INTERVALOS DE CONFIANZA

Selección de actividades resueltas

UOC Universitat Oberta de Catalunya

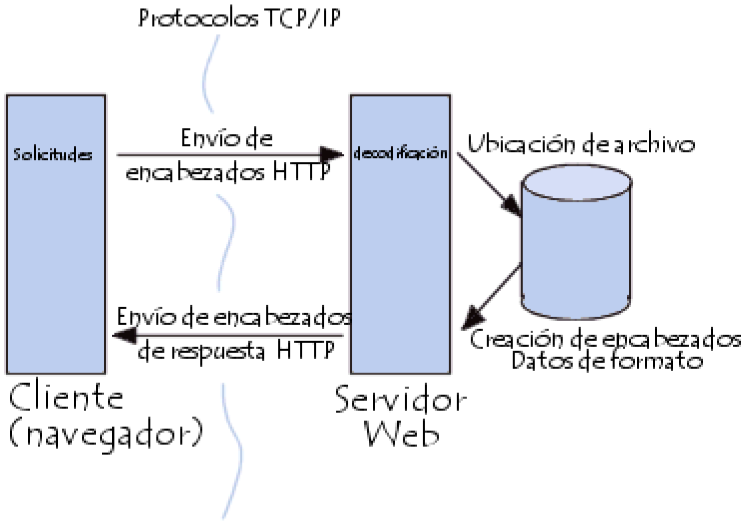
Introducción

En este *módulo*, se pretende conocer y saber calcular las estimaciones puntuales y por intervalo para la media –ya sea conocida o no la desviación estándar poblacional-, así como las estimaciones para la probabilidad de éxito en una binomial. En el caso en que conozcamos todos los elementos de una población, es sencillo calcular todos los parámetros asociados; sin embargo, en la mayoría de casos no será así, y necesitaremos estimar algunos de ellos a partir de los parámetros de la muestra.

En todo proceso industrial es de suma importancia poder lanzar al mercado un producto que tenga un alto grado de satisfacción por parte de los clientes. Para ello es vital que toda la producción, desde la recepción al embalaje y la comercialización, pasando por cada uno de los estadios de fabricación, se encuentren bastante controlados, esto pasa, por ejemplo, con los componentes de los ordenadores. Nuestros objetivos en este módulo y los sucesivos serán el establecimiento de tablas de control, a partir de los intervalos de confianza, y el diseño de controles de aceptación o rechazo, a partir del contraste de hipótesis.

Comunicación entre el navegador y el servidor

La comunicación entre el navegador y el servidor se lleva a cabo en dos etapas:



El navegador realiza una **solicitud HTTP.**
El servidor procesa la solicitud y después envía una **respuesta HTTP.**

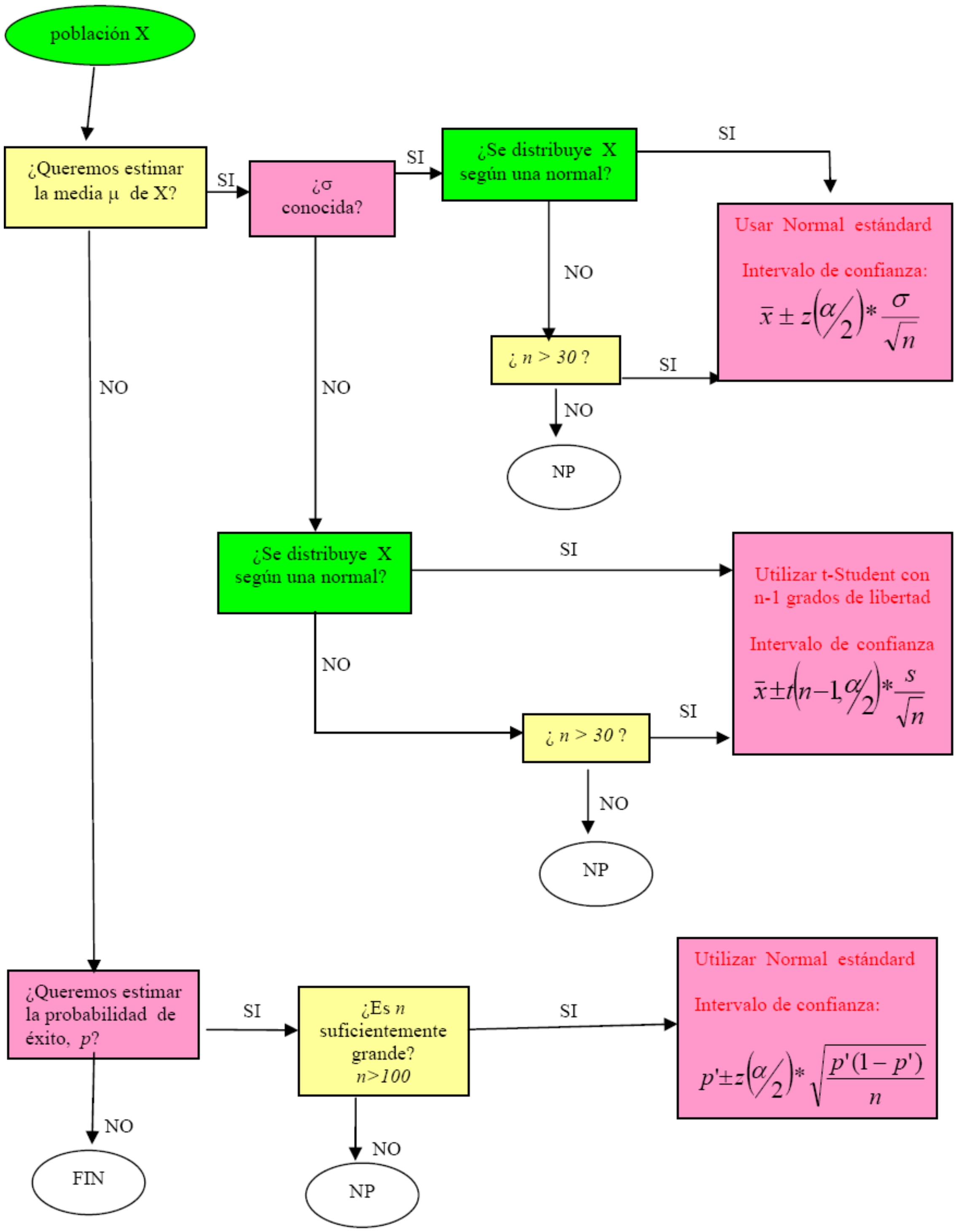
El número de solicitudes se simulan con un proceso Poisson. El tiempo de respuesta se puede modelizar como una distribución normal. Por ejemplo, estamos interesados en medir el tiempo medio de respuesta de un servidor a las peticiones de los usuarios. Para realizar esto medimos en horarios diferentes el tiempo en segundos que hemos tardado en establecer una conexión con nuestro servidor. Los datos obtenidos fueron los siguientes.

