

APLICACIÓN DEL BOOTSTRAP EN MÉTODOS INDIRECTOS DE ESTIMACIÓN EN MUESTREO ESTRATIFICADO

BOOTSTRAP APPLICATION IN INDIRECT ESTIMATION METHODS IN STRATIFIED SAMPLING

¹Jaime C. Porras C.

Resumen

Actualmente, en muchos estudios de investigación de mercado y de opinión se utiliza el diseño de Muestras Complejas. Para resumir los datos de una variable cuantitativa obtenidos mediante este diseño, se puede hacer uso de una medida como la media. El cálculo de esta medida puede depender del uso de variables auxiliares. Es decir, para estimar la media poblacional, se pueden utilizar diferentes estimadores y es necesario determinar cuál de ellos es el más preciso. El error estándar puede ser utilizado como un indicador que ayude a comparar la precisión entre ellos. El presente artículo se centró en la comparación de dos estimadores para la media en muestreo estratificado. Se utilizaron métodos indirectos de estimación. El primero es el estimador clásico por regresión lineal simple, y el segundo es una nueva alternativa a la que se denominó el estimador por regresión lineal simple bootstrap. Este último resulta ser el mejor. La implementación de los estimadores se hizo con ayuda del programa estadístico R.

Palabras clave: Muestreo, Muestras Complejas, Bootstrap, Muestreo Estratificado, Simulación.

Abstract

Currently, many studies of market research and design review Complex Samples used. To summarize the data in a quantitative variable obtained by this design, we can use a measure such as the average. The calculation of this measure may depend on the use of auxiliary variables. That is, to estimate the population mean, you can use different estimators and is necessary to determine which one is the most accurate. The standard error can be used as an indicator to help compare the accuracy between them. This article focused on the comparison of two estimators for the mean in stratified sampling. Indirect estimation methods were used. The first is the classic linear regression estimator, and the second is a new alternative which is called the simple linear regression estimator bootstrap. The latter turns out to be the best. The implementation of the estimates made using the statistical program R.

Key words: Sampling, Complex Samples, Bootstrap, Stratified Sampling, Simulation.

1. Introducción

La elección del adecuado diseño muestral es el punto de partida de muchos estudios de investigación que requieren, posteriormente, un resumen descriptivo o análisis inferencial del conjunto de datos recolectado. Los diseños muestrales simples más utilizados son: el muestreo aleatorio simple, el muestreo aleatorio estratificado, muestreo sistemático replicado y el muestreo por conglomerados. Sin embargo, un diseño muy utilizado, actualmente, para recolectar los datos son las Muestras Complejas. Este tipo de diseño muestral se fundamenta principalmente en la combinación de dos diseños muestrales simples (mencionados anteriormente) y/o el uso de variables auxiliares que puedan ayudar a mejorar la estimación de un parámetro de interés como: la media o la proporción poblacional.

El muestreo aleatorio estratificado se utiliza, principalmente, cuando una población objetivo puede subdividirse en varias subpoblaciones denominadas estratos

o segmentos. De tal manera, que las unidades muestrales dentro de cada estrato deben ser lo más homogéneas posible y entre los estratos deben ser heterogéneas.

Luego de obtener los datos, éstos necesitan ser resumidos mediante una medida descriptiva como la media. La forma más sencilla de estimar la media en el muestreo aleatorio estratificado es haciendo uso de la ponderación de acuerdo al tamaño de cada uno de los estratos. Otra forma de estimación de la media es utilizando una variable auxiliar que se encuentre correlacionada con la variable objetivo de estudio. Esta variable auxiliar es utilizada para ajustar la media en cada estrato con ayuda de los coeficientes de un modelo de regresión lineal simple. Posteriormente, se ponderan las medias ajustas según el tamaño de cada estrato.

Como existen diferentes estadísticos para estimar la media poblacional en un muestreo estratificado, si se desea analizar la precisión de ellos se puede hacer uso de alguna medida de variabilidad como su: varianza,

desviación estándar (error estándar) o coeficiente de variación.

La presente investigación tiene como principal objetivo comparar dos métodos de estimación de la media en muestreo estratificado utilizando una variable auxiliar. La obtención de los errores estándar de los estimadores que se compararán se realizará a través de simulación mediante la elaboración de procedimientos obtenidos con ayuda del programa estadístico R.

