

Muestreo Estratificado.

Problema 1.

Se dispone de una población de 1060 empresas. Se desea estimar el número medio de empleados por empresa. La población está compuesta de 5 estratos definidos a partir del tamaño de la empresa según el número de empleados en clase.

Esta información es conocida a partir de los registros oficiales que no dan el número exacto de empleados, sino el tamaño en clase. Se dispone de un presupuesto que permite encuestar 300 empresas. Se decide realizar un muestreo aleatorio simple en cada estrato según el reparto indicado en la tabla. Sobre cada empresa, se mide la variable Y : "número de empleados" y se calcula la media y la varianza de dicha variable en cada estrato. Escoger un estimador, dar su expresión. Hacer una estimación de Y por punto y por intervalo.

Estrato según tamaño empresa	N_h	\bar{y}_h	s_h^2	n_h
0-9	500	5	1.5	130
10-19	300	12	4.0	80
20-49	150	30	8.0	60
50-499	100	150	100.0	25
500 y más	10	600	2500.0	5
Total	1060			300

Problema 2

Un ayuntamiento quiere estudiar, mediante muestreo, la participación de los habitantes mayores de edad a los actos culturales que organiza. En particular, se preguntará el número de dichos actos a los cuales se participó el año anterior. Se dispone de la información siguiente:

Clase de edad	Población	Desviación-tipo
de 18 a 24 años	5000	2.5
de 25 a 50 años	10000	7
mayores de 50 años	15000	12

Se decide observar una muestra de 500 personas. Dadas las diferencias entre las varianzas intra-estratos, y sabiendo que la participación es muy distinta según el grupo de edad, en particular más importante entre los jóvenes, se decide hacer un muestreo estratificado y entrevistar 100 personas en el primer estrato, 150 en el segundo estrato y 250 en el estrato de los más de 50 años. Se observa en la muestra:

Clase de edad	Media	Desviación-tipo
de 18 a 25 años	8.2	3.2
de 25 a 50 años	3.4	8.3
mayores de 50 años	2.1	11.7

1. ¿Cuál es la variable de interés?
2. Estimar por punto y por intervalo la media del número de actos presenciado en la población estudiada.

Problema 3 (Julio 2007, Ardilly/ Tillé)

Un director de circo tiene 100 elefantes, hembras y machos. Se tiene que estimar el peso total de los elefantes para cruzar el río por barco. El año pasado, el mismo director de circo había pesado a todos sus elefantes y obtenidos los siguientes resultados (en toneladas):

	N_h	Media año pasado \bar{Z}_h	Casi-varianza año pasado ($\sigma_h'^2$)
Machos	60	6	4
Hembras	40	4	2.25

La variable de interés Y es “peso este año”. En el recuadro, se tienen las medias y casi-varianzas poblacionales de la variable Z “*peso el año pasado*” en los dos estratos determinados por el sexo del elefante. Se asume que la varianza de la variable peso no ha variado mucho desde el año anterior, es decir que la varianza de la variable “peso este año”, es decir Y, y la varianza de la variable “peso el año pasado”, es decir Z, son iguales.

1. Calculen la varianza y casi-varianza poblacionales de Z, “*peso el año pasado*”. Se recuerda que:

$$\sigma^2 = \sum_h \frac{N_h}{N} \sigma_h^2 + \sum_h \frac{N_h}{N} (\bar{Z}_h - \bar{Z})^2$$

y que

$$\sigma_h^2 = \frac{N_h - 1}{N_h} \cdot \sigma_h'^2 \qquad \sigma^2 = \frac{N - 1}{N} \cdot \sigma'^2$$

2. El director supone que las dispersiones del peso no evolucionan de manera sensible de un año para otro (este tipo de hipótesis es aquí muy razonable y se encuentra frecuentemente en las encuestas que se repiten de forma regular). Por lo tanto, se considera que la variable *peso este año* (variable Y) y la variable *peso el año pasado* (variable Z) tienen varianzas iguales.

Si el director procede a un elección ASSR de 10 elefantes, cuál será la varianza del estimador del peso total del rebaño?

3. Con un diseño estratificado con reparto proporcional (con un tamaño de muestra de 10 elefantes), Cuál sería el tamaño de la muestra en cada estrato? ¿cuánto valdrá la varianza del estimador del peso medio este año? ¿cuánto valdrá la varianza del estimador del peso total este año?

Problema 4 (Julio 2007, Ardilly/ Tillé)

En una ciudad grande, se estudia la media del número de pacientes que un médico ve a lo largo de un día de trabajo. Se supone que un médico tiene tantos más pacientes que experiencia tiene. Por tanto, se clasifican a los médicos en 3 grupos: principiantes, experiencia media, experiencia alta. Hay 500 médicos en el primer grupo, 1000 en el grupo 2 y 2500 en el grupo 3. Se extrae una muestra ASSR sin reposición de tamaño 200 en cada uno de los 3 grupos y así se estima el número medio de pacientes al día en cada grupo: 10 en el grupo 1, 15 en el 2 y 20 en el 3. Se estima también la varianza del número de pacientes en los 3 grupos: respectivamente, 4, 7 y 10.

- 1 ¿Cómo se llama el diseño de muestra empleado?
- 2 Estimen la media de pacientes que un médico ve al día en la población global de médicos, por punto y por intervalo.