

# **CAPÍTULO 10**

## **ESTIMACIÓN POR PUNTO Y POR INTERVALO**

**1.- ESTIMACIÓN PUNTUAL DE LA MEDIA Y DE LA VARIANZA**

**2.- INTERVALO DE CONFIANZA PARA LA MEDIA**

**3.- INTERVALO DE CONFIANZA PARA LA VARIANZA**

**4.- INTERVALO DE CONFIANZA PARA EL COCIENTE DE VARIANZAS**

**5.- INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS**



## 1.- ESTIMACIÓN PUNTUAL DE LA MEDIA Y DE LA VARIANZA

En este tema se va a hacer un recorrido muy somero por algunos aspectos importantes de la inferencia estadística, como son la estimación por punto y por intervalo y los contrastes de hipótesis paramétricos, pues en *R* las estimaciones puntuales y por intervalo de confianza están íntimamente ligadas con los test de hipótesis estadísticas.

La estimación por punto de la media, es decir la media muestral, se obtiene con la función `mean`. Supongamos que se han medido en 10 días tomados al azar los niveles de cloro del agua que sale de una planta de tratamiento y se han obtenido los valores siguientes (se supone normalidad): 2.2 – 1.9 – 1.7 – 1.6 – 1.7 – 1.8 – 1.7 – 1.9 – 2.0 – 2.0. La estimación puntual de la media es 1.85, como vemos a continuación:

```
> nivclor<-c(2.2,1.9,1.7,1.6,1.7,1.8,1.7,1.9,2,2)
> nivclor
[1] 2.2 1.9 1.7 1.6 1.7 1.8 1.7 1.9 2.0 2.0
> mean(nivclor)
[1] 1.85
```

Los estimadores centrados de la varianza y de la desviación típica son, respectivamente, la cuasivarianza muestral (`var`) y la cuasidesviación típica muestral (`sd`). Calculemos esos valores para los datos del nivel de cloro:

```
> var(nivclor)
[1] 0.03388889
> sd(nivclor)
[1] 0.1840894
```

Como es evidente, se obtiene el mismo valor de la cuasidesviación típica muestral si hacemos

```
> sqrt(var(nivclor))
[1] 0.1840894
```

Como se indicó en el capítulo 8 dedicado al análisis de datos con *R*, mediante la función `var` se obtiene la cuasivarianza muestral y no la varianza muestral. Así mismo, la función `sd` proporciona la cuasidesviación típica muestral en lugar de la desviación típica muestral (*standard deviation*).

## 2.- INTERVALO DE CONFIANZA PARA LA MEDIA

La manera más cómoda, aunque indirecta, de obtener en *R* intervalos de confianza es mediante contrastes de hipótesis, cuestión que no vamos a tratar en este curso con detalle, sino que abordaremos únicamente desde el punto de vista de su relación con las estimaciones por intervalo.

Supongamos, por ejemplo, que en el caso de los datos de los niveles de cloro tratamos de contrastar si el nivel de cloro es 1.9. Para ello, y como los datos proceden de una población normal, lo adecuado es llevar a cabo un contraste t de Student en la forma siguiente:

```
> t.test(nivclor,mu=1.9)

One Sample t-test

data:  nivclor
t = -0.8589, df = 9, p-value = 0.4127
alternative hypothesis: true mean is not equal to
1.9
95 percent confidence interval:
 1.718310 1.981690
sample estimates:
mean of x
 1.85
```

La salida anterior se interpreta del siguiente modo: se ha efectuado un test t de Student para una muestra a partir de los datos `nivclor`. El estadístico muestral, que sigue una distribución t de Student, es igual a -0.8589, los grados de libertad  $10-1=9$  y el valor-p = 0.4127. La hipótesis alternativa es que la media poblacional no es igual a 1.9, por lo que la hipótesis nula será que la media es 1.9. Al obtenerse un valor-p grande (mayor que 0.1) no se puede rechazar la hipótesis nula y se concluye que la media de la población es 1,9. El intervalo de confianza al nivel de confianza del 95% es [1.718310 1.981690] y la estimación de la media es 1,85.