En actuales estudios de investigación de mercado, de opinión o socioeconómicos se utilizan principalmente las muestras complejas como fundamento, basta con revisar las fichas técnicas de los diversos estudios realizados por empresas o instituciones como: Ipsos Apoyo, Arellano o el INEI.

Por ejemplo, para estimar el gasto promedio mensual (en soles) en una ciudad cuyas familias pueden estar clasificadas en tres niveles socioeconómicos, se puede utilizar como variable auxiliar el ingreso mensual (en soles). Es decir, se utiliza un diseño muestral aleatorio Estratificado con L estratos socioeconómicos para estimar el gasto mensual (Y) en función del ingreso mensual (X).

Sin embargo, las muestras complejas no solo son utilizadas en estudios de mercado; pues su aplicación se puede extender a otras áreas como el realizado en Cuba por Cañizares y otros (2001) quien lo utiliza para realizar estimaciones de indicadores de salud. También se encuentra el estudio Álvarez y otros (2012) en el cual utiliza muestras complejas para analizar el sobrepeso y la obesidad en la población peruana en los años 2009-2010.

La evolución de los sistemas de cómputo hace que procedimientos a los que no se les brindaba la debida importancia en décadas pasadas sean muy utilizados actualmente. En ese sentido, el Bootstrap es una técnica estadística que se basa en el uso intensivo de computadoras (Efron y Tibshirani).

El uso de la metodología Bootstrap en las muestras complejas podría incrementar la eficiencia de los estimadores. Esto es lo que se pretende demostrar en el presente artículo, lo cual se llevará a cabo con ayuda del programa R versión 3.1.1

2. Materiales y métodos

MATERIALES

Para realizar la aplicación del presente trabajo de investigación, se requirió de la implementación de algunas funciones (ver anexo) que permitan generar datos con ayuda del programa estadístico R versión 3.1.1 El paquete MASS del R fue utilizado para la generación de datos provenientes de una distribución normal bivariada.

La idea es generar diferentes escenarios o conjuntos de datos (poblaciones) agrupados en L estratos que provengan de una distribución normal bivariada con una correlación significativa entre las variables que lo conforman. Posteriormente, se extraerán muestras aleatorias e independientes de cada estrato las cuales

serán proporcionales al tamaño de la población propuesta.

Para los datos simulados, se considerarán medias significativamente distintas en cada estrato y una ligera homogeneidad de varianzas entre los distintos estratos.

Los datos bivariados que fueron generados tanto de la población como de la muestra tienen la siguiente estructura:

Tabla 1. Estructura de los datos simulados.

Estrato de pertenencia	Variable dependiente (Y)	Variable independiente (X)
------------------------	--------------------------	----------------------------

Fuente: Elaboración propia

En la tabla 2, se presenta una breve descripción de los conjuntos de datos que serán simulados. Para cada escenario propuesto, se indica: el número de estratos utilizados, el tamaño de la población en cada estrato (N_h), el tamaño de la muestra seleccionada al azar de cada estrato (n_h), las medias poblacionales (μ_{yh} y μ_{xh}), varianzas poblacionales (σ_{yh}^2 y σ_{xh}^2), correlación (ρ_{xyh}) y covarianzas (σ_{xyh}) de la variables dependiente e independiente en cada uno de los estratos.

En cada escenario, se estimará el error estándar de los estimadores propuestos y adicionalmente del estimador simple, es decir sin el uso de variable auxiliar (X).

Para la obtención de los resultados con el estimador por regresión lineal simple bootstrap, se utilizaron 200 muestras bootstrap (Efron y Tibshirani).

MÉTODOS

A continuación, se describen los aspectos teóricos del Muestreo Aleatorio Estratificado, el estimador por regresión lineal simple y de los estimadores utilizados en la presente investigación.

- Muestreo aleatorio estratificado

En este diseño de muestreo, una población heterogénea con N unidades $\{U_i\}_{i=1,2,\dots,N}$ se divide en L estratos (no traslapados), los cuales son lo más homogéneos posible. Los estratos $\{U_h\}_{h=1,2,\dots,L}$ son de tamaño N_1, N_2, \dots, N_L y se cumple que $N_1 + N_2 + \dots + N_L$.

