**EVALUACIÓN DEL CAPÍTULO 6 DEL MASTER EN ESTADÍSTICA APLICADA**

**TÉCNICAS DE MUESTREO Y DISEÑO DE ENCUESTAS**

**EJERCICIO PRÁCTICO 1**

**Se plantea la realización de un estudio sobre el hábito de fumar en la población mayor de 65 años en una determinada región compuesta por tres distritos, en donde se ha realizado una encuesta por muestreo.**

**El tamaño muestral se ha establecido en 900 personas, igualmente repartido en cada distrito.**

**En cada distrito se seleccionan 15 secciones censales, empleando el método de Sampford, siendo obviamente el número de personas mayores de 65 años de cada sección censal la variable tamaño.**

**Una vez seleccionadas las correspondientes secciones censales, se reparte uniformemente el tamaño muestral entre todas, y en cada una de ellas se obtiene una muestra de varones y otra de mujeres, empleando muestreo aleatorio simple en cada grupo de sexo, y repartiendo el tamaño de forma proporcional.**

1. **Describir la estructura del diseño muestral utilizado, especificando cada una de las fases y describiendo las variables de diseño utilizadas, es decir, estratos y conglomerados.**

**Fases:**

I ESTUDIO Y PLANTEAMIENTO: DISEÑO GENERAL:

1. **Fase de establecimiento de objetivos: deben de establecerse previamente los objetivos principales y secundarios.** En este caso el objetivo principal es evaluar el hábito de fumar en la población mayor de 65 años. No se especifican la definición de objetivos secundarios ni de objetivos estratégicos( como para el control de calidad en las etapas del estudio)
2. **Fase de estudio de fuentes estadísticas y metodológicas:**

Estudio de las fuentes de información que ya existen relacionadas con la investigación y al análisis de los diseños muestrales y de los métodos de investigación empleados por cada una de las fuentes analizadas.

1. **Fase de planificación de la operación y diseño del plan de muestreo.**
   1. Diseño sobre la información a obtener y sobre el método empleado para recabar la información**. Diseño del cuestionario** y soporte físico para la observación.
      1. Se debe de determinar la naturaleza y el contenido de las preguntas: En este caso es mejor escoger un cuestionario estructurado pues la n es grande con preguntas cerradas dicotómicas ( tipo fuma: si o no) o de elección múltiple ( tipo: nº de cigarrillos: < 5; entre 5 y 10; 10-20 …)
      2. Redacción de las preguntas, determinar las escalas y elección del orden de las preguntas.
      3. Es importante tener en cuenta que son personas mayores de 65 años por lo que han de ser preguntas claras y concisas.
      4. Previamente ha de realizarse una evaluación del cuestionario (Pre-test). Cuando esto es satisfactorio, se puede comenzar el trabajo de campo.
   2. Definición de la **periodicidad** con que se realizará el trabajo y del marco temporal de referencia.
   3. Elección del **método** de observación o de trabajo de campo y del modo de administración del mismo.
      1. En este caso ya que son 900 personas, puede resultar más fácil la entrevista telefónica o un procedimiento mixto ( correo con seguimiento telefónico)
   4. Decisión sobre el tipo de muestreo a implementar. Cálculo del tamaño muestra y su distribución.
   5. Definición y selección de las unidades muestrales de investigación.
   6. Elaboración del plan de seguimiento y de logística.

II TRABAJO DE CAMPO.

III TRATAMIENTO DE LA INFORMACIÓN.

Muestra compleja con:

1. Estratificación: implica el seleccionas muestras independientes dentro de subgrupos independientes. En este caso, el estrato es los mayores de 65 años
2. Conglomerados: implica la selección de grupos de unidades de muestreo. En este caso, las 15 secciones censales en cada distrito.

Estudio multietápico:

1. Primera etapa: selección de una muestra basada en conglomerados: distritos.
2. Segunda etapa: submuestras: de los conglomerados seleccionados ( 15 secciones censales de cada distrito)
3. Tercera etapa: personas mayores de 65 años que viven en cada sección censal, distribuidos por sexo.

Variable tamaño: número de personas mayores de 65 años de cada sección.

n: 900 en 3 distritios 🡪 300 en cada distrito.

