

Intervalos de confianza con STATGRAPHICS

Ficheros empleados: TiempoaccesoWeb.sf3 ; TiempoBucle.sf3;

1. Ejemplo 1: Tiempo de acceso a una página Web

Se desean construir intervalos de confianza para la media μ y la desviación típica σ de la distribución del **tiempo de acceso** a la página web de la UC3M desde un domicilio particular así como desde la universidad. Los intervalos de confianza de los parámetros se construirán usando la información que proporciona una muestra de 55 datos del tiempo (en segundos) que se tarda en acceder a la página web de la UC3M. Las mediciones se hacen desde dos ordenadores: desde un domicilio particular y desde la universidad (fichero TiempoaccesoWeb.sf3)

1.1 Entrada de datos:

Lo primero que hacemos es leer ese fichero de datos.

The image shows the STATGRAPHICS Plus software interface. The 'Open Data File' dialog box is open, displaying a file list with 'TiempoaccesoWeb' selected. The 'Nombre de archivo' field contains 'TiempoaccesoWeb' and the 'Tipo de archivos' dropdown is set to 'SG PLUS Files (*.sf3;*.sfx;*.sf)'. A red arrow points from the 'Open Data File' dialog to a data table titled 'TiempoaccesoWeb.sf3'.

	Ordenador_Ca	Ordenador_Uni
1	6,147	1,165
2	5,833	1,416
3	5,718	1,393
4	6,221	1,407
5	5,722	1,435
6	5,976	1,348
7	5,773	1,287
8	6,138	1,283
9	5,709	1,158
10	6,489	1,633
11	6,023	1,499

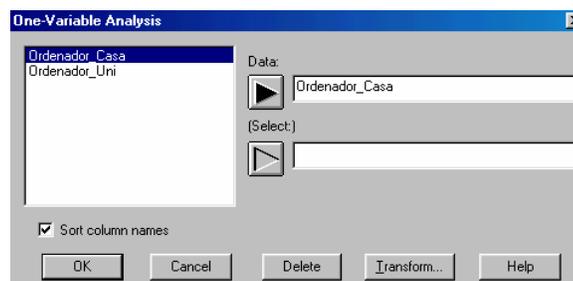
A red box highlights the value '5,976' in the 'Ordenador_Ca' column of row 6. A red arrow points from this box to a text box containing 'Tiempo en segundos'.

1.2. Análisis univariante de la variable

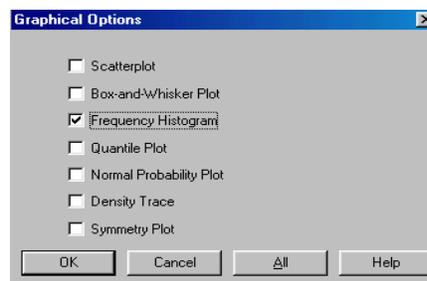
Antes de realizar cualquier análisis conviene hacer una descripción de las variables. Vamos a comenzar analizando la variable tiempo de acceso a la web desde un domicilio particular (variable Ordenador_Casa). El análisis gráfico y numérico de esta variable se hace en: Describe / Numeric Data / One-Variable Análisis



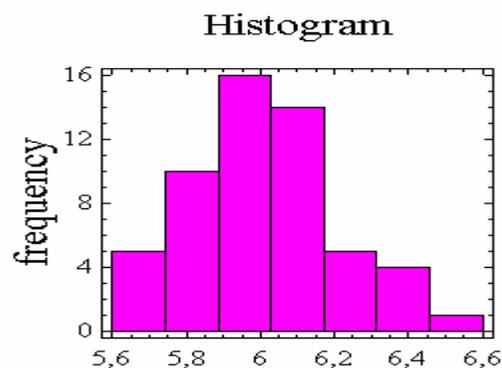
Seleccionamos la variable Ordenador_Casa



Em primer lugar haremos el histograma. Para hacer el histograma de la variable seleccionamos el botón de opciones gráficas  y aparece la ventana:

