



II CEMACYC

II Congreso de Educación Matemática de América Central y El Caribe

29 octubre al 1 noviembre. 2017

Cali, Colombia

REDUMATE

ii.cemacyc.org



CIAEM
CME
desde - since 1961



Pensamiento estadístico en problemas de contexto

Esteban **Ballestero** Alfaro

Escuela de Ciencias Naturales y Exactas, Tecnológico de Costa Rica

Costa Rica

eballestero@gmail.com

Resumen

La forma en la que las personas razonan problemas que involucran datos, define diferentes niveles de pensamiento estadístico que van, desde la idea más simple donde el individuo siente necesidad de los datos para poder explicar la situación, hasta llevar esos datos y darles un tratamiento que tome en cuenta la variabilidad, que construya modelos complejos de estos para obtener conjeturas que ayuden a comprender y tomar decisiones en el contexto.

Este escrito se analiza específicamente los tipos de pensamiento estadístico que mostraron profesores de nivel de maestría al analizar un problema estadístico real, haciendo uso del programa Fathom 2.0.

Palabras clave: Pensamiento Estadístico, Análisis Exploratorio de Datos, Modelos, Variabilidad, Tecnología.

Introducción

Toda idea de razonamiento estadístico, empieza por reconocer la importancia de disponer de información relevante para comprender y resolver algún problema de interés. Muchos son los problemas en los que la estadística encuentra su razón de ser, no obstante, la experiencia de resolver problemas a partir de datos reales dentro del currículo escolar, le daría al estudiante una invaluable experiencia que se ajusta en gran medida a la realidad, no obstante, los profesores encargados de dicha función están quedando a deber. Como indica Arteaga (2011):

Precisamente algunas de las mayores dificultades que observaremos en los futuros profesores es no ser capaces de poner en relación el resultado del análisis estadístico con el contexto del proyecto (p.17).

Se tomó un grupo de estudio conformado por profesores de matemática que cursaban su nivel de maestría en Matemática Educativa sobre los cuáles, se quería identificar los niveles de complejidad sobre el uso de conocimiento estadístico que ellos utilizaban al resolver problemas, tanto con datos simulados como datos reales.

Para este artículo, se seleccionó un ejemplo de actividad no guiada con un problema real de contexto justo con su respectiva base de datos, sirviendo este como situación problema para que los participantes llegaran, luego de un análisis propio, a proponer una solución.

Algunas preguntas surgen de previo: ¿habrá similitud entre las estrategias de resolución del problema empleadas por los estudiantes?, ¿cuáles objetos estadísticos serán seleccionados en estos razonamientos?, ¿qué bondades tecnológicas encontrará el estudiante al usar Fathom 2.0 como instrumento mediador?. Las respuestas a estas interrogantes, se detallan en las siguientes secciones.

Fundamentación

La caracterización del pensamiento estadístico que se hará en este escrito, está basado en el modelo propuesto por Wild & Pfannkuch (1999). Estos autores son los que de una manera más concreta y completa han asumido la tarea de explicar lo que es el pensamiento estadístico. Precedente a su trabajo, es poco lo que se puede encontrar en la literatura que aborde esta temática, a saber, autores como Moore (1997), Snee (1990) realizaron aportes importantes sin mucha profundidad.

Wild & Pfannkuch en su propuesta, explican el pensamiento estadístico a partir de una investigación empírica donde se ven involucrados personas de todos los niveles, es decir estudiantes, profesores, investigadores mediante entrevistas donde se les preguntaba que era para ellos el pensamiento estadístico, además se complementaba con la resolución de problemas. La lista propuesta por Moore guarda cierta similitud con los tipos de pensamiento de Wild y Pfannkuch, pero, la diferencia radica en que la lista de Moore no tiene un sustento empírico, sino, es una percepción subjetiva. Por otra parte Snee, le da central importancia a la variación, pero descuida los demás elementos de la lista, no así con Wild y Pfannkuch.

Wild & Pfannkuch (1999) señalan los fines de la elaboración de este cuadro:

... estamos interesados en desarrollar una estructura para patrones de pensamiento involucrados en la resolución de problemas, estrategias para resolver problemas y la integración de elementos estadísticos en la resolución de problemas. No dirigimos el pensamiento involucrado en desarrollar nueva metodología y teoría estadísticas. Reconocemos que mucho pensamiento estadístico puede tener lugar provechosamente en actividades cotidianas, particularmente en la interpretación de información en los medios y otros reportes (p.3).

