ANÁLISIS ESTADÍSTICO

INTERVALOS DE CONFIANZA

Jorge Fallas
jfallas56@gmail.com

2010

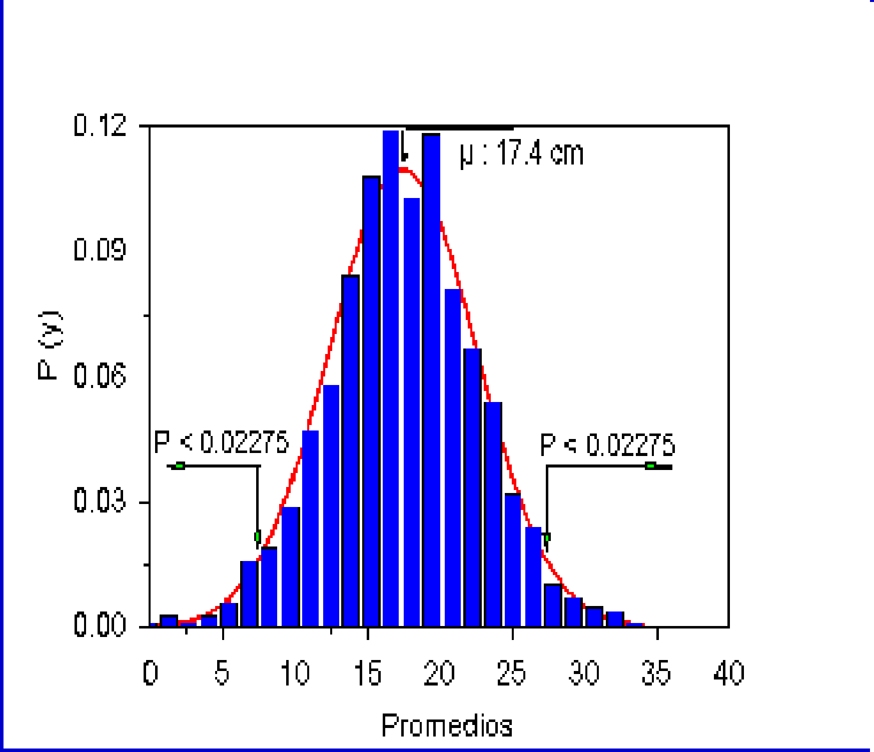
Temario

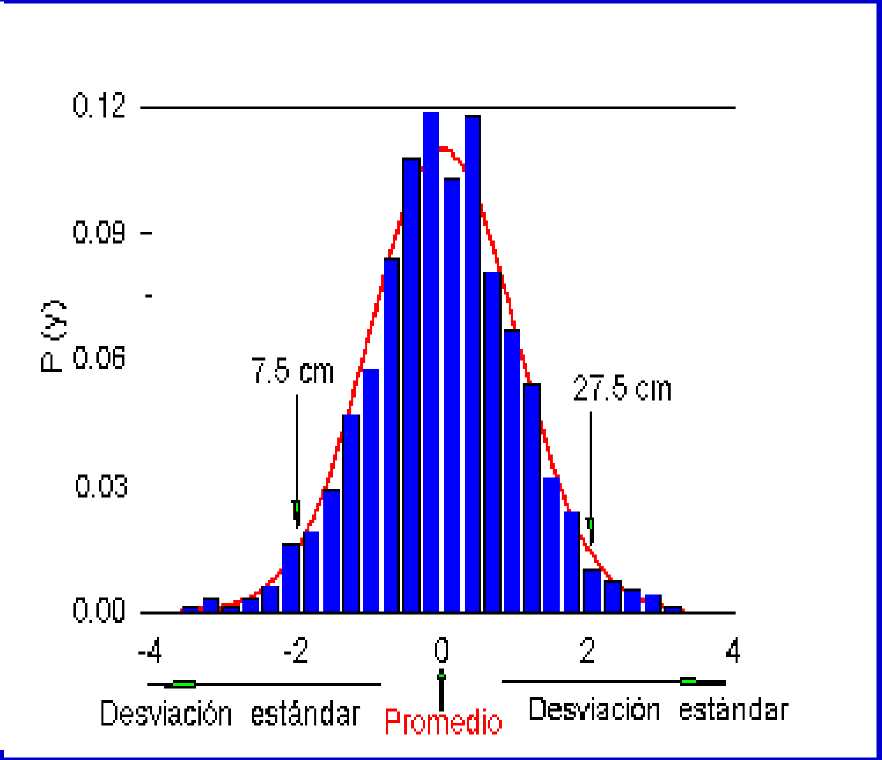
- Estimación puntual
- Distribución muestral de los estimadores
- Estimación por intervalos
- Intervalo para la media
- Intervalo para la varianza y desviación Estándar
- Nivel de confianza
- Uso tablas de Z (distribución normal estandarizada), t de Estudiante, F y Chi-cua-drado
- Pruebas de normalidad
- Herramientas
  - Software
    - XLSTats
    - Instat
    - Remuestreo

