

Curso de Estadística Aplicada a las Ciencias Sociales

Tema 8 Poblaciones y muestras

Fuentes: Manual (tema 19) y Agresti (cap. 2)

Tema 8

Tema 8. Poblaciones y muestras

■ Introducción

- Poblaciones y muestras
- Tipos de muestras
- Azar y estudios experimentales
- Error muestral y otras formas de error
- Resumen

Tema 8

Introducción

- Contenido principal de la estadística (y de este curso): una vez que tenemos unos datos, “manipularlos”, analizarlos, estudiarlos, “explotarlos”....
- Pero antes: ¿de dónde vienen esos datos?
- Esencial: garbage in-garbage out

Tema 8

Introducción

- En particular: cuando no estudiamos a toda la población sino sólo a una muestra
- Condiciones para que las muestras sean “representativas”
- Para que podamos aplicar todo lo que queda por delante en el curso: estadística inferencial

Tema 8

1. Poblaciones y muestras

- La estadística consiste en acumular y analizar datos
- Los objetos de los que uno toma medidas para generar datos son los **sujetos** del estudio: pueden ser individuos, familias, países, ciudades, empresas, instituciones...
- La **población**: conjunto de sujetos sobre el que el estudio quiere saber algo
- Una **muestra**: es un subconjunto de la población sobre el que el estudio toma datos

Tema 8

1. Poblaciones y muestras

- A veces se estudian poblaciones enteras:
 - ◆ Elecciones
 - ◆ Censos
 - ◆ Números “pequeños”: los alumnos de una facultad, los habitantes de una ciudad, los miembros de una asociación...
- Pero otras muchas veces estudiamos muestras

Tema 8

1. Poblaciones y muestras

- ¿Por qué estudiar muestras y no poblaciones?
 - ◆ Razones económicas
 - ◆ Tiempo: la variable cambiaría (ejemplo: paro)
 - ◆ El estudio altera el objeto (investigaciones sanitarias o industriales, biológicas...)
- En todos esos casos: muestra

Tema 8

1. Poblaciones y muestras

- Otra razón para estudiar muestras: existen poblaciones “virtuales” o “conceptuales” que no podemos estudiar
 - ◆ Poblaciones realmente existentes y claramente definidas: “todos los adultos españoles”
 - ◆ Poblaciones *conceptuales*:
 - ★ Todas las personas que ahora o en el futuro puedan padecer insomnio (un estudio de fármacos)
 - ★ Todos los coches que se puedan fabricar de un determinado modelo (en estudio de consumos)
 - ★ Todos los estudiantes de universidad actuales y futuros (en un estudio sobre efectos de un método de estudio)

Tema 8

1. Poblaciones y muestras

- Por todas esas razones: hay que estudiar una muestra, no una población
- Ejemplo: encuestas de opinión pública
 - ◆ La **población**: el conjunto de los adultos españoles (o franceses, o toledanos, o andaluces...)
 - ◆ La **muestra**: 1.000 personas entrevistadas por la encuesta
- Estadística descriptiva: resumir los datos de la muestra
- Estadística inferencial: de la muestra a la población

Tema 8

1. Poblaciones y muestras

- **Estimador (o estadístico)**: número que resume información sobre la **muestra** (Por ejemplo: porcentaje de los encuestados a favor de autorizar la eutanasia en ciertos casos)
- Los estimadores los obtenemos usando la estadística descriptiva, a partir de los datos de la muestra
- **Parámetro**: número que resume información sobre la **población** (Por ejemplo: porcentaje de los españoles a favor de autorizar la eutanasia en ciertos casos).
- Con la estadística inferencial estimamos, a partir de los **estimadores**, que son conocidos, los **parámetros**, que son desconocidos

Tema 8

2. Tipos de muestras

- Preocupación principal al hacer una muestra: que sea “representativa” de la población
- Que sea una “buena” muestra, que se “parezca” a la población
- Los resultados serán “sensatos” (los parámetros se parecerán a los estadísticos)
- Si seguimos ciertas reglas: podremos calcular matemáticamente, a partir del estadístico, el valor aproximado del parámetro

Tema 8

2.1. Muestreo aleatorio simple

- Mecanismo ideal para la mejor muestra posible: el **muestreo aleatorio simple**, muestreo en el que
 - ◆ Cada elemento de la población tiene la misma probabilidad de ser escogido
 - ◆ Todas las posibles muestras del tamaño muestral escogido (n) tienen la misma probabilidad de ser seleccionadas

