

## **Statistics-to-Go**

Módulo 1: Histograma, diagrama de barras, diagrama de dispersión, correlación y regresión lineal.

## **Manual de usuario**

Creadores:

M.C. Héctor Antonio Villa Martínez

Dr. Francisco Javier Tapia Moreno

# 1. Introducción

En la Universidad de Sonora, los estudiantes de las licenciaturas en Matemáticas, Física y Ciencias de la Computación toman un curso obligatorio de introducción a la estadística. El objetivo del curso es enseñar a los estudiantes herramientas estadísticas básicas y familiarizarlos con el análisis estadístico usando software estadístico. Estas herramientas de escritorio corren solamente en computadoras de escritorio o en laptops, lo que implica que los alumnos deben estar en el laboratorio de cómputo o cargando sus laptops.

Por otra parte, tomando en cuenta que en 2009 había casi 80 millones de subscriptores celulares en México<sup>1</sup> y que mundialmente alrededor del 50% de los teléfonos celulares están habilitados para Java<sup>2</sup>, estamos interesados en saber si los teléfonos celulares pueden utilizarse para resolver problemas estadísticos, ayudando a los alumnos a aprender estadística a cualquier hora y en cualquier lugar y liberándolos de tener que cargar una laptop o de tener que estar en un centro de cómputo.

De esta forma, en febrero de 2010 el proyecto Statistics-to-Go fue propuesto a la División de Ciencias Exactas y Naturales de la Universidad de Sonora con el objetivo de producir herramientas para la enseñanza de estadística en teléfonos celulares utilizando Java ME (micro edition). Java ME fue seleccionado porque casi todos los celulares de rango medio y alto están habilitados para Java, lo que significa que los programas escritos en Java ME, conocidos como “midlets”, son portables entre los distintos sistemas operativos móviles.

Este documento presenta el manual de usuario del software correspondiente al módulo 1 del proyecto Statistics-to-Go que cubre los temas de las siguientes herramientas estadísticas: medidas sumarias, histograma, diagrama de barras, diagrama de dispersión, correlación y regresión lineal. El software consiste de una “suite” de cuatro midlets: diagrama de barras, gráfica de dispersión, histograma y medidas sumarias. Por su naturaleza, las herramientas de correlación y regresión lineal son parte del midlet de gráfica de dispersión.

El resto de este manual consta de dos secciones: la sección 2 muestra el proceso de instalación del software en un teléfono celular y la sección 3 describe la utilización de las herramientas estadísticas.

## 2. Instalación

El software se distribuye en dos archivos: modulo1.jad y modulo1.jar. Una vez que estos archivos han sido almacenados en una computadora de escritorio o laptop,

