VI Congresso de Estatística e Investigación de Operacións
Vigo 5-7 de Noviembre de 2003

ESTADÍSTICA CON GEOMETRÍA DINÁMICA

José Alexandre dos Santos Vaz Martins

Escola Superior de Turismo e Telecomunicações
Instituto Politécnico da Guarda

RESUMO

Se pretende mostrar una forma visual y dinámica de abordar algunos conceptos, resultados y deducciones de fórmulas de la Estadística Descriptiva, que tienen una interpretación gráfica y/o geométrica, permitiendo una más fácil asimilación por parte de los alumnos, una mayor facilidad de exposición por parte de los profesores, y una más fructífera interacción entre profesor y alumnos.

INTRODUCÍON

Se pretende con este trabajo presentar una de las potencialidades de la geometría dinámica, al servicio, no de los problemas geométricos puros, pero sí de la aprehensión de conceptos básicos de Estadística Descriptiva.

En ese sentido se intentaron crear situaciones que estimulasen visualmente y facilitasen la adquisición de los conceptos de media aritmética, mediana, moda y el método de los mínimos cuadrados, en el caso de la regresión linear, a través de una interpretación geométrica de los mismos y/o de sus propiedades. También se presentarán aplicaciones con algunos tipos de gráficos para explorarlos y establecer interrelaciones.

Las aplicaciones creadas pueden ser presentadas en cualquier disciplina que introduzca los conceptos referidos.

MEDIDAS DE LOCALIZACIÓN

- Media Aritmética

Para datos no clasificados la expresión para la media aritmética es $\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$.

Por otro lado, la media tiene como característica geométrica que la suma de las distancias de ella a los datos que le son inferiores es igual a la suma de las distancias de ella a los datos que le son superiores:

$$\sum_{i=1}^{n} \left(x_i - \bar{x}\right) = 0$$

Resulta entonces que, geométricamente, la media aritmética puede ser interpretada como la abcisa (*x*) del centro de gravedad del sistema formado por los puntos $x_1, \ldots, x_N$ con pesos unitarios.

Basado en estas dos relaciones entre la media y la geometría, se puede construir una aplicación con dos componentes, que están ínter ligados, y que se explicarán sucintamente a continuación:

- Sin pérdida de generalidad, en una recta están colocados siete puntos, $x_1,\ldots,x_7$, con frecuencia unitaria, y está también colocado el punto M que puede ser movido a lo largo de la recta. Al mover el punto M se observa en la pantalla el valor, vsuma, de la medida del vector suma de los vectores con origen en el punto M y la otra extremidad en cada uno de los puntos, $x_1,\ldots,x_7$. Entonces, se puede procurar, de forma dinámica, la media haciendo que al punto M correr la recta de forma a que el valor vsuma pase a ser nulo.
- Es visible una "balanza" en que la recta, en que están los puntos, sirve de plato de la balanza y el punto M sirve de fiel de la balanza, de tal manera que ésta está en equilibrio cuando M coincide con la media, y se queda proporcionalmente desequilibrada a la vez que M se aleja de la media.

Por fin, la aplicación, incluye el cálculo de la media aritmética, permitiendo que se haga una confirmación analítica del valor.

La visualización de lo que se acabó de referir, intentando dar, de manera simple, una idea sobre la relación entre el "peso" (en última análisis, el centro de gravedad) y la media, puede ser hecha en las tres figuras siguientes:

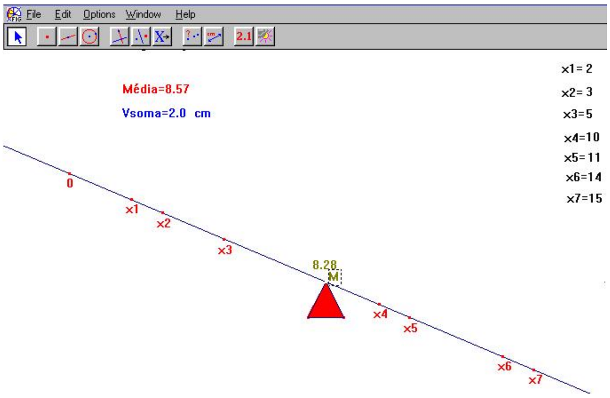


Figura ¡Error!Argumento de modificador desconocido.**: Punto M con abcissa superior al valor de la media**

