

# Haciendo estadística con SPAC-FL<sup>©</sup> y Minitab<sup>©</sup>

Mayo de 2011

Ing. Fernando Tomati  
Director de Contenidos  
HLTnetwork S.A.

---

## HACIENDO ESTADÍSTICA CON SPAC-FL Y MINITAB

El uso de las estadísticas para el análisis de datos disponibles en los procesos industriales o de servicios sigue siendo una materia pendiente en muchas organizaciones. Esto ocurre a pesar de contar hoy día con poderosos softwares que facilitan este tipo de análisis. El problema es “acercar” tanto a profesionales como a operadores y empleados al conocimiento de estas herramientas de una manera didáctica y sencilla.

HLTnetwork cuenta con una amplia experiencia en lograr este tipo de “acercamiento”. Edward Deming (\*), acérrimo defensor del uso de las estadísticas para la toma de decisiones, decía: “no es que las organizaciones no cuenten con información, el problema es no saber qué hacer con ella. Muchas las apilan y, cuando esa pila es muy grande, desechan la mitad de abajo y siguen coleccionando datos”.

Un ingeniero controla un proceso industrial, que genera a diario muchos lotes de un producto (Población de lotes). Para cada lote se mide una característica de calidad, obteniéndose una gran cantidad de resultados numéricos (Población de datos). El ingeniero realiza esta tarea porque a través de los datos numéricos obtenidos se puede evaluar el comportamiento del proceso, que es lo que realmente le interesa. Lo mismo podría decirse de un responsable de la provisión de un servicio o de un administrador de un emprendimiento de tipo comercial.

Los llamados “Green Belts y Black Belts Lean Six Sigma”, por ejemplo, deben conocer los conceptos y saber interpretar los estadísticos. No será necesario que realicen los cálculos por sí mismos (quita tiempo y aumenta las probabilidades de error humano) sino que utilizarán un software que hará los cálculos por ellos. Afortunadamente el progreso logrado por la informática, ha permitido el desarrollo de softwares que hicieron de la estadística una ciencia “alcanzable”.

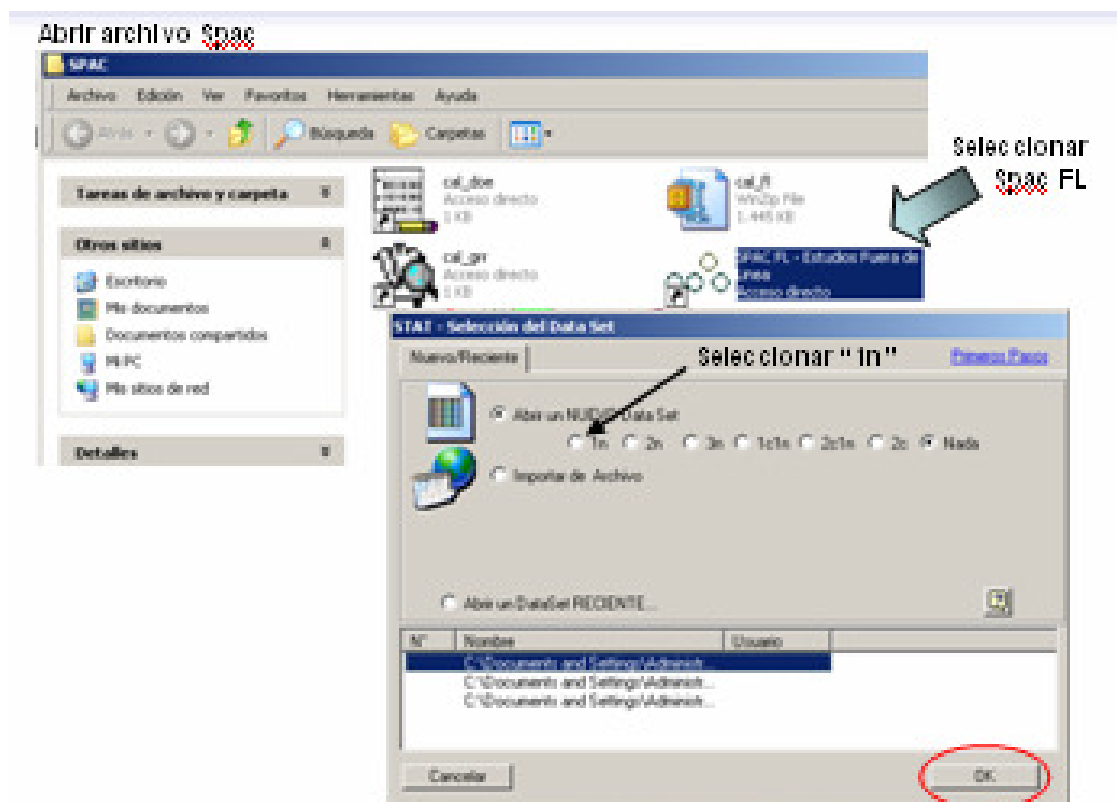
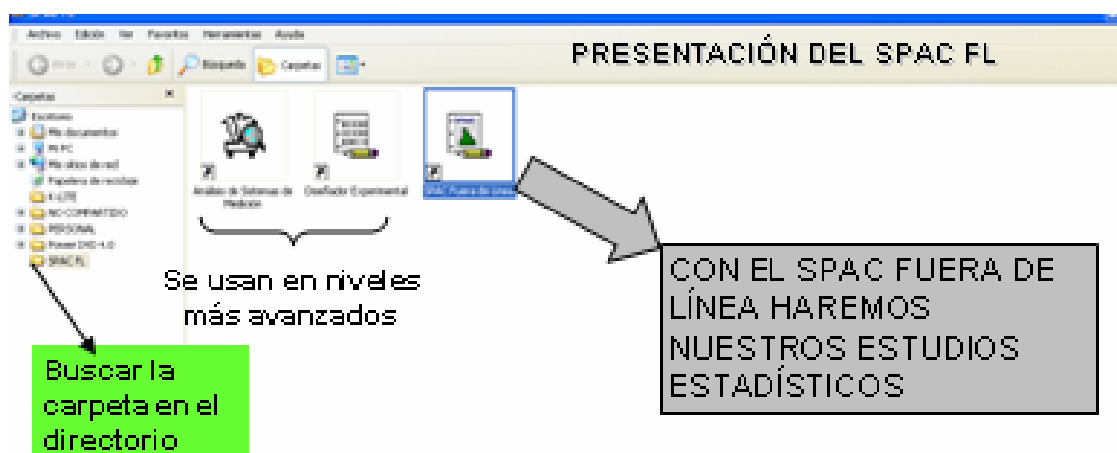
En esta nota presentaremos dos de ellos. Uno de desarrollo argentino (Spac FL) y otro estadounidense (Minitab). Desarrollaremos conceptos básicos para su correcto manejo y daremos algunos ejemplos prácticos.