La muestra estratificada de tamaño n se obtiene seleccionando (de forma independiente) n_h elementos ($h=1,2,\dots,L$) de cada uno de los L estratos en los que se subdivide la población.

Una manera simple de resumir la información general cuando se trabaja con muestreo estratificado es la siguiente:

Tabla 2. Descripción de la estructura de los datos simulados.

Escenario	Estrato	N_h	n_h	μ_{yh}	μ_{xh}	σ^2_{yh}	σ^2_{xh}	ρ_{xyh}	σ_{xyh}
1	1	1000	20	2000	1800	100	80	0.7	63
	2	1000	20	1500	1100	90	70	0.7	56
2	1	1000	20	2000	1800	100	80	0.7	63
	2	2000	40	1500	1100	90	70	0.7	56
3	1	10000	100	2000	1800	100	80	0.7	63
	2	10000	100	1500	1100	90	70	0.7	56
4	1	10000	100	2000	1800	100	80	0.7	63
	2	20000	200	1500	1100	90	70	0.7	56
5	1	1000	20	2000	1800	100	80	0.9	80
	2	1000	20	1500	1100	90	70	0.9	71
6	1	1000	20	2000	1800	100	80	0.9	80
	2	2000	40	1500	1100	90	70	0.9	71
7	1	10000	100	2000	1800	100	80	0.9	80
	2	10000	100	1500	1100	90	70	0.9	71
8	1	10000	100	2000	1800	100	80	0.9	80
	2	20000	200	1500	1100	90	70	0.9	71
9	1	1000	20	2500	2300	100	90	0.7	66
	2	1000	20	2000	1800	95	85	0.7	63
	3	1000	20	1500	1200	90	80	0.7	59
10	1	2000	20	2500	2300	100	90	0.7	66
	2	3000	30	2000	1800	95	85	0.7	63
	3	4000	40	1500	1200	90	80	0.7	59
11	1	10000	100	2500	2300	100	90	0.7	66
	2	10000	100	2000	1800	95	85	0.7	63
	3	10000	100	1500	1200	90	80	0.7	59
12	1	10000	100	2500	2300	100	90	0.7	66
	2	20000	200	2000	1800	95	85	0.7	63
	3	30000	300	1500	1200	90	80	0.7	59
13	1	1000	20	2500	2300	100	90	0.9	85
	2	1000	20	2000	1800	95	85	0.9	81
	3	1000	20	1500	1200	90	80	0.9	76
14	1	2000	20	2500	2300	100	90	0.9	85
	2	3000	30	2000	1800	95	85	0.9	81
	3	4000	40	1500	1200	90	80	0.9	76
15	1	10000	100	2500	2300	100	90	0.9	85
	2	10000	100	2000	1800	95	85	0.9	81
	3	10000	100	1500	1200	90	80	0.9	76
16	1	10000	100	2500	2300	100	90	0.9	85
	2	20000	200	2000	1800	95	85	0.9	81
	3	30000	300	1500	1200	90	80	0.9	76

Fuente: Elaboración propia

Tabla 3. Estructura de resumen de datos en un Muestreo Estratificado.

Estrato	Población	Muestra	Pesos (W_h)	f_h	Media
1	$U_{11}, U_{12}, \dots, U_{1N_1}$	$U_{11}, U_{12}, \dots, U_{1n_1}$	$W_1 = N_1/N$	$f_1 = n_1/N_1$	$\bar{y}_1 = \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
H	$U_{h1}, U_{h2}, \dots, U_{hN_h}$	$U_{h1}, U_{h2}, \dots, U_{hn_h}$	$W_h = N_h/N$	$f_h = n_h/N_h$	$\bar{y}_h = \frac{1}{n_h} \sum_{j=1}^{n_h} y_{hj}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
L	$U_{L1}, U_{L2}, \dots, U_{LN_L}$	$U_{L1}, U_{L2}, \dots, U_{Ln_L}$	$W_L = N_L/N$	$f_L = n_L/N_L$	$\bar{y}_L = \frac{1}{n_L} \sum_{j=1}^{n_L} y_{Lj}$
Total	$\sum_{h=1}^L N_h = N$	$\sum_{h=1}^L n_h = n$	$\sum_{h=1}^L W_h = 1$		

Un estimador de un parámetro poblacional (como la media) puede expresarse como la suma de las estimaciones para el parámetro en los diferentes estratos ponderado por el tamaño de cada estrato.