1. **Hallar las probabilidades de inclusión de cada elemento de la población.**

F = f1 x f2:

f1: probabilidad de inclusión de la sección censal a la que pertenece (Sampford)

f2: probabilidad de inclusión dentro de la sección censal (muestreo aleatorio simple en estratos de sexo con afijación proporcional)

f1:

El método de Sampford es de tipo PiPS, es decir, con probabilidades de inclusión proporcionales al tamaño, de la forma, πi= n \* xi /tx siendo n el tamaño de la muestra, X la variable tamaño, y tx el total de la misma sobre la población.

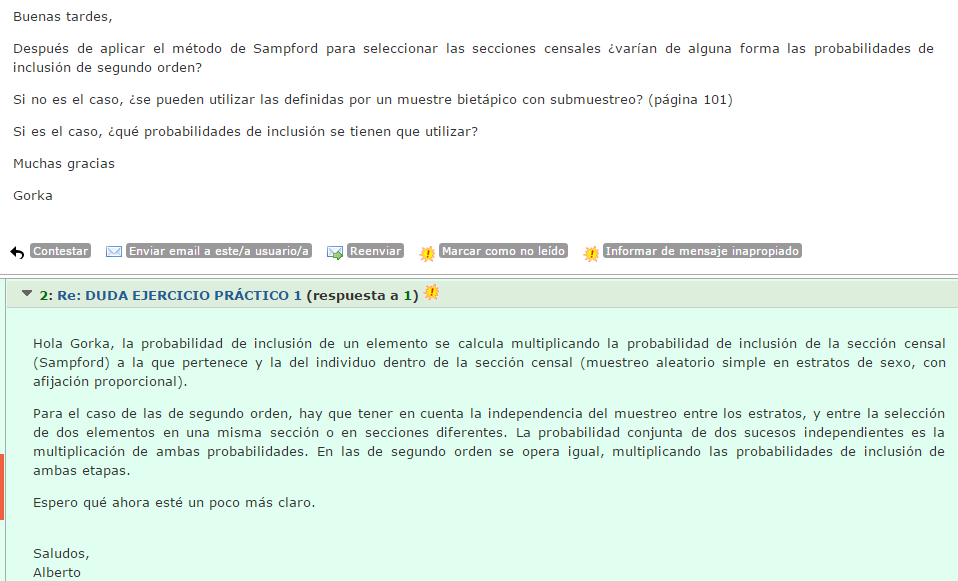
1. Si definimos *πij* como la probabilidad de inclusión de segundo orden derivada de la aplicación del método de Sampford, hallar las probabilidades de inclusión para:

P110 176-180

1. Dos individuos pertenecientes a distritos diferentes.

2. Dos individuos pertenecientes al mismo distrito pero a diferentes secciones censales.

1. Estimar la proporción de mayores de 65 años que tienen el hábito de fumar, suponiendo que la población objeto de estudio en las provincias es de 25.000, 30.000 y 20.000 respectivamente, y que en las muestras obtenidas hemos observado 80, 44 y 62 mayores que fuman.
2. Supongamos que dentro de cada distrito tenemos un hábitat rural y otro urbano, y que existen diferencias significativas en el hábito de fumar dependiendo de dicho entorno. ¿Podría mejorarse el diseño muestral utilizado en este caso? Si así fuera, explicar cómo.



**EJERCICIO PRÁCTICO 2**

Se ha realizado un estudio por muestreo sobre el uso de Internet en 3 provincias de una determinada región. La población se ha estratificado atendiendo a la provincia y al hábitat (rural / urbano).

La muestra se ha seleccionado en dos etapas, siendo las unidades de primera etapa las secciones censales de cada provincia y las de segunda etapa la población residente de 16 y más años.

La selección de secciones se ha efectuado de forma aleatoria y dentro de cada sección censal, se ha obtenido una muestra de forma sistemática de 20 personas, obteniendo por tanto un total de 900 encuestas.

Los datos correspondientes a la población y a la muestra aparecen en la hoja de cálculo “Ejemplo Uso de Internet.xls” incluida en la carpeta Datos del DVD.

Se pide:

1. Obtener las estimaciones del uso de Internet por provincias, hábitat y sexo, atendiendo al diseño muestral establecido.
2. Como puede observarse, la variable sexo no ha sido tenida en cuenta en el diseño de la muestra. Si quisiéramos que las estimaciones reproduzcan adecuadamente la población, tanto por provincias, como por hábitat y sexo, ¿qué procedimiento/s podríamos utilizar?