
TÉCNICAS DE MUESTREO I

Profesor: Ing. Celso Gonzales Ch. Mg.Sc

DEFINICIONES

CONTENIDO

- Introducción
- Tipos de muestreo: Probabilístico y No Probabilístico.
- Conceptos básicos Conceptos básicos en la investigación por muestreo: Población Objetivo y Muestreada
- Muestra, muestra piloto.
- Unidad de análisis, unidad de muestreo
- Marco de muestreo
- Instrumentos de medición
- Margen de error permisible
- Nivel de confianza

INTRODUCCION

La investigación puede ser:

- Prospectiva o retrospectiva
- Transversal y longitudinal
- Descriptiva o Comparativa
- Observación o experimentación

VENTAJAS DEL MUESTREO EN COMPARACIÓN CON LA ENUMERACIÓN COMPLETA.

- **Costo reducido**
- **Mayor rapidez**
- **Más posibilidades**
- **Mayor Exactitud**

DESVENTAJAS DEL MUESTREO EN COMPARACIÓN CON LA ENUMERACIÓN COMPLETA

- **Error de muestreo**
- **Personal especializado**

Objetivo fundamental de la Estadística

Hacer inferencias acerca de una población.

Objetivo de la inferencia

- Describir conjunto de mediciones
- Considerar la forma en que se puede hacer la inferencia acerca de la población.

MUESTREO

Proceso que nos permite la extracción de una muestra a partir de una población.

Hay dos tipos básicos de muestreo:

MUESTREO PROBABILÍSTICO.

En este tipo de muestreo, la probabilidad de aparición en una muestra de cualquier elemento de la población es conocida. Es el único científicamente válido.

MUESTREO NO PROBABILÍSTICO.

Es aquel en el que la selección de los elementos de la muestra no se hacen al azar

CONCEPTOS BÁSICOS EN LA INVESTIGACIÓN POR MUESTREO

- Población** : Conjunto de observaciones o datos obtenidos de una medición, conteo o cualidad de ciertos caracteres de los mismos.
- Población Objetivo:** Es la población que se intenta investigar o la cual se desea información.
- Pob. Muestreada** : Conjunto de elementos que serán muestreados.
- Marco de muestreo:** Relación completa de unidades de muestreo.
- U. de muestreo** : elementos o unidades de las que se trata obtener información.
- U. de análisis** : elementos de los que hay que obtener información.
- Muestra** : conjunto de unidades de muestreo sacadas del marco de muestreo.
- Muestra piloto** :

INSTRUMENTOS DE MEDICION

- Cuestionario
- Guía de entrevista
- Hoja de registro de observaciones
- Guía de moderador del grupo focal,
- Guía de análisis de contenido, etc

NIVEL DE CONFIANZA :

MARGEN DE ERROR PERMISIBLE O PRECISION

Es la variabilidad entre estadísticos calculados a partir de muestras aleatorias del mismo tamaño.

$$d = \pm t_{\left(1-\frac{\alpha}{2}, n-1\right)} s_y$$

EJEMPLO

De la Encuesta Demográfica y de Salud Familiar ENDES 2004:

Hogares de la ciudad de Lima.

Los Hogares particulares y sus miembros:

- Residentes habituales;
- Personas que pernoctaron en la vivienda la noche anterior al día de la entrevista.
- Todas las mujeres de 15 a 49 años de edad, sus hijos menores de 5 años

Cuestionario del Hogar

- **Tamaño y Composición del hogar**
- **Cobertura de Seguro**
- **Condición de Actividad**
- **Nivel Educativo**
- **Supervivencia y residencia de padres**
- **Características de las viviendas**
- **Tenencia del hogar**

Cuestionario Individual

- Antecedentes de la MEF
- Reproducción (Historia de Nacimientos)
- Anticoncepción
- Embarazo, Parto, Puerperio y Lactancia
- Inmunización y Salud
- Nupcialidad
- Preferencias de Fecundidad
- Antecedentes de Cónyuge y Trabajo de la MEF
- SIDA y otras ETS
- Mortalidad Materna
- Violencia Doméstica

Ejercicio

A fin de estudiar la posibilidad de brindar incentivos a la actividad ganadera en un pequeño valle de la sierra potencialmente rico de pastos, con un total de 4000 has. de superficie de labranza y pastos naturales, y agrupadas en 180 unidades agropecuarias(UA), se toma una muestra aleatoria simple de 20 U.A para investigar el número de cabezas de ganado por U.A.

- ¿Cual es la población objetivo?
- Marco de muestreo
- Unidad de Análisis Y de Muestreo

Ejercicio

Se quiere estimar el consumo promedio de agua en un distrito. Analice las siguientes unidades de muestreo:

- Familias individuales
- Unidades habitacionales(casas para una sola familia, edificios de apartamentos, etc.)
- Manzanas del distrito.

Qué usaría usted como marca de muestreo?

Ejercicio

La unidad agraria de una región de la costa desea estimar el área sembrada con maíz amarillo duro dentro de la región.

Sugiera posibles unidades de muestreo y marco de muestreo.

TIPO DE MUESTREO

Probabilístico, autoponderado, estratificado, multietápico e independiente para cada departamento.

MARCO MUESTRAL

Diseño multietápico: marco para cada etapa

UPM: Censo de Población y Vivienda de 1993

USM: y UTM: y Actualización Cartográfica y el Directorio de Viviendas y Establecimientos de 1999.

USO DE LA DISTRIBUCION NORMAL

Sea la población $\{ Y_i: Y_1, Y_2, \dots, Y_N \}$ se calcula S^2 , \bar{Y}
 \bar{Y} : promedio poblacional.

S^2 : variancia poblacional finita.

Con el fin de estimar el promedio poblacional, se extraen muestras posibles de tamaño n .

$$E(\bar{Y}_j) = \bar{Y}$$

$$V(\bar{Y}_j) = \left(\frac{N-n}{N}\right) \frac{S^2}{n}$$

Distribución muestral de la media

Veremos *primero* el caso de que la distribución subyacente sea normal, con media μ y varianza σ^2

La media de la distribución muestral de medias es μ

La varianza de la distribución muestral de medias es σ^2 / n

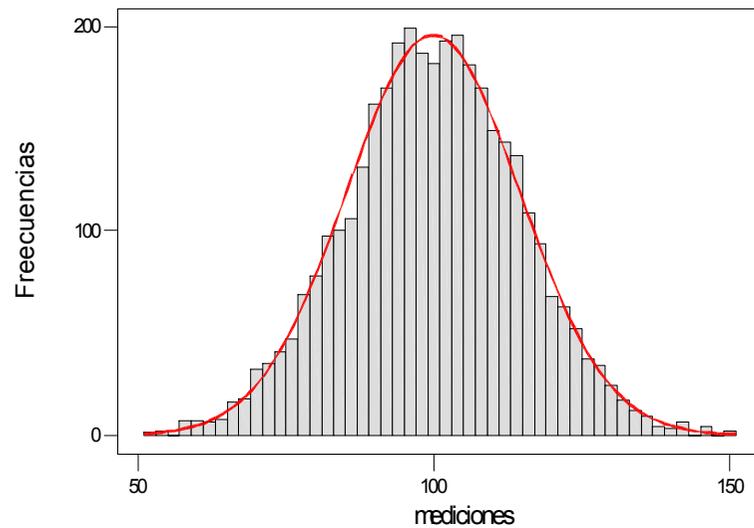
La forma de la distribución muestral de la media es normal.