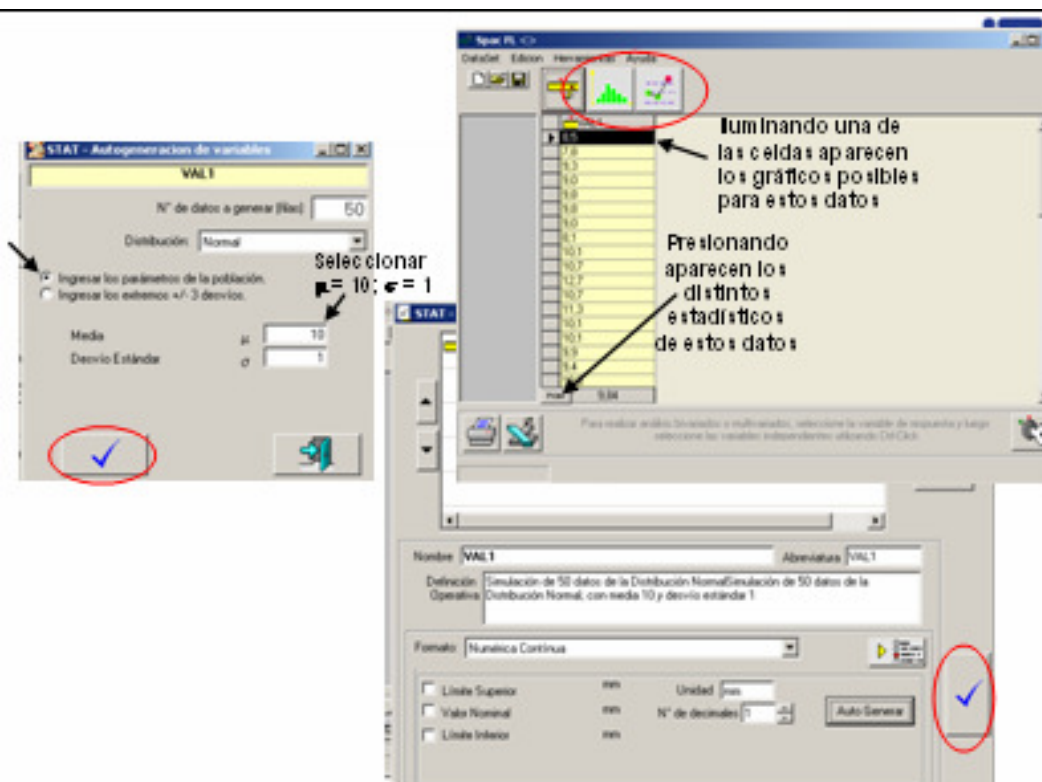
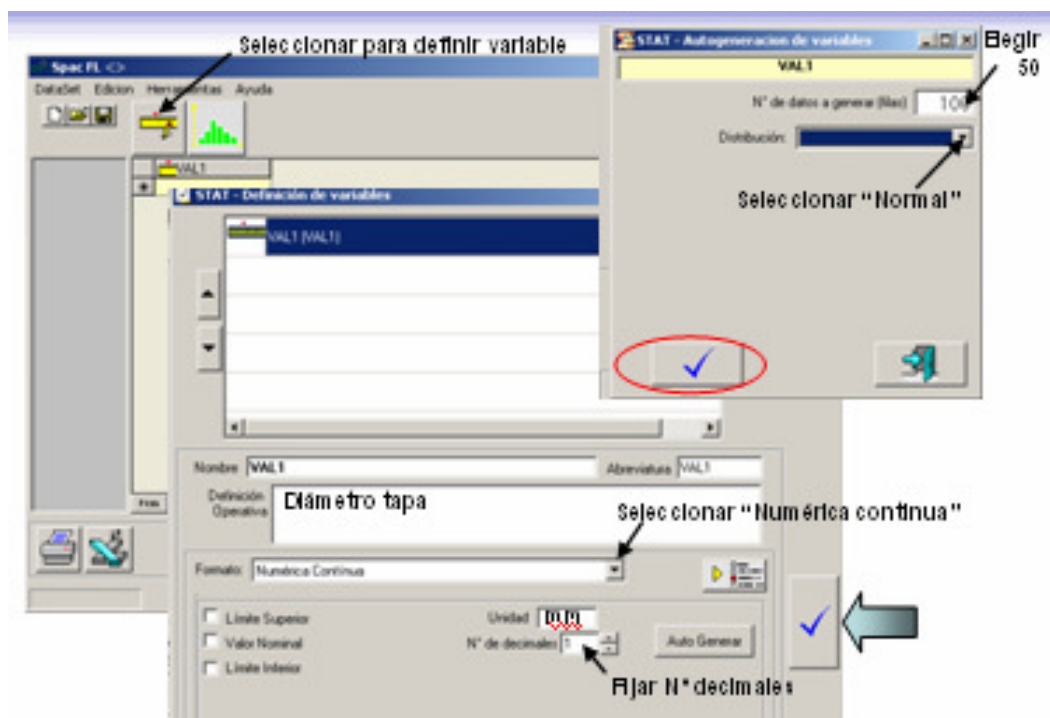
*(\*) William Edwards Deming (14 de octubre de 1900 - 20 de diciembre de 1993). Estadístico estadounidense, profesor universitario, autor de textos, consultor y difusor del concepto de calidad total. Su nombre está asociado al desarrollo y crecimiento de Japón después de la Segunda Guerra Mundial.*