1 , 2 , 0.5 , 4 , 6 , 0.7 , 2.1 , 3 , 2.5 , 4

Suponiendo que el tiempo de respuesta sigue una distribución normal, calculad el intervalo de confianza al 95%.

Mapa conceptual

INTERVALOS DE CONFIANZA



NP significa que tenemos que hacer servir métodos No Paramétricos (fuera del contenido del curso)

Actividades

Actividad 1: Estimación de μ con σ conocida

Generar muestra. TCL. Intervalo de confianza para la media. Nivel de confianza. Software R.

Sea X una v.a. que puede tomar aleatoriamente los valores $0,1,2,\dots,9$, todos con la misma probabilidad. Es inmediato comprobar que su media es $\mu = 4,5$ y su desviación estándar $\sigma=2,87$. En lo que sigue, supondremos que desconocemos el valor de μ, el cual estimaremos, y que conocemos el valor de la desviación estándard $\sigma = 2,87$.

Nuestro primer objetivo será obtener una muestra de 40 observaciones (suficientemente grande como para que, usando el TCL, se cumpla el supuesto de que X se distribuye normalmente) y, a partir de la misma, construir el IC para la media μ a un nivel de confianza del 90%.

Solución

En primer lugar, definimos la variable dígitos que contendrá los dígitos de 0 a 9:

```
> digitos <- 0:9
> digitos
 [1] 0 1 2 3 4 5 6 7 8 9
> |
```

A continuación generamos la muestra de 40 observaciones con valores de 0:9:

```
>
> muestra <- sample(digitos,40,replace=T)
> muestra
 [1] 7 3 6 8 4 8 3 9 7 9 1 0 5 2 3 0 8 3 2 2 2 1 1 5 3 1 3 0 5 6 5 6 2 6 4 0 4 4 2 8
> |
```

A continuación, calcularemos el intervalo de confianza al 90% de confianza para la muestra anterior con desviación estàndard $\sigma = 2,87$. Como R no tiene implementado el intervalo de confianza para la media con desviación estándar conocida, implementamos la función que nos lo calcule:

```
>
> z.test <- function(muestra,desv.tipica, nivel.confianza){izquierda <- mean(muestra)-qnorm(1-
(1-nivel.confianza)/2)*desv.tipica/sqrt(length(muestra)); derecha <- mean(muestra)+qnorm(1-(1-
nivel.confianza)/2)*desv.tipica/sqrt(length(muestra));return(c(izquierda,derecha))}
>
```

A continuación usamos la función anterior para calcular el intervalo de confianza pedido:

```
>
> z.test(muestra,2.87,0.90)
[1] 3.203587 4.696413
>
```