Nota: La desviación estándar de la distribución muestral suele ser denominada: *error estándar* de tal estadístico: por ejemplo, “error estándar de la media”, etc.)

EJEMPLO 1

Distribución muestral de la media.

La línea (en este y sucesivos ejemplos) es una curva normal



Distribución de la población
(dist. Normal): $N = 3600$

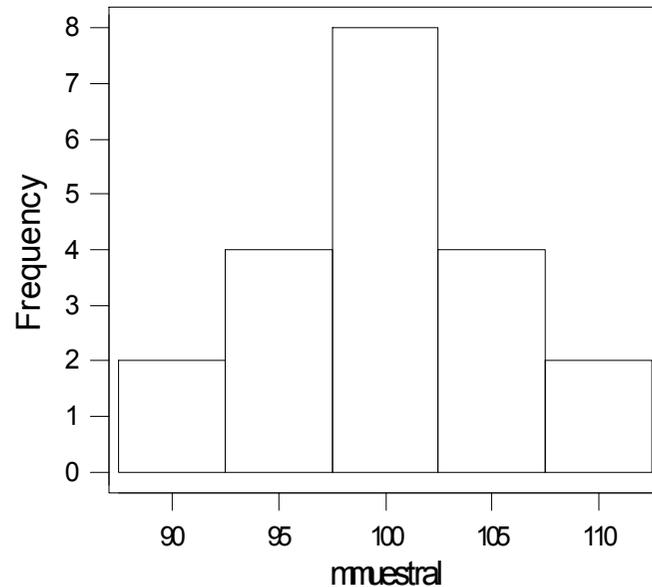
Media=100

Varianza=225

Desv.estándar=15

EJEMPLO 2

Distribución muestral de la media.



Muestras	Promedio
1	109.15
2	94.27
3	101.59
4	99.86
5	103.86
6	92.34
7	108.54
8	95.16
9	104.07
10	103.60
11	92.41
12	97.18
13	98.40
14	101.36
15	98.63
16	98.24
17	102.89
18	97.90
19	100.33
20	97.08

Distribución muestral de la media:

Tamaño de muestra=10

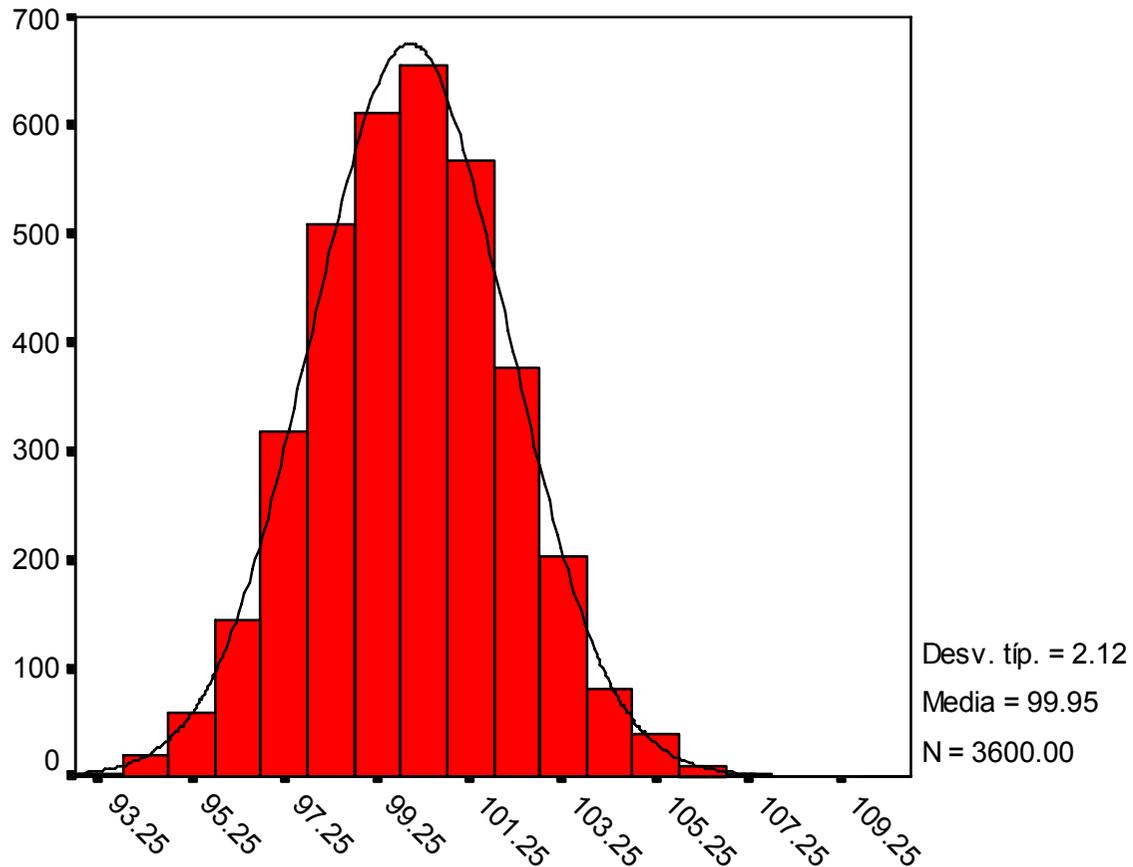
Media=100

(Varianza=22,4375)

Error estandar = $\sqrt{22.4375} = 4.74$

EJEMPLO 3

Distribución muestral de la media.



N50

Distribución poblacional
subyacente (dist. Normal):

Media=100

Desv.estándar =15

Distribución muestral de la
media:

Tamaño de muestra=50

Media=100

(Varianza= $225/50=4.5$)

Desv.estándar=2.12

Distribución muestral de la media.

