Índice

1. Estadística Descriptiva **9**
1.1. Introducción 9
1.2. Representaciones gráficas 9
1.2.1. Representaciones de datos de tipo cualitativo 10
1.2.2. Representaciones de datos de tipo cuantitativo 11
1.3. Medidas de posición 13
1.4. Medidas de dispersión 15
1.5. Distribuciones bidimensionales de frecuencias 19
1.5.1. Ajuste por mínimos cuadrados 21
1.5.2. Precisión del ajuste por mínimos cuadrados 25

2. Modelización y Estimación: La Distribución Normal **29**
2.1. Introducción 29
2.2. La ley de Probabilidad Normal 31
2.3. La distribución t de Student 38
2.4. Estimación de la media poblacional 41
2.5. Estimación de la varianza poblacional: Distribución χ^2 de Pearson 43
2.6. Estimación del cociente de varianzas poblacionales: Distribución F de Snedecor 44

3. Estimación por Intervalos de Confianza **47**
3.1. Introducción 47
3.1.1. Cálculo de Intervalos de Confianza con R 49
3.2. Intervalo de confianza para la media de una población normal . 51
3.3. Intervalo de confianza para la media de una población no necesariamente normal. Muestras grandes 53
3.4. Intervalo de confianza para la varianza de una población normal 56
3.5. Intervalo de confianza para el cociente de varianzas de dos poblaciones normales independientes 57
3.6. Intervalo de confianza para la diferencia de medias de dos poblaciones normales independientes 59

3.7. Intervalo de confianza para la diferencia de medias de dos poblaciones independientes no necesariamente normales. Muestras grandes . . . 61
3.8. Intervalos de confianza para datos apareados . . . 63

4. Contraste de Hipótesis **65**
4.1. Introducción y conceptos fundamentales . . . 65
4.2. Contraste de hipótesis relativas a la media de una población normal . . . 73
4.3. Contraste de hipótesis relativas a la media de una población no necesariamente normal. Muestras grandes . . . 78
4.4. Contraste de hipótesis relativas a la varianza de una población normal . . . 82
4.5. El contraste de los rangos signados de Wilcoxon . . . 86

5. Comparación de Poblaciones **91**
5.1. Introducción . . . 91
5.2. Análisis de la Normalidad . . . 93
5.3. Análisis de la Homocestacidad . . . 95
5.4. Transformaciones Box-Cox . . . 98
5.5. Contraste de hipótesis relativas a la diferencia de medias de dos poblaciones normales independientes . . . 105
5.6. Contraste de hipótesis relativas a la diferencia de medias de dos poblaciones independientes no necesariamente normales. Muestras grandes . . . 111
5.7. El contraste de Wilcoxon-Mann-Whitney . . . 115
5.8. Análisis de la Varianza . . . 117
5.8.1. Comparaciones Múltiples . . . 120
5.9. Contraste de Kruskal-Wallis . . . 123
5.9.1. Contraste χ^2 de homogeneidad de varias muestras . . . 125

6. Modelos de Regresión **127**
6.1. Introducción . . . 127
6.2. Modelo de la Regresión Lineal Simple . . . 128
6.3. Análisis de los residuos . . . 132
6.4. Modelo de la Regresión Lineal Múltiple . . . 133
6.5. Otros Modelos Lineales . . . 136

7. Bibliografía **139**

Capítulo 3

Estimación por Intervalos de Confianza

3.1. Introducción

En el capítulo anterior estudiamos la *Estimación por punto* de las características o *parámetros* de la población que queremos investigar y así dijimos que, si queremos estimar la media μ de una población, debemos utilizar la media $\overline{x}$ de una muestra representativa extraída de la población en estudio. No obstante, raramente la estimación por punto coincidirá exactamente con el parámetro a estimar, es decir, rara vez la media de la muestra seleccionada al azar será tal que $\overline{x} = \mu$. Sin duda, es mucho más interesante realizar la inferencia con un intervalo de posibles valores del parámetro —al que denominaremos *Intervalo de Confianza*—, de manera que, antes de tomar la muestra, el desconocido valor del parámetro se encuentre en dicho intervalo con una probabilidad todo lo alta que deseemos.

