Problemas de *Muestreo*. Diseños con probabilidades desiguales

Problema 1 Extracción sistemática

Se considera una población U compuesta de 6 hogares de respectivas dimensiones, 2, 4, 3, 9, 1 y 2. El tamaño del hogar, variable auxiliar X es el número de personas que le integran. Se selecciona 3 hogares sin reposición con probabilidad proporcional al tamaño.

1. Den las probabilidades de inclusión de primer orden
2. Supongan que el algoritmo sistemático ha llevado a seleccionar los hogares 2 y 3, además del hogar que siempre pertenecerá a la muestra por tener una probabilidad de inclusión igual a 1; den una estimación del tamaño medio de los hogares de la población ¿Os sorprende el resultado?

Problema 2 Cooperativas de consumo.

Se desea conocer el número medio de cooperativas de consumo en activo en una determinada comarca de 10 municipios. Se piensa que dicho número tiene relación estrecha con el número de habitantes, conocido gracias a un censo reciente.

Municipio	**Nº habitantes**
A	12000
B	15000
C	3000
D	40000
E	60000
F	10000
G	10000
H	25000
I	30000
J	120000

Dado dicho conocimiento, escoger un método de selección de la muestra pertinente. Sabiendo que desea una muestra de tamaño igual a 3, y que el número escogido entre 0 y 1, que se debe utilizar para arrancar la selección vale 0.7, decir cuál es la muestra seleccionada.

Problema 3 Extracción sistemática de empresas

Se considera una población U compuesta de 6 empresas, para las cuales se conoce el total de ventas (variable auxiliar *X)*, respectivamente, 40, 10, 8, 1, 0.5, 0.5 millones de euros. Con el objetivo de estimar la paga total de los empleados, seleccionen una muestra con probabilidades desiguales sin reposición de tamaña 3 mediante un algoritmo sistemático; Consideren que 0.83021 ha sido el valor obtenida al seleccionar un valor de una variable uniforme sobre [0,1). ¿Qué pasa si se modifica el orden de la lista.

Problema 4 Extracción sistemática y varianza

Se considera una población U compuesta de 6 unidades. Se conoce los valores de una variable auxiliar x para todas las unidades de la población:

$$X_1=200,\quad X_2=80,\quad X_3=50,\quad X_4=50,\quad X_5=10,\quad X_6=10$$

1. Calculen las probabilidades de inclusión de primer orden, proporciona a x_k para una muestra de tamaño n=4. Consideren que 0.48444 ha sido el valor obtenida al seleccionar un valor de una variable uniforme sobre [0,1]. Seleccionen una muestra con

probabilidades desiguales sin reposición de tamaña 4 mediante un algoritmo sistemático, guardando el orden inicial de la lista.

2. Den la matrice de las probabilidades de inclusión de segundo orden (lista en el orden inicial)
3. Se asume que la variable de interés Y toma los siguientes valores:

$$Y_1=80,\quad Y_2=50,\quad Y_3=30,\quad Y_4=25,\quad Y_5=10,\quad Y_6=5$$

Construyan una tabla que tenga:

- En fila, cada muestra s posible
- En columnas: $p(s)$; $\hat{T}(s)$; $\hat{V}(\hat{T})(s)$

Calculen $E(\hat{T}(s))$ y $E(\hat{V}(\hat{T})(s))$. Comenten los resultados.

Problema 5 Algoritmo de Tillé y varianza

Se considera una población U compuesta de 6 unidades. Se conoce los valores de una variable auxiliar x para todas las unidades de la población:

$$X_1=200,\quad X_2=80,\quad X_3=50,\quad X_4=50,\quad X_5=10,\quad X_6=10$$

1 Calculen las probabilidades de inclusión en la muestra de las unidades para una muestra de tamaño 4.

2 Con el script de la práctica, se ha calculado la matriz de probabilidades de inclusión dobles, para este mismo tamaño de muestra:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	1.0000000	1.0000000	0.83333333	0.83333333	0.16666667	0.16666667
[2,]	1.0000000	1.0000000	0.83333333	0.83333333	0.16666667	0.16666667
[3,]	0.8333333	0.8333333	0.83333333	0.66666667	0.08333333	0.08333333
[4,]	0.8333333	0.8333333	0.66666667	0.83333333	0.08333333	0.08333333
[5,]	0.1666667	0.1666667	0.08333333	0.08333333	0.16666667	0.00000000
[6,]	0.1666667	0.1666667	0.08333333	0.08333333	0.00000000	0.16666667

Asumiendo que la variable de interés Y toma los siguientes valores:

$$Y_1=80,\quad Y_2=50,\quad Y_3=30,\quad Y_4=25,\quad Y_5=10,\quad Y_6=5$$

Construyan una tabla que tenga:

- En fila, cada muestra s posible
- En columnas: $p(s)$; $\hat{T}(s)$; $\hat{V}(\hat{T})(s)$

Calculen $E(\hat{T}(s))$ y $E(\hat{V}(\hat{T})(s))$. Comenten los resultados sabiendo que $V(\hat{T})=110$

Se recuerdan las fórmulas:

$$\hat{T} = \sum_i \frac{y_i}{\pi_i}$$

$$\hat{V}\left(\hat{T}\right)=\frac{1}{2}\sum_{i}\sum_{\substack{j\\ i\neq j}}\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2$$

Problema 6 Productos electrónicos

Se quiere estudiar el consumo medio en productos electrónicos de los municipios de una determinada región. Se decide extraer una muestra de 5 municipios, mediante una extracción proporcional al tamaño. En efecto, se conoce la población de los municipios incluidos en la región, indicada en la siguiente tabla:

Mun.	Efectivos.
1	20
2	30
3	20
4	30
5	20
6	30
7	30
8	20
9	15
10	15
11	25
12	35
13	25
14	35
15	30
16	30
17	20
18	20
19	25
20	25

1.

a) Justificar la elección del método de muestreo

b) Suponiendo que se obtuvo "0.322" al seleccionar un número al azar entre 0 y 1 para arrancar la selección de la muestra, decir qué municipios pertenecen a la muestra.

c) Calcular la probabilidad de pertenecer a la muestra asignada a cada municipio

2.

En los 5 municipios seleccionados, se ha observado el siguiente gasto:

Muestra (según orden de extracción)	**Gastos en productos electrónicos**
1	300000
2	500000
3	400000
4	200000
5	600000

Estimar el consumo medio en productos electrónicos el conjunto de la región estudiada.