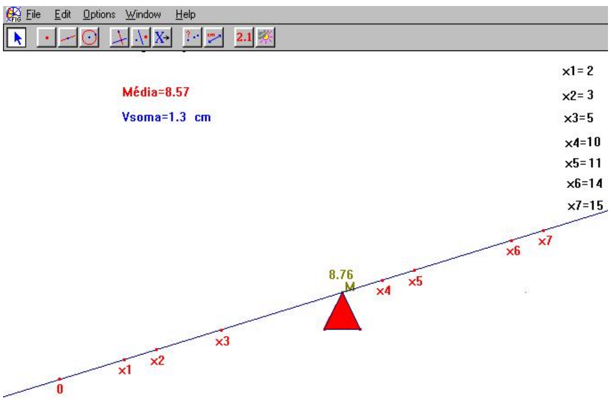


Figura ¡Error!Argumento de modificador desconocido.**: Punto M con abcissa inferior al valor de la media**

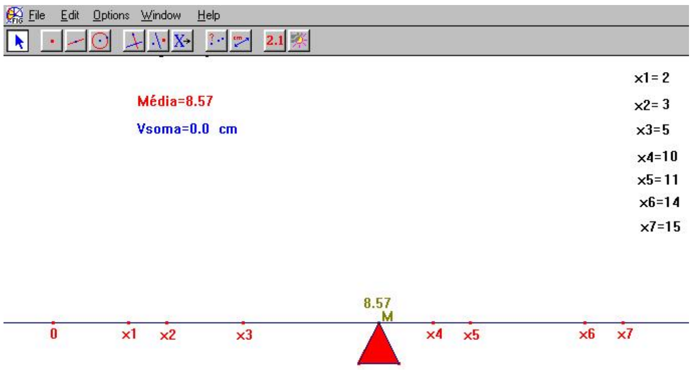


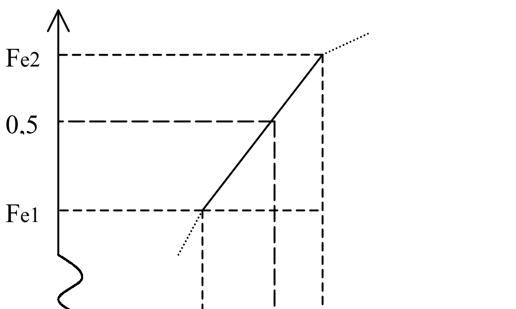
Figura ¡Error!Argumento de modificador desconocido.**: Punto M con abcissa igual al valor de la media**

- Mediana

La mediana es una medida de localización definida por su posición en la sucesión de las observaciones o en la distribución de frecuencias. Así, en el caso de variables continuas, la mediana es la abcisa (*x*) a que corresponde la ordenada 0,5 (50%) en el gráfico de la función acumulada (polígono integral). Está, de esta forma, establecido un proceso gráfico-geométrico de obtención del valor (aproximado) de la mediana para el caso referido.

Por otro lado, en el histograma, en que las frecuencias se pueden interpretar como áreas, la mediana se define como el valor de la variable tal que la ordenada levantada en el punto correspondiente del eje de las abcisas (*x*) divide el área del histograma en dos áreas iguales.

Finalmente, habiendo sido admitida, en el trazado del polígono integral, la hipótesis de que las frecuencias se distribuyen uniformemente dentro de cada clase, o lo que es lo mismo, que las frecuencias acumuladas tienen una variación linear dentro de cada clase, y utilizando semejanza de triángulos, se obtiene la fórmula para el cálculo de la mediana:



$$\boxed{\frac{Me - x_{e1}}{x_{e2} - x_{e1}} = \frac{0.5 - F_{e1}}{F_{e2} - F_{e1}}}$$

$$\boxed{Me = x_{e1} + \frac{F_{e1} - 0.5}{F_{e1} - F_{e2}}\left(x_{e2} - x_{e1}\right)}$$

(o en la notación de la aplicación (Figura 6) $Me = x_A + \frac{BD}{CE}(AC)$, siendo M*e* - $x_A = AB$).

En la aplicación creada se pueden observar, de forma dinámica, las características referidas, o sea:

- Localización gráfica de la mediana, pudiendo hacer experiencias con las frecuencias.