Introducción

- **PARÁMETROS:** son valores desconocidos propios de una población
- **Trabajos empíricos**: sólo es posible obtener una estimación puntual del parámetro (Ej. la media)
- Es necesario indicar el grado de variabilidad muestral del estimador. Esto puede lograrse calculando un intervalo de confianza para el estimador
- Cálculo intervalo de confianza para la media, la varianza y la desviación estándar
- Nivel de confianza (1-α)*100
- Supuestos análisis paramétrico: normalidad
- Opciones de análisis
  - Pruebas paramétricas
  - Remuestreo (Ej. Resampling Stats)
    - http://www.uvm.edu/~dhowell/StatPages/Resampling/Resampling.html

Datos experimentales y distribuciones teóricas





Concepto de distribución de frecuencia y desviación estándar de la media. Estas gráficas ilustran la distribución de frecuencia esperada para una población con un diámetro medio de 17.4cm y una varianza de 25 cm2. Cualquier parcela con un diámetro medio <u>superior a 27.5 cm</u> ó <u>inferior a 7.5 cm</u> se consideraría un <u>evento raro</u> dado la distribución de referencia.

IC media de población con distribución normal y con varianza conocida

- SUPUESTO: La muestra proviene de una población normal con varianza conocida

$$P\left(\bar{X} - Z_{1-\alpha/2} * \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{1-\alpha/2} * \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

***Media* es una variable aleatoria que posee una distribución muestral.**

Amplitud del intervalo de confianza

$$2 * Z_{1-\alpha/2} * \frac{\sigma}{\sqrt{n}}$$

Se utilizan valores estandarizados de la distribución normal (tabla valores Z)

Intervalo es simétrico

Interpretación IC

- **μ pertenece ó no al intervalo y no es posible por lo tanto asignar una probabilidad de 1- α a dicha afirmación**
- **La media es una variable aleatoria y por lo tanto variará de muestra en muestra.**
- **La probabilidad de 1- α se refiere a dicha variación**
- **Para resolver esta limitante se asignará a la media el sufijo "o"; el cual indica que se trata de uno los tantos valores que puede tomar y que por ende el intervalo de confianza calculado es también sólo uno de los tantos posibles que se pueden obtener de la población.**
- **La probabilidad 1- α indica que de los posibles intervalos solo 1- α contendrán μ.**

$$\bar{X}_o \pm Z_{1-(\alpha/2)}\ \sigma/(n)^{1/2}$$

Decimos que tenemos la confianza de que 1- α veces μ estará contenido entre dichos límites.

Interpretación IC

$$C\left(\bar{X} - Z_{1-\alpha/2} * \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{1-\alpha/2} * \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

- Dado que en teoría existe un número infinito de muestras, también existirá un número infinito de intervalos de confianza para μ
- El valor de 100*α representa el porcentaje de veces que en <u>promedio</u> se espera que el intervalo no incluya el valor del parámetro.
- Ejemplo: si α es 0,05 esto significa que de cada 20 intervalos 1 no incluirá a μ. En términos porcentuales significa que 5 de cada 100 intervalos no contendrán a μ.