Así por ejemplo, es mucho más deseable afirmar que la media poblacional μ está entre $\overline{x} - 0'1$ y $\overline{x} + 0'1$, con probabilidad $0'99$, que dando un valor concreto como estimación puntual de μ, el cual es posible que esté muy alejado del verdadero.

Con objeto de aumentar la precisión de la inferencia, serán deseables intervalos de confianza lo más cortos posible.

No obstante, la longitud del intervalo de confianza dependerá de lo alta que queramos sea la probabilidad con la que dicho intervalo —cuyos extremos son aleatorios— cubra a μ y, por tanto, del modelo que elijamos para explicar la variable en estudio. Así por ejemplo si queremos determinar el intervalo de confianza para la media de una población normal de varianza conocida σ, éste será

$$\left[\,\overline{x} - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\ ,\ \overline{x} + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\,\right]$$

en donde $z_{\alpha/2}$ es, como dijimos en el capítulo anterior, el valor de la abscisa de una $N(0,1)$ que deja a su derecha —bajo la función de densidad— un área de probabilidad $\alpha/2$.

Como se ve, la longitud del intervalo de confianza, es decir, la diferencia entre el extremo superior y el inferior,

$$2 \cdot z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}$$

depende de la probabilidad $1-\alpha$ elegida en su construcción, a la que denominaremos *coeficiente de confianza*, y del tamaño muestral (a mayor tamaño muestral n, menor será la longitud del intervalo).

Para un tamaño muestral fijo, cuanto mayor sea el coeficiente de confianza, más grande será $z_{\alpha/2}$ y por tanto, mayor su longitud. Por tanto, antes de construir un intervalo de confianza, habrá que prefijar cuidadosamente el valor del coeficiente de confianza de manera que la *probabilidad con la que confiamos* el intervalo cubra al desconocido valor del parámetro sea alta, pero conservando inferencias válidas.

Así, de poco interés resultará concluir que hay probabilidad $0'999$ de que el intervalo (en metros) $[\overline{x}-2\,,\,\overline{x}+2]$, cubra la estatura media de la población.

Los coeficientes de confianza que se suelen considerar son $0'90$, $0'95$ y $0'99$, aunque esto dependerá del investigador, el cual deberá tener siempre en cuenta los comentarios anteriores. Por ejemplo, una varianza poblacional σ^2 pequeña o un tamaño muestral grande pueden permitir un mayor coeficiente de confianza sin un aumento excesivo de la longitud del intervalo.

Formalmente definimos el intervalo de confianza para un parámetro θ de la siguiente manera.

Definición

Supongamos que X es la variable aleatoria en estudio, cuya distribución depende de un parámetro desconocido θ, y $X_1,..,X_n$ una muestra aleatoria simple de dicha variable.

Si $T_1(X_1,...,X_n)$ y $T_2(X_1,...,X_n)$ son dos estadísticos tales que

$$P\{T_1(X_1,...,X_n) \leq \theta \leq T_2(X_1,...,X_n)\} = 1-\alpha$$

el intervalo

$$[\;T_1(x_1,...,x_n)\;,\;T_2(x_1,...,x_n)\;]$$

recibe el nombre de *Intervalo de Confianza* para θ de *coeficiente de confianza* $1-\alpha$.

Obsérvese que tiene sentido hablar de que, antes de tomar la muestra, el intervalo aleatorio

$$[\, T_1(X_1, ..., X_n) \,,\, T_2(X_1, ..., X_n) \,]$$

cubra al verdadero y desconocido valor del parámetro θ con probabilidad $1-\alpha$ pero, una vez elegida una muestra particular $x_1, ..., x_n$, el intervalo no aleatorio

$$[\, T_1(x_1, ..., x_n) \,,\, T_2(x_1, ..., x_n) \,]$$

cubrirá o no a θ, pero ya no tiene sentido hablar de la probabilidad con que lo cubre.

Es decir, podemos hacer afirmaciones del tipo de que en un $100(1-\alpha)\,\%$ de las veces, el intervalo que obtengamos cubrirá al parámetro, pero nunca de que, por ejemplo, hay probabilidad $1-\alpha$ de que el intervalo de confianza $[1'65 \,,\, 1'83]$ cubra al parámetro, ya que los extremos de este último intervalo —y como siempre el parámetro— son números y no variables aleatorias.