Tema 8

2.1. Muestreo aleatorio simple

- Ejemplo teórico:
 - ◆ Todos los nombres en una lista
 - ◆ Papeleta o bola por cada nombre
 - ◆ Mezcla y extracción de las n bolas o papeletas
- Cumple las dos condiciones

Tema 8

2.1. Muestreo aleatorio simple

- En la práctica:
 - ◆ Numerar todos los sujetos de la población
 - ◆ Obtener lista de números aleatorios
 - ★ En libros: ejemplo, Apéndice B, Tabla 6
 - ★ Por ordenador (Excel)
 - ★ www.random.org
 - ◆ Propiedades:
 - ★ Cada número elegido separadamente
 - ★ Igual probabilidad de ser 0, 1, 2, ... 9
 - ★ Ninguna influencia de unos números en otros
 - ◆ Escoger los n primeros números aleatorios

Tema 8

2.1. Muestreo aleatorio simple

- Ejemplo con EXCEL
 - =aleatorio.entre(número;número) da un número aleatorio entre los valores mínimo y máximo indicados
 - Si número de sujetos en población es 3000 y quiero 30 casos en la muestra
 - 30 celdas =aleatorio.entre(1;3000)
 - Si población=40.000.000 y necesito 1.000 casos para la muestra
 - 1.000 celdas =aleatorio.entre(1;40000000)

2.1. Muestreo aleatorio simple

- Este es el mecanismo ideal
- Estadística inferencial: se basa en este modelo ideal de muestreo aleatorio simple
- Casi todos los métodos de inferencia: suponen que la muestra se ha obtenido por este método
- En la vida real: pocas veces aplicamos este método estrictamente
- Aplicamos más habitualmente otros métodos de muestreo **probabilísticos**

Tema 8

2.2. Otros muestreos probabilísticos

- Métodos de muestreo probabilístico: aquellos en los que es posible calcular la probabilidad de aparición de cada una de las muestras posibles
- El muestreo aleatorio simple es un muestreo probabilístico
- Otros métodos probabilísticos:
 - ◆ Todos los sujetos igual probabilidad de formar parte de la muestra
 - ◆ No todas las muestras posibles (combinaciones de n sujetos) tienen la misma probabilidad

Tema 8

2.2.1. Muestreo sistemático

- Población ordenada en lista numerada
- Dividimos el tamaño de la población (N) por el tamaño muestral (n): obtenemos el **factor de elevación**. Ejemplo: población de 25.000, muestra de 500. Factor de elevación= 50.
- Escogemos al azar (números aleatorios) un sujeto entre el primero y el que ocupa el puesto igual al factor de elevación. Ejemplo: entre los 50 primeros, el 24.
- Sumamos a ese número el factor de elevación hasta completar la muestra: 24, 74, 124, 174, 224, 274...

Tema 8

2.2.1. Muestreo sistemático

- Es más sencillo que muestreo aleatorio simple
- Pero normalmente produce un resultado igual de bueno: se pueden usar los mismos métodos de estadística inferencial
- Todos los sujetos igual probabilidad de estar en la muestra
- No todas las muestras igual probabilidad de aparecer

Tema 8

2.2.1. Muestreo sistemático

- Perfectamente seguro cuando no hay ningún orden en la población
- Peligro de sesgo: si orden en población en forma de ciclos que coincide con el valor del factor de elevación
- Ejemplo:
 - ◆ Estudio de los contenidos de un periódico, población son los periódicos de cada día en uno o más años y factor de elevación 7 o múltiplo de 7. Todos los días seleccionados serán el mismo día de la semana
 - ◆ Igual con otros fenómenos cíclicos: entradas de cine, ventas en restaurantes o tiendas, tráfico, uso de transportes públicos....

Tema 8

2.2.2. Muestreo estratificado

- O muestreo aleatorio estratificado
- Dividimos la población en diferentes grupos, o estratos, y tomamos de cada uno de ellos una muestra aleatoria simple
- Lo hacemos cuando conocemos que la población contiene grupos o subpoblaciones que son homogéneos internamente (estratos), pero sospechamos que son muy diferentes entre ellos respecto a la característica estudiada

Tema 8

2.2.2. Muestreo estratificado

- Muestreo estratificado proporcional
- Ejemplo:
 - ◆ muestra de estudiantes
 - ◆ sospechamos que sexo influye en respuesta a pregunta
 - ◆ Sabemos proporción de hombres y mujeres en población
- Muestreo estratificado proporcional:
 - ◆ Calcular número de hombres y mujeres en la muestra, manteniendo la proporción de la población
 - ◆ Muestra aleatoria simple de mujeres y hombres
- Resultado: “mejor” que muestreo aleatorio simple