Una población: síntesis

$$C\left(\bar{X} - Z_{1-\alpha/2} * \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{1-\alpha/2} * \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

SUPUESTOS
Muestreo aleatorio
Población normal
Se desconoce μ y se conoce el valor de σ

El intervalo indica que existe una confianza de (1-α)*100 de que en <u>promedio</u> el intervalo incluya el valor del parámetro

Ej. De 100 Intervalos de Confianza

100*α **no contendrán a μ**

¿Qué afecta la amplitud IC?

- **Nivel de confianza.** La amplitud del nivel de confianza se reduce al aumentar el valor de α. Usualmente no se acepta un nivel de confianza inferior a un 90% (α=0,10); o sea que sólo estamos dispuestos a errar en 10 de cada 100 intervalos que calculemos. **Nivel de Confianza** DEBE FIJARSE ANTES DE REALIZAR ANÁLISIS

- **Valor de σ: L**a variabilidad de la población puede reducirse modificando la población de interés
- **Tamaño de muestra (n): IC** decrece en función de la raíz cuadrada del tamaño de la muestra; por lo tanto para reducir la amplitud del intervalo en un 50% se requiere cuadruplicar el tamaño de la muestra.

IC media de población con distribución normal, media y varianza desconocida

- El intervalo de confianza para μ para un nivel de confianza de 1-α está dado por:

$$C\left(\bar{X} - Z_{1-\alpha/2} * \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{1-\alpha/2} * \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

El factor t 1-(α/2) es el percentil 100 (1- α/2) de la distribución "t" con n-1 grados de libertad. Dado que podemos seleccionar un número infinito de muestras de tamaño n, existen también un número infinito de intervalos de confianza.

AMPLITUD DEL IC

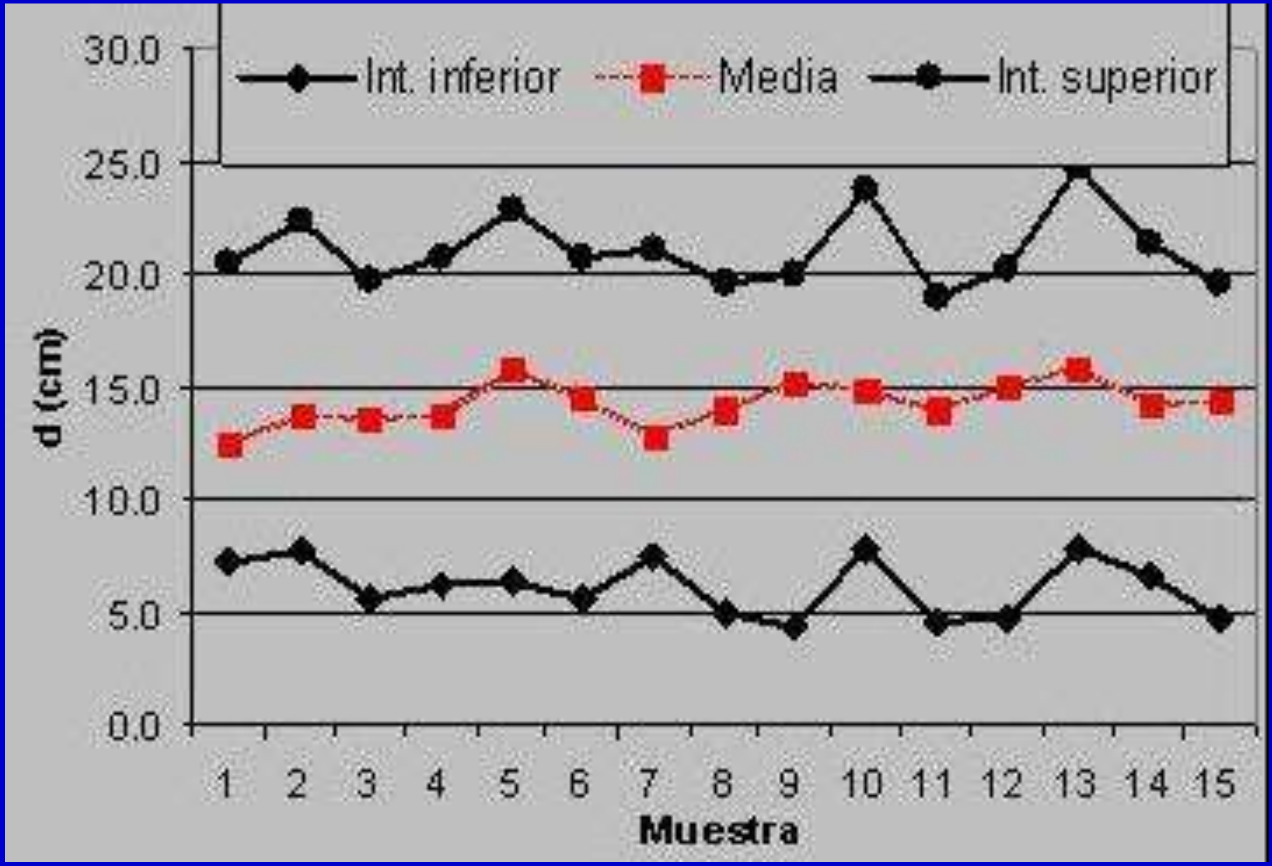
- **Nivel de confianza (1- α)**
- **Valor de la desviación estándar (σ) y**
- **Tamaño de la muestra. (n)**