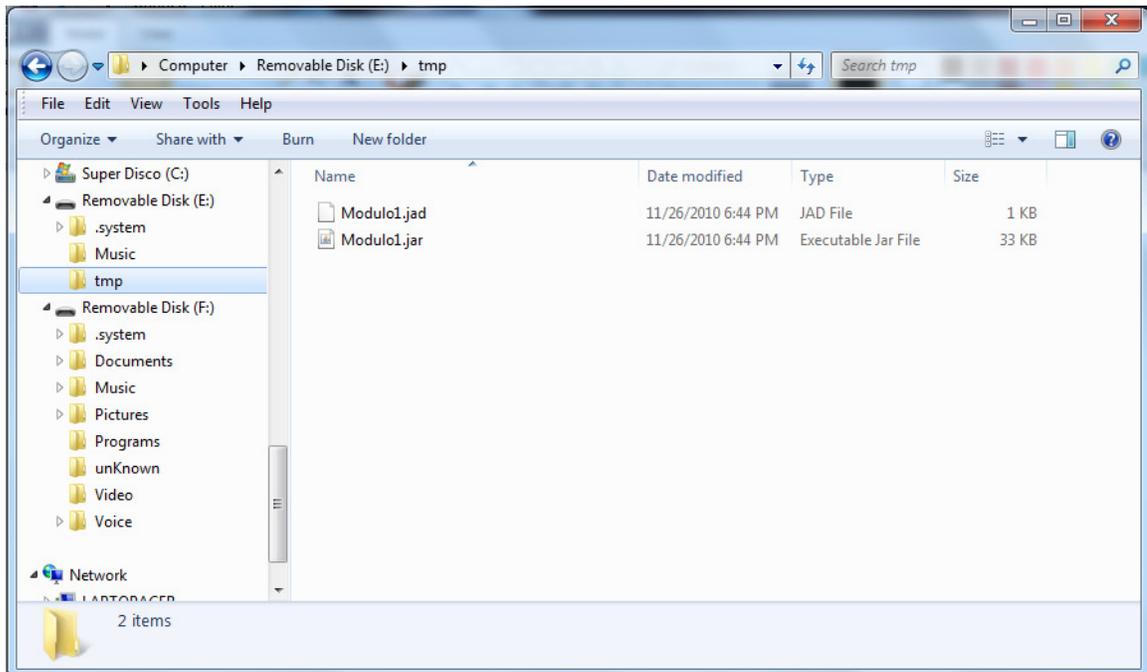
---

<sup>1</sup> <http://www.totaltele.com/view.aspx?ID=445636>

<sup>2</sup> <http://uclue.com/?xq=1180>

el proceso de instalación del software en el teléfono celular depende tanto del sistema operativo de la computadora como del modelo del celular. A manera de ejemplo, se presenta el proceso de instalación en un celular Motorola ROKR Z6 utilizando una computadora con sistema operativo Windows 7.

1. Conectar el celular a la computadora utilizando el cable mini-USB original. Windows 7 reconoce el teléfono y lo monta como 2 drives, que en este ejemplo son E: y F:.
2. Copiar los archivos modulo1.jad y modulo1.jar en cualquier carpeta de algún drive que corresponda al teléfono. En este ejemplo, los archivos se guardan en E:\tmp (ver la siguiente figura).

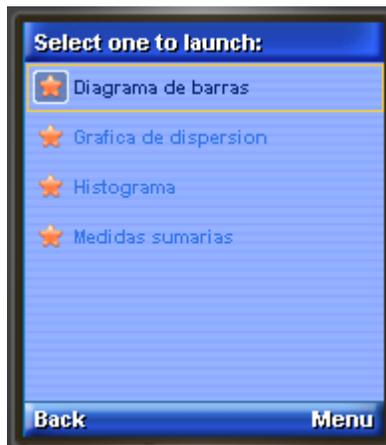


3. Expulsar los drives correspondientes al celular mediante la función de remover hardware que se localiza en el extremo derecho de la barra de tareas de Windows. Desconectar el celular del cable mini-USB.
4. Localizar en el celular los archivos recién copiados. En el Z6 se abre el manejador de archivos (File Manager), se escoge la opción Memory Card y enseguida se selecciona la carpeta tmp.
5. Se ejecuta el archivo modulo1.jad. En el Z6 se posiciona el cursor sobre dicho archivo y se oprime la tecla Fire.
6. El celular despliega las características del programa a instalar y presenta dos opciones: Install y Cancel. Seleccionar Install.
7. El celular avisa que la aplicación no está firmada y presenta dos opciones Continue y Cancel. Seleccionar Continue.
8. El celular pregunta en donde se desea instalar la aplicación, si en la memoria interna del teléfono (Phone) o en la tarjeta externa de memoria

- (Memory Card). Seleccionar Memory Card en caso de que se desee dejar libre la memoria interna. Al programa no le afecta en donde sea instalado.
9. El programa queda instalado y listo para ser usado. En el Z6, los programas instalados por el usuario se localizan en la carpeta de juegos (Games).

### 3. Uso

Al correr el programa, éste presenta una pantalla con el menú principal que permite seleccionar alguno de los cuatro midlets del módulo 1: diagrama de barras, gráfica de dispersión, histograma o medidas sumarias.



Para ejecutar un midlet es necesario mover el cursor con las flechas hasta marcar la opción deseada y oprimir la tecla Fire. Los cuatro midlets presentan una primera pantalla en dónde el usuario puede capturar los datos necesarios para el midlet respectivo.