A continuación generamos 19 muestras aleatorias y las guardamos en la matriz mat_muestra:

```
> mat_muestra <- c();
> for (i in 1:19){mat_muestra <- cbind(mat_muestra,sample(digitos,40,replace=T))}
> mat_muestra
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15] [,16] [,17] [,18] [,19]
 [1,]    1    9    7    4    2    3    2    1    8     7     1     9     4     3     3     1     6     4     9
 [2,]    9    8    3    1    6    2    7    6    8     3     7     0     2     8     0     8     8     7     3
 [3,]    3    0    3    3    7    9    0    9    2     7     8     3     3     9     9     4     9     0     7
 [4,]    2    0    5    5    6    2    4    0    3     8     8     8     5     0     4     5     0     4     8
 [5,]    4    8    5    2    6    6    5    3    6     9     3     8     1     0     9     7     1     3     6
 [6,]    5    7    9    7    0    6    5    2    6     7     3     9     2     3     9     9     4     4     8
 [7,]    6    4    7    3    1    0    5    9    7     2     7     1     8     0     9     1     4     7     9
 [8,]    5    2    7    7    7    1    3    9    5     4     2     4     7     6     6     2     9     8     3
 [9,]    1    7    7    9    6    3    2    7    7     7     1     6     0     3     0     8     1     3     1
[10,]    0    9    8    3    3    2    5    3    8     3     5     0     6     6     6     6     6     4     0
[11,]    7    8    7    8    8    3    1    5    9     0     9     4     1     9     7     0     2     7     0
[12,]    2    4    1    9    5    1    4    2    7     9     5     3     1     0     6     3     5     7     1
[13,]    5    9    8    8    1    4    8    8    4     1     6     8     2     9     1     1     7     8     3
[14,]    9    9    6    4    5    1    6    4    0     0     3     1     2     4     6     6     7     9     5
[15,]    4    5    1    4    5    2    8    2    7     6     9     2     3     9     0     4     2     9     8
[16,]    0    4    3    5    8    9    6    3    6     2     9     3     0     3     8     3     4     6     1
[17,]    9    4    3    2    1    9    1    4    9     9     3     5     9     6     6     5     5     7     4
[18,]    1    2    8    8    2    1    8    2    5     5     5     1     6     0     2     2     9     5     4
[19,]    3    1    3    9    4    6    7    4    5     1     3     2     0     0     1     2     1     5     1
[20,]    0    5    5    8    6    6    5    4    5     5     2     5     6     1     3     3     0     0     2
[21,]    1    0    4    0    4    0    5    2    6     1     0     2     5     7     7     4     4     8     1
[22,]    9    4    4    2    9    4    4    2    9     0     2     2     9     1     2     5     3     4     5
[23,]    9    1    5    7    3    6    7    1    2     8     0     0     0     2     7     6     1     9     9
[24,]    4    9    6    7    1    1    6    8    1     1     5     0     5     3     6     5     1     4     9
[25,]    7    8    9    7    9    6    5    9    2     4     1     6     9     6     7     7     4     2     6
[26,]    9    2    8    9    6    6    9    6    9     3     4     4     5     2     8     2     2     3     9
[27,]    4    0    5    5    1    4    7    2    4     5     7     5     8     8     2     3     9     6     0
[28,]    4    0    2    4    1    8    5    2    7     7     4     3     7     4     6     0     7     4     4
[29,]    5    6    7    8    8    9    6    2    2     2     2     9     2     7     5     4     7     8     1
[30,]    6    1    8    2    4    3    9    6    5     1     9     6     5     9     5     6     2     9     9
[31,]    5    0    2    5    9    6    0    4    1     1     0     8     7     1     0     7     8     0     3
[32,]    9    0    2    2    0    9    4    8    5     1     2     1     3     2     5     4     1     2     4
[33,]    0    7    8    2    9    6    7    5    9     8     5     5     5     2     5     7     9     0     3
[34,]    4    2    6    0    3    6    1    6    4     1     0     6     9     0     9     1     5     6     4
[35,]    6    0    5    1    2    0    5    0    1     7     7     7     7     9     0     2     4     9     3
[36,]    7    0    3    7    0    9    9    9    0     7     1     6     0     1     7     5     1     4     5
[37,]    7    0    4    0    9    5    8    5    5     6     3     8     4     3     9     1     4     9     9
[38,]    4    7    3    4    3    5    4    4    5     3     1     2     0     9     0     4     1     4     7
[39,]    7    4    8    5    4    7    2    8    6     9     2     6     7     7     8     8     9     0     7
[40,]    4    6    5    8    7    6    7    2    8     1     4     5     0     5     5     1     1     3     1
```

A continuación creamos los 19 intervalos de confianza para las medias de las 19 muestras anteriores con σ = *2,87*:

```
> intervalos_confianza <- c()
> for (i in 1:19){intervalos_confianza <- rbind(intervalos_confianza,z.test(mat_muestra[,i],
2.87,0.90))}
> intervalos_confianza
          [,1]     [,2]
 [1,] 3.928587 5.421413
 [2,] 3.303587 4.796413
 [3,] 4.503587 5.996413
 [4,] 4.103587 5.596413
 [5,] 3.778587 5.271413
 [6,] 3.803587 5.296413
 [7,] 4.303587 5.796413
 [8,] 3.703587 5.196413
 [9,] 4.453587 5.946413
[10,] 3.528587 5.021413
[11,] 3.203587 4.696413
[12,] 3.578587 5.071413
[13,] 3.378587 4.871413
[14,] 3.428587 4.921413
[15,] 4.203587 5.696413
[16,] 3.303587 4.796413
[17,] 3.578587 5.071413
[18,] 4.278587 5.771413
[19,] 3.803587 5.296413
```

A continuación calculamos en cuántos de los intervalos anteriores está contenida la media de la variable 4.5:

```
> contador <- 0
> for (i in 1:19){if(intervalos_confianza[i,1]<=4.5 && intervalos_confianza[i,2]>=4.5){contador
<- contador+1}}
> contador
[1] 18
> 
```

Teniendo en cuenta que la media poblacional 4.5 está también en el primer intervalo de confianza calculado, concluimos que está en 19 de los 20 intervalos de confianza.

Actividad 2: Estimación de μ con σ desconocida

Generar muestra. TCL. Intervalo de confianza para la media. Nivel de confianza. Software R.

Sea *X* una v.a. que se distribuye según una normal de media μ y desviación estándar σ desconocidas.

Vamos a generar 20 muestras de la población anterior, cada una de ellas de tamaño 9 (durante la generación tomaremos $\mu = 69$ y $\sigma = 3$) . Para cada una de las muestras, calcularemos un intervalo de confianza para la media poblacional μ a un nivel del 90%.