Distribución poblacional
subyacente (dist. GAMMA):

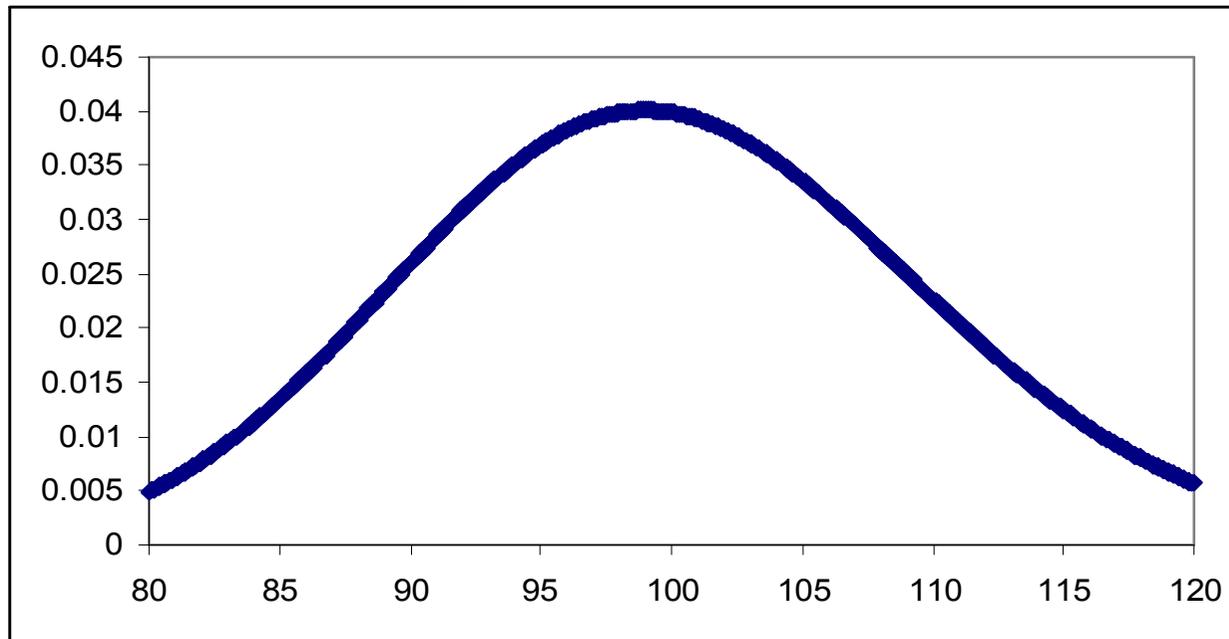
$$\text{Media}=100=\frac{p}{\lambda}=\frac{100}{1}=100$$

$$\text{Varianza}=100=\frac{p}{\lambda^2}=\frac{100}{1^2}=100$$

La distribución GAMMA tiene 2 parámetros:

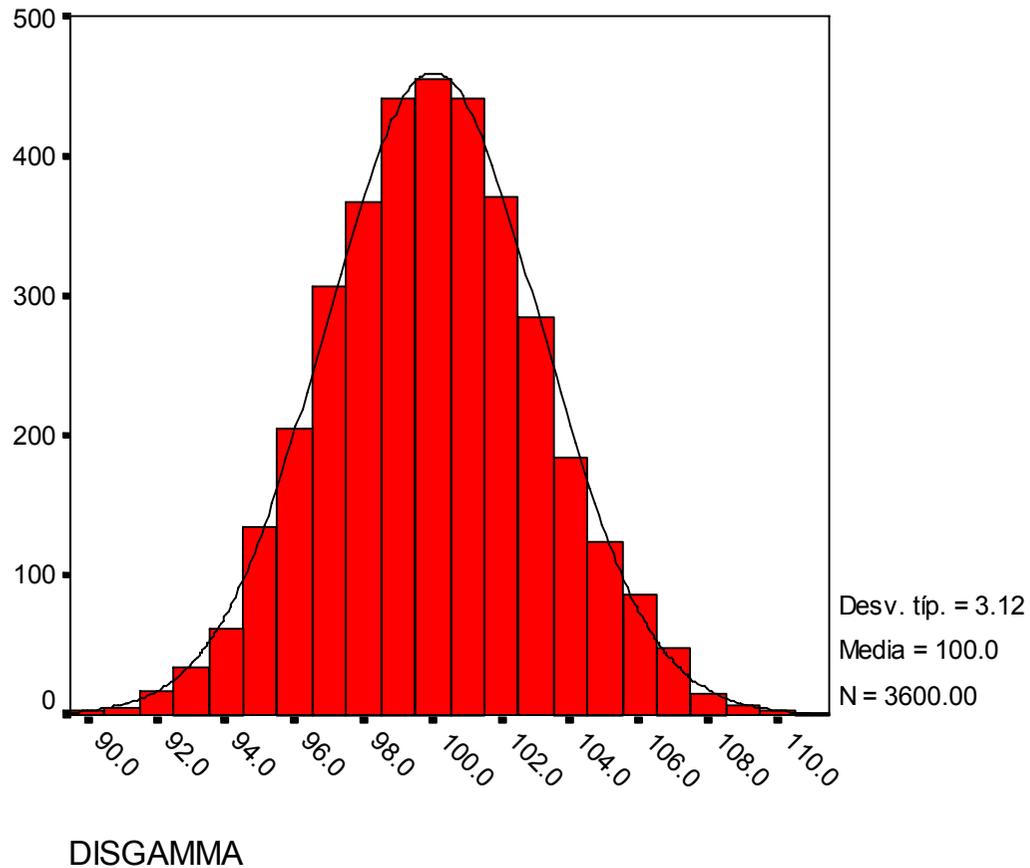
λ que es un parámetro de escala (1)

p que es un parámetro de forma (100)



EJEMPLO 4

Distribución muestral de la media.