## CURVA DE DISTRIBUCIÓN DE FRECUENCIAS - EJERCICIO

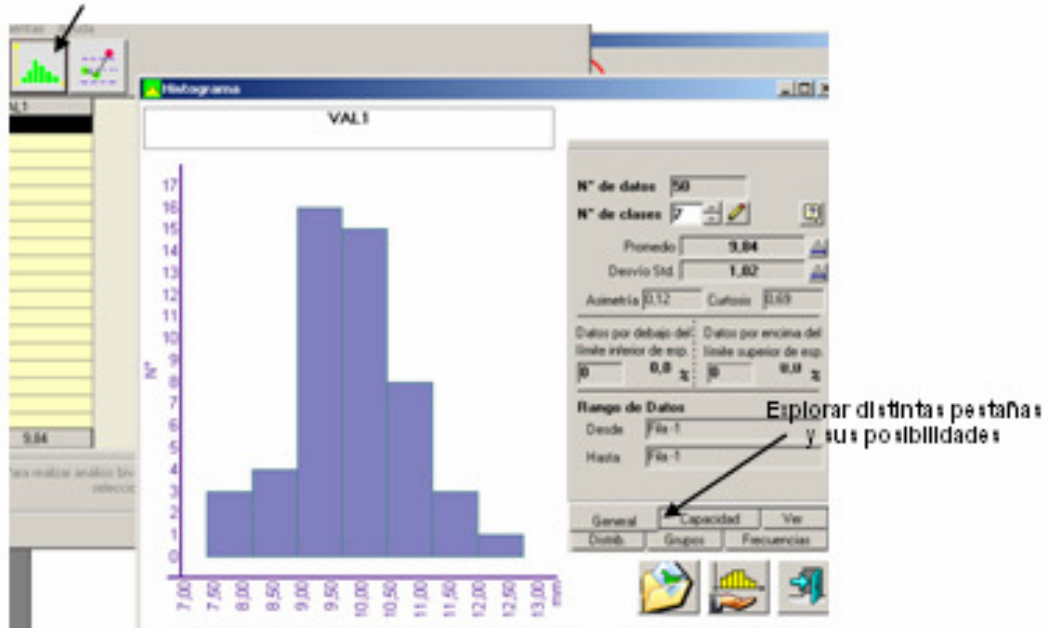
**Objetivo:** Realizar un estudio de distribución normal de datos y su análisis usando **SPAC FL**.

- Utilizando el Spac FL, generar una serie de 50 datos al azar, distribuidos en forma normal.
- Generar el histograma de la distribución e interpretarlo.