Solución

Calculemos 20 muestras de tamaño 9 de una variable normal $\mu = 69$ y $\sigma = 3$.

```
> muestras <- c()
> for (i in 1:20){muestras <- cbind(muestras,rnorm(9,69,3))}
> muestras
         [,1]     [,2]     [,3]     [,4]     [,5]     [,6]     [,7]     [,8]     [,9]
 [1,] 69.71891 69.21977 65.35206 68.25897 66.33439 71.70181 69.90676 70.34068 69.28446
 [2,] 70.42929 71.54618 70.27688 71.62894 68.19223 66.81864 68.14041 75.13066 68.75156
 [3,] 68.54938 69.07140 70.62059 64.66371 68.86249 70.15736 70.32229 69.57313 75.02254
 [4,] 66.29473 69.15491 66.24418 65.39184 72.39279 68.99431 67.88496 64.50666 64.69892
 [5,] 66.38347 65.77956 69.34937 66.80876 71.46085 68.24346 72.50710 62.76231 73.98519
 [6,] 72.93841 76.46294 66.39434 69.66173 68.47846 74.92420 70.18630 64.41641 69.27952
 [7,] 69.54085 67.45604 70.18215 67.06387 64.17030 67.04110 68.52829 72.42352 67.70429
 [8,] 68.63466 67.98968 73.44535 68.62214 72.35766 71.55314 68.14197 65.32012 72.25240
 [9,] 69.99991 68.95395 65.11497 70.65198 76.28786 67.81929 66.62366 69.12474 67.64247
        [,10]    [,11]    [,12]    [,13]    [,14]    [,15]    [,16]    [,17]    [,18]
 [1,] 73.38132 70.03103 68.55508 66.98135 72.86119 67.92750 65.68901 69.76511 67.67795
 [2,] 71.41692 73.42706 69.27451 69.45949 72.06781 67.35463 62.82467 68.44955 68.75801
 [3,] 65.36724 68.55360 70.24772 75.51673 67.56460 70.81360 61.79754 67.87115 73.48145
 [4,] 75.47642 67.46085 62.55452 62.62843 70.17217 68.87932 66.41469 66.38770 71.12577
 [5,] 71.98151 68.11643 70.60297 70.71731 65.51914 66.50217 70.21593 72.36358 65.97355
 [6,] 64.07117 69.18097 68.09890 69.74107 62.51655 69.69751 60.19150 68.37098 68.68557
 [7,] 68.04647 69.42296 71.07857 70.27126 69.72258 72.93417 70.60979 65.12047 63.12270
 [8,] 67.95088 66.26270 72.29687 69.89526 67.83678 74.45650 70.39501 62.76815 67.42736
 [9,] 75.70720 70.92722 70.91762 72.71183 69.41265 72.06661 68.34517 64.44057 66.50045
        [,19]    [,20]
 [1,] 67.63829 71.67919
 [2,] 69.41856 69.41650
 [3,] 69.49081 69.38730
 [4,] 72.20671 64.30236
 [5,] 73.62958 63.59900
 [6,] 64.23456 71.06797
 [7,] 68.76647 67.04820
 [8,] 70.11320 71.91039
 [9,] 72.30662 71.02949
>
```

A continuación calculamos los intervalos de confianza de cada muestra usando la instrucción t.test:

```
> intervalos_confianza <- c()
> for (i in 1:20){intervalos_confianza <-
rbind(intervalos_confianza,t.test(muestras[,i],conf.level=0.90)[[4]])}
> intervalos_confianza
          [,1]     [,2]
 [1,] 67.89486 70.43616
 [2,] 67.63423 71.39564
 [3,] 66.76585 70.34079
 [4,] 66.63677 69.53033
 [5,] 67.57037 72.10453
 [6,] 68.04796 71.34166
 [7,] 68.04894 70.22700
 [8,] 65.59376 70.76141
 [9,] 67.79699 71.89665
[10,] 67.75451 73.00085
[11,] 67.97199 70.55753
[12,] 67.52698 71.05675
[13,] 67.55893 71.97945
[14,] 66.63718 70.62359
[15,] 68.39787 71.74258
[16,] 63.82668 68.72517
[17,] 65.46410 69.09974
[18,] 66.23221 69.93508
[19,] 68.00593 71.50625
[20,] 66.87708 70.77635
```

Por último calculamos en cuántos de los intervalos anteriores está incluida la media poblacional 69:

```
> contador <- 0
> for (i in 1:20){if(intervalos_confianza[i,1]<=69 && intervalos_confianza[i,2]>=69){contador <-
contador+1}}
> contador
[1] 19
>
```

Vemos que en 19 de los 20 intervalos está incluida la media poblacional 69.

Actividad 3: Estudio sobre el funcionamiento del servicio de préstamo interbibliotecario de la UOC.

Semilla. Simulación. Intervalo de confianza para la media. Amplitud de los intervalos. Error muestral. Software R.

A partir de los datos obtenidos en los dos últimos años, los responsables de la biblioteca de la UOC afirman que su servicio de préstamo interbibliotecario (SPI) funciona según un modelo de distribución normal, con media μ = 15 días y desviación típica σ = 3 días (es decir, que cuando un estudiante de la UOC solicita un documento de otra universidad, la variable "tiempo transcurrido desde que se ha solicitado el libro hasta que éste llega" sigue una distribución normal con los parámetros anteriores). Suponiendo cierta la afirmación anterior, se pide:

a) Simular, usando la opción Muestra de una distribución Normal, la solicitud de 9 libros al SPI (observar que el tamaño muestral será pues n = 9). Repetir esta simulación hasta un total de 20 veces y, en cada caso, calcular un intervalo de confianza a nivel del 90% para la media poblacional μ (tiempo que se espera tarde cada libro solicitado, dato que supondremos ahora desconocido). Antes de realizar la simulación, usar la siguiente instrucción para elegir como "semilla" el número 503.