A continuación, primero se describe el uso de la pantalla de datos y luego el funcionamiento de cada uno de los cuatro midlets.

#### 3.1 Pantalla de datos

Como se muestra a continuación, la pantalla de datos contiene un área de texto donde se pueden capturar los datos de entrada.

**Nota:** El título de la pantalla de datos depende de que midlet fue el que la invocó. Por otro lado, la forma del área de texto y la disposición de los comandos en la barra inferior dependen del teléfono celular. Sin embargo, la funcionalidad es la misma.



El usuario puede escribir los datos para el midlet como se describe abajo, o puede seleccionar alguno de los siguientes comandos disponibles:

- a) Salir. Regresa al menú principal.
- b) Leer datos. Recupera los datos que se utilizaron la última vez que el midlet fue ejecutado.
- c) Guardar datos. Almacena los datos actuales. Estos datos estarán disponibles la próxima vez que el midlet sea ejecutado.
- d) Graficar. (Solo disponible en los midlets de diagrama de barras, gráfica de dispersión e histograma). Almacena automáticamente los datos actuales y despliega la herramienta estadística correspondiente.
- e) Resultados. (Solo disponible en el midlet de medidas sumarias). Presenta las medidas sumarias que se detallan más adelante.

Los datos se escriben utilizando el teclado del teléfono celular ya sea en modo alfabético o en modo numérico. La forma de cambiar el modo del teclado depende de cada modelo. En el Z6, por ejemplo, el modo se cambia oprimiendo sucesivamente la tecla #.

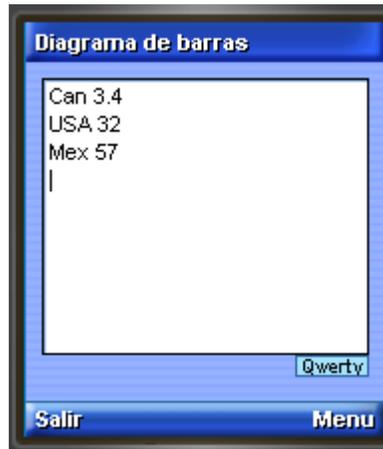
### 3.2 Diagrama de barras

Los datos del diagrama de barras se escriben en parejas separadas por espacios o brincos de línea (es indistinto). El primer dato de la pareja representa una categoría y el segundo dato es un número asociado con esa categoría.

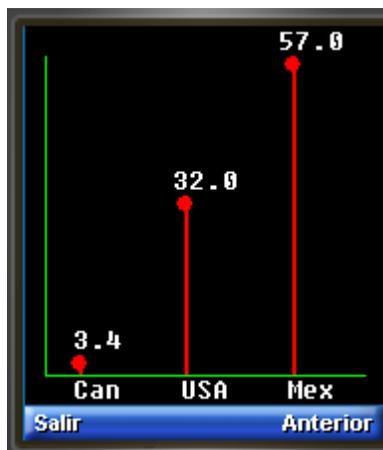
**Ejemplo.** Dibujar un diagrama de barras para los siguientes datos.

<b>País</b>	<b>Densidad poblacional</b>
Canadá	3.4
USA	32
México	57

Los datos se capturan como se muestra a continuación.



Al seleccionar el comando Graficar, se presenta el siguiente diagrama de barras.



En la visualización del diagrama de barras el usuario tiene dos comandos disponibles:

- a) Salir. Regresa al menú principal.
- b) Anterior. Regresa a la pantalla de captura de datos. El usuario puede alterar los datos y desplegar un nuevo diagrama de barras.

### 3.3 Gráfica de dispersión

Los datos de la gráfica de dispersión se escriben en parejas de números separadas por espacios o brincos de línea (es indistinto). La pareja de números representan los valores de dos variables para un conjunto de datos. Por lo general, el primer número de la pareja es el valor de la variable independiente, mientras que el segundo número es el valor de la variable dependiente. En caso

de que no exista una relación de dependencia, entonces el orden se puede intercambiar.