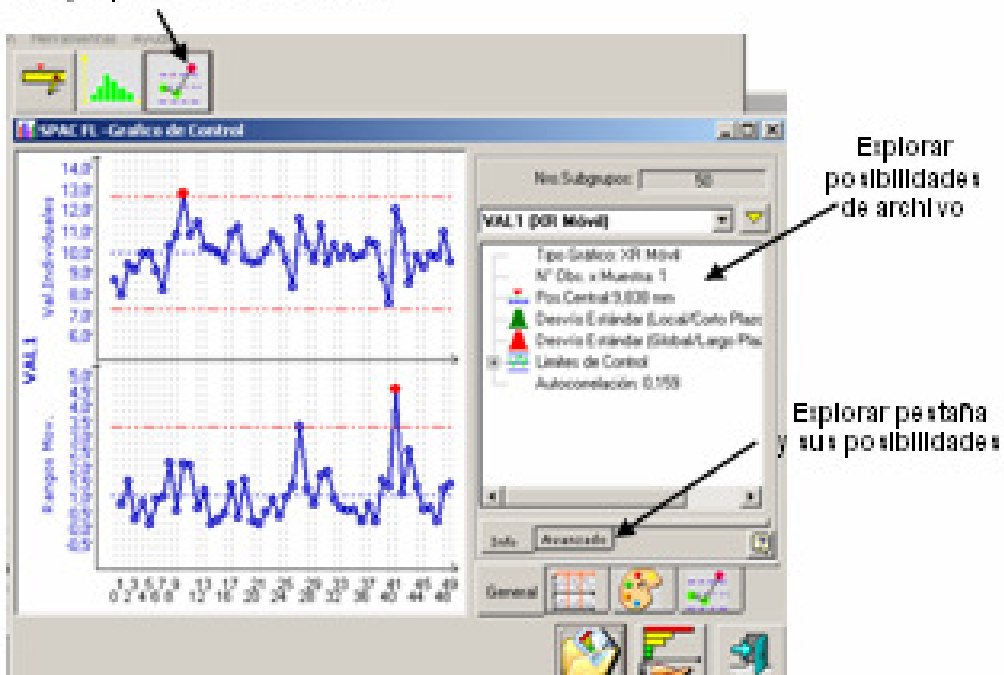




Por ejemplo: Histograma

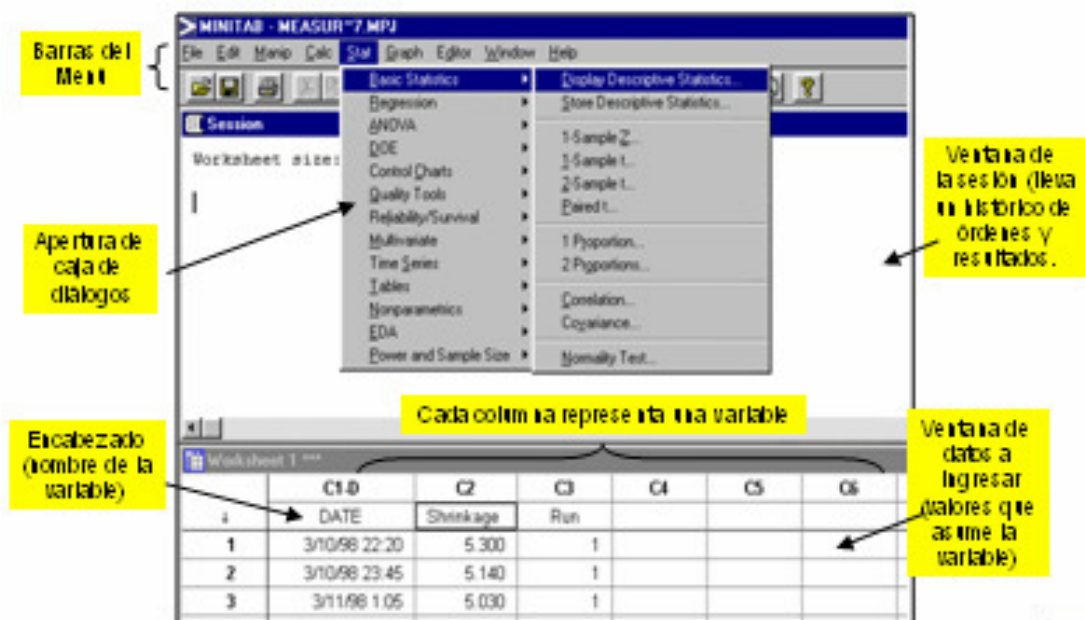


Por ejemplo: Grafico de control



Haremos ahora igual aplicación a un ejemplo utilizando el Software **MINITAB**:

## Uso del Minitab®



**Barra del Menu**

**Apertura de caja de diálogos**

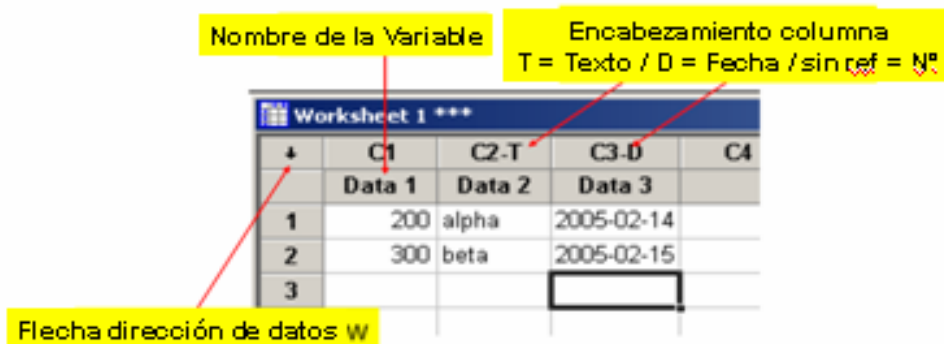
**Encabezado (nombre de la variable)**

**Cada columna representa una variable**

**Ventana de la sesión (lista de historial de órdenes y resultados).**

**Ventana de datos a ingresar (valores que asume la variable)**

	C1-D	C2	C3	C4	C5	C6
1	DATE	Shrinkage	Run			
1	3/10/98 22:20	5.300	1			
2	3/10/98 23:45	5.140	1			
3	3/11/98 1:05	5.030	1			



