

**UNIVERSIDAD DE EL SALVADOR**  
**FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE**  
**DEPARTAMENTO DE MATEMÁTICA**



**Licenciatura en Estadística**

**Control Estadístico del Paquete R**

**Estimación en R: Intervalos de confianza.**

**Alumna:**  
**Martha Yoana Medina Sánchez**

**Fecha de elaboración**  
**Santa Ana - 28 de noviembre de 2015**

## 1. INTERVALOS DE CONFIANZA PARA UNA MEDIA POBLACIONAL.

Antes que nada se definirá que es un intervalo de confianza:

Conjunto de valores que se forman a partir de una muestra de datos de forma que exista la posibilidad de que el parámetro poblacional ocurra dentro de dicho conjunto con una probabilidad específica. La probabilidad específica recibe el nombre de nivel de confianza.

### PRIMER CASO: VARIANZA CONOCIDA ( $\sigma^2$ ).

Un intervalo de confianza se calcula con el empleo de dos estadísticos: la media muestral  $\bar{X}$  y la desviación estándar.

En general, un intervalo de confianza de la media poblacional, cuando se conoce la desviación estándar poblacional, se calcula de la siguiente manera:

$$\bar{X} - Z_{(\alpha/2)} * (\sigma / \sqrt{n}) \leq \mu \leq \bar{X} + Z_{(\alpha/2)} * (\sigma / \sqrt{n})$$

Ejemplo: Los siguientes datos son los pesos en gramos de 20 cajas de cereal. Si el peso de la caja es una variable aleatoria normal con una desviación estándar de 5 gr. Obtenga los intervalos de confianza estimado del 99 % para la media del proceso del llenado

```
Peso <- c( 506, 510, 493, 508, 497, 506, 499, 512, 502, 503, 504, 514,
          505, 505, 509, 496, 480, 499, 560, 478);

Peso

## [1] 506 510 493 508 497 506 499 512 502 503 504 514 505 505 509 496 480
## [18] 499 560 478

# Construyendo una función en R para realizar la prueba de intervalos de
# confianza para varianza conocida.

intervalovarcono <- function(n, sigma, nivel.conf=0.99)
{
  n = length(Peso)
  media = mean(Peso) # Calcula la media
  alfa = (1 - nivel.conf) # Calcula el alfa
  margen_error = qnorm(alfa/2, mean = 0, sd = 1, lower.tail = FALSE)*
    (sigma/sqrt(n))
  print(rbind(media, n, alfa, margen_error))
  LInf = media - margen_error
  LSup = media + margen_error
  print("LInf, LSup")
  print(paste("Intervalo para varianza conocida es: [", round(LInf, 2),
    ",", round(LSup, 2), "]"))
}
```

```
n=length(Peso); sigma=5; nivel.conf=0.99

intervalovarcono(n, nivel.conf)

##           [,1]
## media      504.3000000
## n          20.0000000
## alfa        0.0100000
## margen_error 0.5702132
## [1] "LInf, LSup"
## [1] "Intervalo para varianza conocida es: [ 503.73 , 504.87 ]"
```

## CASO DOS: VARIANZA DESCONOCIDA ( $SIGMA^2$ )

En la mayoría de los casos de muestreo, no se conoce la desviación estándar de la población.

Para encontrar el intervalo de confianza de la media poblacional con *sigma* desconocida.

$$\tilde{X} - Z_{(\alpha/2, gl)} * (\sigma / \sqrt{n}) \leq \mu \leq \tilde{X} + Z_{(\alpha/2, gl)} * (\sigma / \sqrt{n})$$