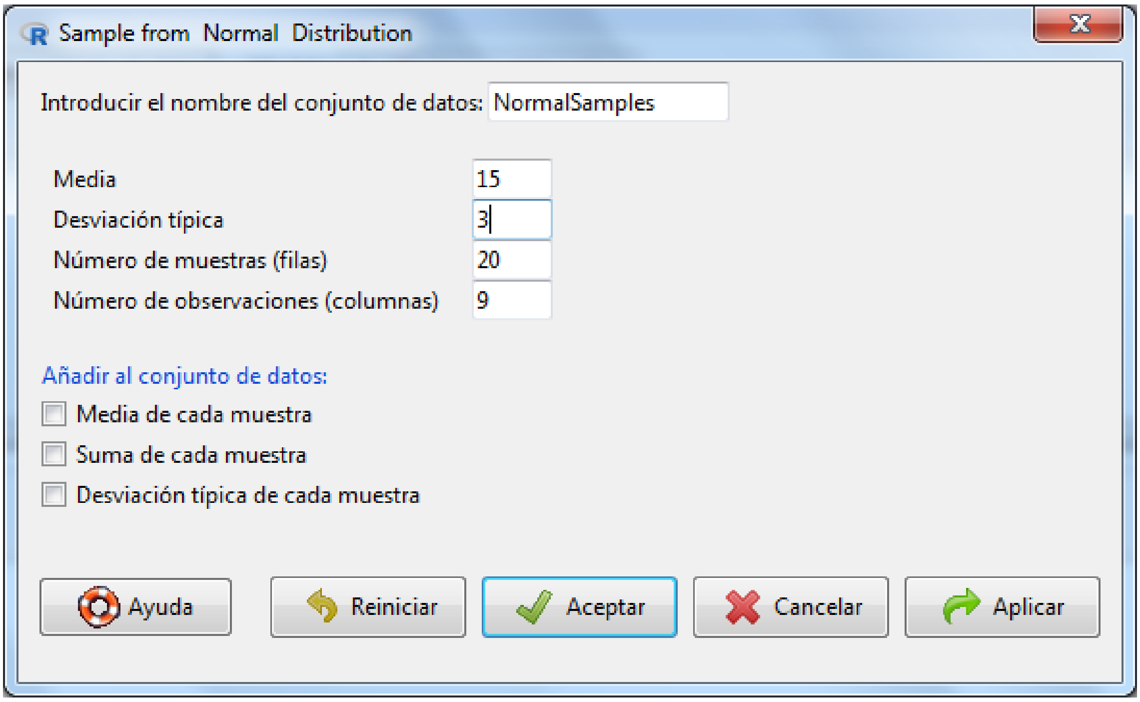
> set.seed(503)

b) ¿Cuántos intervalos contienen el verdadero valor de μ? ¿Cuántos intervalos esperabas que contuviesen dicho valor? ¿Por qué?
c) ¿Tienen todos los intervalos la misma amplitud? ¿Por qué?

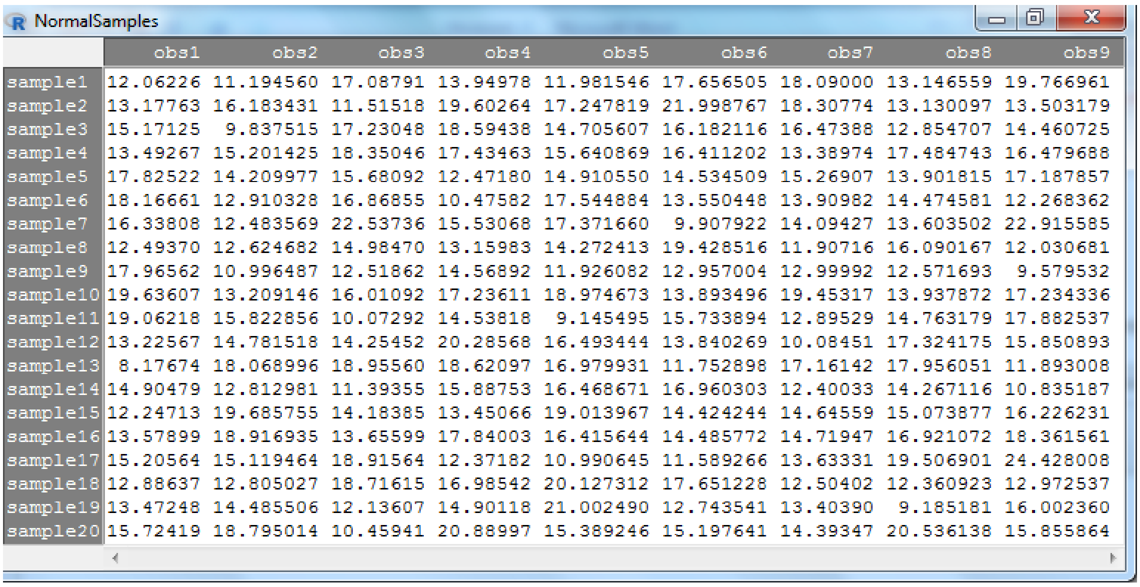
d) Interpreta el valor obtenido para el segundo intervalo, i.e.: ¿qué significa para el usuario que ha solicitado el libro?.

