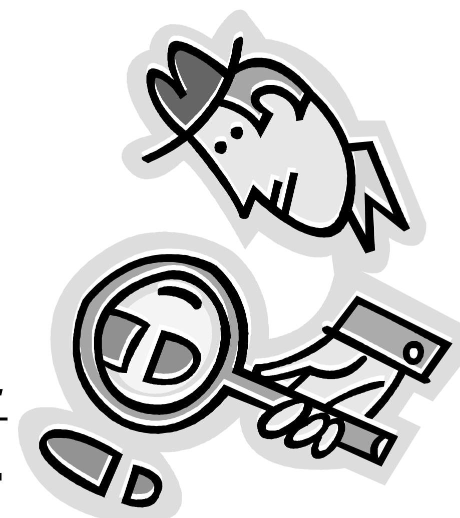
IN627
Investigación de Mercados

Auxiliar Nº4

- Tamaño de muestra
- Test hipótesis



Auxiliar:

- Mauricio Ramírez F.

Auxiliares:

- Gonzalo León (grupos 1 a 4)
- Martín Fuentes (grupos 5 a 8)

Agenda de hoy

I. Repaso de Muestreo

II. Tamaño de muestra

III. Análisis de Datos (test de hipótesis)

IV. Ejercicios

Muestreo Probabilistico

- Tiene como finalidad obtener información sobre una población
- Lineamientos a considerar:
  - Selección de la Población
  - Selección de la Muestra
  - Determinación del tamaño de la Muestra
  - Tratamiento de las “No Respuestas”

Muestreo Probabilistico

- Selección de la muestra:
  - Aleatorio Simple
  - Sistemático
  - Estratificado
  - Por conglomerados

Muestreo No probabilistico

- Métodos No probabilísticos:
  - Por conveniencia
    - Elementos dispuestos a participar
  - De criterio
    - Elementos que al parecer representan
    - Ej.: mercado de prueba, bola de nieve
  - Por cuotas
    - Se elige una cuota y se agregan elementos hasta completarlas

Tamaño de la muestra

- Decisión relacionada con el costo de la investigación
- Conceptos necesarios de considerar:
  - Características de la población
  - Características de la muestra
  - Confiabilidad de la muestra
  - Estimación del intervalo

Tamaño de la muestra

- Características de la población
  - μ = media de la población
  - Sigma^2 = varianza de la población
  - Sigma= desviación estándar de la población
- Características de la muestra

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \text{ media muestral}$$

$$s^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}\left[X_i - \overline{X}\right]^2 = \text{ varianza muestral}$$

Tamaño de la muestra

- Confiabilidad de la muestra
  - Media muestral y varianza muestral varían de muestra en muestra

- Error estándar de la muestra:

$$\sigma_{\overline{X}} = \frac{\sigma_x}{\sqrt{n}}$$

Tamaño de la muestra

- Estimación del intervalo
  - Dado que la media muestral varía, existe un error muestral, entonces se requiere dar una estimación de ese error.
  - La estimación, conociendo sigma es:

$$\bar{X} \pm e.m. = \text{estimacion de intervalo}$$

$$\bar{X} \pm z \cdot \sigma_{\bar{X}} = \bar{X} \pm \frac{z \cdot \sigma}{\sqrt{n}} = \text{estimacion de intervalo}$$

$$\bar{X} \pm 2 \cdot \sigma_{\bar{X}} = \bar{X} \pm \frac{2 \cdot \sigma}{\sqrt{n}} = \text{estimacion de intervalo al 95\%}$$

Tamaño de la muestra

- Estimación del intervalo
  - Si no se conoce sigma se usa "s".