Ejemplo: La Cámara de Comercio de una ciudad se encuentra interesada en estimar la cantidad promedio de dinero que gasta la gente que asiste a convenciones, calculando comidas, alojamiento y entretenimiento por día. De las distintas convenciones que se llevan a cabo en la ciudad, se seleccionaron 16 personas y se les preguntó la cantidad de dinero que gastaba por día. Si se supone que la cantidad de dinero gastada en un día es una variable distribuida normal, obtener los intervalos de confianza estimado del 90 % para la cantidad promedio real.

```
Gasto <- c( 150, 175, 163, 148, 142, 189, 135, 174, 168, 152, 158, 184,
           134, 146, 155, 163);

Gasto

## [1] 150 175 163 148 142 189 135 174 168 152 158 184 134 146 155 163

# Construyendo una función en R para realizar la prueba de intervalos de
# confianza para varianza desconocida.

intervardescono <- function(n, nivel.conf=0.90)
{
  n = length(Gasto)
  media = mean(Gasto) # Calcula la media
  desviacion = sd(Gasto) # Calcula la desviación estándar
  alfa = (1 - nivel.conf) # Calcula el alfa
  gl = n - 1 # Calcula los grados de libertad
  margen_error = qt(alfa/2, gl, lower.tail = FALSE) *
    (desviacion/sqrt(length(Gasto)))
}
```

```

print(rbind(n, media, desviacion, alfa, gl, margen_error))
LInf = media - margen_error
LSup = media + margen_error
print("LInf, LSup")
print(paste("Intervalo para varianza conocida es: [", round(LInf, 2),
            ",", round(LSup, 2), "]"))
}
n=length(Gasto); nivel.conf=0.90

intervardescono (n, nivel.conf)

##           [,1]
## n          16.000000
## media       158.500000
## desviacion   16.415440
## alfa         0.100000
## gl          15.000000
## margen_error  7.194273
## [1] "LInf, LSup"
## [1] "Intervalo para varianza conocida es: [ 151.31 , 165.69 ]"

```

```

# Utilizando la funci\on proporcionada por el R.

Gasto <- c( 150, 175, 163, 148, 142, 189, 135, 174, 168, 152, 158, 184,
           134, 146, 155, 163);

Gasto

## [1] 150 175 163 148 142 189 135 174 168 152 158 184 134 146 155 163

t.test(Gasto, alternative = "two.sided", conf.level = 0.90)

##
## One Sample t-test
##
## data:  Gasto
## t = 38.622, df = 15, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  151.3057 165.6943
## sample estimates:
## mean of x
##      158.5

```

## 2. INTERVALOS DE CONFIANZA PARA UNA PROPORCIÓN.

Para crear el intervalo de confianza de una proporción de población se aplica la fórmula:

$$\hat{p} - Z_{(\alpha/2)} * (\text{sqrt}(p(1-p)/n)) \leq p \leq \hat{p} + Z_{(\alpha/2)} * (\text{sqrt}(p(1-p)/n))$$

Ejemplo: En una año de elecciones presidenciales, los candidatos quieren saber que votarán los votantes de diferentes partes del país. Suponga que se pregunta a 420 posibles votantes del noreste si votarían a un determinado candidato si las elecciones fueran hoy. En esta muestra 233 indicaron que votaría a favor de este candidato. Halle la estimación del intervalo de confianza 95 % del apoyo con que cuenta este candidato en el noreste.

```
# Construyendo una función en R para realizar la prueba de intervalos de  
# confianza para proporción.
```

```
intervaloProp <- function(x, n, nivel.conf = 0.95)  
{  
  pe <- x/n  
  alfa <- (1 - nivel.conf)  
  z <- qnorm(1-alfa/2)  
  SE <- sqrt(pe*(1-pe)/n)  
  print(rbind(pe, alfa, z, SE))  
  LInf <- pe-z*SE  
  LSup <- pe+z*SE  
  print(" ")  
  print(paste("Intervalo para p es: [", round(LInf, 2),  
              ", ", round(LSup, 2), "]" ))  
}  
x=223; n=420; nivel.conf=0.95  
  
intervaloProp(x, n, nivel.conf)  
  
##           [,1]  
## pe    0.53095238  
## alfa  0.05000000  
## z     1.95996398  
## SE    0.02435071  
## [1] " "  
## [1] "Intervalo para p es: [ 0.48 , 0.58 ]"
```

```
# Utilizando la función proporcionada por el R.
```

```
prop.test(223, 420, alternative = "two.sided", conf.level=0.95)  
  
##  
## 1-sample proportions test with continuity correction  
##
```

```
## data: 223 out of 420, null probability 0.5
## X-squared = 1.4881, df = 1, p-value = 0.2225
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
## 0.4819748 0.5793550
## sample estimates:
## p
## 0.5309524
```

### 3. INTERVALOS DE CONFIANZA PARA LA VARIANZA POBLACIONAL

Ejemplo: El gerente de Inlet Square Mall, cerca de Ft. Myers, Florida, desea estimar la cantidad media que gasta los clientes que visitan el centro comercial. Una muestra de 20 clientes revela las siguientes cantidades.