Solución

a) Para hacerlo con R Commander: Distribuciones ->Distribuciones continuas ->Distribución Normal ->Muestra de una distribución Normal.



Se obtiene el siguiente“output”:


NormalSamples

	obs1	obs2	obs3	obs4	obs5	obs6	obs7	obs8	obs9
sample1	12.06226	11.194560	17.08791	13.94978	11.981546	17.656505	18.09000	13.146559	19.766961
sample2	13.17763	16.183431	11.51518	19.60264	17.247819	21.998767	18.30774	13.130097	13.503179
sample3	15.17125	9.837515	17.23048	18.59438	14.705607	16.182116	16.47388	12.854707	14.460725
sample4	13.49267	15.201425	18.35046	17.43463	15.640869	16.411202	13.38974	17.484743	16.479688
sample5	17.82522	14.209977	15.68092	12.47180	14.910550	14.534509	15.26907	13.901815	17.187857
sample6	18.16661	12.910328	16.86855	10.47582	17.544884	13.550448	13.90982	14.474581	12.268362
sample7	16.33808	12.483569	22.53736	15.53068	17.371660	9.907922	14.09427	13.603502	22.915585
sample8	12.49370	12.624682	14.98470	13.15983	14.272413	19.428516	11.90716	16.090167	12.030681
sample9	17.96562	10.996487	12.51862	14.56892	11.926082	12.957004	12.99992	12.571693	9.579532
sample10	19.63607	13.209146	16.01092	17.23611	18.974673	13.893496	19.45317	13.937872	17.234336
sample11	19.06218	15.822856	10.07292	14.53818	9.145495	15.733894	12.89529	14.763179	17.882537
sample12	13.22567	14.781518	14.25452	20.28568	16.493444	13.840269	10.08451	17.324175	15.850893
sample13	8.17674	18.068996	18.95560	18.62097	16.979931	11.752898	17.16142	17.956051	11.893008
sample14	14.90479	12.812981	11.39355	15.88753	16.468671	16.960303	12.40033	14.267116	10.835187
sample15	12.24713	19.685755	14.18385	13.45066	19.013967	14.424244	14.64559	15.073877	16.226231
sample16	13.57899	18.916935	13.65599	17.84003	16.415644	14.485772	14.71947	16.921072	18.361561
sample17	15.20564	15.119464	18.91564	12.37182	10.990645	11.589266	13.63331	19.506901	24.428008
sample18	12.88637	12.805027	18.71615	16.98542	20.127312	17.651228	12.50402	12.360923	12.972537
sample19	13.47248	14.485506	12.13607	14.90118	21.002490	12.743541	13.40390	9.185181	16.002360
sample20	15.72419	18.795014	10.45941	20.88997	15.389246	15.197641	14.39347	20.536138	15.855864

O directamente con instrucciones podemos crear la siguiente función:

```
simulaMuestra = function(n)
{
muestra = rnorm(n, 15, 3)
return(muestra)
```

}

muestras = t(replicate(20, simulaMuestra(9)))

muestras

```
RGui (64-bit)
Archivo  Editar  Visualizar  Misc  Paquetes  Ventanas  Ayuda
          [,1]      [,2]      [,3]      [,4]     [,5]     [,6]      [,7]      [,8]      [,9]
 [1,] 17.15807 13.896756 15.530953 17.308892 15.26998 10.08254 16.690831 16.036518  9.635296
 [2,] 16.20926 13.600783 12.069228 15.399898 14.74390 16.26112 18.436951 15.695946 14.470674
 [3,] 15.28338 14.992000 21.716112 14.563713 15.33494 16.27242 18.422050 14.724297  9.065959
 [4,] 16.07910 11.907225 11.202372 10.879534 10.81254 12.85362 10.303631 13.709882 15.538281
 [5,] 13.62268 15.649587 13.010043 14.383528 12.81245 10.64185 11.100945 12.203227 13.167838
 [6,] 18.75787 19.160468 20.039386 15.723454 12.10328 18.10581 13.522289 11.174795 13.681870
 [7,] 14.44604 15.227186 18.419247 18.091641  9.26158 16.19625 15.584363  8.465311 12.866674
 [8,] 13.02925 14.806184 13.248678 15.191226 13.72278 15.89845 18.397706 11.652181 20.448489
 [9,] 12.17528 14.645865 12.195557 11.801008 12.85162 13.70659 16.355590 12.931368 17.538997
[10,] 13.57766 14.737319 21.022135  8.652686 11.17870 10.41759 13.571439 12.025025 11.292154
[11,] 15.67549 15.306613 10.917133 17.526577 19.46874 10.66296 18.166972 17.582291 11.478573
[12,] 11.14544 15.445133  7.350161 17.570580 10.94774 16.58199 20.797254 13.102481 11.111253
[13,] 13.80866 19.760020 11.781683 15.794511 13.35118 12.57930 14.868105 14.911831 12.226873
[14,] 14.18309 16.084522  9.726441 11.537130 13.35792 14.26862 12.824211 19.929454 16.758628
[15,] 19.97039 13.321667 15.535744 16.200317 18.89162 15.36533 16.975511 16.788065 13.882144
[16,] 17.34222 10.902821 17.486581  8.988329 15.10568 17.82559 18.454483 13.577188 10.451322
[17,] 14.89900 11.104755 11.183211 15.195897 13.44649 15.50845 15.932812 16.526728 14.421064
[18,] 16.05701 11.909202 11.249099 15.154334 13.47409 18.83141 18.540038 10.738885 13.429121
[19,] 18.22591  9.667918 11.522247 11.569999 15.07053 13.32779  8.975293 15.028486 10.244921
[20,] 15.88547 12.660641 12.472583 16.954376 15.57579 19.91660 18.937509 17.600593 19.376826
> 
```