Obsérvese también que el intervalo de confianza es un subconjunto de los posibles valores del parámetro precisamente por ser no aleatorio.

Así mismo mencionemos que cualquier par de estimadores T_1 y T_2 que cumplan la condición impuesta en la definición anterior darán lugar a un intervalo de confianza. Habitualmente éstos serán dos funciones del estimador natural obtenido para cada caso en el capítulo anterior. De hecho, en las siguientes secciones indicaremos cuál es el intervalo de confianza que razonablemente debe utilizarse en cada situación concreta. En muchos casos su obtención se hará utilizando un paquete estadístico y, en otras, aplicando las fórmulas que se indica por lo que incluiremos ejemplos de ambas situaciones.

Recordamos la notación que utilizaremos, tanto en los intervalos de confianza como en el resto del libro: denotaremos por z_p , $t_{n;p}$, $\chi^2_{n;p}$ y $F_{n_1,n_2;p}$, respectivamente, el valor de la abscisa de una distribución $N(0,1)$, t_n de Student, χ^2_n de Pearson y F_{n_1,n_2} de Snedecor, que deja a su derecha —bajo la correspondiente función de densidad— un área de probabilidad p.

3.1.1. Cálculo de Intervalos de Confianza con R

En el capítulo siguiente veremos que el intervalo de confianza de un parámetro se corresponde con la región de aceptación de un test bilateral. Por esta razón se utiliza la misma función de R para obtener intervalos de confianza y test de hipótesis sobre un parámetro.

En concreto, la función de R que nos va a proporcionar los intervalos (y los tests), es la función `t.test`. Con ella vamos a poder determinar los Intervalos de Confianza (y tests) para la media, para datos apareados y para la diferencia de medias, pero no para aquellos casos en los que la varianza, varianzas o medias poblacionales sean conocidas sino para cuando haya que estimarlas a partir de los datos. También queremos advertir que, para poder aplicar esta función, es necesario conocer los datos individualmente ya que no podremos utilizarla cuando sólo conozcamos los valores de las medias o cuasivarianzas muestrales y no los datos de donde éstas proceden.

La función a utilizar en el caso de Intervalos de Confianza es

```
t.test(x, y = NULL, paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

Entrando a describir cada uno de sus argumentos, en primer lugar diremos que los valores que aparecen después del símbolo `=` son los que toma la función por defecto y que, por tanto, no será necesario especificar si son los valores que deseamos ejecutar. En `x` incorporamos los datos de la muestra, si se trata de inferencias para una sola muestra; si se trata de datos apareados o de dos muestras independientes, introduciremos los datos de la segunda muestra en el argumento `y`.

Si especificamos `paired=F` (lo cual no es necesario puesto que es la opción tomada por defecto), estamos es una situación de datos no apareados. Un caso de datos apareados debe especificarse con `paired=T`.

El argumento `var.equal` nos permite indicar qué tipo de situación tenemos en el caso de comparación de dos poblaciones independientes. Si es `var.equal=T` tendremos una situación en la que las varianzas de ambas poblaciones se suponen iguales, y el intervalo será el habitual basado en una t de Student. Si especificamos `var.equal=F` las varianzas de ambas poblaciones no se suponen iguales y, en ese caso, estamos requiriendo un intervalo basado en una t de Student pero en donde los grados de libertad se determina por la aproximación de Welch.

El último argumento permite especificar el coeficiente de confianza, tomándose por defecto el valor $0'95$.

El intervalo de confianza para el cociente de varianzas poblacionales se obtiene con la función