**Nombre de la Variable**

**Encabezamiento columna  
T = Texto / D = Fecha / sin ref = N°**

**Fecha dirección de datos W**

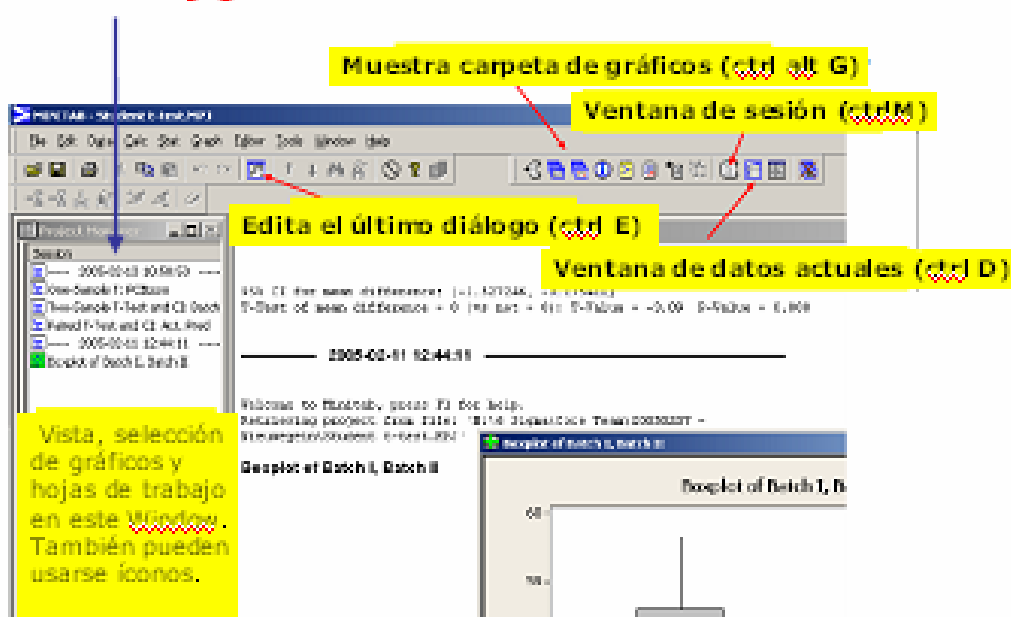
	C1	C2-T	C3-D	C4
	Data 1	Data 2	Data 3	
1	200	alpha	2005-02-14	
2	300	beta	2005-02-15	
3				

Minitab cuenta con un **Project manager**, o manejador de Proyectos, que contiene:

- ✓ **Session folder.** Lista todas las salidas por comandos de la ventana de sesión y todas las salidas gráficas.
- ✓ **History folder.** Lista los comandos que fueron usados en la sesión. Permite volver a ejecutar los comandos copiándolos de *history folder* y pegándolos en el *Comand Line Editor*.
- ✓ **Graph folder.** Lista y visualiza todos los gráficos, permitiendo su manejo, por ejemplo renombrar, eliminar, añadir gráficos al *ReportPad*.

- ✓ **ReportPad folder.** Se utiliza para crear, ordenar y editar informes del Proyecto.
- ✓ **Related Documents folder.** Facilita referencias para un rápido acceso a Proyectos relacionados que no son archivos Minitab, por ejemplo documentos *URL* de Internet.
- ✓ **Worksheet folder.** Muestra un resumen de las columnas, constantes almacenadas, matrices y diseños existentes en la *worksheet* en uso.

### Para datos y gráficos Windows



**Objetivo:** Realizar un estudio de distribución normal de datos y su análisis usando **Minitab**.

- Utilizando el Minitab, generar una serie de datos al azar, distribuidos en forma normal.

Comandos en Minitab:  
 Calc > Random Data > Normal

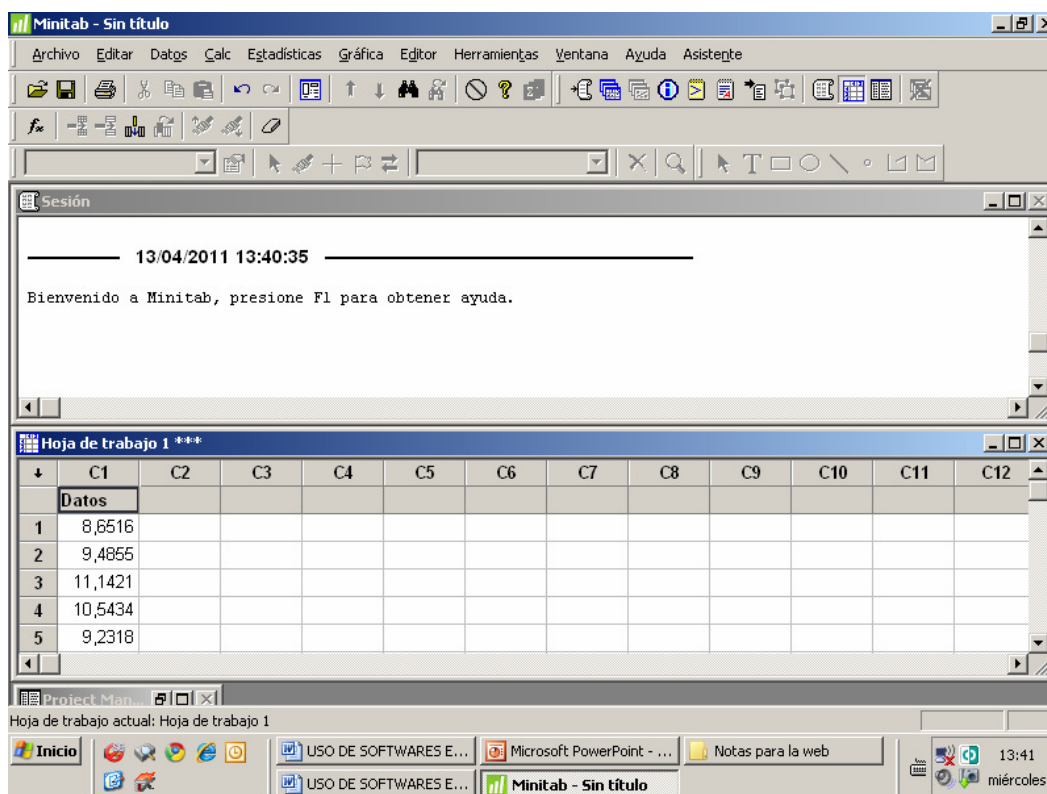
```
Generate 50 rows of data
Store in column(s) C1
Mean 150
Standard deviation 10
```