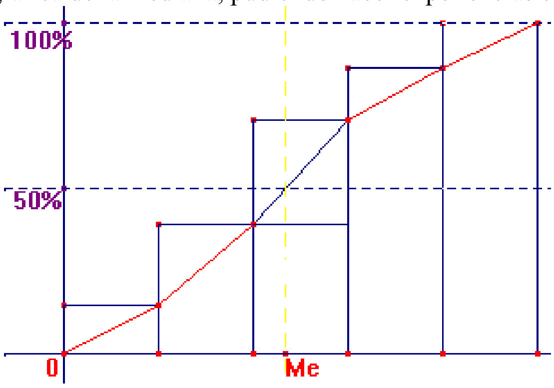


Figura ¡Error!Argumento de modificador desconocido.**: Localización gráfica de la mediana**

- Igualdad, en el histograma, de las áreas a la izquierda y a la derecha de la mediana, además de la visualización de la relación entre el gráfico de barras de las frecuencias acumuladas y el histograma.

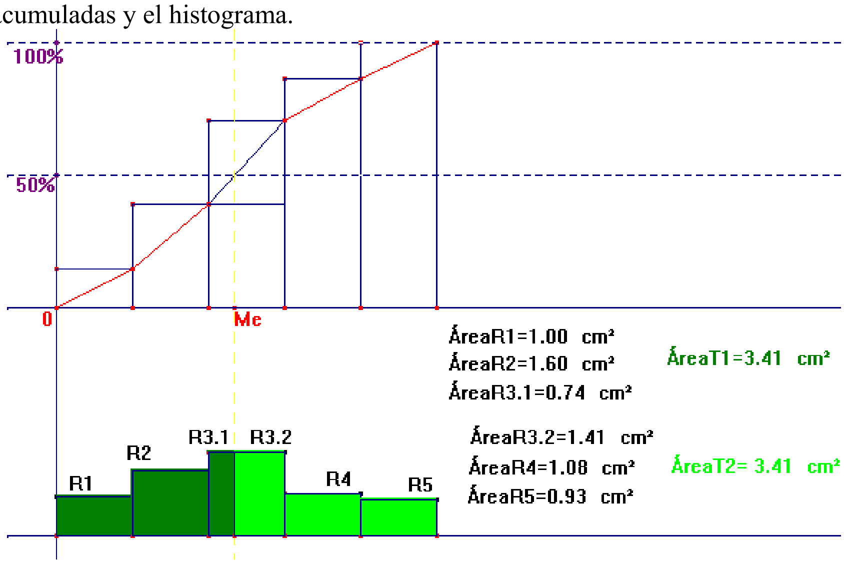


Figura ¡Error!Argumento de modificador desconocido.**:División del histograma por la mediana en dos áreas iguales**

- Semejanza de triángulos que da origen a la fórmula para el cálculo analítico de la mediana.

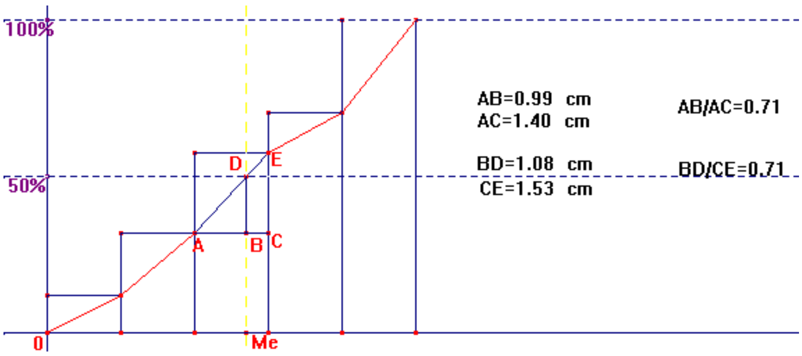


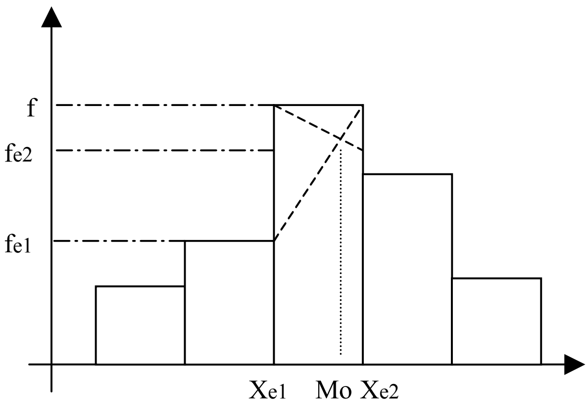
Figura ¡Error!Argumento de modificador desconocido.**: Igualdad de proporciones que conducen a la fórmula del cálculo de la mediana**

- Moda

Para una distribución de frecuencias de una variable continua, a pesar de conocerse la clase modal – clase con mayor frecuencia en clases de igual amplitud – la determinación de la moda causa algunos problemas.

