Universidad Nacional de Colombia-Sede Medellín

Pacial-1 Muestreo Estadístico. **Fecha: 2021-03-24**

Nombre Completo: ______________________________

Firma: ______________________ **C.C** ______________________

1. **fun**

1. **Ejercicio**
   Para los hogares de los empleados de la sección de personal de una entidad bancaría en una ciudad se desea hacer un estudio para estimar el **gasto promedio por hogar** (en S.M.L.V), **el gasto promedio por hogar para hogares con y sin niños** y la **proporción de hogares con niños**. El investigador desea hacer la estimación usando MAS sin reemplazo. Suponga que la tabla que se anexa a continuación reúne los datos asociados a todos los hogares de los empelados (Población).

Código	Gastos	Niños	Código	Gastos	Niños	Código	Gastos	Niños	Código	Gastos	Niños
1	0.89	1	51	1.04	1	101	0.98	1	151	1.17	1
2	1.05	1	52	1.12	1	102	1.14	1	152	0.81	0
3	1.05	1	53	1.07	1	103	1.19	1	153	1.14	1
4	0.75	0	54	1.05	1	104	0.81	1	154	0.93	0
5	0.98	1	55	0.80	0	105	1.00	1	155	1.03	1
6	1.14	1	56	1.01	0	106	1.29	1	156	1.04	1
7	1.19	1	57	0.72	0	107	1.16	1	157	0.93	0
8	1.38	1	58	1.14	1	108	1.14	1	158	1.11	1
9	0.85	0	59	0.93	0	109	0.94	1	159	0.93	0
10	0.99	0	60	0.81	1	110	0.91	1	160	0.91	0
11	0.97	0	61	0.84	0	111	0.84	1	161	1.07	1
12	0.96	1	62	1.01	1	112	1.05	1	162	1.19	1
13	1.01	1	63	0.92	1	113	0.94	1	163	1.20	1
14	1.05	1	64	1.11	1	114	1.22	1	164	1.12	1
15	1.12	1	65	0.88	0	115	0.82	1	165	1.10	1
16	1.04	1	66	1.16	1	116	1.28	1	166	1.11	1
17	0.77	1	67	0.74	1	117	1.03	1	167	0.92	0
18	0.82	0	68	1.11	1	118	0.89	1	168	1.17	1
19	0.81	0	69	1.14	1	119	1.07	0	169	1.06	1
20	0.96	1	70	1.22	1	120	1.03	0	170	0.75	0
21	0.93	0	71	0.87	1	121	1.22	1	171	1.09	1
22	1.16	1	72	1.11	1	122	1.09	1	172	1.18	1
23	1.23	0	73	1.15	1	123	1.21	1	173	1.11	1
24	0.59	1	74	0.63	1	124	1.25	1	174	1.19	1
25	1.09	1	75	0.95	1	125	1.06	0	175	1.00	1
26	1.07	1	76	0.85	1	126	1.00	0	176	0.74	0
27	0.74	0	77	1.19	1	127	1.10	0	177	0.77	0
28	1.07	1	78	1.02	1	128	0.89	0	178	1.19	1
29	0.90	1	79	0.96	1	129	0.98	1	179	0.81	0
30	1.12	1	80	0.93	0	130	0.72	0	180	1.20	1
31	0.68	0	81	1.07	0	131	1.01	0	181	0.80	1
32	1.02	1	82	0.89	1	132	0.93	0	182	0.96	1
33	1.00	1	83	0.99	1	133	0.82	0	183	1.00	1
34	1.22	1	84	1.31	1	134	0.88	0	184	1.10	1
35	0.88	1	85	1.12	1	135	1.22	1	185	1.02	1
36	1.23	1	86	1.16	1	136	1.01	1	186	0.76	0
37	1.01	1	87	1.02	1	137	0.93	1	187	1.01	1
38	1.07	1	88	1.07	1	138	1.02	1	188	0.69	0
39	0.88	1	89	0.85	0	139	1.17	1	189	0.88	0
40	0.98	1	90	0.75	0	140	1.06	1	190	1.02	1
41	1.04	1	91	0.88	1	141	0.97	1	191	0.94	1
42	1.08	1	92	1.03	1	142	0.92	1	192	1.12	1
43	0.72	1	93	0.94	1	143	0.97	1	193	1.00	1
44	0.89	1	94	0.95	1	144	1.34	1	194	0.79	0
45	1.19	0	95	1.18	1	145	0.79	0	195	0.77	0
46	1.02	0	96	0.98	1	146	0.98	1	196	1.05	1
47	1.08	1	97	0.93	1	147	0.84	0	197	1.01	1
48	1.27	1	98	1.07	1	148	0.99	0	198	1.05	1
49	1.12	1	99	0.80	0	149	1.27	1	199	0.74	0