Stat > Basic Statistics > Display Descriptive Statistics

```
Variables C1
Graphs Histogram of data
Boxplot of data
Graphical summary
```

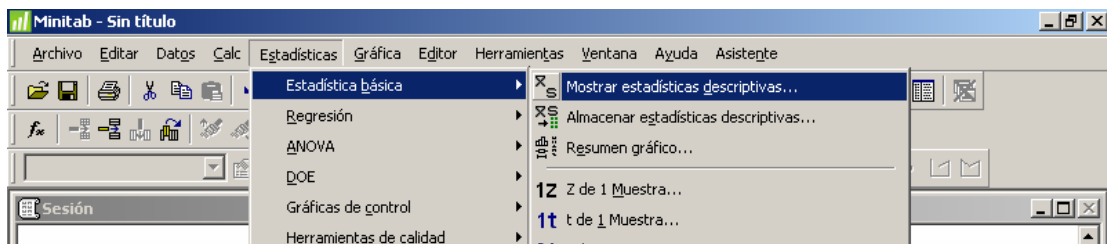
Nota: Si usted quiere eliminar un dato ingresado o carece del mismo para una determinada muestra, debe colocar \* pues de lo contrario Minitab interpretará “cero” si deja la celda en blanco.

Vista de datos generados con un software Minitab en su versión en español:

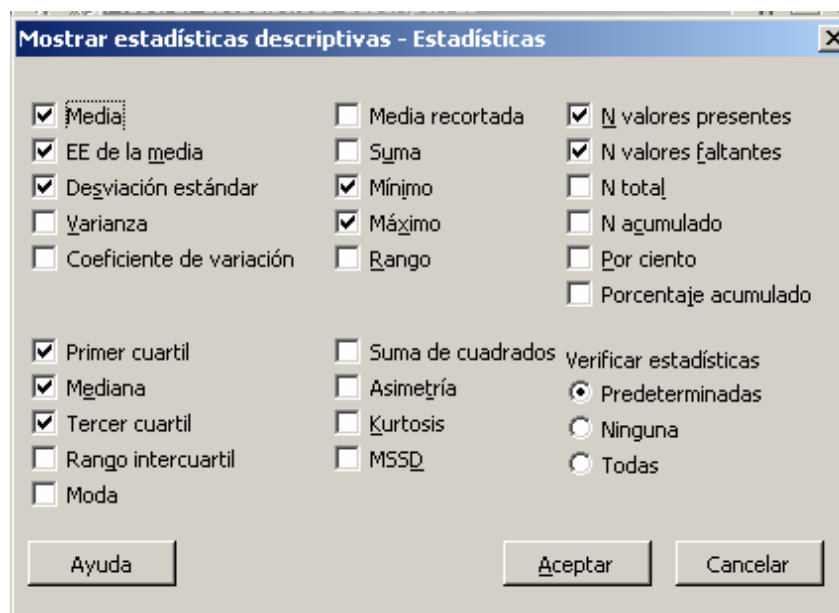
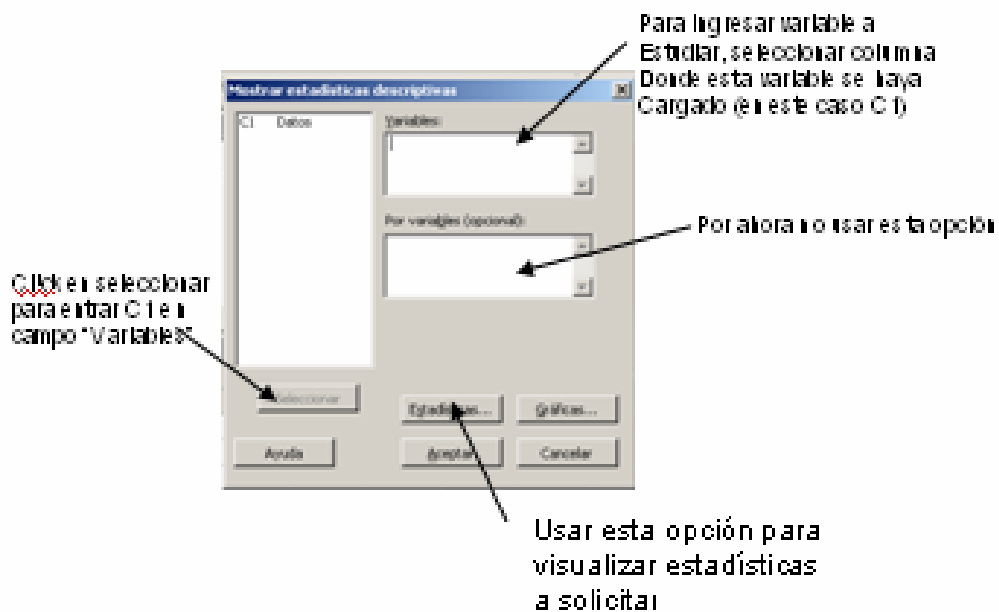