Determine un intervalo de confianza del 95 %.

```
GastoClientes <- c (48.16, 42.22, 46.82, 51.45, 23.78, 41.86, 54.86,
                   37.92, 52.64, 48.59, 50.82, 49.94, 61.83, 661.69,
                   49.17, 61.46, 51.35, 52.68, 58.84, 43.88);
GastoClientes

## [1] 48.16 42.22 46.82 51.45 23.78 41.86 54.86 37.92 52.64 48.59
## [11] 50.82 49.94 61.83 661.69 49.17 61.46 51.35 52.68 58.84 43.88

# Construyendo una función en R para realizar la prueba de intervalos de
# confianza para la varianza poblacional.

intervalovarpobla <- function(n, nivel.conf=0.95)
{
  n=length(GastoClientes)
  desviacion = (sd(GastoClientes)) # varianza
  gl = n - 1 # grados de libertad
  alfa = 1 - nivel.conf # valor de alfa
  X_1 <- qchisq((alfa/2), gl, lower.tail = FALSE)
  X_2 <- qchisq(1 - (alfa/2), gl, lower.tail = FALSE)
  u <- ((n - 1)*desviacion)
  print(rbind(desviacion, gl, alfa, X_1, X_2, u))
  LInf <- u/X_1
  LSup <- u/X_2
  print(" ")
  print(paste("Intervalo para D es: [", round(LInf, 2), ",", round(LSup, 2), "]"))
}
n=length(GastoClientes); nivel.conf=0.95
intervalovarpobla ( n, nivel.conf)
```

```
##           [,1]
## desviacion 137.298685
## gl         19.000000
## alfa       0.050000
## X_1        32.852327
## X_2        8.906516
## u          2608.675007
## [1] " "
## [1] "Intervalo para D es: [ 79.41 , 292.9 ]"
```

#### 4. INTERVALOS DE CONFIANZA PARA DIFERENCIA DE MEDIAS, MUESTRAS INDEPENDIENTES.

##### CASO 1: VARIANZAS CONOCIDAS.

Ejemplo: La Area Chamber of Commerce de Tampa Bay (Florida) quería saber si el salario semanal medio de las enfermeras eran mayor que el de los maestros de escuela. Para esta investigación recopiló la información siguiente sobre las cantidades que ganó la semana pasada una muestra de maestros y enfermeras. Utilice un nivel de significancia de 99 %.

```
Maestros <- c(845, 826, 827, 875, 784, 809, 802, 820, 829, 830, 842, 832);

Maestros

## [1] 845 826 827 875 784 809 802 820 829 830 842 832

Enfermeras <- c(841, 890, 821, 771, 850, 859, 825, 829);

Enfermeras

## [1] 841 890 821 771 850 859 825 829

# Construyendo una funci\on en R para realizar la prueba de intervalos de
# confianza para varianza conocida.

interdifvarcono <- function(n1, n2, nivel.conf=0.99)

{
  n1 = length(Maestros) # Tama\~no de la poblaci\on Maestros
  n2 = length(Enfermeras) # Tama\~no de la poblaci\on Enfermeras
  desviacion1 = (sd(Maestros)) # Desviacion de la poblaci\on Maestros
  desviacion2 = (sd(Enfermeras)) # Desviacion de la poblaci\on Enfermeras
  media1 = (mean(Maestros)) # Media de la poblaci\on Maestros
  media2 = (mean(Enfermeras)) # Media de la poblaci\on Enfermeras
  alfa = (1 - nivel.conf)
  Diferencia = media1 - media2
```

```

margen_error = qnorm(alfa/2, mean = 0, sd = 1, lower.tail = FALSE)*
  sqrt((desviacion1/n1) + (desviacion2/n2))
print(rbind(n1, n2, media1, media2, desviacion1, desviacion2, alfa,
  margen_error))
LInf = Diferencia-margen_error
LSup = Diferencia+margen_error
print("")
print(paste("Intervalo para diferencia: [", round(LInf, 2), ",", round(LSup, 2), "]"))
}

n1 = length(Maestros); n2 = length(Enfermeras); nivel.conf=0.99
interdifvarcono (n1, n2, nivel.conf=0.99)

##           [,1]
## n1      12.000000
## n2       8.000000
## media1   826.750000
## media2   835.750000
## desviacion1 22.839858
## desviacion2 34.404111
## alfa      0.010000
## margen_error 6.415747
## [1] ""
## [1] "Intervalo para diferencia: [ -15.42 , -2.58 ]"