Distribución poblacional
subyacente (dist. GAMMA):

Media=100

Varianza=100

Distribución muestral de la
media:

Tamaño de muestra =10

Media=100

Error estandar=

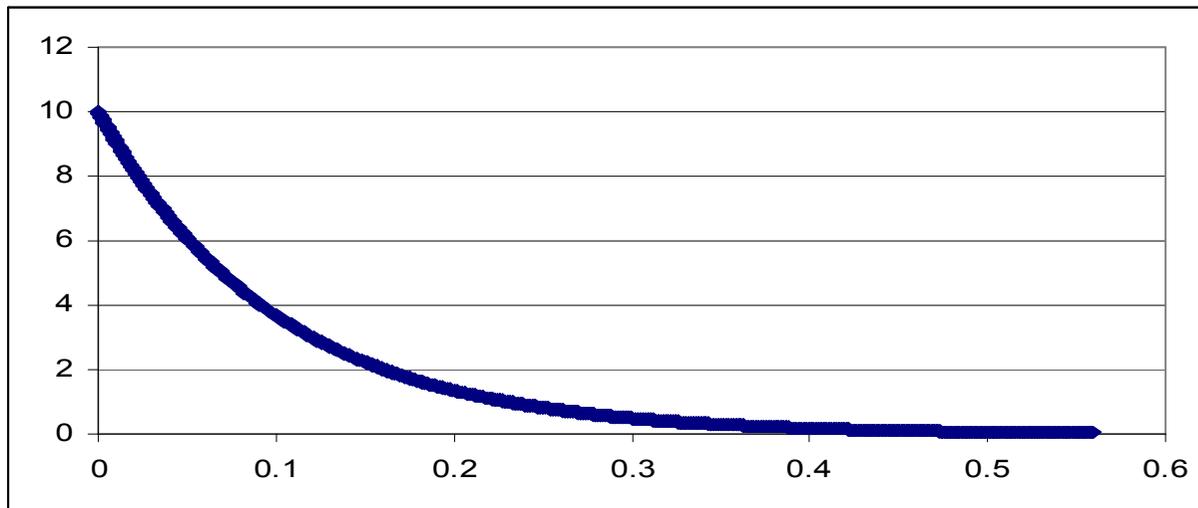
Distribución muestral de la media.

Distribución poblacional (dist. EXPONENCIAL):

$$\text{Media}=0.1=1/\lambda$$

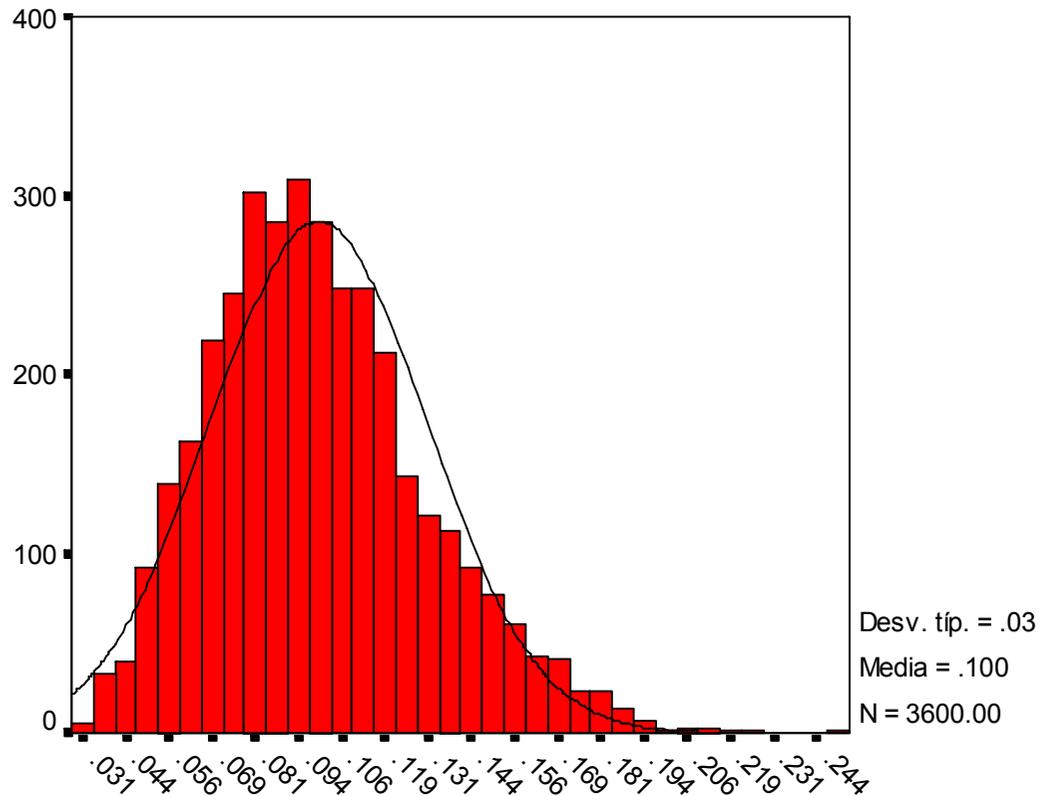
$$\text{Varianza}=0.01=1/\lambda^2$$

La distribución EXPONENCIAL tiene 1 parámetro: λ (en el ejemplo: 10)



EJEMPLO 5a

Distribución muestral de la media.



EXPON10

Distribución poblacional (dist. EXPONENCIAL):

$$\text{Media} = 0.1 = 1/\lambda$$

$$\text{Varianza} = 0.01 = 1/\lambda^2$$

Distribución muestral de la media:

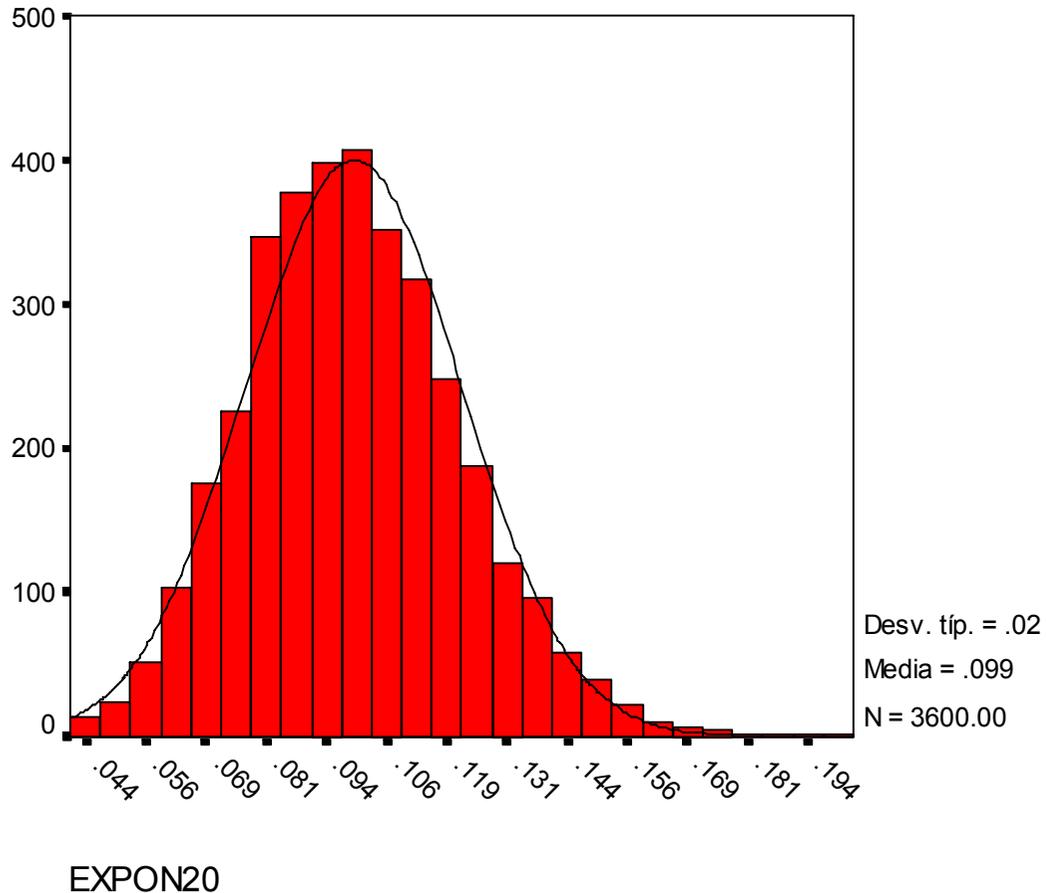
Tamaño de muestra=10

$$\text{Media} = .100$$

Error estándar=

EJEMPLO 5b

Distribución muestral de la media.



Distribución poblacional (dist. EXPONENCIAL):

$$\text{Media} = 0.1 = 1/\lambda$$

$$\text{Varianza} = 0.01 = 1/\lambda^2$$

Distribución muestral de la media:

Tamaño de muestra = 20

Media = .100

(Varianza = $0.01/20 = .0005$)

Desv. Estándar = .022

Observad que la distribución muestral se aproxima más a la normal (al elevar el tamaño muestral).

TRABAJO DE CAMPO: OBTENCIÓN DE INFORMACIÓN

- Probar las hipótesis
- Responder las preguntas formuladas
- Lograr los objetivos propuestos

PROCEDIMIENTO DE RECOLECCION DE INFORMACION

Escribir los procedimientos que utilizará:

- Encuesta a población,
- Entrevistas a profundidad
- Dinámica de grupos focales
- Análisis de contenido, etc.

Cómo y cuándo los aplicará y los instrumentos que utilizará para recopilar la información (cuestionario, guía de entrevista, hoja de registro de observaciones, guía de moderador del grupo focal, guía de análisis de contenido, etc.).

ERROR DE MUESTREO

Es la diferencia entre el valor observado y el valor desconocido de la característica poblacional, que tratamos de estimar.

Formas de Control :

- Uso adecuado del diseño muestral .
- Aumentando n
- Supervisión

**ERRORES DE
NO MUESTREO**

**NO
RESPUESTA**

**RESPUESTA
INEXACTA**

**SESGO DE
SELECCION**



CONFIABILIDAD Y VALIDEZ: REQUISITOS DE LA MEDICIÓN

- **Confiabilidad**

- El instrumento utilizado para medir produce resultados similares en diferentes oportunidades de medición

- **Validez**

- El instrumento utilizado para medir, es adecuado para medir lo que se desea medir (para medir calidad se pregunta sobre calidad de los productos, limpieza, rapidez, cortesía y no la fachada de los comercios)

CUESTIONARIOS

Recomendación



- Determinar las preguntas importantes.
- Esbozar los detalles del cuestionario.
- Diseñar el cuestionario.
- Hacer una prueba preliminar del cuestionario.
- Desarrollar una estrategia para la recolección de datos.
- Elaborar una carta introductoria y enviar cuestionarios.
- Controlar las respuestas.
- Analizar los datos de la encuesta

CUESTIONARIO

Es un conjunto de preguntas diseñadas para generar los datos necesarios, para alcanzar los objetivos del proyecto de investigación.

CUESTIONARIO: GUÍA PARA LA ELABORACIÓN

Tener claridad del problema, objetivos e hipótesis

Conocer las características de la población, objeto de estudio

Buscar cuestionarios previos sobre el mismo tema

Elegir las preguntas, ordenarlas,
probar el cuestionario (prueba piloto)
redactar el cuestionario definitivo (previando el procesamiento)

FACTORES QUE AFECTAN LA CONFIABILIDAD Y VALIDEZ

- La improvisación
- La utilización de instrumentos utilizados en otros contextos y aplicados sin adaptación
- Inadecuación del instrumento para las personas a las que está dirigido
- Condiciones en que se aplica
- Instrucciones deficientes
- Escasa capacitación de quienes aplican el instrumento

PROCESAMIENTO DE DATOS

- Codificación: asignar a cada categoría un número.
- Tabulación: sistematizar resultados cuantitativos en tablas, cuadros, gráficos, mapas.
- Programas para procesamiento de los datos.

CUADROS

Cuadro1: Evolución de la pobreza			
Tipo de Hogares	Año 1980	Año 1991	Año 2000
No pobres	80.2%	73.6%	76.1%
Sólo NBI	12.2%	10.1%	3.1%
Sólo LP	3.9%	10.6%	15.2%
NBI y LP	3.7%	5.7%	5.6%
Total	100%	100%	100%
Fuente: INEI			

TABLAS, CUADROS Y GRÁFICOS

- El Título (precisión de lo que contiene)
- La disposición de las variables
- Encabezados de columnas y filas (cortos y precisos)
- Las unidades que representan las cifras (absolutos, porcentajes/ nuevos soles, euros, dls, etc)
- Columnas de totales
- Notas o explicaciones anexas.
- Fuente de los datos

LECTURA DE TABLAS, CUADROS Y GRÁFICOS

- Título, notas, fuente, unidades en que se expresa, promedios totales, variabilidad entre distintos datos, asociaciones entre variables.

