Solución Tarea 1

Muestreo Estadístico

14 de mayo de 2019

Integrantes del grupo

Nombre	Cédula
Esteban Moreno Rodríguez	1152459914
Luis Felipe Bedoya Martínez	1088015006
Sergio Iván Figueroa Sierra	1031156865
Yuberth Anderson Saavedra Coneo	Campo

Ejercicio

Se tiene una población hipotética de tamaño $N = 7$. Los valores de la variable de interés para los siete elementos u objetos de la población son: 13, 18, 6, 9, 16, 11 y 14.

1) ¿Cuántas muestras posibles sin reemplazo de tamaño $n = 3$ pueden seleccionarse?

Las muestras posibles son las combinaciones de a tres elementos entre los siete sin tener en cuenta su orden.

$$C_n^N = \binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{7!}{3!(7-3)!} = \frac{7!}{3!\,4!} = \frac{(7)(6)(5)\cancel{4!}}{3!\,\cancel{4!}} = \frac{(7)\cancel{6}(5)}{\cancel{3!}_{\;\cancel{6}}} = 35$$

A continuación se presenta un código en R para calcularlo.

```r
N = 7
n = 3
valores <- c(13, 18, 6, 9, 16, 11, 14)
muestras <- combinations(length(valores), n, valores)
nrow(muestras)
```

```
## [1] 35
```

2) Seleccione todas las muestras posibles de tamaño $n = 3$.

Las muestras de tamaño tres extraídas de la población son:

```r
muestras
```

```
##       [,1] [,2] [,3]
##  [1,]    6    9   11
##  [2,]    6    9   13
##  [3,]    6    9   14
##  [4,]    6    9   16
##  [5,]    6    9   18
##  [6,]    6   11   13
##  [7,]    6   11   14
##  [8,]    6   11   16
##  [9,]    6   11   18
## [10,]    6   13   14
## [11,]    6   13   16
## [12,]    6   13   18
## [13,]    6   14   16
```

```
## [14,]    6   14   18
## [15,]    6   16   18
## [16,]    9   11   13
## [17,]    9   11   14
## [18,]    9   11   16
## [19,]    9   11   18
## [20,]    9   13   14
## [21,]    9   13   16
## [22,]    9   13   18
## [23,]    9   14   16
## [24,]    9   14   18
## [25,]    9   16   18
## [26,]   11   13   14
## [27,]   11   13   16
## [28,]   11   13   18
## [29,]   11   14   16
## [30,]   11   14   18
## [31,]   11   16   18
## [32,]   13   14   16
## [33,]   13   14   18
## [34,]   13   16   18
## [35,]   14   16   18
```

3) Calcule la media, la varianza S^2, la varianza S_n^2 y el total muestral de cada una de dichas muestras.

El cálculo de estos estadísticos se realiza mediante las siguientes ecuaciones:

La media muestral es:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

La varianza(cuasivarianza) muestral es:

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$

La varianza(simple) muestral es:

$$s_n^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}$$

El total muestral es:

$$y = \sum_{i=1}^{n} y_i$$

La desviación estándar muestral es:

$$s = \sqrt{s^2}$$

El código en R usado para calcular cada una de las estadísticas anteriores se expone en el siguiente cuadro, y, la tabla que le sigue recopila toda la información calculada.

```r
# Media Muestrales
ym <-apply(muestras, 1, mean)
# Varianzas(cuasivarianzas) muestrales
SS <- apply(muestras, 1, var)
# Varianzas(simples) muestrales
SSn <- rowSums(matrix((c(muestras) - rep(apply(muestras, 1, mean),3))^2,
                      nrow = nrow(muestras), ncol = 3))/n
```

```r
# Desviación estándar ó raíz cuadrada de S^2 (sigma)
S <- apply(muestras, 1, sd)
# Totales muestrales
a <- rowSums(muestras)
```

```
##       Unidades Total MediaMuestral  Varianza VarSimple DesvEstandar
## 1     6  9  11    26      8.666667  6.333333  4.222222     2.516611
## 2     6  9  13    28      9.333333 12.333333  8.222222     3.511885
## 3     6  9  14    29      9.666667 16.333333 10.888889     4.041452
## 4     6  9  16    31     10.333333 26.333333 17.555556     5.131601
## 5     6  9  18    33     11.000000 39.000000 26.000000     6.244998
## 6    6  11  13    30     10.000000 13.000000  8.666667     3.605551
## 7    6  11  14    31     10.333333 16.333333 10.888889     4.041452
## 8    6  11  16    33     11.000000 25.000000 16.666667     5.000000
## 9    6  11  18    35     11.666667 36.333333 24.222222     6.027714
## 10   6  13  14    33     11.000000 19.000000 12.666667     4.358899
## 11   6  13  16    35     11.666667 26.333333 17.555556     5.131601
## 12   6  13  18    37     12.333333 36.333333 24.222222     6.027714
## 13   6  14  16    36     12.000000 28.000000 18.666667     5.291503
## 14   6  14  18    38     12.666667 37.333333 24.888889     6.110101
## 15   6  16  18    40     13.333333 41.333333 27.555556     6.429101
## 16   9  11  13    33     11.000000  4.000000  2.666667     2.000000
## 17   9  11  14    34     11.333333  6.333333  4.222222     2.516611
## 18   9  11  16    36     12.000000 13.000000  8.666667     3.605551
## 19   9  11  18    38     12.666667 22.333333 14.888889     4.725816
## 20   9  13  14    36     12.000000  7.000000  4.666667     2.645751
## 21   9  13  16    38     12.666667 12.333333  8.222222     3.511885
## 22   9  13  18    40     13.333333 20.333333 13.555556     4.509250
## 23   9  14  16    39     13.000000 13.000000  8.666667     3.605551
## 24   9  14  18    41     13.666667 20.333333 13.555556     4.509250
## 25   9  16  18    43     14.333333 22.333333 14.888889     4.725816
## 26 11  13  14    38     12.666667  2.333333  1.555556     1.527525
## 27 11  13  16    40     13.333333  6.333333  4.222222     2.516611
## 28 11  13  18    42     14.000000 13.000000  8.666667     3.605551
## 29 11  14  16    41     13.666667  6.333333  4.222222     2.516611
## 30 11  14  18    43     14.333333 12.333333  8.222222     3.511885
## 31 11  16  18    45     15.000000 13.000000  8.666667     3.605551
## 32 13  14  16    43     14.333333  2.333333  1.555556     1.527525
## 33 13  14  18    45     15.000000  7.000000  4.666667     2.645751
## 34 13  16  18    47     15.666667  6.333333  4.222222     2.516611
## 35 14  16  18    48     16.000000  4.000000  2.666667     2.000000
```

4) Hallar la distribución muestral de las medias y de las varianzas s^2 y s_n^2.

Con la información obtenida en los puntos anteriores hacemos las tablas de frecuencia correspondientes de cada estimador. Conocidas como distribuciones muestrales.

La distribución muestral de las medias $\bar{y}$ es:

$\bar{y}$	$\frac{26}{3}$	$\frac{28}{3}$	$\frac{29}{3}$	10	$\frac{31}{3}$	11	$\frac{34}{3}$	$\frac{35}{3}$	12	$\frac{37}{3}$	$\frac{38}{3}$	13	$\frac{40}{3}$	$\frac{41}{3}$	14	$\frac{43}{3}$	15	$\frac{47}{3}$	16	Total
π_i	$\frac{1}{35}$	$\frac{1}{35}$	$\frac{1}{35}$	$\frac{1}{35}$	$\frac{2}{35}$	$\frac{4}{35}$	$\frac{1}{35}$	$\frac{2}{35}$	$\frac{3}{35}$	$\frac{1}{35}$	$\frac{4}{35}$	$\frac{1}{35}$	$\frac{3}{35}$	$\frac{2}{35}$	$\frac{1}{35}$	$\frac{3}{35}$	$\frac{2}{35}$	$\frac{1}{35}$	$\frac{1}{35}$	1

La distribución muestral de las varianzas (cuasivaranzas) s^2 es:

s^2	$\frac{7}{3}$	4	$\frac{19}{3}$	7	$\frac{37}{3}$	13	$\frac{49}{3}$	19	$\frac{61}{3}$	$\frac{67}{3}$	25	$\frac{79}{3}$	28	$\frac{109}{3}$	$\frac{112}{3}$	39	$\frac{125}{3}$	Total
π_i	$\frac{2}{35}$	$\frac{2}{35}$	$\frac{5}{35}$	$\frac{2}{35}$	$\frac{3}{35}$	$\frac{5}{35}$	$\frac{2}{35}$	$\frac{1}{35}$	$\frac{2}{35}$	$\frac{2}{35}$	$\frac{1}{35}$	$\frac{2}{35}$	$\frac{1}{35}$	$\frac{2}{35}$	$\frac{1}{35}$	$\frac{1}{35}$	$\frac{1}{35}$	1

La distribución muestral de las varianzas (simples) s_n^2 es:

s_n^2	$\frac{14}{9}$	$\frac{8}{3}$	$\frac{38}{9}$	$\frac{14}{3}$	$\frac{74}{9}$	$\frac{26}{3}$	$\frac{98}{9}$	$\frac{38}{3}$	$\frac{122}{9}$	$\frac{134}{9}$	$\frac{50}{3}$	$\frac{158}{9}$	$\frac{56}{3}$	$\frac{218}{9}$	$\frac{224}{9}$	26	$\frac{248}{9}$	Total
π_i	$\frac{2}{35}$	$\frac{2}{35}$	$\frac{5}{35}$	$\frac{2}{35}$	$\frac{3}{35}$	$\frac{5}{35}$	$\frac{2}{35}$	$\frac{1}{35}$	$\frac{2}{35}$	$\frac{2}{35}$	$\frac{1}{35}$	$\frac{2}{35}$	$\frac{1}{35}$	$\frac{2}{35}$	$\frac{1}{35}$	$\frac{1}{35}$	$\frac{1}{35}$	1