```
var.test(x, y, conf.level = 0.95)
```

en donde incorporamos los datos en los argumentos `x` e `y`. De nuevo aquí necesitaremos conocer los datos concretos y no admite esta función la situación de ser las medias poblacionales conocidas.

3.2. Intervalo de confianza para la media de una población normal

Tanto en esta sección como en las siguientes, determinaremos intervalos de confianza de *colas iguales*. Es decir, aquellos tales que, si el coeficiente de confianza es $1-\alpha$, dejan en cada uno de los extremos la mitad de la probabilidad, $\alpha/2$.

En esta sección suponemos que los n datos proceden de una población $N(\mu, \sigma)$, y lo que pretendemos determinar es el intervalo de confianza para la media μ.

Como vimos en la Sección 2.4, en esta situación, tanto si la varianza poblacional σ^2 es conocida como si no lo es, el estimador natural de μ es la media muestral $\overline{x}$.

σ conocida

El intervalo buscado será

$$\left[\, \overline{x} - z_{\alpha/2}\, \frac{\sigma}{\sqrt{n}} \; , \; \overline{x} + z_{\alpha/2}\, \frac{\sigma}{\sqrt{n}} \,\right].$$

σ desconocida

En este caso de que la varianza poblacional sea desconocida, el intervalo de confianza para la media resulta

$$\left[\, \overline{x} - t_{n-1;\alpha/2}\, \frac{S}{\sqrt{n}} \; , \; \overline{x} + t_{n-1;\alpha/2}\, \frac{S}{\sqrt{n}} \,\right]$$

en donde S^2 es la cuasivarianza muestral.

Ejemplo 3.1

Un terapeuta desea estimar, con una confianza del 99 %, la fuerza media de un músculo determinado en los individuos de una población. Admitiendo que las unidades de fuerza siguen una distribución normal de varianza 144, seleccionó una muestra aleatoria de 25 individuos de la población, para la que obtuvo una media muestral de $\overline{x} = 85$.

Como no tenemos los datos observados, en este caso deberemos utilizar las fórmulas anteriores para calcular el intervalo de confianza. En estas condiciones, el intervalo de confianza será

$$\left[\, \overline{x} - z_{\alpha/2}\, \frac{\sigma}{\sqrt{n}} \; , \; \overline{x} + z_{\alpha/2}\, \frac{\sigma}{\sqrt{n}} \,\right] = \left[\, 85 - z_{0'01/2}\, \frac{12}{\sqrt{25}} \; , \; 85 + z_{0'01/2}\, \frac{12}{\sqrt{25}} \,\right]$$

Como es $z_{0'01/2} = z_{0'005}$ es valor de una abscisa de una normal estándar $N(0,1)$ que deja a la derecha un área de probabilidad $0'005$, este valor se calculará, como vimos en la Sección 2.2, ejecutando

```
> qnorm(1-0.005)
[1] 2.575829
```

Por tanto, el intervalo de confianza buscado será,

$$\left[\ 85 - 2'575829\ \frac{12}{\sqrt{25}}\ ,\ 85 + 2'575829\ \frac{12}{\sqrt{25}}\ \right] = [\ 78'82\ ,\ 91'18\ ].$$

Estos cálculos los puede obtener con una calculadora o con R ejecutando

```
> 85-2.575829*12/sqrt(25)
[1] 78.81801

> 85+2.575829*12/sqrt(25)
[1] 91.18199
```

Si, como es más razonable, el terapeuta no supone conocida la varianza poblacional, deberá estimarla con la cuasivarianza muestral de los 25 individuos seleccionados. Si ésta fue $S^2 = 139$, el intervalo de confianza será

$$\left[\ 85 - t_{24;0'01/2}\ \sqrt{\frac{139}{25}}\ ,\ 85 + t_{24;0'01/2}\ \sqrt{\frac{139}{25}}\ \right] = [\ 78'4\ ,\ 91'59\ ]$$

ya que el valor de la abscisa de una t de Student con 24 grados de libertad que deja a la derecha un área de probabilidad $0'01/2 = 0'005$ será (vea la Sección 2.3),

```
> qt(1-0.005,24)
[1] 2.79694
```

y es

```
> 85-2.79694*sqrt(139/25)
[1] 78.40491

> 85+2.79694*sqrt(139/25)
[1] 91.59509
```

Ejemplo 3.2

Una muestra aleatoria de 10 clientes de una farmacia determinada mostró los siguientes tiempos de espera hasta que son atendidos, en minutos:

$$2\ ,\ 10\ ,\ 4\ ,\ 5\ ,\ 1\ ,\ 0\ ,\ 5\ ,\ 9\ ,\ 3\ ,\ 9$$

Determinar un intervalo de confianza, con coeficiente de confianza $0'9$, para el tiempo medio de espera, admitiendo que el tiempo de espera en esa farmacia sigue una distribución normal. Se trata de calcular el intervalo de confianza para la media de una población normal de varianza desconocida que vimos era

$$\left[\overline{x} - t_{n-1;\alpha/2}\,\frac{S}{\sqrt{n}}\;,\;\overline{x} + t_{n-1;\alpha/2}\,\frac{S}{\sqrt{n}}\right].$$

De los datos del enunciado se desprende que es $\overline{x} = 4'8$ y $S = 3'52$, como fácilmente se obtiene con R,