Sea \bar{y}_h la media muestral correspondiente al estrato h , (para $h = 1, 2, \dots, L$). La media poblacional (μ) estimada mediante muestreo aleatorio estratificado es obtenida utilizando la siguiente expresión:

$$\bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h$$

El error estándar de \bar{y}_{st} es estimada por:

$$\hat{EE}(\bar{y}_{st}) = \sqrt{\sum_{h=1}^L W_h^2 \frac{s_h^2}{n_h} (1 - f_h)}$$

Estimador por regresión

El estimador por regresión para la media en un Muestreo Aleatorio Simple estaría dado por:

$$\bar{y}_l = \bar{y} + b_1 (\mu_x - \bar{x})$$

$$b_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}$$

Donde:

\bar{y} : Es el estimador simple de la media para la variable dependiente.

b_1 : Es la pendiente de la recta que relaciona X e Y.

\bar{x} : Es el estimador simple de la media para la variable

independiente.

μ_x : Es la media poblacional de la variable independiente.

El error estándar del estimador \bar{y}_l es:

$$\hat{EE}(\bar{y}_l) = \sqrt{\frac{1-f}{n} (s_y^2 + b_1^2 s_x^2 - 2b_1 s_{xy})}$$

Donde:

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n-1} \left[\sum_{i=1}^n y_i^2 - n \bar{y}^2 \right]$$

El error estándar se puede simplificar reemplazando $s_{xy} = b_1 s_x^2$, obteniendo:

$$\hat{EE}(\bar{y}_l) = \sqrt{\frac{1-f}{n} (s_y^2 - b_1^2 s_x^2)}$$

Estimador de regresión simple (o separado) en muestreo estratificado

Los datos consisten en muestras aleatorias e independientes de pares ordenados $(X_p, Y_p), (X_y, Y_y), \dots, (X_{nh}, Y_{nh})$ provenientes de cada uno de los estratos en estudio ($\forall h=1, 2, \dots, L$).

El estimador estratificado de la media mediante el estimador de regresión es:

$$\bar{y}_{st \ l_s} = \sum_{h=1}^L W_h \bar{y}_{lh}$$

Donde:

$$\bar{y}_{lh} = \bar{y}_h + b_{1h} (\mu_{xh} - \bar{x}_h)$$

El error estándar de este estimador es:

$$\widehat{EE}(\bar{y}_{st\ l_s}) = \sqrt{\sum_{h=1}^L W_h^2 \frac{(1-f_h)}{n_h} (s_{yh}^2 + b_{1h}^2 s_{xh}^2 - 2b_{1h} s_{xyh})} = \sqrt{\sum_{h=1}^L W_h^2 \frac{(1-f_h)}{n_h} (s_{yh}^2 - b_{1h}^2 s_{xh}^2)}$$

Estimador de regresión simple (o separado) en muestreo estratificado con Bootstrap

El Bootstrap es una técnica desarrollada para realizar ciertos tipos de inferencia estadística. Permite simplificar los engorrosos cálculos de la teoría estadística tradicional con el uso de sistemas de cómputo.

Si F es la distribución empírica con probabilidad $1/n$ para cada valor observado $x_i, i=1,2,...,n$. Una muestra bootstrap es definida como una muestra aleatoria de tamaño n obtenida a partir de F , es decir $x^*=(x_1^*, (x_2^*,L, x_K^*))$, $F @ x^*=(x_1^*, (x_2^*,L, x_K^*))$ La notación de asterisco indica que x^* no es el actual conjunto de datos x , pero es aleatorizado o remuestreado a partir de x .