Para Wild & Pfannkuch, toda persona que realiza un proyecto estadístico debe necesariamente pasar por cuatro dimensiones diferentes. La primera es denominada ciclo investigativo, que comprende cinco fases conocidas como *PPDAC*: Problema, plan, datos, análisis y conclusiones; la segunda dimensión hace referencia a los *tipos fundamentales de pensamiento estadístico*: reconocimiento de la necesidad de los datos, transnumeración, variación, conjunto distintivo de modelos y conocimiento del contexto; la tercera dimensión es denominada el *ciclo interrogativo* donde se destacan cinco aspectos: generar que consiste en imaginar y tener ideas brillantes para generar posibilidades como individuo o grupo, buscar información, interpretar, criticar y juzgar; finalmente la dimensión cuatro tiene que ver con características de *la personalidad del investigador* que pueden afectar o iniciar la entrada en un modo de pensamiento: curiosidad y conciencia, compromiso, escepticismo, ser lógico y predisposición a buscar significado más profundo.

Es de interés para la presente investigación considerar los tipos de pensamiento estadístico que se proponen en la *dimensión dos* del marco de Wild & Pfannkuch, para estudiar las estrategias empleadas por los estudiantes al resolver los problemas estadísticos.

Seguidamente se detalla los tipos de pensamiento estadístico de la dimensión 2, propuesta por Wild & Pfannkuch, la cual se utilizará para analizar los datos de esta investigación.

Reconocimiento de la necesidad de los datos: El reconocimiento de lo inadecuado de las experiencias personales y de la evidencia anecdótica que conduce a un deseo de fundamentar las decisiones de los datos deliberadamente recolectados, es un impulso estadístico.

Transnumeración: Se define como transformaciones numéricas hechas para facilitar la comprensión. La transnumeración ocurre cuando encontramos formas de obtener datos (a través de la medida o de la clasificación) que capturan elementos significativos del sistema real. La transnumeración es un proceso dinámico para cambiar representaciones que engendren comprensión

Consideración de la variación: Pensando en lo estadístico, de cualquier manera en el sentido actual, tiene que ver con el aprendizaje y la toma de decisiones bajo incertidumbre. Cualquier discusión seria del pensamiento estadístico debe examinar el papel de la “variación”. Los primeros tres mensajes de la “variación” son: la variación es omnipresente; la variación puede tener serias consecuencias prácticas.

Razonamiento con modelos estadísticos: Todo pensamiento usa modelos. La principal contribución de la estadística al pensamiento ha sido su propio conjunto distintivo de modelos o estructuras para pensar acerca de ciertos aspectos de la investigación en forma genérica.

Integrando el conocimiento del contexto y conocimiento estadístico: Las materias primas con que trabaja el pensamiento estadístico son el conocimiento estadístico, el conocimiento del contexto y la información en los datos. El pensamiento mismo es la síntesis de estos elementos para producir implicaciones, perspicacias y conjeturas.

Estos cinco tipos de pensamiento estadístico serán estudiados a partir de los razonamientos que los profesores manifestaron cuando resolvieron un problema aplicado (ver sección de resultados). Claro está, que la tecnología juega un papel importante dentro del proceso de análisis de los estudiantes, no obstante será un tema a ampliar en otro momento.

Metodología.

Para esta investigación se contó con la participación de 8 estudiantes de nivel de maestría quienes ya laboraban por profesores de Matemática a nivel medio y medio-superior, poseían conocimientos suficientes en el manejo de paquetes computacionales básicos y con formación en probabilidad y estadística consecuencia de su formación profesional de base.

Este proceso contempla el desarrollo de una serie de actividades semanales con una duración de dos horas cada una durante 10 semanas, utilizando como instrumento mediador el programa Fathom 2.0 sobre el cual, los participantes fueron parte de un plan de instrumentación donde aprendieron el uso y bondades que ofrece dicho paquete computacional. La escogencia de dicho programa computacional se debe a su versatilidad de manejar variadas formas de representación y dinamismo para trabajar los datos, así como su robusta herramienta de cálculo.

Las actividades fueron estructuradas en dos bloques: las primeras fueron actividades guiadas donde los estudiantes recibían una guía didáctica de trabajo que detallaba los pasos a seguir para la construcción de objetos y la realización de cálculos, desde luego, no hubo sugerencias para el análisis de resultados. Por otra parte el segundo bloque, estaba constituido por dos actividades no guiadas, donde a los estudiantes se les entregaba una situación contextualizada a resolver a partir de una base de datos reales, pero, donde la estrategia de solución era de libre elección para cada uno de ellos.

Cada una de las sesiones de trabajo guiadas, constaba de un pre-test que era aplicado al inicio y con la finalidad de identificar el nivel de conocimiento del estudiante a priori sobre los temas a trabajar en dicha sesión, así como la percepción alrededor de estos a partir de preguntas cuidadosamente planificadas para tal fin.