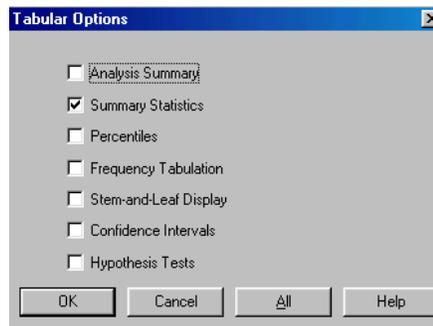


Marcamos la opción Frequency Histogram para que nos muestre el histograma



Lo importante en este histograma es que la variable Ordenador_Casa tiene una distribución parecida a la normal: es bastante simétrica y con forma de campana. La hipótesis de normalidad es importante para calcular intervalos de confianza. Por ejemplo, sólo podremos hacer intervalos de confianza para la varianza de un población basados en la distribución chi-cuadrado si la población es normal.

Para obtener las medidas características de la muestra (estadísticos) seleccionamos el botón TABULAR OPTIONS  y aparece la ventana



Marcamos la opción SUMMARY STATISTICS para obtener un resumen estadístico. La información que se muestra es la siguiente

```
Summary Statistics for Ordenador_Casa

Count = 55
Average = 6,01835
Median = 6,023
Mode =
Variance = 0,0385763
Standard deviation = 0,196409
Standard error = 0,0264837
Minimum = 5,704
Maximum = 6,489
Range = 0,785
Skewness = 0,353707
Std. skewness = 1,0709
Kurtosis = -0,22173
Std. kurtosis = -0,33566
Coeff. of variation = 3,2635%
```

Esta tabla muestra el resumen estadístico para la variable Ordenador_Casa. Incluye las medidas de tendencia central, medidas de variabilidad y medidas de forma. Entre los valores obtenidos se encuentran la media y varianza muestrales que son estimaciones puntuales de la media y la varianza poblacionales. Es decir, tenemos que en esta muestra, las estimaciones 'puntuales' de los parámetros que nos interesan son

$$\hat{\mu} = 6.018$$

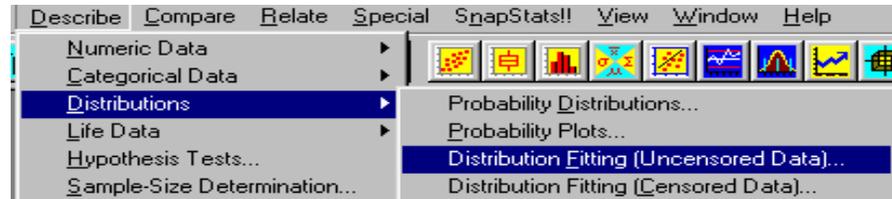
$$\hat{\sigma}^2 = 0.0386$$

Nuestro objetivo es obtener estimaciones 'por intervalos' de esos parámetros.

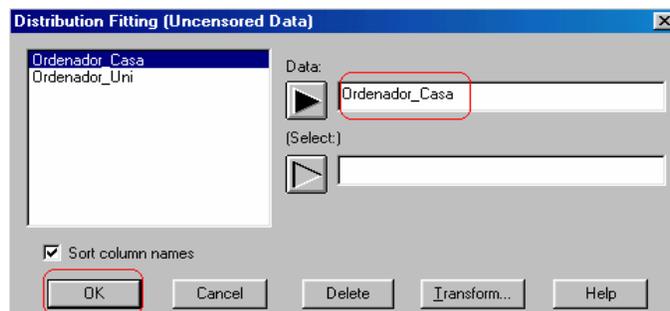
Además del histograma similar a la normal, observamos que los coeficientes de asimetría y curtosis que aparecen en el resumen estadístico son bastante pequeños por lo que la variable es bastante simétrica y con forma de campana.