El estimador bootstrap de $EE_F(\hat{\theta})$, el error estándar de una estadística $\hat{\theta}$, es el estimador plug-in que usa la función de distribución empírica F en lugar de la desconocida distribución F . Específicamente, el estimador bootstrap de $\hat{\theta}$ es definido por:

$$EE_{F^*}(\hat{\theta}^*)$$

Los pasos para la estimación del error estándar mediante bootstrap son:

- a) Seleccionar B muestras bootstrap independientes x^{*1}, x^{*2},K, x^{*3} , cada una compuesta por n valores obtenidas con reemplazo de x . Para estimar el error estándar, el valor de B podría estar comprendido en el rango de 25-200.
- b) Evalúe las replicaciones bootstrap correspondiente a

$$\widehat{EE}(\bar{y}^*_{st\ l_s}) = \sqrt{\sum_{h=1}^L W_h^2 \frac{(1-f_h)}{n_h} (s_{yh}^{*2} + b_{1h}^2 s_{xh}^{*2} - 2b_{1h} s_{xyh}^*)} = \sqrt{\sum_{h=1}^L W_h^2 \frac{(1-f_h)}{n_h} (s_{yh}^{*2} - b_{1h}^2 s_{xh}^{*2})}$$

4. Resultados

La precisión de los estimadores se evaluó mediante su error estándar. Para la obtención de los errores estándar, se implementaron funciones en R (ver anexo).

Asimismo, para no favorecer a ninguno de los estimadores evaluados se realizaron 100 repeticiones para cada uno de los escenarios propuestos y se promediaron los errores estándar en cada uno de ellos.

Luego de realizar las corridas respectivas, se obtuvieron los resultados que se resumen en la tabla 3.

cada muestra bootstrap

$$\hat{\theta}^*(b) = s(x^{*b}) \quad b=1,2,...,B$$

c) Estime el error estándar $se_F(\hat{\theta})$ para la muestra de mediante la desviación estándar de las B replicaciones

Donde
$$se_B = \sqrt{\frac{\sum_{b=1}^B \hat{\theta}^*(b) - q^*(\hat{\theta})}{B-1}}$$

$$q^*(\hat{\theta}) = \frac{\sum_{b=1}^B \hat{\theta}^*(b)}{B}$$

En el caso de modelos de regresión lineal simple, se remuestrea datos bivariados y en cada una de ellas se estima los coeficientes del modelo. Por lo tanto, el estimador estratificado de la media mediante el estimador de regresión es:

$$\bar{y}^*_{st\ l_s} = \sum_{h=1}^L W_h \bar{y}^*_{lh}$$

Donde:

$$\bar{y}^*_{lh} = \bar{y}^*_h + b_{1h}^* (\bar{x}_h - \bar{x}^*_h)$$

El error estándar de este estimador es:

Tabla 4. Errores Estándar de los estimadores evaluados

Escenario	Estimadores		
	Simple	Regresión Lineal Simple	Regresión Lineal Simple Bootstrap
1	1.528	1.106	1.078
2	1.247	0.938	0.921
3	0.686	0.507	0.503
4	0.554	0.416	0.415
5	1.529	0.796	0.782
6	1.231	0.673	0.664
7	0.678	0.362	0.361
8	0.557	0.309	0.308
9	1.129	0.791	0.770
10	1.168	0.834	0.813
11	0.519	0.372	0.369
12	0.418	0.299	0.298
13	1.137	0.502	0.486
14	1.158	0.499	0.487
15	0.520	0.231	0.228
16	0.417	0.185	0.184

Fuente: Elaboración propia

Los estimadores que utilizan una variable auxiliar (X) presentaron una mayor precisión (menor error estándar) en comparación al estimador simple (no utiliza una variable auxiliar). Esto debido a que en los datos existía una correlación significativa entre la variable de interés (Y) con la variable auxiliar.

Se puede apreciar que en todos los escenarios, los errores estándar obtenidos con el método de regresión lineal simple bootstrap son menores al de los otros métodos propuestos. Esto indica que este estimador es el más preciso. Este resultado es lo que se pretendía demostrar en la presente investigación.

No existe una gran diferencia en los resultados de los errores estándar obtenidos con el estimador de regresión lineal simple y regresión lineal simple bootstrap. Esto quizás se deba a que los datos generados provienen de una distribución normal bivariada; y, bajo esta condición, el estimador de regresión lineal simple funciona muy bien. Sin embargo, la metodología bootstrap es utilizada en datos cuya distribución no necesariamente es conocida. Por ello, se deja abierta la posibilidad de evaluar a este estimador con datos provenientes de otras distribuciones.