- Finalmente se debe especificar el error muestral que se desea y el nivel de confianza deseado:

$$e.m. = \text{error de la muestra}$$

$$\frac{z \cdot \sigma}{\sqrt{n}} = e.m. \rightarrow n = \frac{z^2 \sigma^2}{(e.m.)^2}$$

Tamaño de la muestra

- Media:
  - Error = $z*S = z*((N-n/N)(s^2/n))^{\wedge}½$
  - $n=(Nz^2s^2)/(Ne^2+z^2s^2)$
- Totales:
  - Error = $z*S_Y = z*(N(N-n)(s^2/n))^{\wedge}½$
  - $n = (N^2z^2s^2)/(e^2+Nz^2s^2)$
- Proporción:
  - Error = $z*S_P = z*((N-n/N-1)(pq/n))^{\wedge}½$
  - $n = (z^2pqN)/(e^2(N-1)+z^2pq)$

Test de hipótesis

- Para proporciones de una muestra

- Para proporciones de 2 muestras independientes

- Para proporciones para más de 2 muestras independientes

Test de hipótesis

- Para la media de una muestra
- Para las medias de 2 muestra
- Para medias con más de 2 muestras
- Para la varianza de 2 muestras

Test de hipótesis

- Muestras relacionadas
  - (la misma muestra en distintos momentos del tiempo)
- Para media entre 2 muestras relacionadas
- Para proporciones entre 2 muestras relacionadas

Ejercicio 1

- **Pregunta 2B (C1 – Prim 2004)**

- Cuando W.Sabor aún era asesor de Kike, realizó una encuesta de opinión para evaluar el apoyo que el rey Kike tenía entre la población de MCC antes de agrupar las cuatro diosas.

- Ahora el rey Kike le ha pedido a usted, como su asesor personal, que determine si efectivamente la agrupación de diosas que usted sugirió y que el rey de MCC hizo efectiva tuvo un efecto positivo en el apoyo mostrado por la población hacia él. Para esto le ha encargado aplicar nuevamente la encuesta de opinión (igual a la utilizada por W.Sabor) sobre los mismos súbditos que la contestaron la primera vez. Los resultados de ambas encuestas se muestran en la **Tabla Nº3.**

Ejercicio 1

Tabla Nº3: Apoyo (valor = 1) o no apoyo (valor = 0)
antes y después de realizar la agrupación de diosas

n	Actitud Previa (yi1)	Actitud Posterior (yi2)
1	1	0
2	1	0
3	1	1
4	1	1
5	1	1
6	0	1
7	0	1
8	0	1
9	0	0
10	0	0
Proporción apoyo	50%	60%
desv. Est.	0,53	0,52

Hint 1: Suponga que todas las estimaciones requieren un nivel de confianza del 95%.

Ejercicio 1

- Determine el test de hipótesis a aplicar sobre los resultados de las encuestas realizadas (explicite cuál es el que se debe utilizar) y plantee claramente: hipótesis nula, hipótesis alternativa y valor del parámetro de comparación (test).
- Calcule el valor del test escogido en la parte 1 y establezca si la agrupación y asignación de los conjuntos de diosas particulares a cada ciudad tuvo o no un efecto positivo estadísticamente significativo en el apoyo al rey Kike entre los pobladores de MCC.

Ejercicio 1

- Respuesta:

Hipótesis nula: H0: no hubo efecto positivo: $p2 \leq p1$
Hipótesis alternativa: H1: si hubo efecto positivo: $p2 > p1$

El test que corresponde utilizar es el de comparación de proporciones relacionadas:

$$z = \frac{p_2 - p_1}{\left[\frac{b + c - \left\{ (b-c)^2 / n \right\}}{n(n-1)} \right]^{1/2}}$$

		Antes		
		Sí	**No**	**Total**
Después	**Sí**	a	b	a+b
	No	c	d	c+d
	Total	a+c	b+d	a+b+c+d

Cabe señalar que se trata de un test unidireccional., $Zc = Z1\text{-}\alpha 1$.

La condición de rechazo se da si: $Z > Zc$

Ejercicio 1

- **Respuesta:**

De la tabla Normal, se encuentra el Zc = 1.64, correspondiente al 95% de confianza.

