{\sigma^2}{\varepsilon^2} = 1.96^2 \times \left(\frac{7.32}{2.4}\right)^2 = 35.736 \approx 36$$

En consecuencia, si el tamaño de la muestra es de 36, se puede tener un 95% de confianza en que μ difiera en menos de 2.4 Kg de x .

Ejemplo

Una empresa eléctrica fabrica focos que tienen una duración aproximadamente normal con una desviación estándar de 40 horas. ¿De qué tamaño se necesita una muestra si se desea tener 96% de confianza que la media real esté dentro de 10 horas de la media real?

Solución

Datos

$\varepsilon=10$

$n=?$

$\sigma=40$

$IC=96\ \%$

$$n = z^2 \frac{\sigma^2}{\varepsilon^2} = 2.055^2 \times \left(\frac{40}{10}\right)^2 = 67.56 \approx 68$$

Fórmula

$$n = z^2 \frac{\sigma^2}{\varepsilon^2}$$

$$IC = 0.5 - (0.96/2)$$

$$= 0.02$$

z=2.055

¿Qué pasaría si en lugar de tener un error de estimación de 10 horas sólo se requiere un error de 5 horas?

$$n = z^2 \frac{\sigma^2}{\varepsilon^2} = 2.055^2 \times \left(\frac{40}{5}\right)^2 = 270.27 \approx 270$$

Se puede observar que el tamaño de la muestra se incrementa, pero como beneficio de ello se obtiene una estimación más exacta.

Ejemplo

Solución

Datos

$\varepsilon=10$

$n=?$

$N=300$

$\sigma=40$

$IC=96\ \%$

Fórmula

$$n=\frac{(z\sigma)^2 N}{\varepsilon^2(N-1)+(z\sigma)^2}$$

$$IC=0.5-(0.96/2)$$
$$=0.02$$

z=2.055

Suponga que en el ejercicio anterior se tiene una población de 300 focos, y se desea saber de que tamaño debe de ser la muestra. El muestreo se realizará sin reemplazo.

$$n=\frac{(z\sigma)^2 N}{\varepsilon^2(N-1)+(z\sigma)^2}=\frac{(2.055\times 40)^2\times 300}{10^2(300-1)+(2.055\times 40)^2}=55.29$$

Si se tiene una población finita de 300 focos sólo se tiene que extraer de la población una muestra sin reemplazo de 56 focos para poder estimar la duración media de los focos restantes con un error máximo de 10 horas.

Cálculo del tamaño de la muestra para estimar una proporcion.

Si deseamos saber que tan grande se requiere que sea una muestra para asegurar que el error al estimar ***P*** sea menor que una cierta cantidad específica ε.

$$\varepsilon = z\frac{\sigma}{\sqrt{n}} = z\frac{\sqrt{pq}}{\sqrt{n}} = z\sqrt{\frac{pq}{n}}$$

Que al elevarla al cuadrado y despejar ***n*** se obtiene :

$$n = z^2\frac{pq}{\varepsilon^2} = \left(\frac{z}{\varepsilon}\right)^2 pq$$

Esta fórmula es <u>engañosa</u>, ya que debemos utilizar **p** para determinar el tamaño de la muestra, pero ***p*** se calcula a partir de la muestra (p=x/n).

Existen ocasiones en las cuales se tiene una idea del comportamiento de la proporción de la población y ese valor se puede sustituir en la fórmula, pero si no se sabe nada referente a esa proporción entonces se tienen dos opciones:

1

- Tomar una muestra preliminar mayor o igual a 30 para proporcionar una estimación de P. Después con el uso de la fórmula se podría determinar de forma aproximada cuántas observaciones se necesitan para proporcionar el grado de precisión que se desea.

2

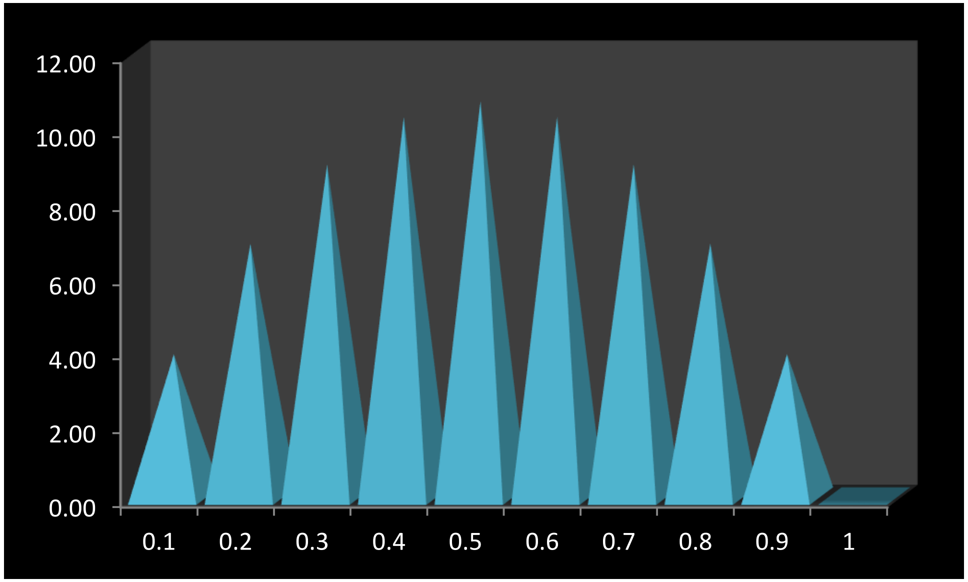
- Tomar el valor de p como 0.5 ya que sustituyendo este en la fórmula se obtiene el tamaño de muestra mayor posible.