Dado que la metodología bootstrap es considerada un método no paramétrico, el estimador propuesto puede ser utilizado para obtener intervalos de confianza en el cual el estimador de regresión lineal simple no puede ser utilizado por limitaciones en el cumplimiento de supuestos.

5. Conclusiones

El presente trabajo constituye un aporte en la investigación sobre la comparación de estimadores en muestras complejas específicamente en muestreo estratificado utilizando métodos indirectos de estimación.

Se ha demostrado que al estimar la media con el método de regresión lineal simple bootstrap se obtiene un menor error estándar es decir es más preciso.

Se ha trabajado con una correlación para toda la muestra, pero los programas están diseñados de tal manera que se pueda considerar una correlación distinta para cada estrato. Esto podría ser de interés para analizar si el estimador propuesto sigue siendo el mejor bajo esas condiciones.

Para el estimador bootstrap propuesto, se utilizaron 200 muestras bootstrap en cada iteración, sin embargo no es complicado realizar una ligera modificación a los programas presentados para que se puedan utilizar diferentes tamaños de muestra bootstrap.

Se han presentado 16 escenarios distintos, sobre todo haciendo variar los tamaños de muestra en los estratos; sin embargo los programas elaborados permiten realizar las estimaciones para otros escenarios donde se pueden considerar por ejemplo varianzas significativamente diferentes entre los estratos es.

El estimador de regresión lineal simple bootstrap

puede ser utilizado cuando los datos no necesariamente provienen de una distribución normal bivariada.

El programa estadístico R es muy útil para implementar nuevos procedimientos de estimación, debido a que los programas estadísticos comerciales no cuentan con ellos.

6. Literatura citada

Álvarez, D. Sánchez, J. Gómez, G. y Tarqui, C. 2012. Sobrepeso y obesidad: Prevalencia y determinantes sociales del exceso del peso en la población peruana (2009-2010). Revista Peruana de Medicina Experimental y Salud Pública. Vol.29, n.3, pp. 303-313. Instituto Nacional de Salud. Lima

Cañizares, M. Barroso, I. Alfonso, A. García, R., Alfonso, K. Chang de la Rosa, M. Bonet, M. y León, E. 2004. Estimaciones usadas en diseños muestrales complejos: aplicaciones en la encuesta de salud cubana del año 2001. <http://www.scielosp.org/pdf/rpss/v15n3/a06v15n3.pdf>.

Cochran, W (1999) Técnicas de Muestreo. México: Editorial Continental.

Efron, B. and Tibshirani, R. (1993): An Introduction to the Bootstrap. Chapman & Hall.

Gutiérrez, H. Estrategias de Muestreo: Diseño de encuestas y estimación de parámetros (2009). Universidad Santo Tomás. Colombia.

Martínez, W.L, & Martínez, A. (2002). Computational Statistics handbook with MATLAB. Chapman and Hall/CRC, Boca Raton. Florida.

Pérez, C. 2009. Técnicas de Muestreo Estadístico: Teoría, práctica y aplicaciones informáticas,

Sharon L. Lohr. 2000. Muestreo Diseño y Análisis Internacional Editorial Thompson.

Sheaffer, J., Mendenhall, A., & Ott, R. (1987). Muestreo estadístico. Editorial Latinoamericana. México.

R version 3.1.1 Copyright © 2014. The R Foundation for Statistical Computing.

7. ANEXO

Función pobla

Descripción: La función a la que se denominó "pobla" permite simular una población normal bivariada compuesta por L estratos.

Argumentos:

Esta función necesita como argumentos:

tp: El tamaño de la población en cada uno de los estratos.

medias: La matriz de medias bivariada para cada uno de los estratos.

Covarianzas: La matriz de covarianzas bivariada para cada uno de los estratos

Valor:

La función brinda como resultado una matriz bivariada

que incluye una columna que indica a que estrato pertenece la observación bivariada generada.

```
pobla<-function(tp,medias, covarianzas)
{
  L<-length(tp)
  cod<-rep(1:L,tp)
  inicial<-matrix(0,1,2)
  for (i in 1:L)
  {
    pobla<-mvrnorm(tp[i],medias[i,], covarianzas[(2*i-1):(2*i),])
    inicial<-rbind(inicial,pobla)
  }
  poblaf<-inicial[-1,]
  poblaf<-as.matrix(cbind(cod,poblaf))
  return(poblaf)
}
```

Función selecc

Descripción: La función a la que se denominó “selecc” permite seleccionar una muestra aleatoria de la población previamente simulada.