**Ejemplo.** Dibujar una gráfica de dispersión para los siguientes datos de estaturas y pesos promedio de mujeres estadounidenses entre 30 y 39 años<sup>3</sup>.

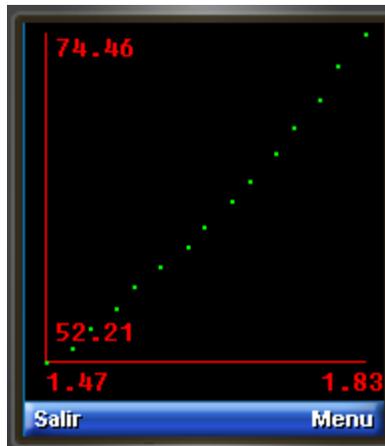
Altura (mts)	Peso (kg)
1.47	52.21
1.50	53.12
1.52	54.48
1.55	55.84
1.57	57.20
1.60	58.57
1.63	59.93
1.65	61.29
1.68	63.11
1.70	64.47
1.73	66.28
1.75	68.10
1.78	69.92
1.80	72.19
1.83	74.46

Los datos se capturan como se muestra a continuación.



Al seleccionar el comando Graficar, se presenta la siguiente gráfica de dispersión.

<sup>3</sup> Fuente: [http://en.wikipedia.org/wiki/Ordinary\\_least\\_squares#Example\\_with\\_real\\_data](http://en.wikipedia.org/wiki/Ordinary_least_squares#Example_with_real_data)



En la visualización de la gráfica de dispersión el usuario tiene cuatro comandos disponibles:

- a) Salir. Regresa al menú principal.
- b) Anterior. Regresa a la pantalla de captura de datos. El usuario puede alterar los datos y desplegar una nueva gráfica de dispersión.
- c) Regresión. Calcula la ecuación de regresión lineal, junto con otras estadísticas (ver sección 3.4.1), y despliega la recta ajustada.
- d) Correlación. Calcula el coeficiente de correlación de Pearson y otras estadísticas relacionadas (ver sección 3.4.2).

### 3.3.1 Regresión lineal

Las estadísticas de regresión lineal que se calculan son las siguientes:

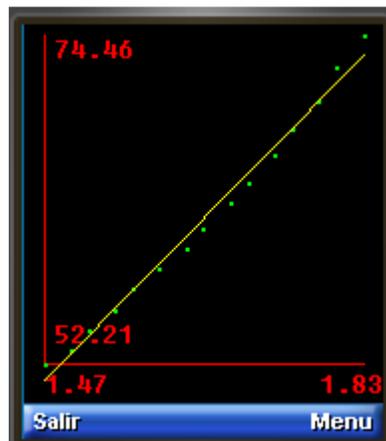
- 1) Ecuación de la recta ajustada.
- 2) Errores standard:
  - a. Error standard del estimado.
  - b. Error standard de la pendiente  $m$ .
  - c. Error standard de la ordenada al origen  $b$ .
- 3) Coeficiente de determinancia.
- 4) Significancia de la pendiente.



En esta pantalla el usuario puede escoger entre dos comandos.

- a) Salir. Regresa al menú principal.
- b) Gráfica. Regresa al diagrama de dispersión y despliega la recta ajustada.

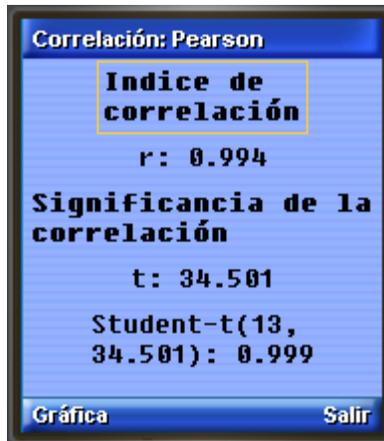
Al seleccionar el comando Gráfica se despliega el diagrama de dispersión con la recta ajustada.



### 3.3.2 Correlación

Las estadísticas de correlación que se calculan son las siguientes:

- 1) Índice de correlación de Pearson.
- 2) Significancia de la correlación.