Código	Gastos	Niños	Código	Gastos	Niños	Código	Gastos	Niños	Código	Gastos	Niños
50	0.92	1	100	1.07	0	150	0.81	0	200	1.27	1

Las descripción de las columnas de la tabla anterior corresponden a:

Columna 1: Código del Hogar, *Columna 2*: Gasto en salarios mínimos legales vigentes S.M.L.V,

Columna 3: Niños en el Hogar: **1**-sí hay niños en el hogar y **0**-sí no hay niños en el hogar.

(a) Determine el tamaño de muestra necesario para realizar la estimación del gasto promedio por hogar para la población en estudio con un límite en el error de estimación de B=0.06 S.M.L.V. y con un nivel de confianza del: 99 %.

Si necesita algunas estimaciones iniciales para poder responder a esta pregunta, utilice la siguiente muestra piloto:

Códigos de hogares de la Muestra piloto: 3, 65, 60, 24, 105, 136, 67, 9, 95, 69

.

b) Utilizando la muestra de hogares cuyos códigos se dan a continuación, realizar la estimación del gasto promedio por hogar para la población en estudio junto con su respectivo intervalo de confinaza del 99 %.

.

Códigos de hogares de la Muestra dada para realizar las estimaciones de los puntos (b), (c) y (d): 143, 177, 138, 80, 164, 57, 8, 101, 26, 35, 173, 161, 85, 44, 103, 95, 194, 157, 1, 39, 54, 167, 96, 58, 153, 113, 156, 4, 180, 200, 104, 124, 37

.

c) Con la muestra dada en (b) realice la estimación del gasto promedio por hogar para los hogares Sin Niños de la población en estudio junto con su intervalo de confianza del 99 % .

.

d) Con la muestra dada en (b) realice la estimación del porcentaje de hogares Con Niños para la población en estudio junto con su intervalo de confianza del 99 %.

.

Valores cíticos de la tabla Normal Estándar Z

α	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
$\alpha/2$	0.005	0.01	0.015	0.02	0.025	0.03	0.035	0.04	0.045	0.05
$1-\alpha/2$	0.995	0.99	0.985	0.98	9.975	0.97	0.965	0.96	0.955	0.95
$Z_{1-\alpha/2}$	2.575829	2.326348	2.17009	2.053749	1.959964	1.880794	1.811911	1.750686	1.695398	1.644854

Para los valroes críticos de la $t_{\alpha/2;n-1}$, lo pueden buscar en R con el código: qt(1-α/2,n-1)

Solución

(a) **El tamaño de muestra requerido es de:** $n = 48.0207947 \approx 49$.

Para encontrar este tamaño de muestra se puede utilizar la expresión:

$$n = \frac{N\sigma^2}{(N-1)\frac{B^2}{Z^2_{\alpha/2}} + \sigma^2} = \frac{N\sigma^2}{(N-1)D + \sigma^2},$$

$$\text{donde:} \quad D = \frac{B^2}{Z^2_{\alpha/2}},$$

con:

B=Límite para error de estimación permitido=0.06,

$$Z_{1-\alpha/2} = \text{Percentil de la normal estándar} = Z_{1-\alpha/2} = Z_{0.995} = 2.575829,$$

$$\sigma^2 = S^2 = \text{Varianza de la muestra piloto} = 0.03411667,$$

y $N = 200$.