El cálculo del estadístico Z, se muestra en la siguiente tabla:

Tabla D: Calculo del estadístico Z

n	10
p1	50%
p2	60%
a	3
b	2
c	3
d	2
X = ((b-c)^2)/n	0,1
Y=[(b+c-X)/(n(n-1))]	0,054444444
Z=(p2-p1)/(Y^0,5)	**0,428571429**

Claramente, **Zc>Z** por lo tanto se encuentra dentro de la zona de aceptación de H0, y no hay efecto positivo estadísticamente significativo en el apoyo hacia el rey Kike

Ejercicio 2

- **PARTE 3B (C1 – Prim 2004)**
- Si Starbucks ha definido un consumo mínimo de 290 tasas de café al día para definir que un mercado es atractivo y así tomar la decisión de abrir una nueva tienda, ¿cuál debería ser el tamaño <u>mínimo</u> de la muestra utilizada por Starbucks para determinar el consumo diario de tasas de café en este mercado? Considere que este mercado está compuesto por 400 personas.
- Por simplicidad suponga que se trataría de un muestreo aleatorio simple.
- **Hint 2:** Utilice el muestreo de la Tabla Nº4 como muestra piloto del estudio.
- **Hint 3:** Suponga que todas las estimaciones son con un nivel del 95% de confianza.

Ejercicio 2

Tabla Nª4

i	Tasas de café consumidas diariamente por persona
1	0.25
2	0.50
3	0.00
4	0.75
5	0.25
6	0.00
7	0.25
8	1.25
9	0.50
10	1.25
11	0.75
12	0.50
13	1.00
14	2.00
15	0.75
16	0.75
17	0.50
18	0.00
19	0.75
20	0.25

Ejercicio 2

- Respuesta:

Como **n** = 20 de la prueba piloto es menor que 50 se debe utilizar el parámetro de una t-student de 1-α/2 y n-1 grados de libertad **(2,43).** Con esto se puede calcular el error piloto y comprobar que el consumo mínimo está dentro del intervalo de confianza y por lo tanto para el cálculo del **n** final se puede usar el parámetro de una normal de 1-α/2 (1,96).

Otra forma de hacerlo es suponer de inmediato normalidad y verificar si el **"n"** final es mayor que 50 con lo que se validaría la suposición.

Se debe utilizar el error estándar del total al cuadrado:

$$Error = e = K * \hat{S}_Y = K * \sqrt{N * (N - n) * (s^2 / n)} \Rightarrow n = \frac{N^2 K^2 s^2}{e^2 + N K^2 s^2}$$

Ejercicio 2

- Respuesta:

De los datos de la muestra piloto se obtiene:

Media	**0,6125**	
Varianza	**0,246546053**	
N	**400**	
Consumo min.	**290**	
Consumo estimado	**245**	**Media * 400**
e = C min. - C estimado	**45**	**290-245**
K = Z(1-α/2)	**1,96**	**se asume normalidad**
n	**63,04089326**	**≈ 64**

Dado que el valor de n es mayor que 50, se puede concluir que fue correcto haber asumido normalidad.

Ejercicio 3

- **Pregunta 2 (C1 – Prim 2003)**

- Soprole desea medir cuál es el efecto de una campaña publicitaria, de 1 mes de duración, para incentivar el consumo diario de leche en la población de Santiago. Para este propósito, la empresa midió el consumo promedio diario de litros de leche por persona en la población justo antes y después de poner al aire la campaña publicitaria, a través de una muestra aleatoria simple de 20 personas en cada ocasión.
- Una de las preguntas de filtro del cuestionario utilizado en las mediciones permitía que sólo contestaran la encuesta aquellas personas que no hubieran participado en un estudio de mercado en los últimos 60 días.
- Ahora bien, Soprole lo ha contratado a usted como <u>Analista de Estudios</u> para que determine si efectivamente la campaña publicitaria que salió al aire tuvo un efecto positivo en el consumo diario promedio de leche por persona en la población, a partir de las encuestas realizadas.

Ejercicio 3

- 1. Determine el test de hipótesis a aplicar sobre los resultados de las encuestas realizadas (explicite cuál es el que se debe utilizar) y plantee claramente: hipótesis nula, hipótesis alternativa y valor del parámetro de comparación (test).

- 2. Calcule el valor del test escogido en la parte 1 y establezca si la publicidad utilizada tuvo o no un efecto positivo estadísticamente significativo en el consumo promedio diario de leche por persona entre los santiaguinos.