```
> x<-c(2,10,4,5,1,0,5,9,3,9)
> mean(x)
[1] 4.8
> sd(x)
[1] 3.521363
```

Por tanto, como además es $t_{n-1;\alpha/2} = t_{9;0'05} = 1'833$ ejecutando

```
> qt(1-0.05,9)
[1] 1.833113
```

el intervalo de confianza solicitado será

$$\left[\overline{x} - t_{n-1;\alpha/2}\,\frac{S}{\sqrt{n}}\;,\;\overline{x} + t_{n-1;\alpha/2}\,\frac{S}{\sqrt{n}}\right] = \left[4'8 - 1'833\,\frac{3'52}{\sqrt{10}}\;,\;4'8 + 1'833\,\frac{3'52}{\sqrt{10}}\right] =$$

$$= [\,2'76\,,\,6'84\,].$$

Si queremos obtener el intervalo directamente con R, ejecutaríamos

```
> t.test(x,conf.level=0.9)
        One Sample t-test

data:  x
t = 4.3105, df = 9, p-value = 0.00196
alternative hypothesis: true mean is not equal to 0
90 percent confidence interval:
 2.758732 6.841268                                          (1)
sample estimates:
mean of x
      4.8
```

obteniendo en `(1)` el mismo intervalo que antes.

3.3. Intervalo de confianza para la media de una población no necesariamente normal. Muestras grandes

Si el tamaño de la muestra es lo suficientemente grande (digamos mayor que 30 datos), el intervalo de confianza se basará siempre en una normal, sea o no conocida la varianza de la población y procedan o no los datos de una normal. En concreto,

Si σ **es conocida** el intervalo de confianza para μ de coeficiente de confianza $1-\alpha$ será

$$I = \left[\ \overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\ ,\ \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\ \right]$$

y si σ **es desconocida**

$$I = \left[\ \overline{x} - z_{\alpha/2}\frac{S}{\sqrt{n}}\ ,\ \overline{x} + z_{\alpha/2}\frac{S}{\sqrt{n}}\ \right]$$

siendo, como antes, S la cuasidesviación típica muestral.

Ejemplo 3.3

Los siguientes datos son valores de actividad (en micromoles por minuto por gramo de tejido) de una cierta enzima observada en el tejido gástrico de 35 pacientes con carcinoma gástrico

0'360	1'185	0'524	0'870	0'356	2'567	0'566
1'789	0'578	0'578	0'892	0'345	0'256	0'987
0'355	0'989	0'412	0'453	1'987	0'544	0'798
0'634	0'355	0'455	0'445	0'755	0'423	0'754
0'452	0'452	0'450	0'511	1'234	0'543	1'501

El histograma de estos datos (Figura 3.1) muestra claramente una fuerte asimetría a la derecha, lo cual sugiere que los valores de actividad no siguen una distribución normal.
No obstante, al ser el tamaño muestral bastante grande la media muestral $\overline{x}$ sí sigue una distribución normal. Es decir, si hiciéramos un histograma en el que representáramos los valores obtenidos por la media muestral en un gran número de muestras, éste tendría forma acampanada aunque, como ocurre en este caso, la variable poblacional no siga una distribución normal.
El intervalo de confianza a utilizar será

$$I = \left[\ \overline{x} - z_{\alpha/2}\frac{S}{\sqrt{n}}\ ,\ \overline{x} + z_{\alpha/2}\frac{S}{\sqrt{n}}\ \right]$$

el cual, para un coeficiente de confianza del 95 % es igual a

$$I = \left[\ 0'753 - 1'96\sqrt{\frac{0'2686}{35}}\ ,\ 0'753 + 1'96\sqrt{\frac{0'2686}{35}}\ \right] = [\ 0'5813, 0'9247\ ].$$

Si queremos resolver este ejemplo con R, primero introducimos los datos ejecutando `(1)`, un histograma suyo, obtenido ejecutando `(2)` y que aparece en la Figura 3.1 muestra una fuerte asimetría a la derecha, lo cual sugiere que los valores de actividad no siguen una distribución normal.