1.3. Análisis de la normalidad de la variable

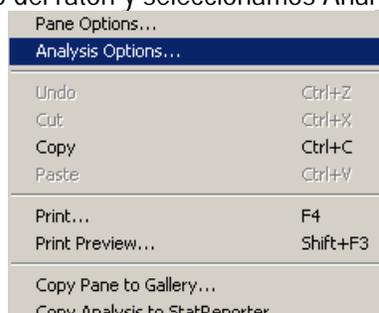
Vamos a realizar un contraste de normalidad mediante el test de la Chi-cuadrado, que nos refuerce nuestra hipótesis de que la población es normal. Seleccionamos DESCRIBE / DISTRIBUTIONS / DISTRIBUTIONS FITTING



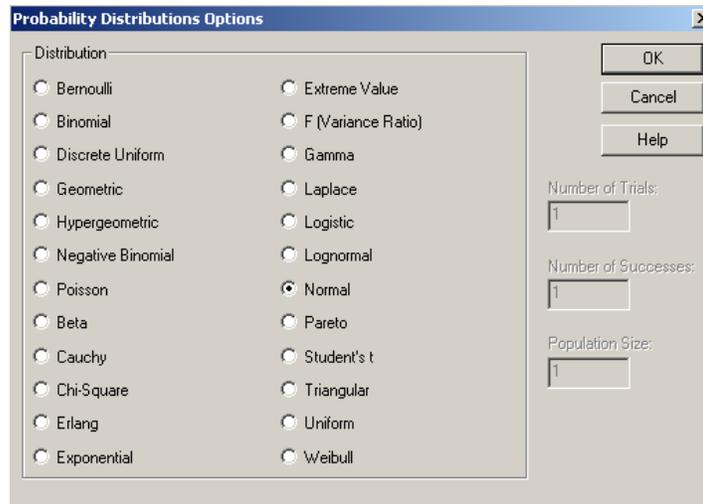
Aparece entonces la ventana



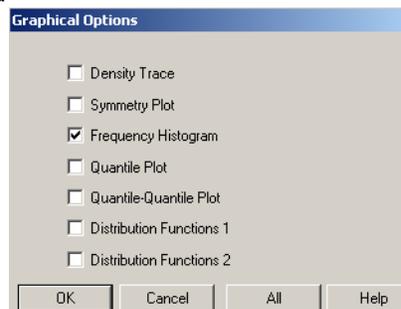
Introducimos en el campo DATA la variable Ordenador_Casa y le damos a OK. Para seleccionar el ajuste a una normal pulsamos el botón derecho del ratón y seleccionamos Analysis Options



En la ventana que aparece seleccionamos la distribución Normal (es la opción por defecto).

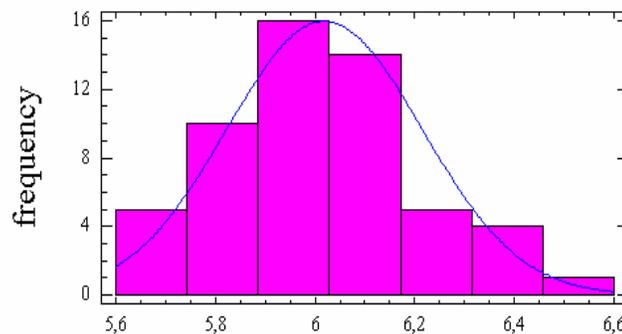


Antes de hacer el test de la chi-cuadrado, lo más recomendable es visualizar el ajuste. En las opciones gráficas seleccionamos el Histograma

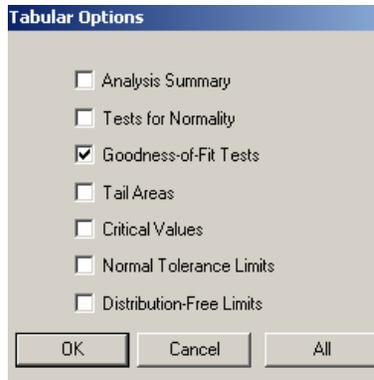


y obtenemos

Histogram for Ordenador_Casa



donde vemos que la curva normal presenta un bien ajuste con nuestros datos. Para ver el test de la Chi-cuadrado seleccionamos en la Opciones Tabulares la opción Goodness-of-Fit



y obtenemos la siguiente ventana de resultados

Goodness-of-Fit Tests for Ordenador_Casa

Chi-Square Test					
	Lower Limit	Upper Limit	Observed Frequency	Expected Frequency	Chi-Square
at or below	5,80867	5,80867	10	7,86	0,58
	5,80867	5,90719	7	7,86	0,09
	5,90719	5,98299	5	7,86	1,04
	5,98299	6,0537	12	7,86	2,18
	6,0537	6,1295	7	7,86	0,09
	6,1295	6,22803	8	7,86	0,00
above	6,22803		6	7,86	0,44