En esta pantalla el usuario puede escoger entre dos comandos.

- a) Salir. Regresa al menú principal.
- b) Gráfica. Regresa al diagrama de dispersión.

### 3.4 Histograma

Los datos del histograma se escriben en parejas de números separados por espacios o por brincos de línea (es indistinto). El primer número de la pareja es el límite inferior del rango y el segundo número es el valor asociado al rango. La última pareja puede estar completa o incompleta dependiendo de si el último rango es abierto o cerrado. A continuación se muestran dos ejemplos de esta situación.

**Ejemplo 1.** Dibujar un histograma para los siguientes datos:

Rango	Valor
$0 \leq x < 5$	16
$5 \leq x < 10$	12
$10 \leq x < 15$	22
$15 \leq x < 20$	21

Note que el último rango es cerrado, por lo tanto, los datos se capturan como se muestra a continuación.



El último renglón consta de un solo número que representa el límite superior del último rango.

Al seleccionar el comando Graficar, se presenta el siguiente histograma.



**Ejemplo 2.** Dibujar un histograma para los siguientes datos:

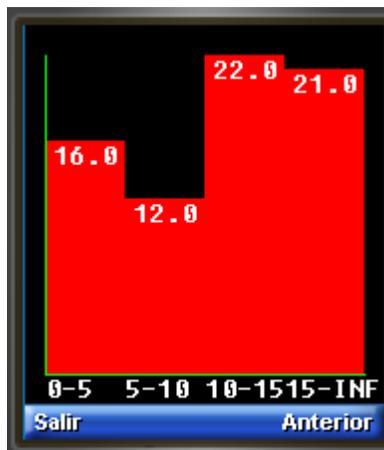
Rango	Valor
$0 \leq x < 5$	16
$5 \leq x < 10$	12
$10 \leq x < 15$	22
$15 \leq x$	21

Note que el último rango es abierto, por lo tanto, los datos se capturan como se muestra a continuación.



El último renglón representa el límite inferior del último rango y su valor asociado. Al no haber más datos, el midlet asume que el último rango es abierto.

Al seleccionar el comando Graficar, se presenta el siguiente histograma.



En la visualización del histograma el usuario tiene dos comandos disponibles:

- a) Salir. Regresa al menú principal.
- b) Anterior. Regresa a la pantalla de captura de datos. El usuario puede alterar los datos y desplegar un nuevo histograma.

### 3.5 Medidas sumarias

Los datos de medidas sumarias es una secuencia de números separados por espacios o por brincos de línea (es indistinto).

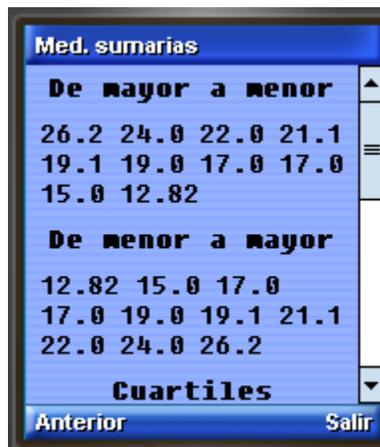
Al oprimir el comando de Resultados, primero se presentan los datos ordenados de mayor a menor y de menor a mayor y enseguida se calculan las siguientes medidas sumarias:

- a) Cuartiles.
- b) Déciles.
- c) Moda.
- d) Máximo y mínimo.
- e) Promedio.
- f) Mediana.
- g) Desviación standard de la muestra y de la población.
- h) Varianza de la muestra y de la población.

**Ejemplo.** Calcular las medidas sumarias de la secuencia 24, 17, 26.2, 17, 19, 21.1, 12.82, 15, 22, 19.1.



Al seleccionar el comando Resultados, se presenta la siguiente pantalla.



En esta pantalla el usuario tiene dos comandos disponibles:

- a) Salir. Regresa al menú principal.
- b) Anterior. Regresa a la pantalla de captura de datos. El usuario puede alterar los datos y calcular de nuevo las medidas sumarias.