Para calular S^2-se usa la muestra piloto dada por:

	c(1:10)	Gastos	Niños
3	1	1.05	1
65	2	0.88	0
60	3	0.81	1
24	4	0.59	1
105	5	1.00	1
136	6	1.01	1
67	7	0.74	1
9	8	0.85	0
95	9	1.18	1
69	10	1.14	1

de donde $\bar{y} = 0.925$ y por lo tanto:

$$S^2 = \frac{1}{n_0 - 1}\sum_{i=1}^{n_0}\Big(y_i - \bar{y}\Big)^2 = \frac{1}{10-1}\sum_{i=1}^{10}\Big(y_i - \bar{y}\Big)^2$$

$$= \frac{1}{9}\sum_{i=1}^{10}\Big(y_i - \bar{y}\Big)^2$$

$$= \frac{1}{9}\left[\Big(1.05 - 0.925\Big)^2 + \cdots\cdots\cdots + \Big(1.14 - 0.925\Big)^2\right]$$

$$= \frac{1}{9}\Big(0.015625 + 0.002025 + \cdots\cdots\cdots + 0.046225\Big)$$

$$\hat{\sigma}^2 = S^2 = 0.03411667$$

De donde reemplazando los respectivos valores se tiene que:

$$n = \frac{N\sigma^2}{(N-1)\frac{B^2}{Z^2_{\alpha/2}} + \sigma^2} = \frac{200 * 0.034117}{(200-1)\frac{(0.06)^2}{(2.575829)^2} + 0.034117}$$

$$= \frac{200 * 0.034117}{(200-1)\frac{0.0036}{6.634897} + 0.034117}$$

$$n = 48.0207947 \approx 49.$$

(b) **Las respuestas de este punto son:**

$\hat{\mu} = \bar{y} = 1.0118182$ con un IC del 99 % dado por: $(Li\ ,\ Ls) = (0.945961\ \ ,\ \ 1.077675)$.

Veamos como se hallan esta respuestas.

Primero se halla la estimación del gasto promedio por hogar utilizando la muestra asignada dada por:

	c(1:nrow(datos_muestra))	Gastos	Niños
143	1	0.97	1
177	2	0.77	0
138	3	1.02	1
80	4	0.93	0
164	5	1.12	1
57	6	0.72	0
8	7	1.38	1
101	8	0.98	1
26	9	1.07	1
35	10	0.88	1
173	11	1.11	1
161	12	1.07	1
85	13	1.12	1
44	14	0.89	1
103	15	1.19	1
95	16	1.18	1
194	17	0.79	0
157	18	0.93	0
1	19	0.89	1
39	20	0.88	1
54	21	1.05	1
167	22	0.92	0
96	23	0.98	1
58	24	1.14	1
153	25	1.14	1
113	26	0.94	1
156	27	1.04	1
4	28	0.75	0
180	29	1.20	1
200	30	1.27	1
104	31	0.81	1
124	32	1.25	1
37	33	1.01	1

El n-utilizado es: $n = 33$, y el gasto promedio estimado para los hogares del barrio en estudio es de:

$$\hat{\mu} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{33}\sum_{i=1}^{33} y_i$$

$$= \frac{1}{33}\Big[0.97 + 0.77 + \cdots\cdots\cdots + 1.25 + 1.01\Big]$$

$$\bar{y} = 1.011818$$

Ahora para hallar el Intervalo de Confianza, primero hallamos el Límite Para el Error de

Estimación, B, definido como:

$$B = c\sqrt{\widehat{Var}[\bar{y}]} = c\sqrt{\left(\frac{N-n}{N}\right)\frac{S^2}{n}},$$

con $c = Z_{1-\alpha/2}$ o $c = t_{1-\alpha/2\ ;\ n-1}$, para este caso se utiliza la Z , es decir: $c = 2.575829$ y