Chi-Square = 4,43589 with 4 d.f. P-Value = 0,350217

Como vemos que el histograma con la curva superpuesta tiene un buen ajuste y dado que el p-valor obtenido es superior a 0.05, utilizando un nivel de significación habitual del 5% no podemos rechazar la hipótesis de que la variable proceda de una distribución normal.

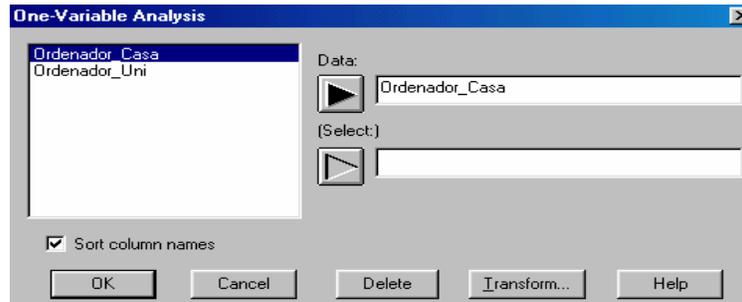
Como podemos asumir normalidad en la variable Ordenador_Casa, podemos calcular los intervalos de confianza para la media y para la desviación típica. Si no pudiésemos asumir que la variable Ordenador_Casa fuese normal, no podríamos utilizar los intervalos de confianza de la desviación típica que proporciona el Statgraphics, al estar basados en la normalidad. Si la muestra fuese suficientemente grande (>30) todavía podríamos utilizar los intervalos de confianza de la media aunque la variable no fuese normal (¿por qué?). Por último, si la variable no fuese normal y la muestra fuese pequeña, no podríamos tampoco usar los intervalos de confianza para la media.

1.4. Intervalos de confianza

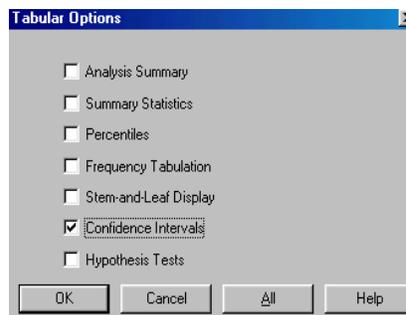
Para realizar los intervalos de confianza para μ y para σ seleccionamos: Describe / Numeric Data / One-Variable Análisis



Seleccionamos la variable Ordenador_Casa



Con el botón TABULAR OPTIONS  aparece la ventana



Marcamos la opción Confidence Intervals y obtenemos la siguiente información:

Confidence Intervals for Ordenador_Casa

```
-----
95,0% confidence interval for mean: 6,01835 +/- 0,0530968 [5,96525;6,07144]
95,0% confidence interval for standard deviation: [0,165349;0,241944]
```

Si calculamos el intervalo de confianza para μ aplicando la fórmula correspondiente obtenemos:

$$\bar{x} \pm t_{\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}} = 6,01835 \pm 1,9944 \cdot \frac{0,196409}{\sqrt{55}} = [5,96525 ; 6,071]$$

que coincide con el valor del intervalo de confianza para la media proporcionado por el Statgraphics. Por tanto, el tiempo medio de acceso a la web será un valor que estará, con una confianza del 95% entre 5.96 y 6.07 segundos.