PROCEDIMIENTO

- Selecciona una muestra aleatoria de tamaño n
- Calcula media y desviación estándar
- Dado que a partir de este momento μ y σ son constantes no podemos decir que μ se encuentra en el intervalo con una probabilidad 1-α; sino que tenemos una confianza de 1- α de que μ estará contenida en el intervalo que hemos calculado

$$\bar{X}o - t\,1\text{-}(\alpha/2) * So / n^{0,5} < \mu < \bar{X}o + t\,1\text{-}(\alpha/2) * So / n^{0,5}$$

Ejemplo

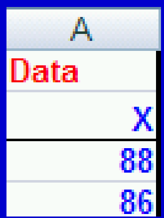


Intervalos de confianza (95%) para 15 muestras de tamaño 10

IC para la media en XLSTats

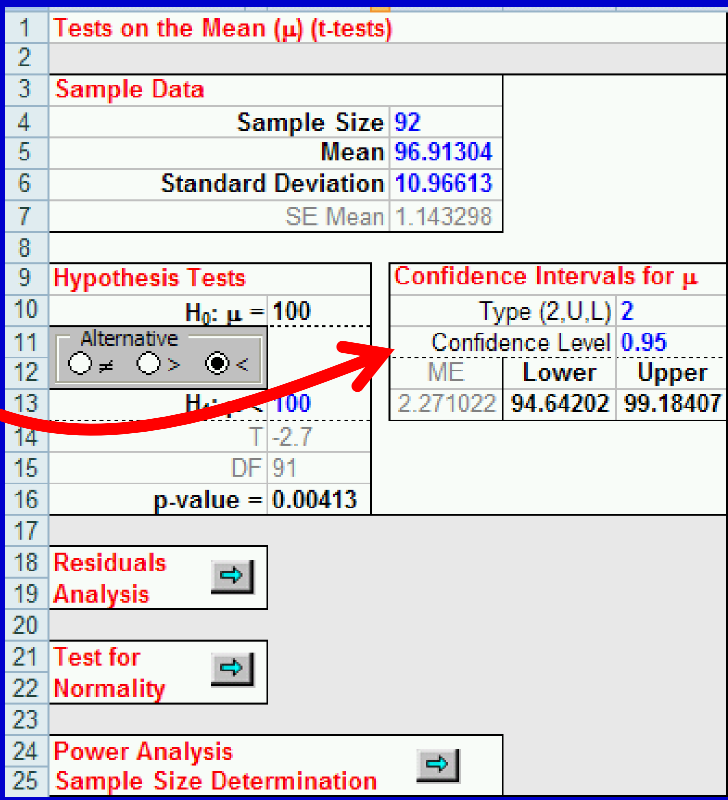
•Seleccione 1Num del menú principal

•Inserte o digite sus datos en la hoja de Excel



•Seleccione la hoja *Test*

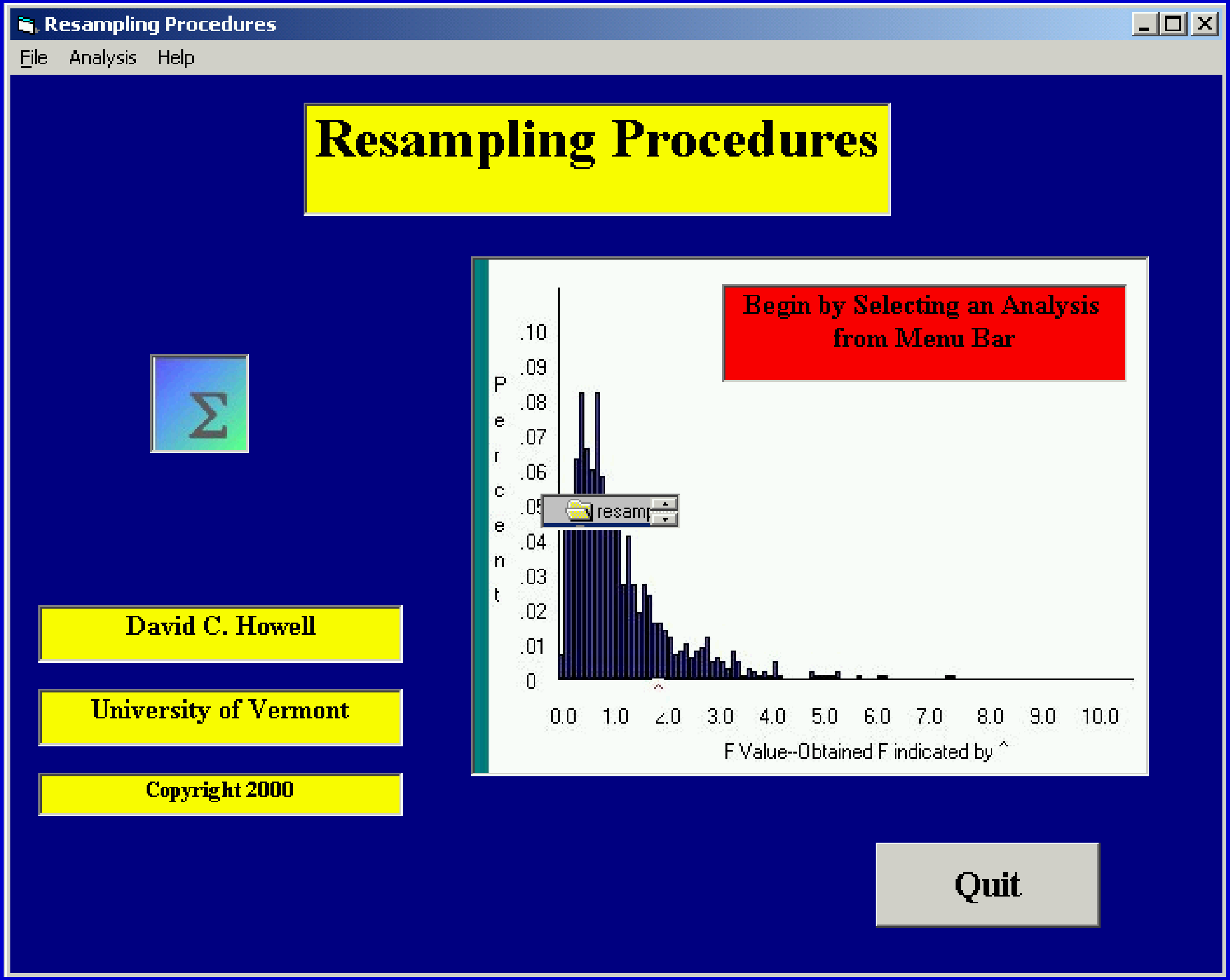
• En la ventana de diálogo usted puede definir el nivel de confianza deseado (Ej. 90%, 95%, 99%)