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = \frac{1}{33-1}\sum_{i=1}^{33}\left(y_i - \bar{y}\right)^2$$

$$= \frac{1}{32}\sum_{i=1}^{33}\left(y_i - \bar{y}\right)^2$$

$$= \frac{1}{32}\left[\left(0.97 - 1.011818\right)^2 + \left(0.77 - 1.011818\right)^2 + \cdots\cdots\cdots + \left(1.01 - 1.011818\right)^2\right]$$

$$= \frac{1}{32}\left(0.00174876 + 0.05847603 + \cdots\cdots\cdots + 0.000003305785\right)$$

$$S^2 = 0.02583409$$

y para B se tiene que:

$$B = c\sqrt{\widehat{Var}[\bar{y}]}$$

$$= c\sqrt{\left(\frac{N-n}{N}\right)\frac{S^2}{n}}$$

$$= (2.575829)\sqrt{\left(\frac{200-33}{200}\right)\frac{0.02583409}{33}}$$

$$= (2.575829)\sqrt{0.0006536808}$$

$$B = 0.06585669$$

y por último el IC del 99 % está dado por:

$$\bar{y} \pm B \Longleftrightarrow 1.0118182 \pm 0.06585669 \Longleftrightarrow \left(\ 0.945961\ ,\ 1.077675\right)$$

(c) **Las respuesta a este punto son las siguientes:**

El gasto promedio estimado para los hogares con niños es de:

$\hat{\mu}_1 = \bar{y}_1 = 1.0607692$ con un IC del 99 % dado por: $(Li\ ,\ Ls) = (\ 0.992063\ ,\ 1.129475\ )$.

Similarmente, el gasto promedio estimado para los hogares sin niños es de:

$\hat{\mu}_2 = \bar{y}_2 = 0.83$ y un IC del 99 % dado por: $(Li, Ls) = ($ 0.708367 , 0.951633 $)$.

Veamos como se hallan esta respuestas.

Primero se halla la estimación del gasto promedio por hogar para las familias con hijos utilizando la muestra asignada dada por:

	c(1:nrow(datos_muestra))	Gastos	Niños
143	1	0.97	1
177	2	0.77	0
138	3	1.02	1
80	4	0.93	0
164	5	1.12	1
57	6	0.72	0
8	7	1.38	1
101	8	0.98	1
26	9	1.07	1
35	10	0.88	1
173	11	1.11	1
161	12	1.07	1
85	13	1.12	1
44	14	0.89	1
103	15	1.19	1
95	16	1.18	1
194	17	0.79	0
157	18	0.93	0
1	19	0.89	1
39	20	0.88	1
54	21	1.05	1
167	22	0.92	0
96	23	0.98	1
58	24	1.14	1
153	25	1.14	1
113	26	0.94	1
156	27	1.04	1
4	28	0.75	0
180	29	1.20	1
200	30	1.27	1
104	31	0.81	1
124	32	1.25	1
37	33	1.01	1

Para esta muestra se tiene que: $n_1 = 26$ y $n_0 = 7$, los cual corresponde a los tamaños de las dos sub-poblaciones o dominios en la muestra que corresponden a hogares con y sin niños respectivamente.

La subpoblación de hogares con niños es:

	c(1:n1)	Gastos	Niños
143	1	0.97	1
138	2	1.02	1
164	3	1.12	1
8	4	1.38	1
101	5	0.98	1
26	6	1.07	1
35	7	0.88	1
173	8	1.11	1
161	9	1.07	1
85	10	1.12	1
44	11	0.89	1
103	12	1.19	1
95	13	1.18	1
1	14	0.89	1
39	15	0.88	1
54	16	1.05	1
96	17	0.98	1
58	18	1.14	1
153	19	1.14	1
113	20	0.94	1
156	21	1.04	1
180	22	1.20	1
200	23	1.27	1
104	24	0.81	1
124	25	1.25	1
37	26	1.01	1

luego el gasto promedio estimado para los hogares con niños está dado por:

$$\hat{\mu}_1 = \bar{y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_i = \frac{1}{26}\sum_{i=1}^{26} y_i$$

$$= \frac{1}{26}\left[0.97 + 1.02 + \cdots\cdots\cdots + 1.25 + 1.01\right]$$

$$\bar{y}_1 = 1.0607692$$

Ahora para hallar el Intervalo de Confianza, primero hallamos el Límite Para el Error de Estimación, B_1, definido como:

$$B_1 = c\sqrt{\widehat{Var}[\bar{y}_1]} = c\sqrt{\left(\frac{N_1 - n_1}{N_1}\right)\frac{S_1^2}{n_1}},$$

con $c = Z_{\alpha/2}$ o $c = t_{1-\alpha/2\ ;\ n_1-1}$, para este caso se utiliza la t , $c = 2.7874358$ y

$$S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}\left(y_i - \bar{y}\right)^2 = \frac{1}{26-1}\sum_{i=1}^{26}\left(y_i - \bar{y}\right)^2$$

$$= \frac{1}{25}\sum_{i=1}^{26}\left(y_i - \bar{y}\right)^2$$

$$= \frac{1}{25}\left[\left(0.97 - 1.060769\right)^2 + \left(1.02 - 1.060769\right)^2 + \cdots\cdots\cdots + \left(1.01 - 1.060769\right)^2\right]$$

$$= \frac{1}{25}\left(0.008239053 + 0.00166213 + \cdots\cdots\cdots + 0.002577515\right)$$

$$S_1^2 = 0.01924738$$

y para B_1 se tiene que:

$$B_1 = c\sqrt{\widehat{Var}[\bar{y}_1]}$$

$$= c\sqrt{\left(\frac{N_1 - n_1}{N_1}\right)\frac{S_1^2}{n_1}}$$

$$= (2.7874358)\sqrt{\left(\frac{145-26}{145}\right)\frac{0.01924738}{26}}$$

$$= (2.7874358)\sqrt{0.0006075434}$$

$$B_1 = 0.06870582$$

y por último el IC del 99 % está dado por:

$$\bar{y}_1 \pm B_1 \iff 1.0607692 \pm 0.06870582 \iff \left(\ 0.992063\ ,\ 1.129475\right)$$

La subpoblación de hogares sin niños es:

	c(1:n2)	Gastos	Niños
177	1	0.77	0
80	2	0.93	0
57	3	0.72	0
194	4	0.79	0
157	5	0.93	0
167	6	0.92	0
4	7	0.75	0

luego el gasto promedio estimado para los hogares sin niños está dado por:

$$\hat{\mu}_2 = \bar{y}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} y_i = \frac{1}{7}\sum_{i=1}^{7} y_i$$

$$= \frac{1}{7}\Bigg[0.77 + 0.93 + \cdots\cdots\cdots + 0.92 + 0.75\Bigg]$$

$$\bar{y}_2 = 0.83$$

Ahora para hallar el Intervalo de Confianza, primero hallamos el Límite Para el Error de Estimación, B_2, definido como:

$$B_2 = c\sqrt{\widehat{Var}[\bar{y}_2]} = c\sqrt{\left(\frac{N_2 - n_2}{N_2}\right)\frac{S_2^2}{n_2}},$$

con $c = Z_{\alpha/2}$ o $c = t_{1-\alpha/2\ ;\ n_2-1}$, para este caso se utiliza la t , $c = 3.707428$ y

$$S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}\Big(y_i - \bar{y}\Big)^2 = \frac{1}{7-1}\sum_{i=1}^{7}\Big(y_i - \bar{y}\Big)^2$$

$$= \frac{1}{6}\sum_{i=1}^{7}\Big(y_i - \bar{y}\Big)^2$$

$$= \frac{1}{6}\Bigg[\Big(0.77 - 0.83\Big)^2 + \Big(0.93 - 0.83\Big)^2 + \cdots\cdots\cdots + \Big(0.75 - 0.83\Big)^2\Bigg]$$

$$= \frac{1}{6}\Big(0.0036 + 0.01 + \cdots\cdots\cdots + 0.0064\Big)$$

$$S_2^2 = 0.008633333$$

y para B_2 se tiene que:

$$B_2 = c\sqrt{\widehat{Var}[\bar{y}_2]}$$

$$= c\sqrt{\left(\frac{N_2 - n_2}{N_2}\right)\frac{S_2^2}{n_2}}$$

$$= (3.707428)\sqrt{\left(\frac{55 - 7}{55}\right)\frac{0.008633333}{7}}$$

$$= (3.707428)\sqrt{0.001076364}$$

$$B_2 = 0.1216332$$

y por último el IC del 99 % está dado por:

$$\bar{y}_2 \pm B_2 \Longleftrightarrow 0.83 \pm 0.1216332 \Longleftrightarrow \Big(\ 0.708367\ ,\ 0.951633\Big)$$