También aparece el intervalo de confianza para la desviación típica. A partir de él obtendremos, elevando al cuadrado cada término, el intervalo de confianza para la varianza [0,0273402 , 0,0585368]. Si calculamos el intervalo de confianza para σ^2 aplicando la fórmula correspondiente obtenemos

$$\frac{(n-1) \cdot \hat{\sigma}^2}{\chi^2_{n-1, 1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1) \cdot \hat{\sigma}^2}{\chi^2_{n-1, \frac{\alpha}{2}}} \Rightarrow \frac{54 \cdot 0.0385763}{\chi^2_{54, 0.975}} \leq \sigma^2 \leq \frac{54 \cdot 0.0385763}{\chi^2_{54, 0.025}}$$

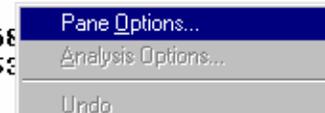
$$\Rightarrow \frac{2,0831202}{76,19555} \leq \sigma^2 \leq \frac{2,0831202}{35,586446} \Rightarrow 0,0273391 \leq \sigma^2 \leq 0,058536$$

que coincide con el intervalo de confianza proporcionado por el Statgraphics. Por tanto el tiempo de acceso a la web la podemos aproximar a una variable aleatoria normal $N(\mu, \sigma^2)$ donde, con un 95% de confianza la varianza será un valor que estará entre 0.0273 y 0.058.

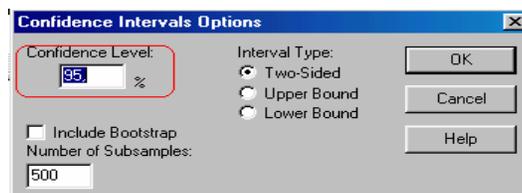
Si queremos otro nivel de confianza diferente, nos situamos sobre la salida anterior, pulsamos el botón derecho del ratón y elegimos la opción PANE OPTIONS,

Confidence Intervals for Ordenador_Casa

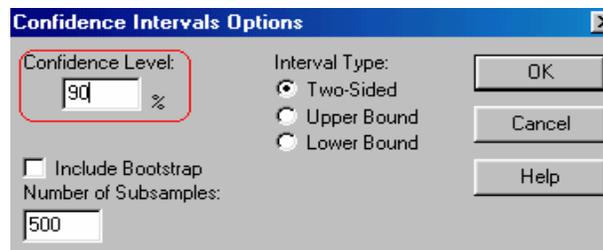
95,0% confidence interval for mean: 6,01835 +/- 0,0530968
 95,0% confidence interval for standard deviation: [0,1653



obtendremos la caja de diálogo CONFIDENCE INTERVALS OPTIONS, cuyo campo CONFIDENCE LEVEL permite elegir el nivel de confianza (la información de esta ventana puede variar con la versión del Statgraphics).



Vamos a calcular los intervalos de confianza del 90% para la media y la desviación típica