Tema 8

2.2.2. Muestreo estratificado

- Muestreo estratificado no proporcional
- Ejemplo: estudio comparativo de acceso a servicios de salud entre personas que viven en municipios pequeños, medianos y grandes
 - ◆ Si muestreo aleatorio simple: muy poca gente de municipios pequeños
 - ◆ Muestra muy pequeña → error muestral grande → imposible hacer inferencia y comparación
 - ◆ “Solución”: muestreo estratificado con muestras del mismo tamaño de personas que viven en municipios pequeños, medianos y grandes

Tema 8

2.2.3. Muestreo por conglomerados

- A veces muestreo aleatorio simple, sistemático o estratificado no es posible
- Requieren listas (totales o por estratos)
- En muchos casos: esas listas no existen (o no son accesibles legalmente)
- Pero sí existen listas de “grupos heterogéneos de sujetos”, o conglomerados
- Hacemos muestreo aleatorio de conglomerados
- Dentro de los conglomerados elegidos: todos los elementos, o muestreo aleatorio simple

Tema 8

2.2.3. Muestreo por conglomerados

- Ejemplo: estudio sobre estudiantes universitarios españoles
- No hay lista de todos los estudiantes, ni por estratos
- Pero sí: lista de universidades y facultades
- Muestreo por conglomerados:
 - ◆ Muestreo aleatorio simple de universidades
 - ◆ Idem de facultades
 - ◆ Idem de grupos
 - ◆ Dentro del grupo (ya hay lista): todos, o muestreo aleatorio simple

Tema 8

2.2.3. Muestreo por conglomerados

- Solución muy práctica cuando conglomerados definidos geográficamente: enorme reducción costes extracción datos (viajes, tiempo, etc...)
- Diferencia con estratos:
 - ◆ Estratos son homogéneos internamente; interesa conocer diferencias entre estratos
 - ◆ Conglomerados son heterogéneos internamente; no interesa particularmente diferencias; es sólo un medio de tomar datos más económico y simple

Tema 8

2.2.3. Muestreo por conglomerados

- Requisitos: los conglomerados lo más heterogéneos posibles (como la población) internamente; muy parecidos entre sí
- Esto nunca es del todo así
- Sobre todo conglomerados geográficos: gente igual vive junta (barrios, ciudades)
- Resultados: más error muestral que muestra aleatoria simple
- Métodos inferencia: diferentes

Tema 8

2.2.4. Muestreo polietápico

- Combinación de varios métodos de muestreo
- Para poblaciones complejas
- Diferentes pasos en la selección de la muestra que usan diferentes métodos
- Ejemplo:
 - ◆ Municipios de Castilla-La Mancha son conglomerados, pero heterogéneos entre sí
 - ◆ Hacer “estratos de conglomerados” por tamaño de la población, o por actividad económica dominante
 - ◆ Seleccionar aleatoriamente, dentro de cada estrato, un número de municipios (muestreo estratificado por conglomerados)

Tema 8

2.2.4. Muestreo polietápico

- ◆ Dentro de cada municipio, las manzanas son conglomerados
- ◆ Distribuir las manzanas de cada municipio en estratos por niveles de renta, u otro indicador conocido
- ◆ Hacer muestreo aleatorio de manzanas en cada estrato de cada municipio (otra vez: **muestreo estratificado por conglomerados**)
- ◆ En cada manzana hacer un **muestreo sistemático** de casas
- ◆ En cada casa hacer un **muestreo aleatorio simple** de los individuos residentes en la casa

Tema 8

2.3. Muestreos no probabilísticos

- Aquellos en los que no es posible calcular la probabilidad de las diferentes muestras
- **NO ES POSIBLE** aplicar métodos de estadística inferencial cuando usamos estos muestreos
- Típico ejemplo: **muestra voluntaria**
 - ◆ Cupón en revista, que pide contestación por correo
 - ◆ Oyentes de programa de radio o televisión, a los que se pide que llamen a un teléfono
 - ★ Doble distorsión: el programa y el sentimiento intenso sobre el tema
 - **NO** es una muestra representativa: es una muestra **sesgada**.
 - **AUNQUE LLAMEN CIENTOS DE MILES DE PERSONAS!!!!**