Argumentos:

Esta función necesita como argumentos:

poblacion: Es la matriz de datos que contiene a la población bivariada simulada

tm: El vector que indica el tamaño de muestra a seleccionar en cada uno de los estratos.

Valor:

La función brinda como resultado una matriz muestral bivariada que incluye una columna que indica a que estrato pertenece cada observación seleccionada.

```
selecc<-function(poblacion,tm)
{
  Ns<-as.numeric(table(poblacion[,1]))
  L<-length(Ns)
  n<-sum(tm)
  var<-dim(poblacion)[2]
  inicial<-poblacion[sample(1:Ns[1],tm[1],FALSE),]
  for (i in 2:L)
  {
    indices<-sample(cumsum(Ns[i-1])+1:cumsum(Ns[i]),t
m[i],FALSE)
    muestra<-poblacion[indices,]
    inicial<-rbind(inicial,muestra)
  }
  return(inicial)
}
```

Función estimacion.s

Descripción: La función a la que se denominó “estimacion.s” realiza la estimación simple de la media en un muestreo estratificado.

Argumentos:

Esta función necesita como argumentos:

poblacion: Es la matriz que contiene a la población simulada.

muestra: Es la matriz que contiene a la muestra aleatoria extraída de la población simulada.

Valor:

Brinda como resultados la media y el error estándar en un muestreo estratificado

```
estimacion.s<-function(poblacion,muestra)
```

```
{
  Nh<-as.numeric(table(poblacion[,1]))
  nh<-as.numeric(table(muestra[,1]))
  Wh<-Nh/sum(Nh)
  fh<-nh/Nh
  ybarh<-as.numeric(by(muestra[,2],muestra[,1],mean))
  yvarh<-as.numeric(by(muestra[,2],muestra[,1],sd))^2
  mest<-sum(Wh*ybarh)
  eeest<-sqrt(sum(Wh^2*yvarh*(1-fh)/nh))
  return(list(mest=mest,eeest=eeest))
}
```

Funciones pendiente, pendienteb, mediasb, varb

Descripción: La funciones denominadas “pendiente”, “pendienteb”, “mediasb” y “varb”, sirven como funciones auxiliares que permitirán el cálculo de la media y el error estándar en las dos funciones de estimación final de la media.

Argumentos:

Esta función necesita como argumento:

datos: Es el vector de datos a utilizar.

Valor:

Brinda como resultados estimaciones de la pendiente, pendiente bootstrap, media bootstrap y varianza bootstrap en cada uno de los estratos.

```
pendiente<-function(datos)
{
  datos<-as.matrix(datos)
  b<-as.numeric(lm(datos[,2]~datos[,3])$coef[2])
  return(b)
}

pendienteb<-function(datos)
{
  datos<-as.matrix(datos)
  n<-nrow(datos)
```

```
betas<-ncol(datos)
betasboot<-matrix(0,200,betas)
for (i in 1:200)
{
  indices<-sample(1:n,n,TRUE)
  betasboot[i,<-lm(datos[indices,2]~datos[indices,3])$
  coe
}
estboot<-apply(betasboot,2,mean)[2]
return(estboot)
}
```

```
meanb<-function(datos)
{
  n<-length(datos)
  estadistico<-rep(0,200)
  for (i in 1:200)
  {
    estadistico[i]<-mean(sample(datos,n,TRUE))
  }
  mediab<-mean(estadistico)
  return(mediab)
}
```

```
varb<-function(datos)
{
  n<-length(datos)
  estadistico<-rep(0,200)
  for (i in 1:200)
  {
    estadistico[i]<-var(sample(datos,n,TRUE))
  }
  variab<-mean(estadistico)
  return(variab)
}
```

Funciones estimacion.mi y estimacion.mib

Descripción: Las funciones denominadas “estimacion.mi” y “estimacion.mib” permiten el cálculo de la media y su error estándar con el método de regresión lineal simple y el método de regresión lineal bootstrap.