CUADROS

Cuadro1: Evolución de la pobreza			
Tipo de Hogares	Año 1980	Año 1991	Año 2000
No pobres	80.2%	73.6%	76.1%
Sólo NBI	12.2%	10.1%	3.1%
Sólo LP	3.9%	10.6%	15.2%
NBI y LP	3.7%	5.7%	5.6%
Total	100%	100%	100%

Fuente: INEI

Título

Variable independiente

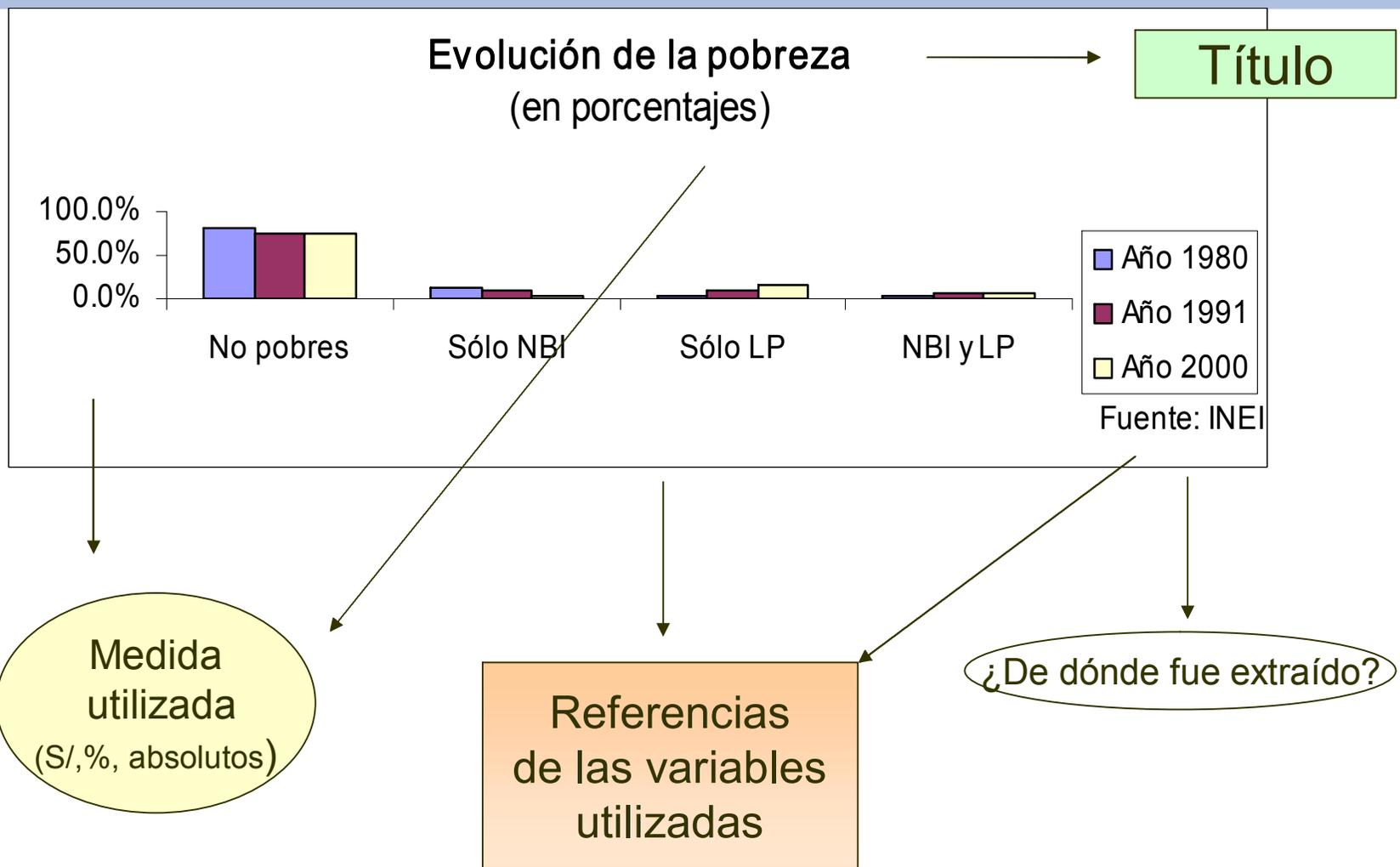
Variable dependiente

TOTALES

Medida utilizada
(\$, %, absolutos)

¿De dónde fue extraído?

GRÁFICOS



REDACCIÓN DEL INFORME

- El problema al que se refiere el estudio.
- Procedimientos de investigación
- Resultados alcanzados.
- Conclusiones
- Recomendaciones
- Reflexiones personales

PRINCIPIOS ÉTICOS

- Objetividad
- Integridad
- Respeto y protección de los sujetos
- Confidencialidad
- Compromiso con la verdad
- Reconocimiento a los colaboradores
- Transparencia económica

FICHA TECNICA

Ámbito	Nacional.
Universo	Población de ambos sexos de 18 años y más
Marco muestral	Padrón de habitantes de 2006
Tamaño muestral	Diseñada: 1501 entrevistas Realizada: 1490 entrevistas
Error muestral	Para un nivel de confianza del 95,5% (dos sigmas), y $P=Q$, el error es de $\pm 2,58\%$ para el conjunto de la muestra en el supuesto de muestreo aleatorio simple.
Diseño	Polietápico, estratificado por conglomerados, con selección de las unidades primarias de muestreo (municipios) y de las unidades secundarias (secciones) con probabilidad proporcional a su tamaño, y de las unidades últimas (individuos) por rutas aleatorias y cuotas de sexo y edad. Los estratos se han formado por el cruce de las 17 localidades autónomas con el tamaño de hábitat, dividido en 7 categorías: menor o igual a 2.000 habitantes; de 2.001 a 10.000; de 10.001 a 50.000; de 50.001 a 100.000; de 100.001 a 400.000; de 400.0001 a 1.000.000; y más de 1.000.000 habitantes. Los cuestionarios se han aplicado mediante entrevista personal en los domicilios.
fecha de realización	Del 21 al 31 de mayo de 2006.
Trabajo de campo	Centro de Investigaciones Sociológicas (CIS)

VICIO O SESGO EN EL MUESTREO (B)