Comandos a utilizar: Estadísticas>Estadística básica>Mostrar estadísticas descriptivas:



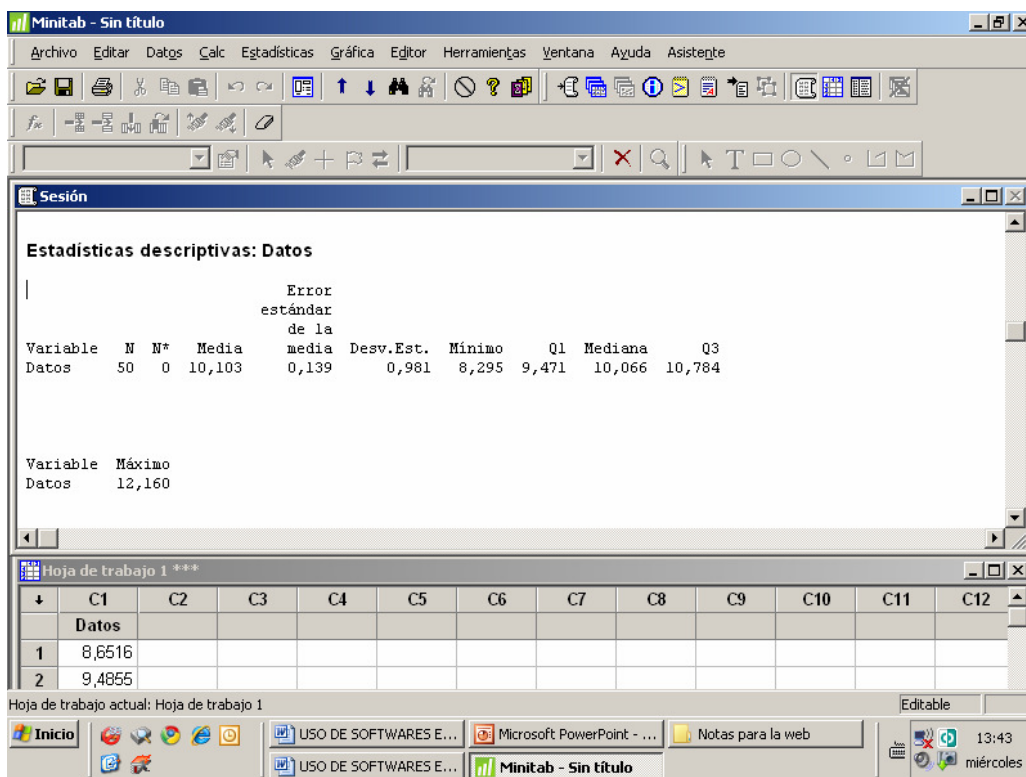


Aparecerá la siguiente caja de diálogo:



Por “default” aparecerán pre-seleccionados los estadísticos tildados (pueden suprimirse o adicionarse opciones).

Salida en el campo de “Sesión” al “Aceptar”:



The screenshot shows the Minitab interface with a window titled 'Sesión' displaying the following descriptive statistics for the variable 'Datos':

Variable	N	N*	Media	Error estándar de la media	Desv.Est.	Mínimo	Q1	Mediana	Q3
Datos	50	0	10,103	0,139	0,981	8,295	9,471	10,066	10,784

Below this, the maximum value is shown:

Variable	Máximo
Datos	12,160

The bottom window shows a worksheet with columns C1 to C12. The first two rows of data are:

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
Datos												
1	8,6516											
2	9,4855											

### Interpretación de los datos:

N y N\* proveen información sobre el número de observaciones en un conjunto de datos, donde:

- N es el número de valores presentes.
- N\* es el número de valores faltantes. Si no existen valores faltantes, no se muestra esta estadística.

Las demás estadísticas mostradas se basan sólo en los valores presentes.

La media (también denominada promedio) es una medida de dónde se encuentra el centro de la distribución de los datos. Es simplemente la suma de todas las observaciones dividida por el número de observaciones.

Los valores extremos ejercen una fuerte influencia sobre la media.

La mediana (denominada también el segundo cuartil o el percentil 50) es el punto medio del conjunto de datos: la mitad de las observaciones se ubica por encima de éste y la otra mitad se ubica por debajo de éste. Se determina al clasificar los datos y hallar el número de observación  $[N + 1] / 2$ . Si existe un número par de observaciones, la mediana se extrapola como el valor intermedio entre el valor de los números de observación  $N / 2$  y  $[N / 2] + 1$ .

La mediana es menos sensible a los valores extremos que la media. Por lo tanto, la mediana se utiliza con más frecuencia en lugar de la media cuando los datos contienen valores atípicos o son asimétricos.