Argumentos:

Esta función necesita como argumento:

poblacion: Es la matriz que contiene a la población simulada.

muestra: Es la matriz que contiene a la muestra aleatoria extraída de la población simulada.

Valor:

Brinda como resultados la media y el error estándar

utilizando métodos indirectos de estimación.

```
estimacion.mi<-function(poblacion,muestra) {
  Nh<-as.numeric(table(poblacion[,1]))
  nh<-as.numeric(table(muestra[,1]))
  Wh<-Nh/sum(Nh)
  fh<-nh/Nh
  muxh<-as.numeric(by(poblacion[,3],poblacion[,1],mean))
  ybarh<-as.numeric(by(muestra[,2],muestra[,1],mean))
  xbarh<-as.numeric(by(muestra[,3],muestra[,1],mean))
  yvarh<-as.numeric(by(muestra[,2],muestra[,1],sd))^2
  xvarh<-as.numeric(by(muestra[,3],muestra[,1],sd))^2
  betah<-as.numeric(by(muestra[,2:3],muestra[,1],pendiente))
  ybarhl<-ybarh+betah*(muxh-xbarh)
  mestl<-sum(Wh*ybarhl)
  eeestl<-sqrt(sum(Wh^2*(1-fh)*(yvarh-betah^2*xvarh)/nh))
  return(list(mestl=mestl,eeestl=eeestl))
}
```

```
estimacion.mib<-function(poblacion,muestra)
{
  Nh<-as.numeric(table(poblacion[,1]))
  nh<-as.numeric(table(muestra[,1]))
  Wh<-Nh/sum(Nh)
  fh<-nh/Nh
  muxh<-as.numeric(by(muestra[,3],muestra[,1],mean))
  ybarhb<-as.numeric(by(muestra[,2],muestra[,1],mean))
  xbarhb<-as.numeric(by(muestra[,3],muestra[,1],mean))
  yvarhb<-as.numeric(by(muestra[,2],muestra[,1],var))
  xvarhb<-as.numeric(by(muestra[,3],muestra[,1],var))
  betahb<-as.numeric(by(muestra[,2:3],muestra[,1],pendiente))
  ybarhbl<-ybarhb+betahb*(muxh-xbarhb)
  mestlb<-sum(Wh*ybarhbl)
  eeestlb<-sqrt(sum(Wh^2*(1-fh)*(yvarhb-betahb^2*xvarhb)/nh))
  return(list(betahb=betahb,mestlb=mestlb,eeestlb=eeestlb))
}
```

Función repeti

Descripción: La función “repeti” permite obtener el cálculo de la media y su error estándar para “r” repeticiones de los procedimientos de estimación en estratificación de la media y su error estándar con los métodos: simple, de regresión lineal simple y de regresión lineal bootstrap

Argumentos:

Esta función necesita como argumento:

r: Número de repeticiones que se quiere realizar los procedimientos.

tp: El vector que contiene los tamaños de las poblaciones de cada estrato.

medias: El vector de medias de la población a ser simulada

covarianzas: La matriz varianza covarianza de la población a ser simulada.

tm: El vector de los tamaños de muestra a extraer de cada estrato.

Valor:

Brinda como resultados la media y el error estándar utilizando métodos indirectos de estimación.

```

repeti<-function(r,tp,medias,covarianzas,tm){
ees<-matrix(0,r,3)
est<-matrix(0,r,3)
for (i in 1:r) {
  poblac<-pobla(tp,medias,covarianzas)
  muestr<-selecc(poblac,tm)
  ees[i,1]<-estimacion.s(poblac,muestr)$eeres
  est[i,1]<-estimacion.s(poblac,muestr)$mest
  ees[i,2]<-estimacion.mi(poblac,muestr)$eeestl
  est[i,2]<-estimacion.mi(poblac,muestr)$mestl
  ees[i,3]<-estimacion.mib(poblac,muestr)$eeestlb
  est[i,3]<-estimacion.mib(poblac,muestr)$mestlb
}
resul1<-colMeans(ees)
resul2<-colMeans(est)
return(list(estima=resul2,eesta=resul1))
}
    
```