Ahora encontramos los intervalos de confianza pedidos, teniendo en cuenta que, en esta ocasión, conocemos el valor de la desviación típica (usaremos pues la normal estándard):

```
simulaMuestra = function(n)
{
muestra = rnorm(n, 15, 3)
intervalo = mean(muestra) + c(-1, 1) * qnorm(0.95) * 2/sqrt(n)
return(intervalo)
}

intervalos = t(replicate(20, simulaMuestra(9)))

intervalos
          [,1]     [,2]
 [1,] 12.85752 15.05066
 [2,] 14.69750 16.89064
 [3,] 12.89831 15.09145
 [4,] 13.83820 16.03134
 [5,] 13.01925 15.21239
 [6,] 13.81254 16.00568
 [7,] 14.72735 16.92049
 [8,] 13.94888 16.14202
 [9,] 12.93980 15.13294
[10,] 14.58121 16.77435
[11,] 13.35476 15.54790
[12,] 14.57377 16.76691
[13,] 14.30084 16.49398
[14,] 12.01026 14.20340
[15,] 13.94177 16.13491
[16,] 14.98144 17.17458
[17,] 16.16821 18.36134
[18,] 13.61003 15.80317
```

[19,] 13.94433 16.13747
[20,] 12.92045 15.11358

b) Observar que 18 de los 20 intervalos contienen el valor $\mu = 15$. De hecho, era de esperar que aproximadamente el 90% de los intervalos encontrados (90% de 20 es justamente 18) contuviesen este valor.

c) La longitud de todos los intervalos anteriores debe ser la misma. El motivo es que la longitud de cada intervalo viene dada por 2*E, donde E es el error muestral máximo ($E = z(\alpha/2) * \sigma/\sqrt{n} = 3{,}29$).

d) Según los valores obtenidos en la simulación el usuario podría afirmar, a un nivel de confianza del 90%, que el libro solicitado tardará entre 14,70 y 16,89 días.

Actividad 4: Estudio sobre la piratería de películas.

TCL. Tamaño muestral. Intervalo de confianza para la media y proporción. Nivel de confianza. Software R.

Un estudio de las SAES sobre películas intenta determinar el nivel de piratería que tienen los alumnos de una determinada universidad. Se encuestaron a 50 individuos elegidos de forma aleatoria de los cuales 28 disponían alguna película pirata. El resto no tenían ninguna película pirata.

a) Obtened un intervalo de confianza al 95% para el porcentaje de individuos que tienen películas pirata. ¿Se puede admitir que la mitad de los individuos tienen películas pirata?

b) De que tamaño debería ser la muestra para reducir a la mitad la longitud del intervalo?

Suponiendo normalidad de la distribución correspondiente y sabiendo que el número de películas pirata que tenían los 28 que lo admitieron es:

43, 48, 37, 94, 105, 38, 55, 45, 80, 77, 51, 34, 53, 63, 14, 42, 38, 24, 70, 93, 102, 25, 28, 34, 36, 50, 36, 62,

c) Hallad un intervalo de confianza del 90% para el número medio de películas pirata de entre los que tienen películas de estas.

d) Interpretad el intervalo de confianza.

Solución

a) Con la instrucción:

```
> prop.test(28,  50,  alternative='two.sided',  p=.5, conf.level=.95, correct=FALSE)
```

```
        1-sample proportions test without continuity correction

data:  28 out of 50, null probability 0.5
X-squared = 0.72, df = 1, p-value = 0.3961
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.4230603 0.6883780
sample estimates:
```

```
  p
0.56
```

Así, el intervalo de confianza al 95% para la proporción es (0.423; 0.688). A partir del intervalo de confianza se observa que la hipótesis de que la mitad de usuarios dispongan de películas pirata es correcta.

b) La longitud del intervalo es $2 * z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. Así, hemos de multiplicar por 4 el tamaño de la muestra para reducir a la mitad de la longitud. Necesitaremos una muestra de tamaño 200.

c) Usaremos la t-student ya que no conocemos sigma. Introducimos los datos en la variable V1 y vamos a R Comander: Estadísticos > Medias y le damos a Test t para una muestra y elegimos la variable: V1, marcamos Media poblacional ¡mu0 Hipótesis nula: mu = 0.0 y nivel de confianza: .90

```
> t.test(peliculas$V1, alternative='two.sided', mu=0.0, conf.level=.90)
```

Obtenemos el “output”:

```
One Sample t-test

data:  peliculas$V1
t = 4.905, df = 27, p-value = 3.932e-05
alternative hypothesis: true mean is not equal to 0
90 percent confidence interval:
44.84993 60.65007
sample estimates:
mean of x
   52.75
```

Así, el intervalo de confianza al 90% para la media es (44.85; 60.65).

d) Indica que de cada 100 muestras distintas que cogiéramos –y que daría lugar a 100 intervalos distintos- en el 90% de los casos, el intervalo obtenido contendría la media poblacional.

Actividad 5: Estudio sobre la proporción de hogares que tienen ordenador y banda ancha.

TCL. Intervalo de confianza para la media y proporción. Nivel de confianza. Software R.