```
> x<-c(0.360,1.185,0.524,0.870,0.356,2.567,0.566,                    (1)
+ 1.789,0.578,0.578,0.892,0.345,0.256,0.987,
+ 0.355,0.989,0.412,0.453,1.987,0.544,0.798,
+ 0.634,0.355,0.455,0.445,0.755,0.423,0.754,
+ 0.452,0.452,0.450,0.511,1.234,0.543,1.501)
```

```
> hist(x,prob=T)                                                          (2)
```

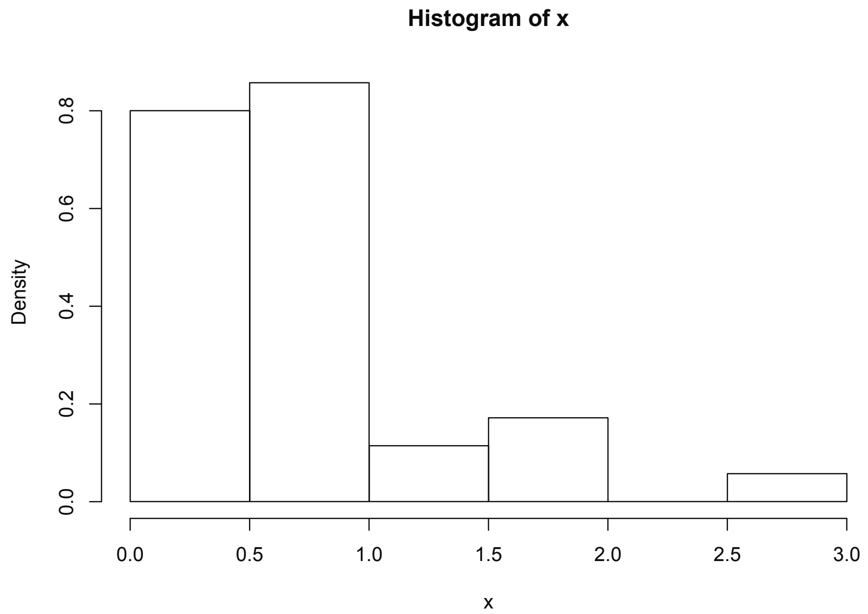

Figura 3.1 : Histograma del Ejemplo 3.3

Si queremos determinar el intervalo de confianza para la media (de una población no necesariamente normal, muestras grandes), de coeficiente de confianza $0'95$, ejecutaríamos (3), obteniendo el intervalo en (4).

```
> t.test(x)                                                               (3)

