

Nombre \_\_\_\_\_ Fecha \_\_\_\_\_


## Actividad NUMB3RS: Solitarios

En "Para llevar" Amita y Charlie ayudan al FBI a encontrar transferencias bancarias desde los Estados Unidos a México. Sin embargo, hay tanto tráfico bancario que es difícil hallar la transferencia específica. Charlie explica que "es cuestión de usar un modelo de optimización específico para un objetivo, algo llamado "Detección de Solitarios". Hemos estado buscando "a lo loco", como cuando uno tira una red a un río y lo atrapa todo. En cambio la Detección de Solitarios es específica para cada objetivo. Es como pescar con caña en un río de datos, escogiendo tu lugar según el comportamiento de los peces, escogiendo la carnada correcta; un método que nos permite lanzar la caña exactamente donde queremos y atrapar exactamente el tipo de pez que necesitamos".

Un solitario es un dato que difiere significativamente de los demás datos de su grupo. Un punto que se encuentre lejos quizá no sea representativo y puede distorsionar los resultados. Ahora bien, el reto es cuantificar qué es "significativamente diferente". Abajo se dan definiciones y exploraciones para dos tipos de análisis: un diagrama de caja y bigotes y un diagrama de dispersión.

### Diagrama de caja y bigotes

Considera los siguientes datos, que son los montos de veinte transferencias bancarias: \$50, \$900, \$1,000, \$1,210, \$1,250, \$1,300, \$1,300, \$1,500, \$1,500, \$1,500, \$1,500, \$1,850, \$2,000, \$2,000, \$2,100, \$2,500, \$2,500, \$2,800, \$3,800, \$6,000.

1.
  - a. Usa una calculadora gráfica para construir un histograma de los datos. Para tener una buena ventana de datos, presiona **ZOOM** y selecciona **9:ZoomEstad**. Para fijar un buen ancho de las barras de estos datos, presiona **WINDOW** y configura **Xescl** para un valor cercano a 200.
  - b. Un solitario es un punto que parece significativamente lejano de los otros puntos. ¿Hay algún punto que parece ser un solitario?
2.
  - a. Cambia el diagrama a un diagrama de caja y bigotes (selecciona el icono  en la pantalla del editor **GRÁF ESTAD**).
  - b. ¿Cuál es el valor de Q1? (Usa **TRACE** para hallar el valor de Q1).
  - c. ¿Cuál es el valor de Q3?
  - d. Calcula la amplitud intercuartílica (IWR), que se define como  $Q3 - Q1$ .

Un dato que es  $1.5 \times (IQR)$  unidades por encima de Q3 ó  $1.5 \times (IQR)$  unidades por debajo de Q1 se considera un *solitario*. Los dos valores  $Q1 - 1.5 (IQR)$  y  $Q3 + 1.5 (IQR)$  se llaman *cercas*.

3.
  - a. Encuentra las dos cercas para el grupo de datos.
  - b. ¿Qué punto(s) son solitarios para cada grupo de datos?

- c. Verifica tu respuesta a 3b haciendo el diagrama de caja y bigote modificado (selecciona el icono  en la pantalla del editor **GRÁF ESTAD**).

Examina los datos sin el solitario. Copia los datos en una lista nueva y elimina el solitario.

4. a. Haz un diagrama de caja y bigote modificado con el grupo de datos ajustados. Mira los dos diagramas en la misma pantalla.
- b. Determina la mediana, la media, la moda y el IQR para cada conjunto de datos. Todos estos valores son *medidas de tendencia central*.
- c. ¿Cómo se comparan los dos conjuntos de datos? ¿Qué medidas de tendencia central son iguales y cuáles son diferentes?

### Diagramas de dispersión

Los datos de la tabla siguiente muestran el número de puntos anotados y de minutos jugados por seis miembros de los Lakers de Los Ángeles en la temporada 2004–05. (Los datos provienen del sitio Web Illuminations en:

<http://illuminations.nctm.org/ActivityDetail.aspx?ID=146>.)

Jugador	Puntos anotados	Minutos jugados
Kobe Bryant	1,819	2,689
Caron Butler	1,195	2,746
Chucky Atkins	1,115	2,903
Lamar Odom	975	2,320
Chris Mihm	735	1,870
Jumaine Jones	577	1,830

5. a. Ingresa los datos en una calculadora gráfica y genera un diagrama de dispersión. ¿Parece haber algún dato solitario?