Hay varias fórmulas empíricas para localizar la moda dentro de la clase modal, y por supuesto que estos son valores aproximados que carecen de una interpretación cautelosa y cuidadosa, por lo que es de gran importancia conocer mejor el "funcionamiento" de las fórmulas.

Existe un proceso geométrico para la obtención del valor de la moda, que se basa en el principio de que, dentro de la clase modal, la moda se debe encontrar más próxima de la clase adyacente con mayor frecuencia, y que usa la semejanza de triángulos para obtener la fórmula de cálculo:

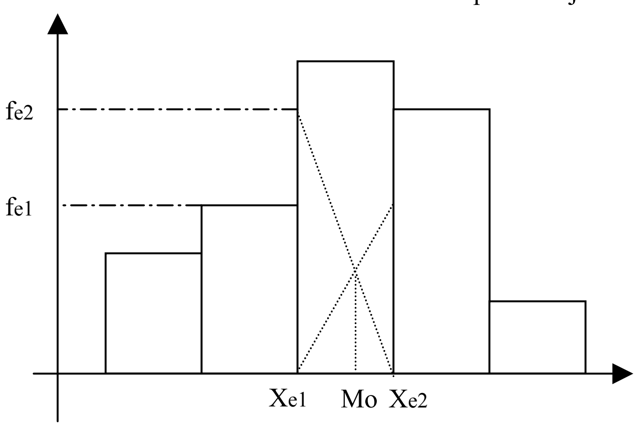


$$\frac{Mo - x_{e1}}{x_{e2} - x_{e1}} = \frac{f - f_{e1}}{(f - f_{e1}) + (f - f_{e2})}$$

$$Mo = x_{e1} + \frac{f - f_{e1}}{(f - f_{e1}) + (f - f_{e2})}(x_{e2} - x_{e1})$$

(o en la notación de la aplicación $\frac{\overline{AMo}}{\overline{AB}} = \frac{\overline{F1F2}}{\overline{F!F2} + \overline{F3F4}} \longrightarrow Mo = x_A + \frac{\overline{F1F2}}{\overline{F1F2} + \overline{F3F4}}\overline{AB}$).

Otra de las fórmulas, puede que sea la más conocida y usada, es la fórmula de King, que se basa en el mismo principio de la fórmula anterior, y su expresión analítica es también obtenida a través de relaciones válidas por semejanza de triángulos:



$$\frac{Mo - x_{e1}}{x_{e2} - x_{e1}} = \frac{f_{e2}}{f_{e1} + f_{e2}}$$

$$Mo = x_{e1} + \frac{f_{e2}}{f_{e1} + f_{e2}}(x_{e2} - x_{e1})$$

(o en la notación de la aplicación $\frac{\overline{CMok}}{\overline{CD}} = \frac{\overline{CF3}}{\overline{DF1}+\overline{CF3}} \longrightarrow Mok = x_C + \frac{\overline{CF3}}{\overline{DF1}+\overline{CF3}}\overline{CD}$).

Fue creada una aplicación con el objetivo de comprobar las relaciones mencionadas, y también visualizar dinámicamente los dos procesos de obtención del valor de la moda y su comparación, tornando más fácil su percepción. Una visualización estática puede ser hecha a través de la figura siguiente:

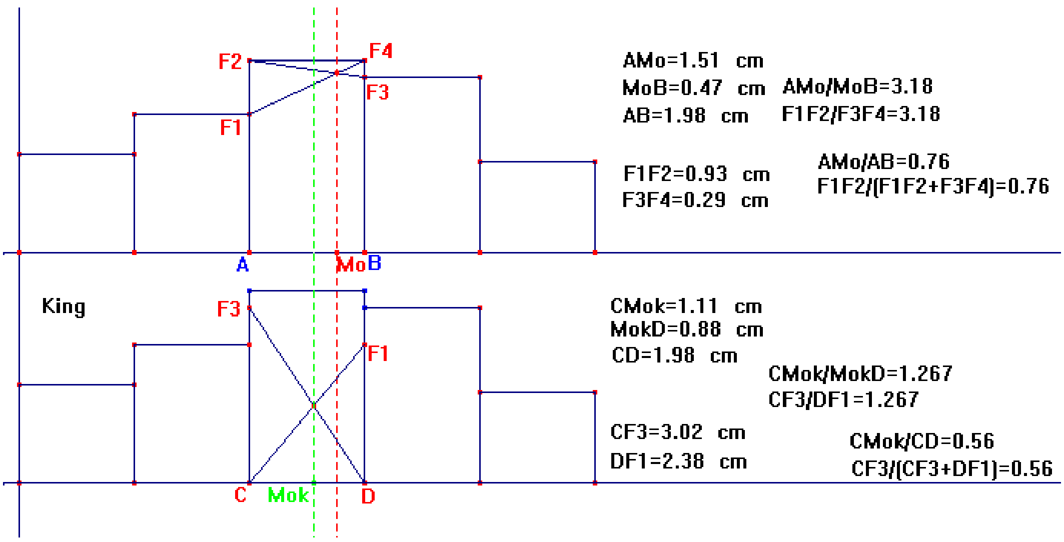


Figura ¡Error!Argumento de modificador desconocido.**: Obtención de la Moda -> Proceso Analítico vs Proceso de King**

RECTA DE REGRESIÓN

Dado un diagrama de dispersión con puntos (xi,yi) sugiriendo la existencia de regresión linear, tiene gran interés obtener, a partir de los datos, la ecuación de una recta que pueda ser tomada como una buena estimativa de la línea de regresión. De esta forma se puede describir formalmente la relación entre "y" y "x", previendo también el valor de "y" cuando sea dado un valor de "x". Puesto esto, se procura una ecuación de la recta de la forma y*=a+bx .

Siendo yi*=a+bxi , se definen los errores, o desvíos, como di=yi-yi*.

Para determinar la recta referida es necesario obtener una estimativa de sus parámetros, "a" y "b". Para tal, el principio base es, en general, el de que los desvíos deben ser tan pequeños cuanto posible. En ese sentido, el método más frecuentemente usado es el método de los mínimos cuadrados en que la determinación de los parámetros es hecha de manera a que la suma de los cuadrados de los desvíos sea mínima, e. d.,

$$\sum_{i=1}^{N} d_i^2 = \sum_{i=1}^{N}\left(y_i - y_i^*\right)^2 = \sum\left(y_i - \left(a + bx_i\right)\right)^2$$

es mínima.

En la aplicación creada (Figura 8) se puede, de una forma propicia a la interactividad, procurar ajustar una recta, a través de translaciones o rotaciones (con relación al punto de intersección entre la recta y el eje de los y´s), de manera a minimizar la suma del cuadrado de los desvíos, Somadi2. Simultáneamente, se pueden observar los desvíos (geométrica y numéricamente) y obtener la ecuación de la recta, además de poder eventualmente hacerse un estudio comparado con los resultados de la regresión linear obtenidos analíticamente y también verificar que la recta tiene que pasar por el punto ($\bar{x}, \bar{y}$).

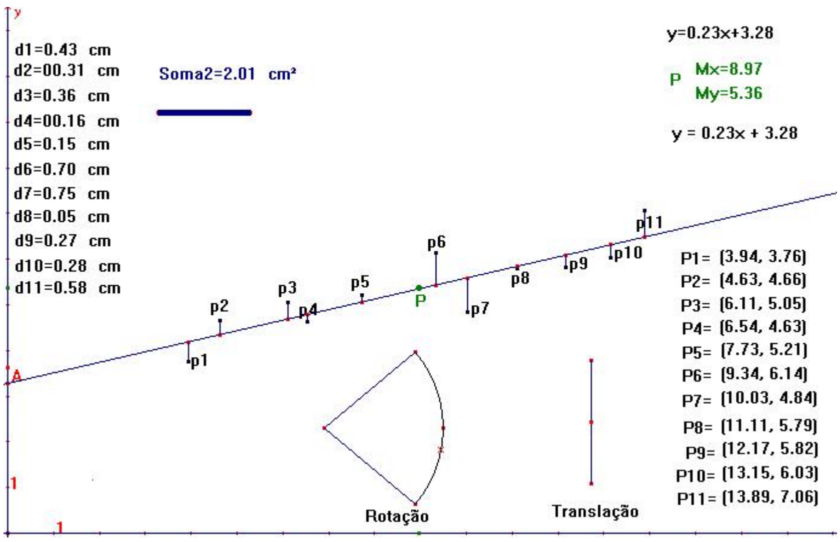


Figura ¡Error!Argumento de modificador desconocido.**: Ejemplo de regresión**

GRÁFICOS

En el campo de los gráficos, otras aplicaciones pueden ser hechas, como por ejemplo las que muestren, de dos perspectivas diferentes y en el caso de intervalos de igual amplitud, la igualdad entre las áreas del histograma y del polígono de frecuencias.