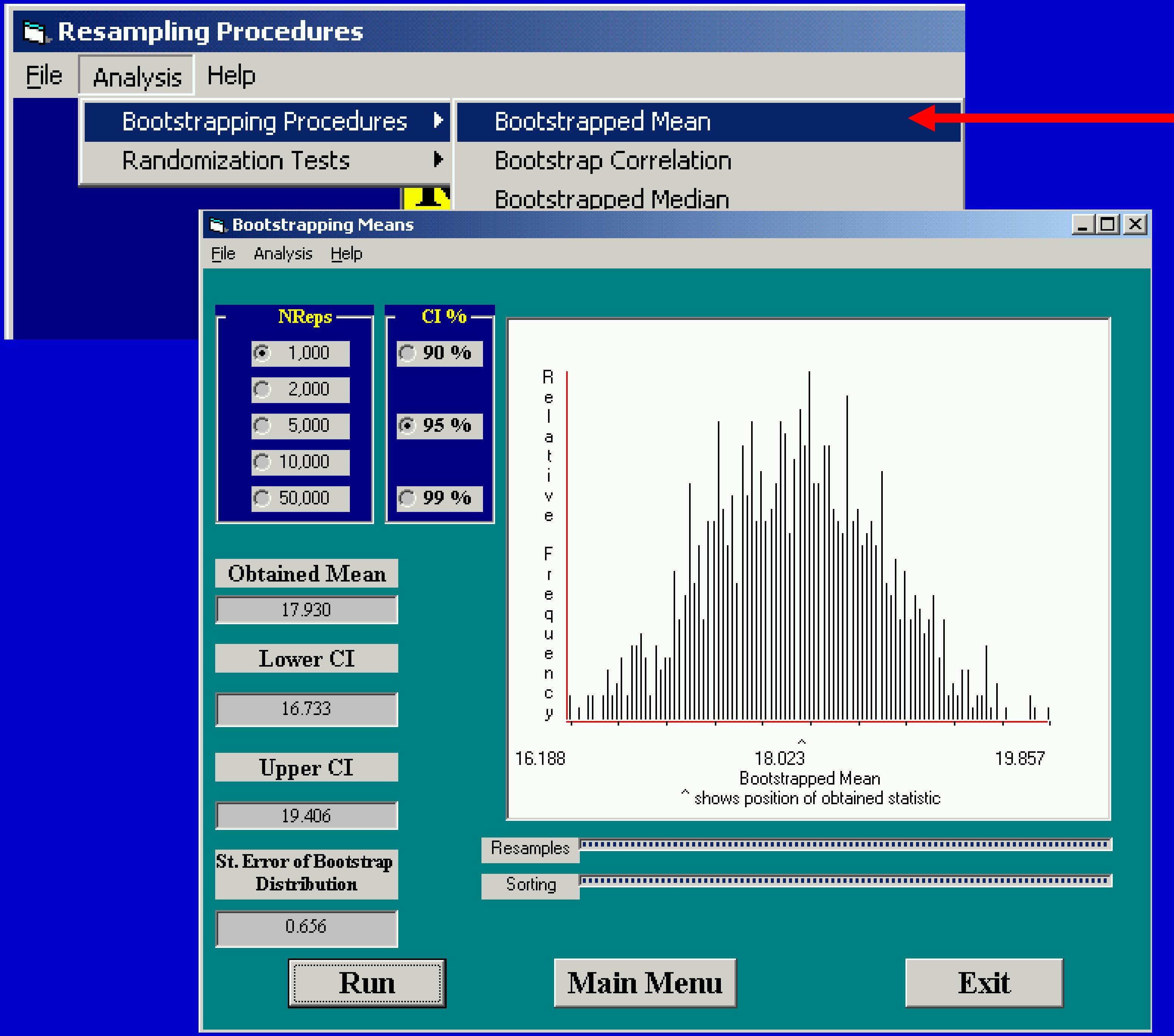


IC 95.0% para μ: 94.64 - 99.18 cm

Se espera que 95 de cada 100 Intervalos contengan μ

Intervalos de confianza en “Resampling Stats”