Haz un regresión lineal de los datos. Presiona **STAT** y en el menú **CALC** selecciona **4:RegLin(ax+b)**. Ingresa (el nombre de la primera lista, el nombre de la segunda lista), luego **ENTER**. Una regresión da la ecuación de una recta que se aproxima a los datos. El valor de  $r$ , conocido como el *coeficiente de correlación*, indica cuán bien representa los datos la ecuación de regresión. Cuanto más se acerque  $|r|$  a 1, mejor será la representación.

- b. ¿Cuál es el *coeficiente de correlación* para los datos?

Elimina un dato y halla el nuevo coeficiente de correlación. Devuelve el dato eliminado, elimina otro dato y calcula de nuevo. Repite esto para cada dato. Un dato que afecte considerablemente el coeficiente de correlación se considera un solitario.

- c. Después de probar todos los datos, ¿qué dato afecta más el coeficiente de correlación? ¿Es el mismo dato que escogiste en la pregunta 5a?

***El objeto de esta actividad es dar a los estudiantes un vistazo breve y sencillo de un tema matemático muy extenso. TI y NCTM lo invitan a usted y a sus estudiantes a aprender más sobre este tema con las extensiones que se ofrecen abajo y con su propia investigación independiente.***

## Extensiones

### Para el estudiante

- La eliminación de un solitario afecta el valor de las mediciones estadísticas. Cuando se identifica un solitario, el estadístico debe analizar las circunstancias para determinar si resulta razonable eliminarlo del conjunto de datos. Considera la siguiente actividad en: **<http://illuminations.nctm.org/ActivityDetail.aspx?ID=151>**.

Elige "Densidad de Población" del menú de Conjuntos de Datos y nota que el mapa está casi en blanco, lo que indica que todos los estados tienen valores bajos y similares. El Distrito de Columbia tiene una densidad de población de 8,370 personas por milla cuadrada, que es significativamente mayor que cualquiera de los 50 estados. En este caso, tiene sentido eliminar el DC como un solitario, porque el DC es una ciudad, no un estado. Después de eliminar el DC, Nueva Jersey parece solitario, pero, ¿es válido eliminarlo? Aunque Nueva Jersey es "semejante a una ciudad" en que su parte sur es un suburbio de Filadelfia y su parte norte un suburbio de Nueva York, no deja de ser un estado y probablemente debe quedar entre los datos que se están considerando.

- Mientras que muchas fórmulas matemáticas son aceptadas universalmente, no hay acuerdo en cuanto al método de calcular Q1 y Q3 para un diagrama de caja y bigote. Todos los métodos aceptados aportan resultados similares, aunque pueden variar ligeramente para el mismo conjunto de datos. Busca libros, calculadoras o programas que usen diferentes métodos y compara sus resultados para el mismo conjunto de datos. ¿Puedes formar un conjunto de datos tal, que un dato sea solitario en un método y no en otro?
- Repite la actividad de regresión lineal con datos de los Pistones de Detroit en la temporada 2004–05. Visita: **<http://www.nba.com/pistons/stats/2004/index.html>** para encontrar la información de los jugadores. Investiga a Ben Wallace como solitario y considera cómo sus resultados se diferencian de los de Kobe Bryant.
- Hay muchos y tipos de regresiones y correlaciones diferentes. La regresión lineal calculada con las calculadoras gráficas TI-83 Plus/TI-84 Plus es la *recta de mínimos cuadrados*, la ecuación que minimiza la suma de los cuadrados de las distancias entre cada dato y la recta. La correlación es la correlación del momento del producto de Pearson. Para ver otra actividad NUMB3RS que explora esta correlación, visita **<http://education.ti.com/exchange>** y busca "7723".

### Recursos adicionales

- Otro ejemplo de solitarios en conjuntos de datos se encuentra en: **<http://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm>**.
- Para ver el efecto que produce sumar datos a una línea de regresión de un diagrama de dispersión, dale un vistazo al applet interactivo en: **<http://www.stat.sc.edu/~west/javahtml/Regression.html>**.