$$B = E(\bar{\theta}) - \theta$$

Causas

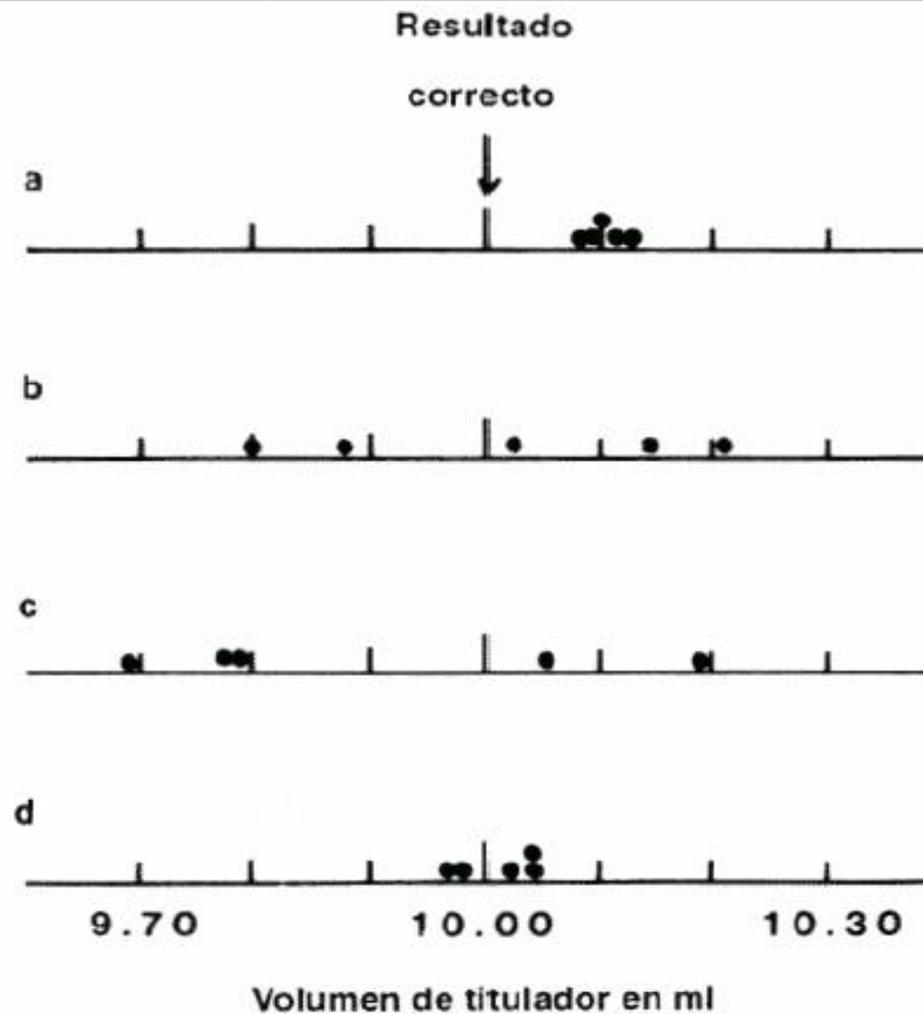
- Errores de medición
- Falta de respuesta

ERROR CUADRATICO MEDIO

$$E C M (\bar{\theta}) = E (\bar{\theta} - \theta)^2$$

$$E C M (\bar{\theta}) = V (\bar{\theta}) + B$$

EXACTITUD Y PRECISIÓN



Exactitud y precisión; representación gráfica de los datos de la tabla 1.1. En (a) los datos son precisos pero inexactos; en (b) exactos pero imprecisos; en (c) inexactos e imprecisos, y en (d) exactos y precisos.

Tipos de Muestreo probabilístico

1. Muestreo aleatorio simple
2. Muestreo sistemático
3. Muestreo estratificado
4. Muestreo por conglomerados

MUESTREO ALEATORIO SIMPLE

Es un método de selección de n unidades en un conjunto de N de tal modo que cada uno de las muestras distintas tenga la misma probabilidad de ser seleccionadas .

VENTAJAS:

- Sencillo y fácil comprensión
- Cálculo de medidas estadísticas.

DESVENTAJAS

- Requiere de un marco de muestreo.
- Requiere de muestra grande.

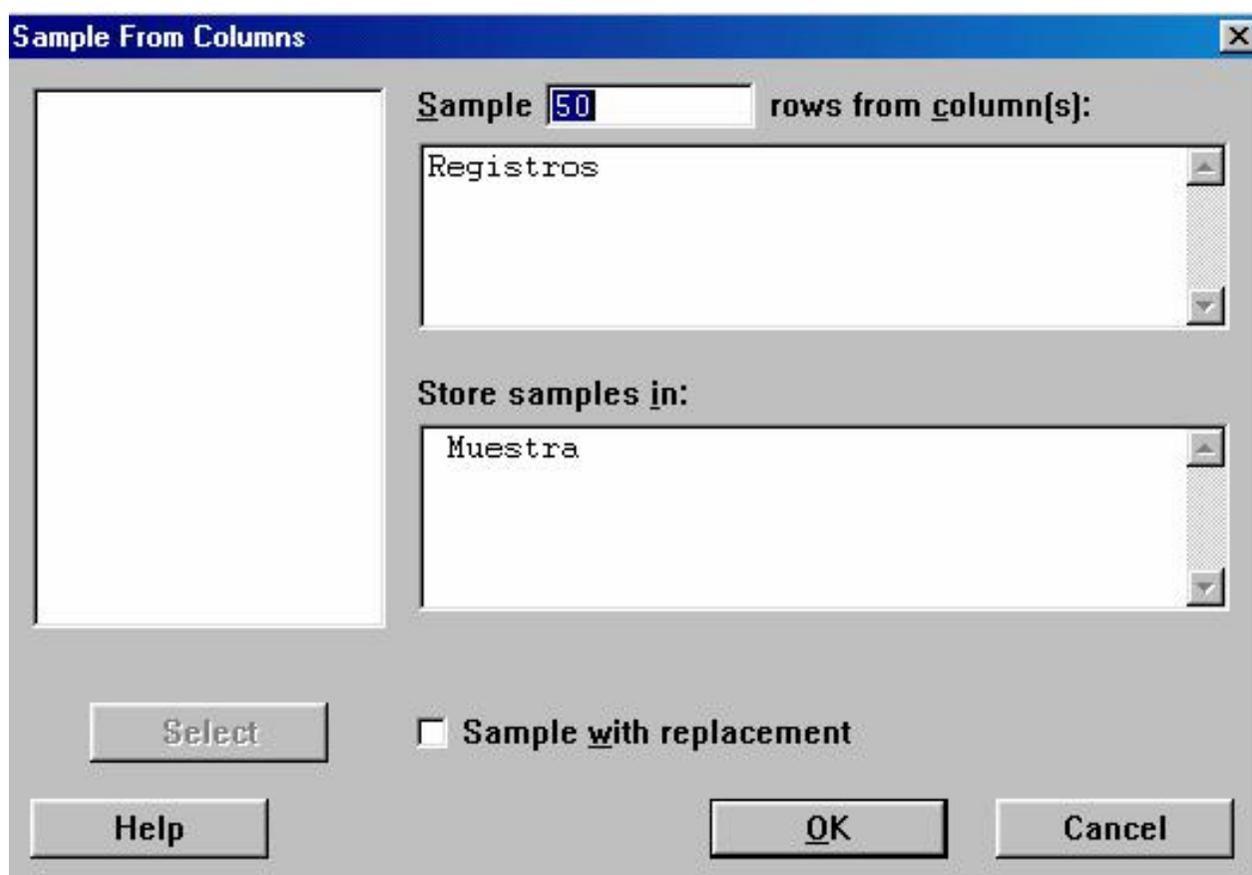
USO DE LOS NUMEROS ALEATORIOS.