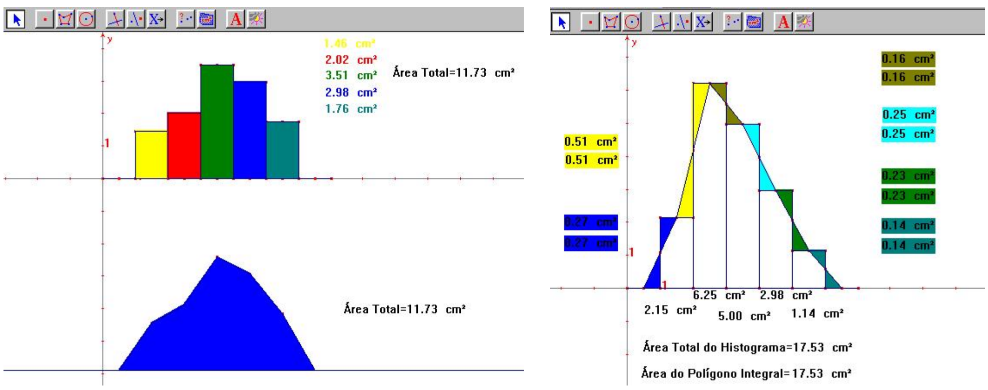


Figura ¡Error!Argumento de modificador desconocido.**: Igualdad de la área del histograma e del polígono de frecuencias**

Otro ejemplo es una aplicación que muestre la relación entre un gráfico de barras y el gráfico sectorial correspondiente, presentando los ángulos respectivos y verificando que existe, por ejemplo, una relación constante entre las áreas de las barras y los ángulos respectivos, tal como se puede observar en la figura siguiente:

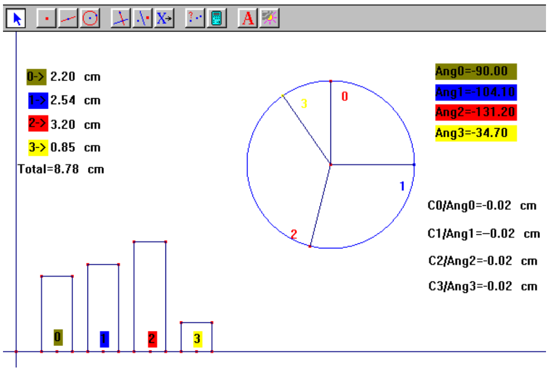


Figura ¡Error!Argumento de modificador desconocido.**: Relación entre el gráfico de barras y el gráfico sectorial**

Finalmente, otra aplicación nos puede mostrar la obtención de los cuartiles usando el histograma y el polígono de frecuencias relativas acumuladas, además de hacer la construcción del diagrama de extremos y cuartiles, dando una visión dinámica del concepto, a veces tan difícil, como es el de los cuartiles.

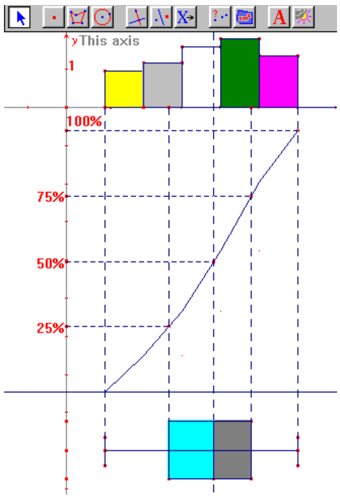


Figura ¡Error!Argumento de modificador desconocido.**: Relación entre histograma, polígono integral y diagrama de cuartiles y extremos**

CONCLUSIÓN

Los ejemplos presentados, que cualquier persona con conocimientos mínimos de utilización de software de geometría dinámica consigue implementar, sin necesidad de programar, habrán cumplido los objetivos iniciales si despertaron interés y si, a través del potencial de la geometría dinámica, pudieran ser considerados, no sólo, como ejemplos versátiles, capaces de estimular y facilitar la asimilación, interpretación y comprensión de algunos conceptos básicos de estadística descriptiva y de algunas de sus propiedades, sino también como siendo capaces de promover una mayor interactividad en ambiente de clase facilitando y mejorando el trabajo del profesor.

Hay, con certeza, innumeras posibilidades de explotación de estas aplicaciones pudiéndose profundizarlas, mejorarlas y/o añadirles otras potencialidades, teniendo como motivación la curiosidad, la imaginación y la voluntad. ¡Aquí queda el desafío!