Tema 8

2.3. Muestreos no probabilísticos

- Otro ejemplo: muestreo **“de calle”**: entrevistador se planta en una esquina y entrevista a gente que pasa
- Muestra sesgada: lugar, hora, día de la semana, proceso de “selección” por el entrevistador de a quién parar...
- Otro ejemplo: **Muestreo de conveniencia**:
 - empresa que encuesta a sus clientes para conocer las opiniones de los compradores de un producto;
 - sindicato que encuesta a sus afiliados para conocer opiniones de los trabajadores
- Todos estos ejemplos: **error o sesgo de selección**
- **NO** se pueden aplicar métodos de estadística inferencial
- **NO** son muestras representativas

Tema 8

3. Azar y estudios experimentales

- Algunas veces (poco en ciencias sociales, más en otras ciencias): estudios experimentales
- Experimento: comparar respuestas de los sujetos en alguna variable, bajo diferentes condiciones. A esas condiciones las llamamos **tratamientos**
- Ejemplos: medicina sobre enfermedad, método de estudio sobre rendimiento académico, forma de funcionamiento de cárcel sobre reincidencia...

Tema 8

3. Azar y estudios experimentales

- Papel del azar en estudios experimentales es doble
 - ◆ Si es posible: el conjunto de la muestra, escogida al azar respecto a la población
 - ◆ Submuestras (reciben dos tratamientos distintos; o una recibe tratamiento y la otra no): también al azar
- Efecto: en todos los otros factores “no controlados” los dos grupos serán similares

Tema 8

4. El error muestral y otras formas de error

4.1. Error muestral

- Aunque la muestra sea la más perfecta posible (muestreo aleatorio simple)
- La muestra escogida es una concreta cuyos estimadores no son idénticos a los parámetros de la población
- Ejemplo: dos encuestas para las mismas elecciones, realizadas al mismo tiempo, con muestra de igual tamaño. ¿Obtienen exactamente el mismo dato en el porcentaje de voto al partido A?

Tema 8

4.1. Error muestral

- Error muestral: la diferencia entre el valor del estimador en la muestra y el valor del parámetro en la población
- Por definición: no lo sabemos
- Pero los métodos de estadística inferencial nos permiten calcular el error muestral más probable (ejemplo: para 1.000 entrevistas, en torno a $\pm 3\%$, con un 95,5% de confianza) (Temas 9 y ss)
- Siempre y cuando sea muestreo aleatorio simple, o al menos muestreo probabilístico

Tema 8

4.2. Otras formas de error o sesgo

- Además del error muestral y el sesgo de selección (muestras no probabilísticas)
- Otras razones por las que los estimadores en la muestra pueden diferir de los parámetros en la población
- Sesgo por no respuesta, cuando algunos sujetos se niegan a cooperar. Si son muchos, o claramente diferentes según características: peligro
- Intentar tomar información incluso de los que no responden: evitar patrones

Tema 8

4.2. Otras formas de error o sesgo

- Pero también pueden existir **sesgos de respuesta**, cuando los entrevistados responden, pero no lo hacen de manera sincera
 - ◆ Pueden mentir para ocultar una idea que saben que es socialmente poco aceptada (racismo)
 - ◆ Pueden decir lo que creen que agrada al entrevistador (encuesta de deportes)
 - ◆ Pueden variar la respuesta según sexo o raza del entrevistador
 - ◆ Se aburren, se cansan

Tema 8

4.2. Otras formas de error o sesgo

- La formulación de las preguntas afecta a las respuestas (candidatos USA “demócrata” y “republicano”)
- El orden de las preguntas (encuesta deportes; encuesta sobre USA y URSS: ¿Debería USA dejar a periodistas URSS venir y contar lo que quisieran?)
- ¿Qué hacer?: cautela, cautela, cautela...

Tema 8

Resumen: conceptos principales

- Sujeto, población, muestra
- Estimador, parámetro
- Muestreos no probabilísticos: sesgo de selección
- Muestreo aleatorio simple
- Muestreo sistemático
- Muestreo estratificado

Tema 8

Resumen: conceptos principales (2)

- Muestreo por conglomerados
- Muestreo polietápico
- Error muestral
- Sesgo de no respuesta
- Sesgos de respuesta

Tema 8

Ejercicios recomendados

- Del manual: 19.3 a 19.7
- Ejercicios de exámenes
 - ◆ Feb04, Jul04: 2
 - ◆ Feb05, Jun05: 8, 9
 - ◆ Feb06: 6
 - ◆ Jun06: 7
 - ◆ Feb07, Jul07, Ene08, Jun08: 6