Si sólo queremos obtener el intervalo de confianza hacemos

```
> t.test(nivclor)$conf
[1] 1.718310 1.981690
attr(,"conf.level")
[1] 0.95
```

Para obtener el intervalo de confianza al nivel 99% hacemos

```
> t.test(nivclor,conf.level=0.99)$conf
[1] 1.660814 2.039186
```

La fórmula con la que se calculan los intervalos anteriores es

$$(\bar{x} \pm t_{n-1;\alpha/2} S / \sqrt{n})$$

A continuación comprobaremos como mediante la aplicación de esta fórmula podemos obtener los mismos resultados.

```
> xraya<-mean(nivclor)
> s<-sd(nivclor)
> 1-0.05/2;1-0.01/2
```

```

[1] 0.975
[1] 0.995

> ICal95<-c(xraya-
qt(0.975,9)*s/sqrt(10),xraya+qt(0.975,9)*s/sqrt(10
))
> ICal95
[1] 1.718310 1.981690

> ICal99<-c(xraya-
qt(0.995,9)*s/sqrt(10),xraya+qt(0.995,9)*s/sqrt(10
))
> ICal99
[1] 1.660814 2.039186

```

### 3.- INTERVALO DE CONFIANZA PARA LA VARIANZA

Para construir un intervalo de confianza para la varianza de la población de donde proceden los datos del nivel de cloro utilizamos la fórmula correspondiente ( $S$  es la *cuasidesviación típica*):

$$\left[ \frac{(n-1)S^2}{\chi_{n-1;\alpha/2}^2}, \frac{(n-1)S^2}{\chi_{n-1;1-\alpha/2}^2} \right]$$

```

> # Redondeamos los valores a 4 decimales
>
round(c(9*var(nivclor)/qchisq(0.975,9),9*var(nivcl
or)/qchisq(0.025,9)),4)
[1] 0.0160 0.1129

>
round(c(9*var(nivclor)/qchisq(0.995,9),9*var(nivcl
or)/qchisq(0.005,9)),4)
[1] 0.0129 0.1758

```

Como  $R$  no tiene implementado directamente el intervalo de confianza anterior, cabe la posibilidad de crear una pequeña rutina, que puede ser almacenada para posteriores ocasiones, con objeto de poder calcular de una forma automática los valores anteriores. Para ello, en un editor de texto creamos las instrucciones que se detallan más abajo; posteriormente, y una vez introducidos los valores de  $x$  y  $\alpha$ , seleccionamos todo, lo copiamos y lo pegamos en  $R$ , obteniendo los resultados requeridos:

```

# Intervalo de confianza para la varianza de una
población normal con media desconocida
# Introducir los valores que aparecen entre las
líneas de puntos: x (datos) y alfa (nivel de
significación)

```

```
# .....
x<-c( )
alfa<-
# .....
CI<-round(c((length(x)-1)*var(x)/qchisq(1-
alfa/2,length(x)-1),(length(x)-
1)*var(x)/qchisq(alfa/2,length(x)-1)),4)
CI
```

Se sugiere al lector que ponga en práctica la rutina anterior para los casos  $\alpha=0.05$  y  $\alpha=0.01$ ; obtendrá los intervalos de confianza [0.0160;0.1129] y [0.0129 0.1758], respectivamente.

Podríamos actuar de una forma alternativa, más simple, creando directamente en *R* un *objeto* (función) que calcule el intervalo de confianza, objeto que podemos almacenar para usos posteriores:

```
> CI.para.la.varianza<-function(x,alfa)
round(c((length(x)-1)*var(x)/qchisq(1-
alfa/2,length(x)-1),(length(x)-
1)*var(x)/qchisq(alfa/2,length(x)-1)),4)

> CI.para.la.varianza(x=nivclor,alfa=0.05)
[1] 0.0160 0.1129
> CI.para.la.varianza(x=nivclor,alfa=0.01)
[1] 0.0129 0.1758
```

#### 4.- INTERVALO DE CONFIANZA PARA EL COCIENTE DE VARIANZAS

Para desarrollar este epígrafe vamos a considerar el siguiente ejemplo: Se quiere saber si existen diferencias significativas en la facturación de dos tiendas de joyería de una misma cadena. Para ello se eligieron al azar 11 días en los que se contabilizaron las ventas en la joyería A y otros 10 días en la joyería B. Los datos obtenidos (se supone normalidad) son:

Ventas A	1320:1495:990:1250:1290:1900:1500:1100:1250:1100:1930
Ventas B	1110:1405:985:1290:1300:1705:1200:1105:1150:1210:

Para comprobar si se pueden considerar iguales o no las varianzas (cociente de varianzas igual a 1 o distinto de 1) debemos hacer un contraste mediante la función `var.test`:

```
> ventasA<-
c(1320,1495,990,1250,1290,1900,1500,1100,1250,1100
,1930)
> ventasB<-
c(1110,1405,985,1290,1300,1705,1200,1105,1150,1210
)
```

```
> var.test(ventasA,ventasB)
```

```
F test to compare two variances
```

```
data:  ventasA and ventasB
F = 2.387, num df = 10, denom df = 9, p-value =
0.2061
alternative hypothesis: true ratio of variances is
not equal to 1
95 percent confidence interval:
 0.6021958 9.0204701
sample estimates:
ratio of variances
      2.387023
```

La salida anterior se interpreta del siguiente modo: se ha efectuado un test para el cociente de varianzas de dos poblaciones normales a partir de los datos **ventasA** y **ventasB**. El estadístico muestral, que sigue una distribución F de Snedecor, es 2.387, los grados de libertad son  $11-1=10$  y  $10-1=9$  y el valor-p = 0,2061. La hipótesis alternativa es que el cociente de varianzas poblacionales no es igual a 1 (o sea, las varianzas son distintas), por lo que la hipótesis nula será que el cociente de varianzas es 1 (varianzas iguales). Al obtenerse un valor-p grande (mayor que 0.1) no se puede rechazar la hipótesis nula y se concluye que el cociente de varianzas es 1; es decir, que las varianzas son iguales. El intervalo de confianza al nivel de confianza del 95% es [0.6021958;9.0204701] que, como se ve, incluye al 1. La estimación del cociente de varianzas es 2.387023.

Si sólo queremos obtener el intervalo de confianza hacemos

```
> var.test(ventasA,ventasB)$conf
[1] 0.6021958 9.0204701
attr(,"conf.level")
[1] 0.95
```

## 5.- INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS

Si queremos contrastar la igualdad de medias de dos poblaciones, por ejemplo si en el caso de las dos joyerías tratamos de ver si se vende lo mismo en ambos establecimientos, debemos hacer un contraste de hipótesis sobre la diferencia de medias t de Student, ya que los datos son normales. Teniendo en cuenta el resultado del test anterior debemos indicarle a *R* que las varianzas son iguales (**var.equal=T**), si bien ésta es la opción por defecto:

```
> t.test(ventasA,ventasB,var.equal=T)
```

```
Two Sample t-test
```

```

data:  ventasA and ventasB
t = 1.1224, df = 19, p-value = 0.2757
alternative hypothesis: true difference in means
is not equal to 0
95 percent confidence interval:
 -111.5503  369.5503
sample estimates:
mean of x mean of y
   1375     1246

```

Los resultados anteriores se interpretan del siguiente modo: se ha efectuado un test para la diferencia de medias de dos poblaciones normales a partir de los datos **ventasA** y **ventasB**, siendo iguales las varianzas poblacionales. El estadístico muestral, que sigue una distribución t de Student, es 1.1224, los grados de libertad son  $11+10-2=19$  y el valor-p = 0.2757. La hipótesis alternativa es que la diferencia de medias poblacionales no es igual a 0 (o sea, las medias son distintas), por lo que la hipótesis nula será que la diferencia de medias es 0 (medias iguales). Al obtenerse un valor-p grande (mayor que 0.1) no se puede rechazar la hipótesis nula y se concluye que la diferencia de medias es 0; es decir, que las medias son iguales y, por tanto, que se vende lo mismo en la joyería A que en la joyería B. El intervalo de confianza, al nivel de confianza del 95%, para la diferencia de medias es [-111.5503;369.5503] que, como se ve, incluye al 0. Las estimaciones de las medias son 1375 y 1246.

Si únicamente queremos obtener el intervalo de confianza hacemos

```

> t.test(ventasA,ventasB,var.equal=T)$conf
[1] -111.5503  369.5503
attr(,"conf.level")
[1] 0.95

```