Definimos la media de la distribución de un estimador $\widehat{\theta}$, como:

$$E[\widehat{\theta}] = \sum_{i=1}^{v} \widehat{\theta}_i \pi_i$$

donde: v el número total de valores distintos tomados por el estimador, $\widehat{\theta_i}$ es la i-ésima estimación diferente de θ y π_i la probabilidad de que el estimador tome el valor $\widehat{\theta}$. Esta probabilidad es igual a la frecuencia relativa de las estimaciones.

Este resultado será usado en los siguientes puntos.

5) Compruebe que la media de las medias muestrales es igual a la media poblacional, ie. verifíque que la media muestral es un estimador insesgado de la media poblacional.

$$E[\bar{y}] = \sum_{i=1}^{n} \bar{y}_i \pi_i = \mu = \frac{\sum_{i=1}^{N} Y_i}{N}$$

```
# Media de las medias muestrales
sum(dmym$Freq*as.double(as.character(dmym$ym)))
```

```
## [1] 12.42857
```

```
# Media poblacional
mean(valores)
```

```
## [1] 12.42857
```

6) Compruebe que la media de las varianzas muestrales s^2 es igual a la varianza poblacional; ie. verifíque que la varianza muestral s^2 es un estimador insesgado de la varianza poblacional.

$$E[s^2] = \sum_{i=1}^{n} s^2 \pi_i = \sigma^2 = Var[Y] = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})^2}{N-1}$$

```
# Media de las varianzas(cuasivarianzas) muestrales
sum(dmSS$Freq*as.double(as.character(dmSS$SS)))
```

```
## [1] 16.95238
```

```
# Varianza poblacional
var(valores)
```

```
## [1] 16.95238
```

7) Compruebe que la media de las varianzas muestrales s_n^2 no es igual a la varianza poblacional; ie. verifíque que la varianza muestral s_n^2 es un estimador sesgado de la varianza poblacional. Hallar el sesgo de dicho estimador.

$$E[s_n^2] = \sum_{i=1}^{n} s_n^2 \pi_i \neq \sigma^2 = Var[Y] = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}{N-1}$$

El sesgo se define como:

$$B[\hat{\theta}] = E[\hat{\theta}] - \theta$$

```
# Media de las varianzas(simples) muestrales
sum(dmSSn$Freq*as.double(as.character(dmSSn$SSn)))
```

```
## [1] 11.30159
```

```
# Varianza poblacional
var(valores)
```

```
## [1] 16.95238
```

```
# Valor de B
mean(SSn) - sum((c(valores) - mean(valores))^(2))/n
```

```
## [1] -22.60317
```

8) Compruebe que la media de los totales muestrales es igual al total poblacional; ie. verifíque que el total muestral es un estimador insesgado del total poblacional.

Falta corregir este punto Falta corregir este punto Falta corregir este punto

```
# Media de los totales muestrales
sum(dma$Freq*as.double(as.character(dma$a)))
```

```
## [1] 37.28571
```

```
mean(a)
```

```
## [1] 37.28571
```

```
# Total poblacional
sum(valores)
```

```
## [1] 87
```

9) Seleccione una de todas las muestras posibles mediante M.A.S y realice las estimaciones de la media poblacional y del total poblacional, junto con sus respectivos $B = L.E.E$. Interprete los resultados.

Falta escribir la interpretación de los resultados Falta escribir la interpretación de los resultados Falta escribir la interpretación de los resultados

```
# Selección de la fila de la matrix
MAS <- sample(1:35, 1, replace = FALSE, prob = rep(1/35,35))
MAS
```

```
## [1] 11
```

```
uMAS <- muestras[MAS,]
uMAS
```

```
## [1]  6 13 16
```

```
mean(uMAS)
```

```
## [1] 11.66667
```

```
sum(uMAS)
```

```
## [1] 35
```

```r
2*sqrt(var(uMAS))
```

```
## [1] 10.2632
```

```r
N*2*sqrt(var(uMAS))
```

```
## [1] 71.84242
```

10) Probar:

$$Var[\bar{y}] = (1 - \frac{n}{N})\frac{\sigma^2}{n}$$

```r
# Varianza de la media muestral
sum(((as.double(as.character(dmym$ym))-mean(valores))^2)*dmym$Freq)
```

```
## [1] 3.229025
```

```r
# Varianza de la media muestral en el M.A.S
(1-(n/N))*(var(valores)/n)
```

```
## [1] 3.229025
```