- 3. Si la empresa ha definido un consumo mínimo de leche de 0.9 litros diarios promedio por persona para tomar la decisión de abrir una nueva planta embotelladora de leche, ¿cuál debería ser el tamaño mínimo de la muestra utilizada por Soprole para determinar el consumo diario promedio de leche por persona en la población de Santiago (aprox. 6,000,000 de hab.)?

- Por simplicidad suponga que se trataría de un muestreo aleatorio simple.

Ejercicio 3

Hint 1: Utilice el muestreo realizado después de la campaña publicitaria como muestra piloto del estudio.

Hint 2: Suponga que todas las estimaciones son con un nivel del 95% de confianza.

	Litros de leche consumidos diariamente por persona	
i	Antes de la campaña	Después de la campaña
1	0.25	1.25
2	0.50	1.50
3	0.00	2.00
4	0.75	1.75
5	0.25	1.50
6	0.00	0.25
7	0.25	2.00
8	1.25	1.25
9	0.50	0.00
10	1.25	0.50
11	0.75	1.25
12	0.50	0.50
13	1.00	0.50
14	2.00	0.00
15	0.75	0.25
16	0.75	0.75
17	0.50	1.00
18	0.00	1.25
19	0.75	1.75
20	0.25	1.00

Tabla 2: Muestra de consumo antes y después de la campaña.

Ejercicio 3

Respuesta:

Las hipótesis nula y alternativa vienen dadas por:

- *H*0: Los consumos medios de leche por persona no tuvieron un efecto positivo (no aumento) con la aplicación de la campaña ($\mu\ ant < \mu\ des$).
- *H*1: Los consumos medios de leche por persona tuvieron un efecto positivo (aumento) con la aplicación de la campaña ($\mu\ ant < \mu\ des$).

Dados que existen dos muestras y la intersección de los participantes en cada una de ellas es nula, se debe aplicar un Test para la media de dos muestras independientes:

$$t = \frac{\overline{x}_{\text{des}} - \overline{x}_{\text{ant}}}{s_{\overline{x}_{\text{des}} - \overline{x}_{\text{ant}}}} \qquad s_{\overline{x}_{\text{des}} - \overline{x}_{\text{ant}}} = \sqrt{\frac{s^2_{\text{des}}}{n_{\text{des}}} + \frac{s^2_{\text{ant}}}{n_{\text{ant}}}} \qquad t_c = t_{1-\alpha, n_{\text{ant}} + n_{\text{des}} - 2}$$

Notar que el test es unidireccional y por tanto en las tablas debe buscarse en las filas correspondientes a $1 - \alpha$ y no $1 - \alpha/2$

Ejercicio 3

- **Respuesta:**

De los datos se tiene que:

	antes	despues
$\overline{x}$	0.61	1.01
s	0.50	0.64
n	20	20

Con esto, el valor del test viene dado por:

$$
\begin{aligned}
t &= \frac{\overline{x}_{\text{des}} - \overline{x}_{\text{ant}}}{s_{\overline{x}_{\text{des}} - \overline{x}_{\text{ant}}}} \\
&= \frac{1{,}01 - 0{,}61}{0{,}18} \\
&= 2{,}21
\end{aligned}
$$

El test es unidireccional por lo que debemos buscar el puntaje a partir del cual la probabilidad acumulada alcanza un 95% en una tabla de una t-student de 38 grados de libertad, el que vale 2.02. Luego, rechazamos la hipótesis nula aceptando la alternativa.

Ejercicio 3

- Respuesta:

Del enunciado tenemos que:

- (0.5 puntos) $N = 6.000.000$
- (0.5 puntos) $e = \overline{x}_{\text{des}} - 0{,}9 = 0.11$
- (0.5 puntos) s =0.64
- (0.5 puntos) $K = 1.96$ (2.5 % en cada cola de la normal: recordemos que el valor de K es tal que el intervalo de confianza tenga la validez estadistica requerida y que por definición los intervalos de confianza son simétricos)

El tamaño de la muestra n viene dado por:

$$\begin{aligned} n &= \frac{NK^2s^2}{Ne^2 + K^2s^2} \\ &= \lceil 124{,}76 \rceil \\ &= 125 \end{aligned}$$