La desviación estándar (Desv.Est.) es una medida de cuánto se desvían las observaciones en una muestra con respecto a la media. Es análoga a una distancia promedio (independiente de la dirección) con respecto a la media. La desviación estándar es la medida de dispersión más comúnmente notificada. Sirve también como un estimado de la dispersión en la población más amplia de la cual se toma una muestra. Al igual que la media, la desviación estándar es muy sensible a los valores extremos.

Si los datos están normalmente distribuidos (en nuestro ejemplo lo están porque lo solicitamos así cuando generamos los datos), la desviación estándar y la media se pueden utilizar para determinar qué proporción de las observaciones se ubica dentro de cualquier rango dado de valores. Por ejemplo, 95% de los valores en una distribución normal se ubica dentro de + 1.96 desviaciones estándar de la media.

El error estándar de la media no se utiliza con frecuencia como una estadística descriptiva, pero es importante en las pruebas de hipótesis. Es un estimado de la dispersión que usted observaría en la distribución de medias de la muestra, si continuó tomando muestras de la población las cuales tenían el mismo tamaño.

El error estándar de la media es la desviación estándar dividida por la Raíz cuadrada de N.

Una de las maneras más sencillas de evaluar la dispersión en sus datos consiste en comparar el mínimo y el máximo. El mínimo es el valor más pequeño en un conjunto de datos y el máximo es el valor más grande.

Mínimo y máximo se utilizan para calcular el rango, que es una estadística utilizada para describir la dispersión en los conjuntos reducidos de datos. El rango es simplemente el máximo menos el mínimo. Tenga en cuenta que el rango es muy sensible a valores extremos.

Exactamente 25% de sus datos son menores que el primer cuartil (Q1, también denominado percentil 25). Es igual al valor de los datos en la posición  $(N + 1) / 4$ . Si este número de posición no es un entero, Minitab extrapola entre las dos observaciones en cualquier lado de esa posición.

Exactamente el 75% de sus datos es menor que el tercer cuartil (Q3, también denominado percentil 75). Es igual al valor de los datos en la posición  $3(N + 1) / 4$ . Si este número de posición no es un entero, Minitab extrapola entre las dos observaciones en cualquiera lado de esa posición.

Q1 y Q3 se utilizan frecuentemente para calcular el rango intercuartil (IQR), que es otro estadístico utilizado para describir la dispersión. El IQR es el rango del 50% central de los valores y se calcula con la fórmula  $Q3 - Q1$ . El IQR es relativamente insensible a los valores extremos.

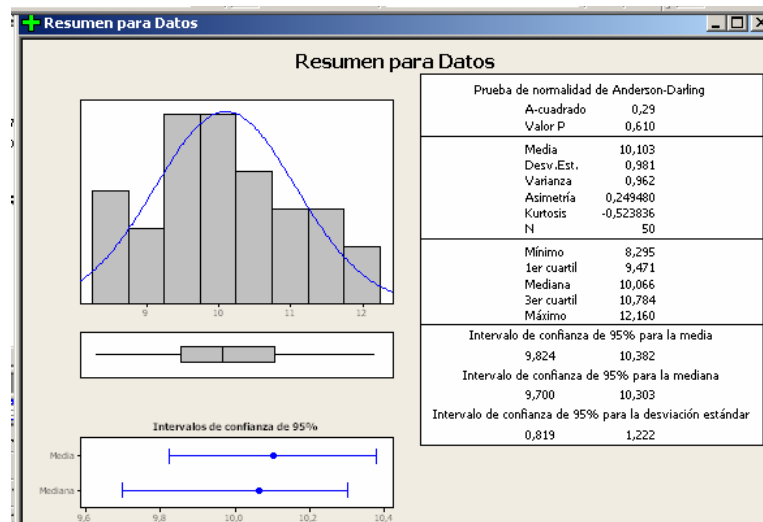
## HISTOGRAMA DE DATOS

Un histograma muestra datos que se han resumido en intervalos. Se puede utilizar para evaluar la simetría o la asimetría de los datos. Para construir un histograma, el eje horizontal se divide en intervalos iguales, y se dibuja una barra vertical en cada intervalo para representar su frecuencia (el número de valores que se ubican dentro del intervalo).

Para visualizar el histograma de los datos de nuestro ejemplo, deberemos utilizar los comandos: Estadísticas>Estadística básica>Resumen gráfico

Completar, como antes la siguiente caja de diálogo:

Por “default” aparecerá “Nivel de confianza” = 95,0 %. Aceptar y se verá:



### Interpretación:

Usted puede utilizar un histograma de los datos sobrepuestos con una curva normal para evaluar la normalidad de sus datos. Una distribución normal es simétrica y tiene forma de campana, como lo indica la curva. (Observe que la curva se trunca donde traspasa los límites de la gráfica). Comúnmente es difícil evaluar la normalidad con muestras pequeñas. La determinación de normalidad la da el estadístico de Anderson-Darling. Un “Valor P” mayor a 0,05 indicará que la distribución de los datos es normal (como en este caso, donde Valor P = 0,610).

Para profundizar sobre estos temas será necesario recurrir al comando de “Ayuda” de los softwares, o a expertos entrenados en su uso, como lo son los llamados “Green Belts” o “Black Belts” de la metodología “Lean Six Sigma”.