```

## CASO 2: VARIANZAS DESCONOCIDAS PERO IGUALES (MUESTRAS PEQUEÑAS)

Ejemplo: La ciudad de laguna Beach opera dos estacionamientos públicos. El de Ocean Drive tiene capacidad para 125 automóviles, y el Río Rancho, para 130. Los planeadores urbanos consideran tanto aumentar el tamaño de los estacionamientos como cambiar la estructura de las tarifas. Para iniciar, la oficina de planeación desea conocer el número de automóviles que hay en los estacionamientos en diversas horas del día. Se encarga a un funcionario de planeación principiante la tarea de visitar los dos estacionamientos a horas aleatorias del día y la tarea para contar el número de vehículos en ellos. El estudio se realizó durante un periodo de un mes. A continuación se presenta el número de automóviles en los estacionamientos durante 25 visitas al estacionamiento Ocean Drive y 28 al Río Rancho. Utilice el nivel de significancia de 95 %.

```

OceanDrive <- c(89, 115, 93, 79, 113, 77, 51, 75, 118, 105, 106, 91,
  54, 63, 121, 53, 81, 115, 67, 53, 69, 95, 121, 88, 64);

OceanDrive

## [1] 89 115 93 79 113 77 51 75 118 105 106 91 54 63 121 53 81
## [18] 115 67 53 69 95 121 88 64

RioRancho <- c(128, 110, 81, 126, 82, 114, 93, 40, 94, 45, 84, 71, 74,

```



```

          92, 66, 69, 100, 114, 113, 107, 62, 77, 80, 107, 90, 129,
          105, 124);

RioRancho

## [1] 128 110 81 126 82 114 93 40 94 45 84 71 74 92 66 69 100
## [18] 114 113 107 62 77 80 107 90 129 105 124

# Construyendo una función en R para realizar la prueba de intervalos de
# confianza para varianzas desconocidas iguales.

interdifvarcono <- function(n1, n2, nivel.conf=0.95)
{
  n1 = length(OceanDrive) # Tamaño de la población OceanDrive
  n2 = length(RioRancho) # Tamaño de la población RioRancho
  varianza1 = (var(OceanDrive)) # Desviación de la población OceanDrive
  varianza2 = (var(RioRancho)) # Desviación de la población RioRancho
  media1 = (mean(OceanDrive)) # Media de la población OceanDrive
  media2 = (mean(RioRancho)) # Media de la población RioRancho
  Diferencia = media1 - media2
  alfa = (1 - nivel.conf)
  gl = (n1 + n2 - 2)
  p = ((n1 - 1)*varianza1) + ((n2 - 1)*varianza2)
  S = p/gl
  margen_error = qt(alfa/2, gl, lower.tail = FALSE)*sqrt(S*(1/n1) + (1/n2))
  print(rbind(n1, n2, media1, media2, varianza1, varianza2, alfa, margen_error))
  LInf <- Diferencia - margen_error
  LSup <- Diferencia + margen_error
  print("")
  print(paste("Intervalo para diferencia varianzas desconocidas: [", round(LInf, 2),
              ",", round(LSup, 2), "]"))
}

n1=length(OceanDrive); n2=length(RioRancho); nivel.conf=0.95
interdifvarcono (n1, n2, nivel.conf=0.95)

##           [,1]
## n1         25.000000
## n2         28.000000
## media1      86.240000
## media2      92.035714
## varianza1   549.106667
## varianza2   581.739418
## alfa        0.050000

```

```
## margen_error    9.563145
## [1] ""
## [1] "Intervalo para diferencia varianza desconocida: [ -15.36 , 3.77 ]"
```