Ejemplo

Se desconoce el valor de P, por lo que se utilizarán diferentes valores y se sustituirán en la formula para observar los diferentes tamaños de muestras. El nivel de confianza que se utilizará es del 95% (z=1.96) con un error de estimación de 0.30.

p	n
0.1	3.84
0.2	6.83
0.3	8.96
0.4	10.24
0.5	10.67
0.6	10.24
0.7	8.96
0.8	6.83
0.9	3.84
1	0.00

$$n = z^2 \frac{pq}{\varepsilon^2} = \left(\frac{z}{\varepsilon}\right)^2 pq$$



como se puede observar en la grafica anterior cuando ***P*** vale 0.5 el tamaño de la muestra alcanza su máximo valor.

Para el caso de que se tenga una población finita y un muestreo sin reemplazo, el error de estimación se convierte en:

$$\varepsilon = z\sqrt{\frac{pq}{n}}\sqrt{\frac{N-n}{N-1}}$$

Que al despejar n se obtiene:

$$n = \frac{z^2 pqN}{\varepsilon^2 (N-1) + z^2 pq}$$

Ejemplo

En una muestra aleatoria de 500 familias que tienen televisores en la ciudad de México, se encuentra que 340 están suscritas a HBO. ¿Qué tan grande se requiere que sea una muestra si se quiere tener 95% de confianza de que la estimación de P esté dentro de 0.02?

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-4.0	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00002	0.00002	0.00002	0.00002
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
-3.1	0.00097	0.00094	0.00090	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3.0	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00103	0.00100
-2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
-2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
-2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
-2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
-2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
-2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
-2.3	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
-2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
-2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
-2.0	0.02275	0.02222	0.02169	0.02118	0.02067	0.02018	0.01970	0.01923	0.01876	0.01831
-1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
-1.7	0.04456	0.04363	0.04272	0.04181	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
-1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551
-1.5	0.06681	0.06552	0.06425	0.06301	0.06178	0.06057	0.05938	0.05821	0.05705	0.05592
-1.4	0.08076	0.07927	0.07780	0.07636	0.07493	0.07353	0.07214	0.07078	0.06944	0.06811
-1.3	0.09680	0.09510	0.09342	0.09176	0.09012	0.08851	0.08691	0.08534	0.08379	0.08226
-1.2	0.11507	0.11314	0.11123	0.10935	0.10749	0.10565	0.10383	0.10204	0.10027	0.09852
-1.1	0.13566	0.13350	0.13136	0.12924	0.12714	0.12507	0.12302	0.12100	0.11900	0.11702
-1.0	0.15865	0.15625	0.15386	0.15150	0.14917	0.14686	0.14457	0.14231	0.14007	0.13786
-0.9	0.18406	0.18141	0.17878	0.17618	0.17361	0.17105	0.16853	0.16602	0.16354	0.16109
-0.8	0.21185	0.20897	0.20611	0.20327	0.20045	0.19766	0.19489	0.19215	0.18943	0.18673
-0.7	0.24196	0.23885	0.23576	0.23269	0.22965	0.22663	0.22363	0.22065	0.21769	0.21476
-0.6	0.27425	0.27093	0.26763	0.26434	0.26108	0.25784	0.25462	0.25143	0.24825	0.24509
-0.5	0.30853	0.30502	0.30153	0.29805	0.29460	0.29116	0.28774	0.28434	0.28095	0.27759
-0.4	0.34457	0.34090	0.33724	0.33359	0.32997	0.32635	0.32276	0.31917	0.31561	0.31206
-0.3	0.38209	0.37828	0.37448	0.37070	0.36692	0.36317	0.35942	0.35569	0.35197	0.34826
-0.2	0.42074	0.41683	0.41293	0.40904	0.40516	0.40129	0.39743	0.39358	0.38974	0.38590
-0.1	0.46017	0.45620	0.45224	0.44828	0.44433	0.44038	0.43644	0.43250	0.42857	0.42465
-0.0	0.50000	0.49601	0.49202	0.48803	0.48404	0.48006	0.47607	0.47209	0.46811	0.46414