Se obtienen los siguientes resultados

Confidence Intervals for Ordenador_Casa

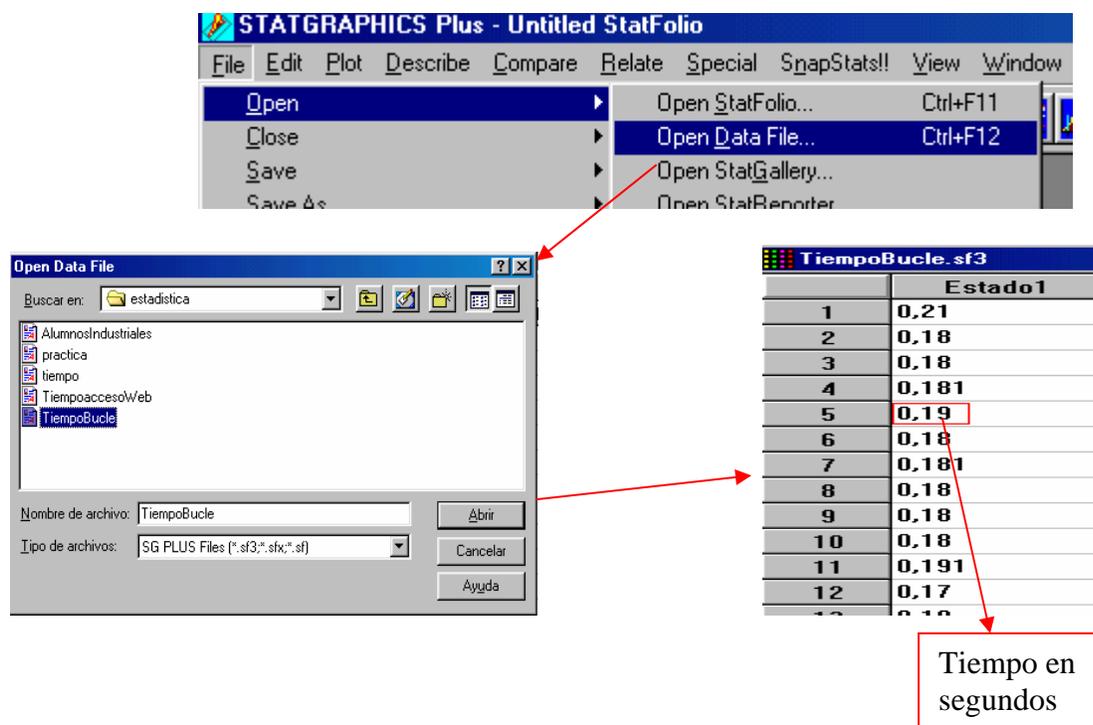
90,0% confidence interval for mean: 6,01835 +/- 0,0443223 [5,97402;6,06267]
 90,0% confidence interval for standard deviation: [0,169914;0,233777]

2. Ejemplo 2: Tiempo de ejecución de un Bucle

Vamos a considerar ahora el fichero TiempoBucle que indica el tiempo en segundos de ejecución de un programa de Matlab bajo distintas circunstancias. En cada circunstancia, la ejecución se repite 200 veces. Se desean construir intervalos de confianza para la media y la varianza **poblacionales** del tiempo que tarda el programa en el Estado1

2.1 Entrada de datos:

Lo primero que hacemos es leer ese fichero de datos.



The image shows the STATGRAPHICS Plus interface. The 'Open Data File' dialog box is open, showing the file 'TiempoBucle' selected. The data table 'TiempoBucle.sf3' is displayed with the following data:

	Estado1
1	0,21
2	0,18
3	0,18
4	0,181
5	0,19
6	0,18
7	0,181
8	0,18
9	0,18
10	0,18
11	0,191
12	0,17

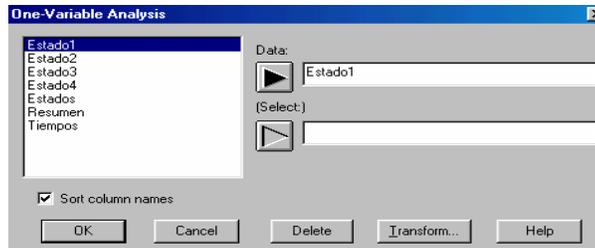
A red box highlights the value 0,19 in the table, with a red arrow pointing to a text box labeled 'Tiempo en segundos'.

2.2. Análisis univariante de la variable

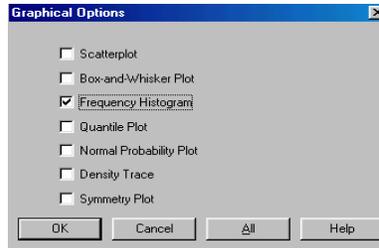
Vamos a comenzar analizando la variable tiempo de ejecución de un bucle en el Estado 1. El análisis gráfico y numérico de esta variable se hace en: Describe / Numeric Data / One-Variable Análisis



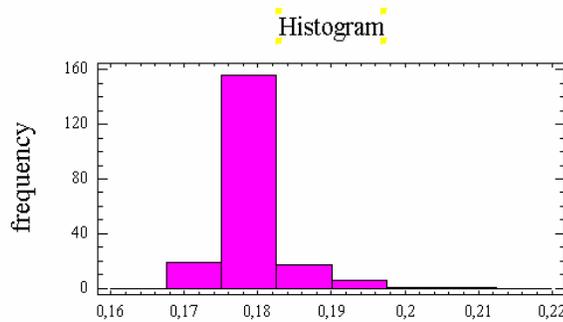
Seleccionamos la variable Estado 1



Para hacer el histograma de la variable seleccionamos el botón de opciones gráficas  y aparece la ventana:

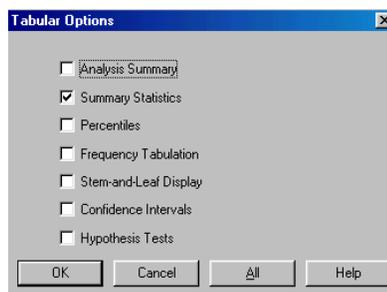


Marcamos la opción Frequency Histogram para que nos muestre el histograma



Podemos observar en el histograma que se trata de una distribución con una sola concentración en torno a 0,18 segundos y hay una asimetría positiva. Además, la variable parece mucho más apuntada que una campana. No parece que esta variable sea muy normal.

Para obtener la información necesaria para nuestros datos, seleccionamos el botón TABULAR OPTIONS  y aparece la ventana



Marcamos la opción SUMMARY STATISTICS para obtener un resumen estadístico. La información que se muestra es la siguiente

```
Summary Statistics for Estado1  
  
Count = 200  
Average = 0,180715  
Median = 0,18  
Mode = 0,18  
Variance = 0,0000273907  
Standard deviation = 0,00523362  
Standard error = 0,000370072  
Minimum = 0,17  
Maximum = 0,21  
Range = 0,04  
Skewness = 1,0937  
Stnd. skewness = 6,31445  
Kurtosis = 5,87382  
Stnd. kurtosis = 16,9563  
Coeff. of variation = 2,89606%
```

Esta tabla muestra el resumen estadístico para la variable tiempo de ejecución de un Bucle en Estado 1. Incluye las medidas de tendencia central, medidas de variabilidad y medidas de forma. Entre los valores obtenidos se encuentran la media y varianza muestrales que son estimaciones puntuales de la media y la varianza poblacionales. Las estimaciones 'puntuales' son

$$\hat{\mu} = 0.1807$$

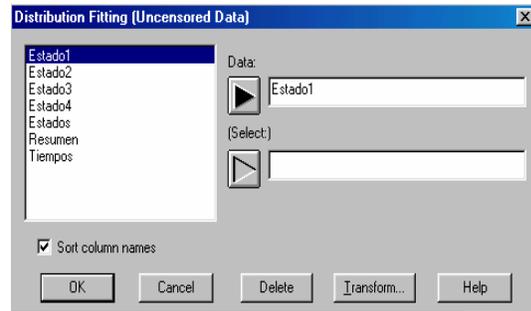
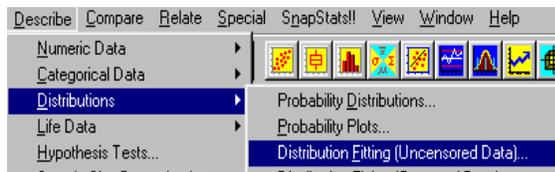
$$\hat{\sigma}^2 = 0.000274$$

Nuestro objetivo es estimar los parámetros utilizando intervalos de confianza.

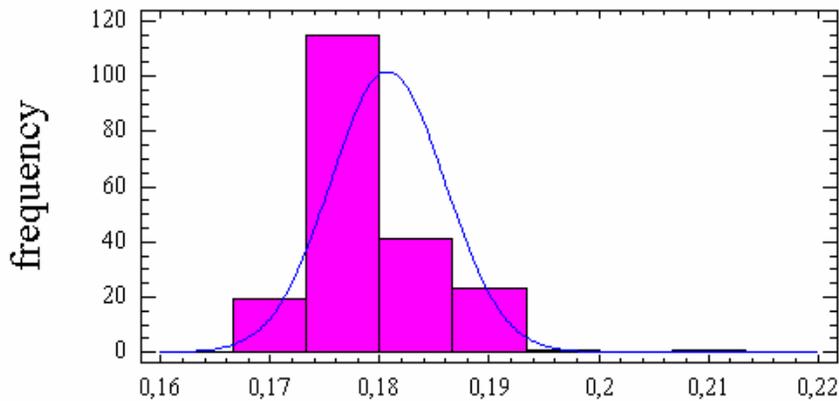
2.3. Análisis de la normalidad de la variable