El requisito fundamental de una tabla de número aleatorios es que cada uno de los dígitos(0, 1, 2, 3, 4, 5, 6, 7, 8, 9) tenga la misma probabilidad de aparecer en la tabla, es $p=1/10$

- Tablas
- Función Random

Suponga que existen 1000 registros de pacientes de cierto hospital, de los cuales se seleccionara una muestra aleatoria de $n=50$.

Calc>random data



Registros de pacientes que seran incluidos en la muestra

54	335	493	672	835
70	349	516	688	856
105	352	531	743	880
115	387	535	766	883
134	402	539	767	901
229	408	558	779	918
241	416	588	814	934
246	439	627	818	960
264	483	641	821	964
309	487	670	828	985

Una financiera realizará un estudio, con fines de evaluación de crédito, de los solicitantes de créditos hipotecarios a nivel nacional, con tal fin se recurrió a la base de datos de la oficina central, obteniéndose una lista numerada de 2600 personas solicitantes de crédito. Se ha determinado que serán 40 solicitantes los seleccionados para la realización del estudio; el encargado del proyecto desea determinar, en base a una lista numerada proporcionada por la oficina central, cuales serán los solicitantes que se han de seleccionar.

Describa brevemente cómo utilizaría una tabla de números aleatorios para obtener una muestra de 40 solicitantes de créditos hipotecarios.

- a) ¿Cuántos dígitos por solicitante debe leer?

-
- b) ¿Cómo decidiría en dónde empezar la tabla?
 - c) ¿Qué efectos, de hacerlo, provocaría en su planteamiento el hecho de que los solicitantes de crédito fueron numerados sólo por pares (2,4,.....)?
 - d) Realice la selección de los 40 solicitantes haciendo uso del muestreo aleatorio simple. Usando la tabla de números aleatorios de la columna 7.

Propiedades de los estimadores

T.1: la media muestral es un estimador insesgado de \bar{Y}

Es decir: $E(\bar{y}) = \bar{Y}$

Corolario 1: \hat{Y} es un estimador insesgado del total de la población

T.2: La varianza de la media muestral de una muestra aleatoria simple es:

$$V(\bar{y}) = \frac{S^2}{n} \left(\frac{N-n}{N} \right)$$

Propósitos del error estándar

- i. Comparar la precisión
- ii. Estimar el tamaño de muestra
- iii. Estimar la precisión.

Estimación del error estándar a partir de una muestra

T. : En una muestra aleatoria simple s^2 es una estimación insesgada de S^2

DONDE:

$$s^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}$$

$$S^2 = \frac{\sum (Y_i - \bar{Y})^2}{N-1}$$

Error estándar del Promedio

$$S_{\bar{y}} = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n-1} \left(\frac{1-f}{n} \right)}$$

Error estándar del Total

$$S_{\bar{y}} = N \sqrt{\frac{\sum(y_i - \bar{y})^2}{n-1} \left(\frac{1-f}{n} \right)}$$

Intervalos de confianza

Promedio $IC(\bar{Y}) = \bar{y} \pm ts_y$

Total $IC(Y) = N\bar{y} \pm Nts_y$

DONDE:

$$\bar{y} = \frac{\sum y_i}{n}$$

$$S^2 = \frac{\sum (y_i - \bar{y})^2}{N-1}$$

$$s_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1} \left(\frac{1-f}{n} \right)}$$

EJEMPLO 6

Supongamos que tengamos una población de 5000 usuarios de energía eléctrica, y que tenemos un listado con sus nombres. Si queremos elegir 20 usuarios, lo que necesitamos es que el computador elija al azar a 20 usuarios de esos 5000.

Solución:

Generar números aleatorios entre 0 y 5000.

Usando la función Excel: `aleatorio()*5000`

Los usuarios seleccionados son:

119	1593	1997	3376
278	1674	2706	3748
291	1734	2744	4245
711	1854	3131	4340
945	1902	3322	4786

EJEMPLO 7

Se tomó una muestra aleatoria simple de 20 usuarios de un distrito de Lima, para determinar el consumo promedio de energía eléctrica(kw-h).

Usuario	Consumo	Usuario	Consumo	Usuario	Consumo	Usuario	Consumo
1	110	6	128	11	127	16	123
2	131	7	108	12	148	17	132
3	110	8	117	13	107	18	109
4	132	9	128	14	129	19	115
5	137	10	122	15	127	20	116

Estimar el consumo promedio, error estándar y coeficiente de variación

EJEMPLO 8

Para estudiar la densidad de siembra de un cultivo (número de plantas por hectárea) de una población compuesta por mil parcelas se elabora el marco de muestreo y se elige una muestra simple aleatoria de 15 parcelas a partir de la columna 22 y fila 2 de la tabla de número aleatorios.

MARCO DE
MUESTREO:



1	21	41	961	981
2	22	42		962	982
3	23	43		963	983
4	24	44		964	984
5	25	45		965	985
6	26	46		966	986
7	27	47		967	987
8	28	48		968	988
9	29	49		969	989
10	30	50		970	990
11	31	51		971	991
12	32	52		972	992
13	33	53		973	993
14	34	54		974	994
15	35	55		975	995
16	36	56		976	996
17	37	57		977	997
18	38	58		978	998
19	39	59		979	999
20	40	60	980	1000

TABLA DE NUMEROS ALEATORIOS

Columna	1234	5678	9012	3456	7890	1234	5678	9012	3456	7890
1	1089	8719	2272	1359	3328	14	6773	1278	2761	3550
2	9385	7902	5034	6723	3835	6978	7084	3992	5857	2377
3	6934	8660	311	2979	95	2647	8299	5163	73	7788
4	52	1007	4866	6497	4138	8144	294	2906	316	4810

Las variables a investigar son:

X_i : Número de plantas por hectárea de la parcela i -ésima

Y_i : Si el número de plantas es menor o igual que 500($Y_i=1$). Si es mayor($Y_i=0$)

Estimar los siguientes parámetros: densidad media, total de plantas y proporción de parcelas con de 500 plantas.

Parcela	X_i	Y_i
073	52	1
084	125	1
095	12	1
163	581	0
299	858	0
311	16	1
377	747	0
647	806	0
660	762	0
788	770	0
857	546	0
934	132	1
978	228	1
979	311	1
992	932	0

EJEMPLO 9

El auditor de una compañía está interesado en estimar el porcentaje de facturas con error en la facturación, para lo cual se toma una muestra aleatoria de 50 usuarios y se observó el número de facturas con error en la facturación de 10. Determinar el tamaño de muestra con un error de estimación del 5 %. Suponga que el número de usuarios es 10000