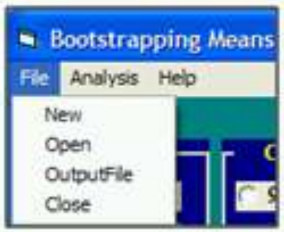


Resampling Procedures
File
Analysis
Help
Bootstrapping Procedures
Randomization Tests
Bootstrapped Mean
Bootstrap Correlation
Bootstrapped Median
Bootstrapping Means
File Analysis Help
NReps
1,000
2,000
5,000
10,000
50,000
CI %
90 %
95 %
99 %
Obtained Mean
17.930
Lower CI
16.733
Upper CI
19.406
St. Error of Bootstrap Distribution
0.656
Relative Frequency
16.188
18.023
19.857
Bootstrapped Mean
^ shows position of obtained statistic
Resamples
Sorting
Run
Main Menu
Exit

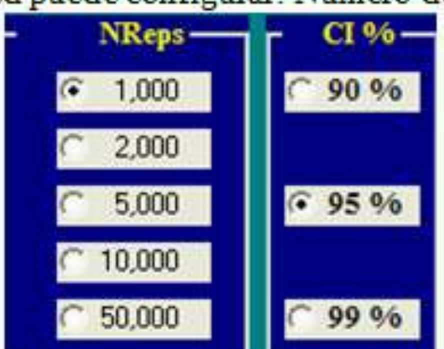
Intervalos de confianza en “Resampling Stats”

1. Haga un doble clic sobre el archivo Resampling.exe
2. Seleccione del menú de análisis la opción: *Bootstrapped Mean*.

3. Habla el archivo recién creado. *File, open*.



4. Usted puede configurar: Numero de repeticiones (NReps) y el nivel de confianza (CI%)



5. Haga un clic sobre **Run** para realizar los cálculos.

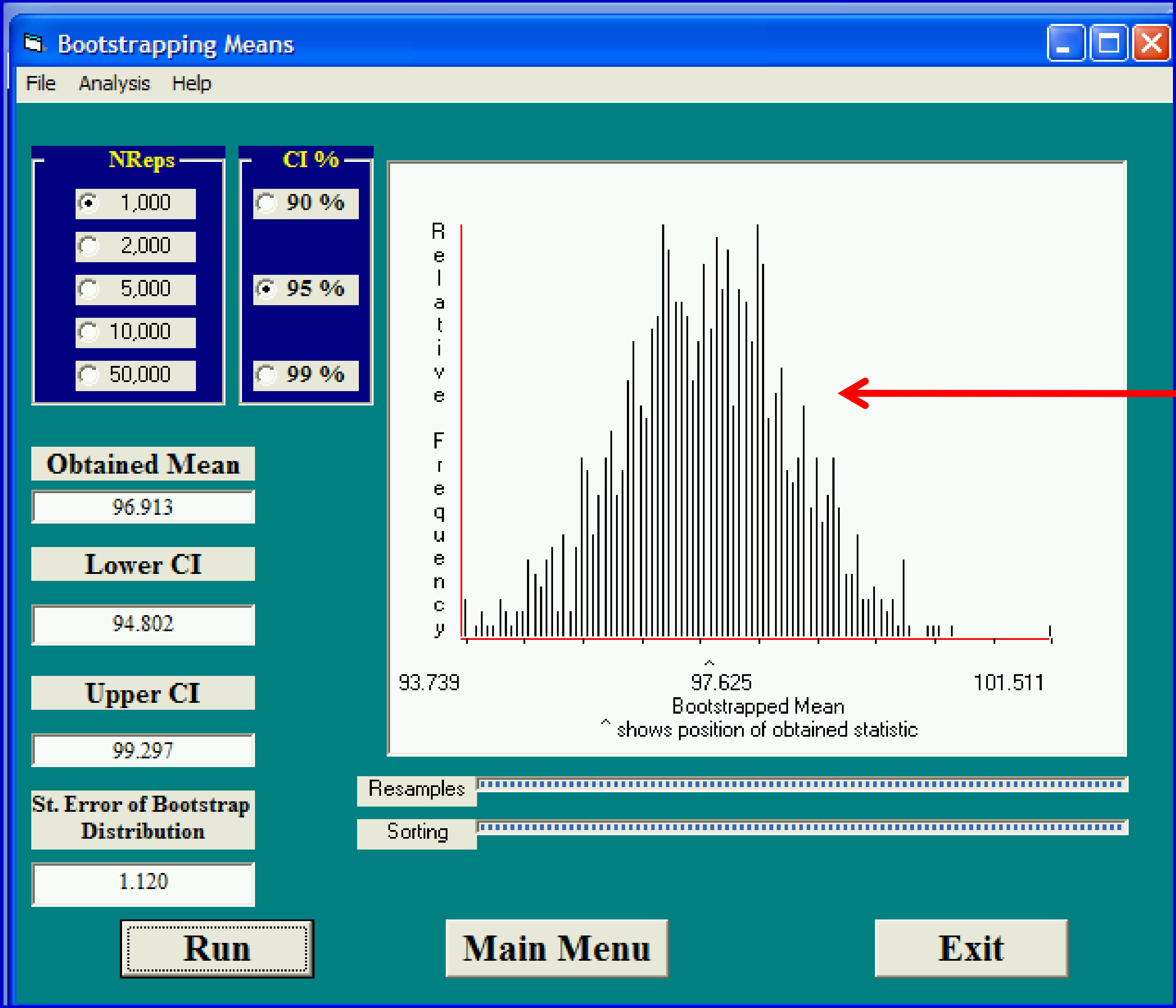
El método se puede utilizar con muestras pequeñas (N < 20) y no requiere el supuesto de normalidad.