Para poder construir intervalos de confianza nos interesa saber si nuestra variable es o no normal (¿por qué?). Podemos observar que el histograma anterior no tiene similitud con la función de densidad de una distribución normal. Además, los coeficientes de asimetría (Skewness) y curtosis (Kurtosis) son muy elevados, indicando que la variable tiene asimetría positiva y un apuntamiento más acusado que la normal. Nuestra variable parece que se aleja mucho de la normal.

Para terminar el ejercicio de comparación de nuestros datos con la distribución normal, realizamos el test de la chi-cuadrado. Vamos a ver si los datos permiten tratar a nuestra población como una normal con el test de la Chi-cuadrado.



Histogram for Estado1



Goodness-of-Fit Tests for Estadol

Chi-Square Test

	Lower Limit	Upper Limit	Observed Frequency	Expected Frequency	Chi-Square
at or below	0,174327	0,174327	19	22,22	0,47
	0,174327	0,176713	0	22,22	22,22
	0,176713	0,178461	0	22,22	22,22
	0,178461	0,179984	0	22,22	22,22
	0,179984	0,181446	156	22,22	805,28
	0,181446	0,182969	0	22,22	22,22
	0,182969	0,184717	0	22,22	22,22
	0,184717	0,187103	0	22,22	22,22
above	0,187103		25	22,22	0,35

Chi-Square = 939,429 with 6 d.f. P-Value = 0,0

Como vemos que el histograma con la curva superpuesta no tiene un buen ajuste y dado que el p-valor obtenido es inferior a 0.05, utilizando el nivel de significación habitual del 5% podemos rechazar que la variable tiempo que tarda el programa en el Estado1 proceda de una distribución normal.

Como estamos ante una muestra grande, tenemos que, por el teorema central del límite la media muestral, seguirá una distribución normal independientemente de cómo sea la distribución de la variable. Por lo tanto,

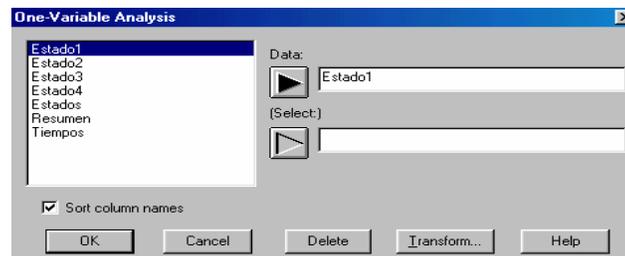
aunque la variable Estado1 no sea normal, podemos calcular el intervalo de confianza para la media, pero no para la desviación típica.

2.3. Intervalos de confianza

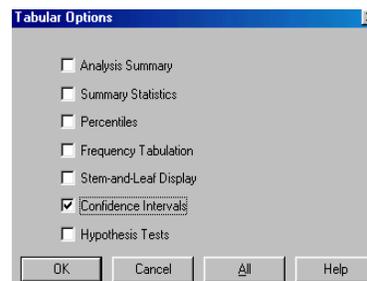
Para obtener el intervalo de confianza para μ seleccionamos Describe / Numeric Data / One-Variable Análisis



Seleccionamos Estado 1



Con el botón TABULAR OPTIONS  aparece la ventana



Marcamos la opción Confidence Intervals y obtenemos la siguiente información:

Confidence Intervals for Estadol

95,0% confidence interval for mean: 0,180715 +/- 0,000729768 [0,179985;0,181445]

95,0% confidence interval for standard deviation: [0,00476603;0,00580376]

Por tanto, el tiempo medio de ejecución de un programa de Matlab en el Estado 1 será un valor que estará, con una confianza del 95% entre 0.17998 y 0.18144 segundos.