En la siguiente pantalla de datos, podemos encontrar las variables LLAR_ORD (proporción de hogares que tienen ordenador), LLAR_BA (proporción de hogares que tienen banda ancha).

↓	C1	C2	C3	C4
	LLAR_ORD	LLAR_BA	USU_ORD	USU_COR
1	64,6188	39,6013	58,9384	45,4938
2	70,6249	45,6342	64,3950	47,4064
3	64,2484	43,5412	60,4109	48,6097
4	71,2663	33,7878	52,4718	35,7512
5	57,6435	32,1987	50,8143	36,3483
6	64,7489	44,0721	61,5918	43,5046
7	66,7538	40,5427	56,4004	40,8341
8	60,4349	41,1113	61,7787	45,5469
9	60,0854	24,4543	51,2101	33,9002
10	58,7004	42,2022	57,7720	42,7404
11	63,9532	51,1428	65,7906	50,6901
12	69,3859	41,2327	61,8521	44,8027
13	65,0687	49,9265	58,7896	49,5121
14	60,9454	40,1733	55,2561	40,1754
15	57,5605	32,8864	62,7563	39,3537
16	77,3762	41,3325	60,6667	45,9460
17	59,2103	45,6849	63,6107	50,0219
18	64,2766	32,5023	53,0574	37,4951
19	67,4305	45,7729	60,4322	42,6967
20	72,1898	41,6095	62,6842	47,4671
21	75,7838	47,3866	65,0358	50,2751
22	63,6186	26,5791	54,3564	36,6763
23	61,9027	30,9928	59,2544	43,6985
24	65,3377	43,6540	61,1756	53,0282
25	74,9900	34,9094	51,7311	36,2705
26	65,1809	33,6252	56,6723	38,1024
27	65,3551	36,0900	60,0942	41,2713
28	68,4741	47,8383	62,4837	49,5373
29	59,2229	35,9583	56,1627	42,2148
30	61,2693	30,5968	50,1595	35,9874
31	62,2441	39,0514	56,7887	41,2898
32	64,6091	42,3303	60,1923	45,2172
33	67,4023	38,2239	62,5300	46,0961
34	65,2089	38,4087	62,6382	45,2434
35	62,7509	37,3973	54,9593	37,0856
36	71,8195	41,1555	65,9579	47,5231
37	56,7604	23,3730	39,5549	25,8601
38	63,3384	34,6439	67,8957	45,6673
39	58,2328	34,3720	56,5405	41,8765
40	64,2704	48,2378	63,2496	50,1087
41	62,5030	46,8367	62,3430	48,9784

Ahora se pide

1. Encontrar un intervalo de confianza al 90% para la media de cada una de estas variables; discutid en base a los intervalos si las variables tienen un comportamiento similar o no.

2. Imaginad que vamos a otra población y descubrimos que de 200 hogares, 95 tienen banda ancha. Encontrad un intervalo de confianza al 95% para la proporción de hogares con banda ancha y comentad si se puede decir que en la mitad de los hogares hay banda ancha.

Solución

1. R Comander: Estadísticos > Medias y le damos a Test t para una muestra y elegimos la variable: LLAR_ORD, marcamos Media poblacional ¡mu0 Hipótesis nula: mu = 0.0 y nivel de confianza: .90

> t.test(hogares$ LLAR_ORD, alternative='two.sided', mu=0.0, conf.level=.90)

Obtenemos el “output”:

```
One Sample t-test
data:  hogares$LLAR_ORD
alternative hypothesis: true mean is not equal to 0
90 percent confidence interval:
63.469 66.131
sample estimates:
mean of x
   64.800
```

Así, el intervalo de confianza al 90% para la media es (63.47; 66.13).

Para LLAR_BA tenemos:

```
One Sample t-test
data:  hogares$LLAR_BA
alternative hypothesis: true mean is not equal to 0
90 percent confidence interval:
37.28 40.83
sample estimates:
mean of x
   39.05
```

Aquí el intervalo es (37.28; 40.83). Los dos intervalos son disjuntos y por lo tanto las medias de las variables serán diferentes, con un nivel de confianza del 90%.

2. Con la instrucción:

> prop.test(95, 200, alternative='two.sided', p=.5, conf.level=.95, correct=FALSE)

```
        1-sample proportions test without continuity correction
data:  95 out of 200, null probability 0.5
X-squared = 0.5, df = 1, p-value = 0.4795
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.4069162 0.5440260
sample estimates:
    p
0.475
```

Con lo que es cierto que puede ser del 50% la proporción de hogares con banda ancha (0.5 está dentro del intervalo).

Direcciones de interés

http://www.unalmed.edu.co/~estadist/confinterval/intervalconf.htm
Características y applet del concepto de Intervalo de confianza.

http://e-stadistica.bio.ucm.es/mod_intervalos/intervalos_applet.html
Características y applet del concepto de Intervalo de confianza.

http://e-stadistica.bio.ucm.es/mod_contraste/contraste_applet.html
Applet sobre contraste de hipótesis para muestras independientes.

http://kitchen.stat.vt.edu/~sundar/java/applets/
Aplicaciones estadísticas con JAVA.

http://www.udc.es/dep/mate/recursos.html
Selección de recursos en Internet para la enseñanza-aprendizaje de la Estadística.

http://www.uoc.edu/in3/e-math/docs/Estimacion_IC.pdf
Math-block del proyecto e-math sobre estimación e intervalos de confianza con teoría y ejemplos con y sin Minitab.