5. Haga un clic sobre **Run** para realizar los cálculos.
6. Resultado.

   A continuación se muestran los resultados del análisis.

Media estimada: 96.91 cm
Límite inferior del IC: 94.80 cm
Límite superior del IC: 99.29 cm
Error estándar de la distribución: 1.12 cm

Intervalos de confianza en “Resampling Stats”



Distribución muestral de la media

IC varianza y desviación estándar para muestra proveniente de distribución normal

- Varianza muestral (S^2) es un estimador insesgado y de varianza mínima de la varianza de la población normal.
- Desviación estándar (S) es un estimador sesgado de σ. El sesgo se reduce rápidamente al aumentar el tamaño de la muestra (n).

$$\frac{n-1 * S^2}{X^2 1-(\alpha/2)} < \sigma^2 < \frac{n-1 * S^2}{X^2(\alpha/2)}$$

IC VARIANZA
DIST. CHI CUADRADO
CON N-1 GRADOS LIBERTAD

IC DESVIACION ESTÁNDAR

$$\left[\frac{n-1 * S}{X^2 1-(\alpha/2)} \right]^{0,5} < \sigma < \left[\frac{n-1 * S2}{X^2(\alpha/2)} \right]^{0,5}$$

- EL IC para la desviación estándar se obtiene extrayendo la raíz cuadrada de los tres términos de IC para la varianza
- Recuerde que el cálculo del IC para la desviación estándar asume que los datos provienen de una distribución normal, y a diferencia del IC para la media, en este caso una violación de este supuesto se traduce en un IC incorrecto.

FLUJOGRAMA

- 1. Seleccionar muestra
- 2. Calcular media y desviación estándar
- 3. Probar por normalidad de los datos
- 4. Si los datos no cumplen con el supuesto de normalidad debe transformarlos
- 5. Seleccione nivel de confianza
- 6. Calcule intervalos de confianza. Seleccione software a utilizar.
- 7. Interprete los intervalos. ¿Cuáles son los argumentos estadísticos y biológicos o de otra índole que le permiten explicar los resultados obtenidos? Ej. tamaño de muestra, muestreo sesgado, nivel de confianza utilizado, condiciones ambientales no normales (Ej. El Niño, La Niña)
- ¿Cual sería su recomendación final (acción)? ¿Otro estudio?

Recuerde la estadística es solo una herramienta!!!!