### CASO 3: VARIANZAS DESCONOCIDAS PERO DIFERENTES (MUESTRAS PEQUEÑAS)

Ejemplo: El centro médico Grand Starnd Family se diseño para atender emergencias médicas menores de los habitantes del área de Myrtle Beach. Hay dos instaaciones, una en Little River Area y la otra en Murrells Intel. El departamento de control de calidad desea comparar los tiempos de espera medios de los pacientes en las dos ubicaciones. Las muestras de los tiempos de espera, en minutos. Suponga que las desviaciones estándares de las poblaciones no son iguales. Con un nivel de significacia de 90 %.

```
LittleRiver <- c(31.73, 28.77, 29.53, 22.08, 29.47, 18.60, 32.94, 25.18, 29.82, 26.49);
LittleRiver

## [1] 31.73 28.77 29.53 22.08 29.47 18.60 32.94 25.18 29.82 26.49

MurrellsIntel <- c(22.93, 23.92, 26.92, 27.20, 26.44, 25.62, 30.61, 29.44, 23.09, 23.10, 26.69);
MurrellsIntel

## [1] 22.93 23.92 26.92 27.20 26.44 25.62 30.61 29.44 23.09 23.10 26.69
## [12] 22.31

# VARIANZAS DESCONOCIDAS PERO DIFERENTES (MUESTRAS PEQUEÑAS)
# Creando nuestra propia funci'on

interdes <- function(n1, n2, nivel.conf=0.90)
{
  n1 = length(LittleRiver) # Tamaño de la poblaci'on Cadiz
  n2 = length(MurrellsIntel) # Tamaño de la poblaci'on MurrellsIntel
  varianza1 = (var(LittleRiver)) # Varianza de LittleRiver
  varianza2 = (var(MurrellsIntel)) # Varianza de MurrellsIntel
  Media1 = (mean(LittleRiver)) # Media de LittleRiver
  Media2 = (mean(MurrellsIntel)) # Media de MurrellsIntel
  Diferencia = Media1 - Media2
  alfa = 1 - nivel.conf
  m1 = (n1 - 1)
  m2 = (n2 - 1)
  w = (varianza1/n1) + (varianza2/n2)
  w1 = (w^2)
  k = ((varianza1/n1)^2/m1) + ((varianza2/n2)^2/m2)
  v = (w1/k) # Grados de libertad
  r = (varianza1/n1) + (varianza2/n2)
  margen_error = qt(alfa/2, v, lower.tail = TRUE)*(sqrt(r))
}
```

```
print(rbind(n1, n2, Media1, Media2, varianza1, varianza2, alfa, margen_error))
LInf <- Diferencia-margen_error
LSup <- Diferencia+margen_error
print("")
print(paste("Intervalo para diferencia varianza desconocida diferentes: [",
            round(LInf, 2), ",", round(LSup, 2), "]"))

}

n1 = length(LittleRiver); n2 = length(MurrellsIntel); nivel.conf=0.90
interdes (n1, n2, nivel.conf=0.95)

##              [,1]
## n1          10.000000
## n2          12.000000
## Media1       27.461000
## Media2       25.689167
## varianza1    19.717477
## varianza2     7.206572
## alfa         0.050000
## margen_error -3.434532
## [1] ""
## [1] "Intervalo para diferencia varianza desconocida diferentes: [ 5.21 , -1.66 ]"
```