(d) **Las respuestas para este puntos son las siguientes:**

La proporción estimada de hogares con niños en el barrios de estudio es de:

$\hat{P} = \hat{p} = \frac{a}{n} = \frac{26}{33} = 0.7878788 = 78.7878788$ % con un IC del 99 % dado por:

$(Li \ , \ Ls) = (\ 0.617778 \ \ , \ \ 0.95798 \ )$.

Para el número total de hogares con niños en el barrio de estudio se tienen las siguientes estimaciones:

$\hat{A} = N\hat{p} = 200 * 0.7878788 = 157.5757576 = 158$ con un IC del 99 % dado por:

$(Li \ , \ Ls) = (\ \ 123.5556 \ \ , \ \ 191.596 \ )=(\ \ 124 \ \ , \ \ 192 \ )$.

Veamos como se hallan esta respuestas.

Primero se halla la proporción estimada de hogares con niños en el barrios de estudio utilizando la muestra asignada dada por:

	c(1:nrow(datos_muestra))	Gastos	Niños
143	1	0.97	1
177	2	0.77	0
138	3	1.02	1
80	4	0.93	0
164	5	1.12	1
57	6	0.72	0
8	7	1.38	1
101	8	0.98	1
26	9	1.07	1
35	10	0.88	1
173	11	1.11	1
161	12	1.07	1
85	13	1.12	1
44	14	0.89	1
103	15	1.19	1
95	16	1.18	1
194	17	0.79	0
157	18	0.93	0
1	19	0.89	1
39	20	0.88	1
54	21	1.05	1
167	22	0.92	0
96	23	0.98	1
58	24	1.14	1
153	25	1.14	1
113	26	0.94	1
156	27	1.04	1
4	28	0.75	0
180	29	1.20	1
200	30	1.27	1
104	31	0.81	1
124	32	1.25	1
37	33	1.01	1

El n-utilizado es: $n = 33$, y la proporción estimada de hogares con niños en el barrio de estudio es de:

$$\hat{P} = \hat{p} = = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{33}\sum_{i=1}^{33} y_i$$

$$= \frac{1}{33}\Bigg[1 + 0 + \cdots\cdots\cdots + 1 + 1\Bigg]$$

$$\hat{p} = 0.787879$$

Ahora para hallar el Intervalo de Confianza, primero hallamos el Límite Para el Error de Estimación, B, definido como:

$$B = c\sqrt{\widehat{Var}[\hat{p}]} = c\sqrt{\left(\frac{N-n}{N}\right)\frac{\hat{p}(1-\hat{p})}{n-1}},$$

con $c = Z_{\alpha/2}$ o $c = t_{1-\alpha/2\ ;\ n-1}$, para este caso se utiliza la Z , $c = 2.575829$.

Para B se tiene que:

$$B = c\sqrt{\widehat{Var}[\hat{p}]}$$

$$= c\sqrt{\left(\frac{N-n}{N}\right)\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= (2.575829)\sqrt{\left(\frac{200-33}{200}\right)\frac{(0.7878788)(0.2121212)}{32}}$$

$$= (2.575829)\sqrt{0.004360939}$$

$$B = 0.1701011$$

y por último el IC del 99 % para P está dado por:

$$\hat{p}\ \pm B \iff 0.787879 \pm 0.1701011 \iff \Bigg(\ 0.617778\ ,\ 0.95798\Bigg)$$

Para el número total de hogares con niños en el barrio de estudio se tienen las siguientes estimaciones:

$$\hat{A} = N\hat{p} = (200)(0.787879) = 157.5757576 = 158$$

y para el Intervalo de Confianza se tine que:

$$N(\hat{p}\ \pm\ B) \iff (200)(0.787879\ \pm\ 0.1701011) \iff \Bigg(\ 123.5556\ ,\ 191.596\Bigg) \iff (124\ ,\ 192)$$