Práctica 3: Inferencia estadística

Marcial-Antonio Barajas Martín. Grupo A1

1. **Utilizando como semilla (seed) el número generar un vector que contenga una secuencia de 100 números que sigan una distribución normal con media 5 y desviación típica 2. Determinar su media y su desviación típica**

Este apartado es de resolución inmediata con el comando rnorm(100,mean=5, sd=2). El comando nos va a generar un vector de 100 números aleatorios con una media teórica de 5 y una desviación estándar de 2. Usando los comandos mean y sd en el vector generado podemos ver que:

> mean(muestra)

[1] 5.180812

> sd(muestra)

[1] 1.825632

Lo cual se ajusta bastante bien con solo 100 valores. Como curiosidad, para una muestra de 10000 valores tenemos:

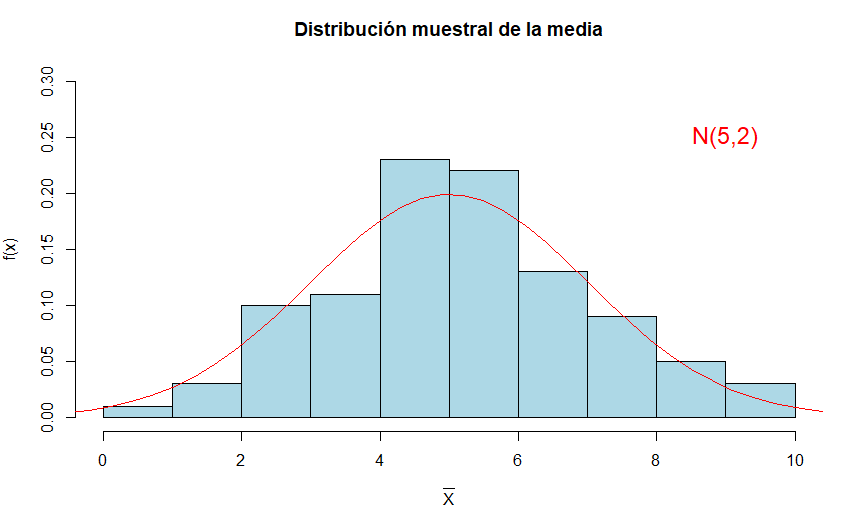
> mean(muestra)

[1] 4.995257

> sd(muestra)

[1] 1.997273

1. **Representar el histograma del vector anterior, superponiendo en otro color la curva con la distribución teórica (normal) asumida**



En esta representación vemos como el histograma de frecuencias se ajusta razonablemente bien a la distribución normal que tiene media 5 y desviación 2 representada por la línea roja.

**c. Obtener un intervalo de confianza del 90% para la media poblacional sin hacer la suposición de que la aproximación normal es razonable de dos formas:**

i. programando el estadístico correspondiente;

ii. directamente mediante un comando de R.

Para el apartado i) Hacemos los cálculos mediante la fórmula de aproximación para una muestra de n>30 y desconocida. Para ello usamos la fórmula:

Ejecutando el código en R tenemos:

> alpha\_05=qnorm(0.05,lower.tail=F)

> center=mean(muestra)

> l1=center + alpha\_05\*sd(muestra)/sqrt(length(muestra))

> l2=center - alpha\_05\*sd(muestra)/sqrt(length(muestra))

> cat("[",l2,",",l1,"]")

[ 4.880522 , 5.481102 ]

Mientras que si ejecutamos el comando t.test de esta muestra.

> t.test(muestra)

One Sample t-test

data: muestra

t = 28.378, df = 99, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

4.818567 5.543057

sample estimates:

mean of x

5.180812

Vemos que el valor del intervalo de confianza considerando la desviación de la población sin que se parezca a una distribución normal está ligeramente más acotado y es más pequeño que el valor de la distribución normal. Esto se debe a que, en una cantidad alta de datos, el valor de la desviación de la muestra es un estimador bastante aceptable de

1. **Obtener un intervalo de confianza del 90% para la media poblacional haciendo la suposición de que la aproximación normal es razonable. Interpretar el resultado.**

En este caso, usando una distribución t de student, a mayor cantidad de grados de libertad, vemos que los valores de esta distribución van aproximándose a una distribución normal, por lo que la aproximación es razonable. El intervalo de confianza en este caso se queda como:

Introduciendo esto en R con los valores de la muestra obtenemos:

> n=length(muestra)

> center=med

> t\_0.5=qt(0.05, df=n-1,lower.tail=F)

> width05=t\_0.5\*sd(muestra)/sqrt(n)

> cat("[",center-width05,",",center+width05,"]")

[ 4.877686 , 5.483938 ]

> t.test(muestra, conf.level=0.90)

One Sample t-test

data: muestra

t = 28.378, df = 99, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 0

90 percent confidence interval:

4.877686 5.483938

sample estimates:

mean of x

5.180812

Entonces vemos que, efectivamente, el valor de la distribución t de student con tantos grados de libertad se aproxima de manera razonable a una aproximación normal.

1. Obtener un intervalo de confianza del 95% para la desviación típica poblacional.

Sabemos que el intervalo de confianza de nivel (1-alfa) para la varianza de una distribución es:

Con esta fórmula, si la introducimos en R podemos hallar:

> nc=95

> lim=(1-nc/100)/2

> chi1=qchisq(lim, df=length(muestra)-1, lower.tail=F)

> chi2=qchisq(lim, df=length(muestra)-1)

> l1=sqrt((length(muestra)-1)\*var(muestra)/chi1)

> l2=sqrt((length(muestra)-1)\*var(muestra)/chi2)

> cat("Intervalo del ", nc, "%: [",l1,",",l2,"]","\n")

Intervalo del 95 %: [ 1.602917 , 2.120791 ]

La curiosidad de este intervalo frente al de los otros valores del intervalo de confianza para la muestra, es que el valor de la desviación estándar está comprendido dentro del intervalo pero no está en el medio, se ve que de 1.825632-1.602917 =0.2053463 frente a que hay desde 2.120791-1.825632= 0.295159