#### CASO 4: VARIANZAS DESCONOCIDAS PERO IGUALES (MUESTRAS GRANDES)

Ejemplo: La productividad en el trabajo depende de muchos factores, tales como el salario, las presentaciones, el ambiente laboral, etc. Por tal motivo una empresa decide someter a prueba dos modelos de producción. Para decidir qué modelo utilizar; se midieron los tiempos de producción de los empleados en cada modelo. Asuma que las desviaciones estándar de las poblaciones son iguales y utilice un nivel de significancia de 0.99.

```
ModeloA <- c(5.0, 5.1, 4.2, 4.5, 4.8, 6.0, 5.2, 5.5, 5.9, 6.3);
ModeloA

## [1] 5.0 5.1 4.2 4.5 4.8 6.0 5.2 5.5 5.9 6.3

ModeloB <- c(5.5, 4.4, 5.4, 4.8, 6.8, 5.5, 6.5, 5.2, 5.7, 5.6, 5.9, 6.0);
ModeloB

## [1] 5.5 4.4 5.4 4.8 6.8 5.5 6.5 5.2 5.7 5.6 5.9 6.0

# Construyendo una función en R para realizar la prueba de intervalos de
# confianza para varianza desconocida iguales.

interdifvarconop <- function(n1, n2, nivel.conf=0.99)
```

```

{
  n1 = length(ModeloA) # Tamaño de la población ModeloA
  n2 = length(ModeloB) # Tamaño de la población ModeloB
  varianza1 = (var(ModeloA)) # Desviación de la población ModeloA
  varianza2 = (var(ModeloB)) # Desviación de la población ModeloB
  media1 = (mean(ModeloA)) # Media de la población ModeloA
  media2 = (mean(ModeloB)) # Media de la población ModeloB
  Diferencia = media1 - media2
  alfa = (1 - nivel.conf)
  p = ((varianza1/n1) + (varianza2/n1))
  margen_error = qnorm(alfa/2, mean = 0, sd = 1, lower.tail = FALSE)*(sqrt(p))
  print(rbind(n1, n2, media1, media2, varianza1, varianza2, alfa, Diferencia,
              margen_error))
  LInf <- Diferencia-margen_error
  LSup <- Diferencia+margen_error
  print("")
  print(paste("Intervalo para diferencia varianza desconocida: [", round(LInf, 2),
              ",", round(LSup, 2), "]"))
}

n1 = length(ModeloA); n2 = length(ModeloB); nivel.conf=0.99
interdifvarconop (n1, n2, nivel.conf=0.99)

##           [,1]
## n1         10.0000000
## n2         12.0000000
## media1        5.2500000
## media2        5.6083333
## varianza1      0.4561111
## varianza2      0.4371970
## alfa           0.0100000
## Diferencia    -0.3583333
## margen_error   0.7698706
## [1] ""
## [1] "Intervalo para diferencia varianza desconocida: [ -1.13 , 0.41 ]"

```

## 5. INTERVALOS DE CONFIANZA PARA LA DIFERENCIA DE DOS PROPORCIONES.

Ejemplo: Una cadena de hoteles grande intenta decidir si debe convertir más de sus habitaciones de no fumar. El año pasado, en una muestra aleatoria de 400 huéspedes, 166 pidieron cuartos de no fumar. Este año, 250 huéspedes en una muestra de 380 prefirieron que no se fumara en su habitación. Utilice un nivel de confianza de 95 %

```
# Construyendo una función en R para realizar la prueba de intervalos de  
# confianza para varianza conocida.
```

```
intervaloProp <- function(x1, x2, n1, n2, nivel.conf = 0.95)
```

```
{
```

```
  p1 = x1 / n1
```

```
  p2 = x2 / n2
```

```
  q1 = 1 - p1
```

```
  q2 = 1 - p2
```

```
  alfa = 1 - nivel.conf
```

```
  margen_error = qnorm(alfa/2, mean = 0, sd = 1, lower.tail = FALSE)*(sqrt((p1*q1/n1) + (p2*q2/n2)))
```

```
  print(rbind(p1, p2, q1, q2, alfa, margen_error))
```

```
  extrIzq = (p1 - p2) - margen_error
```

```
  extrDer = (p1 - p2) + margen_error
```

```
  print("")
```

```
  print(paste("Intervalo para diferencia proporciones: [",  
              round(extrIzq, 2), ",", round(extrDer, 2), "]""))
```

```
}
```

```
n1=400; n2=380; x1=166; x2=250; nivel.conf=0.95
```

```
intervaloProp(x1, x2, n1, n2, nivel.conf = 0.95)
```

```
##           [,1]
```

```
## p1      0.41500000
```

```
## p2      0.65789474
```

```
## q1      0.58500000
```

```
## q2      0.34210526
```

```
## alfa     0.05000000
```

```
## margen_error 0.06787322
```

```
## [1] ""
```

```
## [1] "Intervalo para diferencia proporciones: [ -0.31 , -0.18 ]"
```