        One Sample t-test

data:  x
t = 8.5953, df = 34, p-value = 4.842e-10
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.5749635 0.9310365                                                      (4)
sample estimates:
mean of x
    0.753
```

El intervalo que obtenemos con R, $[0'5749\,,\,0'9310]$ es algo diferente del que se obtuvo anteriormente debido a que antes se utilizaba la aproximación normal para la determinación de los cuantiles $z_{1-\alpha/2}$ y $z_{\alpha/2}$, mientras que aquí se utilizan los correspondientes de la distribución t de Student. Lo correcto sería lo que hicimos más arriba, pero a medida que n aumenta, apenas habrá diferencia entre ambos.

3.4. Intervalo de confianza para la varianza de una población normal

Dada una muestra aleatoria simple $X_1, ..., X_n$ de una población $N(\mu, \sigma)$, vamos a determinar el intervalo de confianza para σ^2 , distinguiendo dos casos según sea desconocida o no la media de la población μ.

μ desconocida

El intervalo de confianza buscado será

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

con S^2 la cuasivarianza muestral.

μ conocida

En este caso, el intervalo de confianza será

$$I = \left[\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi^2_{n;\alpha/2}} , \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi^2_{n;1-\alpha/2}} \right].$$

Ejemplo 3.1 (continuación)

Si el terapeuta del Ejemplo 3.1 quiere determinar un intervalo de confianza para la varianza de la variable en estudio, éste será

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

que para un coeficiente de confianza del 99 % proporciona los valores

$$I = \left[\frac{24 \cdot 139}{45'56} , \frac{24 \cdot 139}{9'886} \right] = [\, 73'22, 337'45 \,].$$

Obsérvese que para un tamaño muestral tan pequeño como el que tenemos, el intervalo de confianza al 99 % determinado resulta poco informativo, al tener éste una longitud muy grande.

El correspondiente al 90 %

$$I = \left[\frac{24 \cdot 139}{36'42} , \frac{24 \cdot 139}{13'85} \right] = [\, 91'6, 240'9 \,]$$

tampoco resulta mucho más informativo, perdiendo éste, además, parte del *grado de confianza* que el primero poseía. Una de las causas es que, habitualmente, estaremos interesados en estimar la desviación típica y no la varianza, puesto que ésta viene expresada en unidades al cuadrado lo que distorsiona en parte el resultado. El intervalo de confianza para la desviación típica será el de extremos la raíz cuadrada del correspondiente de la varianza. Así por ejemplo, el intervalo correspondiente al 90 % será

$$I = [\, \sqrt{91'6} \,, \sqrt{240'9} \,] = [9'57 \,, 15'52].$$

3.5. Intervalo de confianza para el cociente de varianzas de dos poblaciones normales independientes

Supondremos que $X_1, ..., X_{n_1}$ e $Y_1, ..., Y_{n_2}$ son dos muestras de tamaños n_1 y n_2 extraídas respectivamente de dos poblaciones independientes $N(\mu_1, \sigma_1)$ y $N(\mu_2, \sigma_2)$.

μ_1 y μ_2 conocidas

En este caso, el intervalo de colas iguales es

$$I = \left[\frac{n_2 \sum_{i=1}^{n_1} (X_i - \mu_1)^2 \Big/ \sum_{j=1}^{n_2} (Y_j - \mu_2)^2}{n_1 \cdot F_{n_1, n_2; \alpha/2}} \;,\; \frac{n_2 \sum_{i=1}^{n_1} (X_i - \mu_1)^2 \Big/ \sum_{j=1}^{n_2} (Y_j - \mu_2)^2}{n_1 \cdot F_{n_1, n_2; 1-\alpha/2}} \right].$$

μ_1 y μ_2 desconocidas

Si las medias poblacionales son desconocidas y las muestras proporcionan cuasivarianzas muestrales S_1^2 y S_2^2 respectivamente, el intervalo de confianza que se obtiene es

$$I = \left[\frac{S_1^2/S_2^2}{F_{n_1-1, n_2-1; \alpha/2}} \;,\; \frac{S_1^2/S_2^2}{F_{n_1-1, n_2-1; 1-\alpha/2}} \right].$$

Ejemplo 3.4

Con objeto de estudiar la efectividad de un agente diurético, se eligieron al azar 11 pacientes, aplicando a 6 de ellos dicho fármaco y un placebo a los 5 restantes.
La variable observada en esta experiencia fue la concentración de sodio en la orina a las 24 horas, la cual dio los resultados siguientes:

Diurético :	$20'4$	$62'5$	$61'3$	$44'2$	$11'1$	$23'7$
Placebo :	$1'2$	$6'9$	$38'7$	$20'4$	$17'2$	

Supuesto que las concentraciones de sodio, tanto en la población a la que se aplicó el diurético $X_1 \rightsquigarrow N(\mu_1, \sigma_1)$ como a la que se aplicó el placebo $X_2 \rightsquigarrow N(\mu_2, \sigma_2)$, siguen distribuciones normales, en la determinación de un intervalo de confianza para la diferencia de medias poblacionales, veremos que, al ser las muestras pequeñas, necesitamos decidir si las varianzas poblacionales σ_1^2 y σ_2^2 pueden considerarse iguales o no.
Con este propósito se determina un intervalo de confianza para el cociente de dichas varianzas,

$$I = \left[\frac{S_1^2/S_2^2}{F_{n_1-1, n_2-1; \alpha/2}} \;,\; \frac{S_1^2/S_2^2}{F_{n_1-1, n_2-1; 1-\alpha/2}} \right]$$

que resulta ser, para un coeficiente de confianza del 95 %,