Este artículo pretende mostrar los resultados una de las actividades no guiadas que formaron parte de este trabajo y que fueron estructuradas dentro de la corriente de análisis de datos llamada Análisis Exploratorio de Datos (AED). Al respecto, Batanero (2001), señala las siguientes características educativas relacionadas con AED: Posibilidad de generar situaciones de aprendizaje referidas a temas de interés al alumno, fuerte apoyo en representaciones gráficas y no necesita una teoría matemática compleja

Para comprender el contexto de los resultados, tome en cuenta que las actividades no guiadas guardan elementos especiales desde el punto de vista del pensamiento estadístico con el apoyo de la tecnología. Los estudiantes en esta etapa ya han desarrollado ciertos esquemas de uso de la computadora y han venido razonando estadísticamente de manera paralela con el trabajo de las actividades guiadas, entonces, el hecho enfrentarse a un problema abierto permite percibir como los ellos se apoyan con la tecnología, es decir, como implementa los esquemas de uso de la computadora y como razona dentro de cada uno de los tipos de pensamiento estadístico.

Otro aspecto que se puede observar es si el trabajo con las actividades guiadas, que tienen la finalidad de mostrarle al sujeto las virtudes del programa y el potencial de los modelos, ha ayudado a desarrollar sus estrategias de análisis y por ende, evolucionar el conocimiento estadístico a niveles más competentes.

Análisis de resultados.

Un problema sobre agricultura con datos reales sobre la precipitación pluvial y las temperaturas mínimas y máximas registradas por semana a lo largo de 38 años de investigación en la comunidad de Mante, perteneciente al estado de Tamaulipas, México, fue el problema propuesto.

El objetivo en esta fase de la investigación, era observar cómo el estudiante, a partir de la instrucción previa recibida como se explicó en la sección anterior, podría seleccionar, articular y construir diferentes recursos estadísticos para realizar sus razonamientos que lo llevaran a proponer una solución al problema, para luego hacer una caracterización dentro de los tipos de pensamiento estadístico. Seguidamente se detalla el caso:

En los archivos de Excel *Mante.xls* se presentan los datos correspondientes a la precipitación promedio y temperatura promedio por semana de la comunidad de El Mante en el estado de Tamaulipas. Estos promedios están basados en datos recolectados durante 38 años en esta región.

Usted debe determinar cuál es la época más apropiada para que se realice la zafra. Para resolver este ejercicio debe contemplar las siguientes restricciones propuestas el investigador Ángel Martínez Garza en 1978:

1. La lluvia y temperatura de la zona cultivada con caña, influye en la concentración de azúcar de la caña cosechada.
2. Cuando hay mucha variación de temperatura un mes antes de la cosecha, se produce un efecto positivo en el rendimiento del azúcar.
3. La lluvia registrada en un periodo de dos meses antes de la cosecha reducen el rendimiento del azúcar.
4. En el caso de la región de El Mante, la duración de la zafra es aproximadamente de 14 semanas.

Este ejercicio es de solución múltiple, es decir, los estudiantes tienen la posibilidad de proponer diferentes repuestas siempre es que satisfagan con las restricciones preestablecidas. En un inicio se mostraron confusos, intentando tomar acciones aisladas y desorganizadas, producto de que las actividades previas fueron guiadas. Parte de la inseguridad, que es normal en una experiencia nueva, consistía en cómo hacer un uso inteligente del conocimiento estadístico y esquemas de uso del programa que ya poseían.

Marcela hizo un análisis basado en los diagramas de caja, aunque no necesariamente para ver la dispersión de los datos, sino que aprovecha esa cualidad que tiene un diagrama de caja de mostrar los datos de forma ordenada, pues su único interés es encontrar las variaciones de temperatura máximas y las precipitaciones mínimas.

Ella propone que la zafra debe darse entre el 26 de febrero al 3 de junio, a partir de la información de los gráficos que se exponen a continuación:

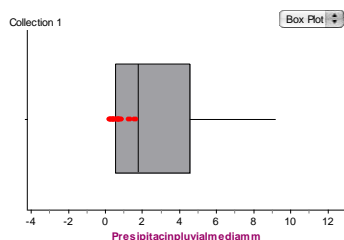


Figura 1: Precipitación pluvial por semana en la región del Mante

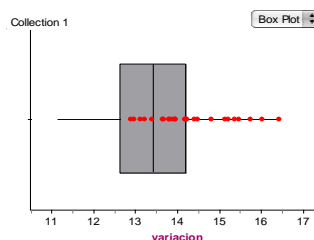


Figura 2: Variación de la temperatura semanal de la región del Mante

Fathom tiene la virtud de que en el diagrama de caja se puede seleccionar los puntos que representan datos y automáticamente se seleccionarán todas las celdas en tablas, o puntos o barras en gráficos, que dependan del mismo conjunto de datos, es decir, el mismo punto en diferentes representaciones donde esté en uso, por lo que puede decirse que se da una codependencia entre los objetos del computador.

La menor precipitación la podemos ver en el primero y segundo cuartil del diagrama de caja correspondiente a las precipitaciones. Note que al seleccionar los datos que están en esa región, se seleccionan respectivamente esos mismos datos en el diagrama de caja de la temperatura, hecho que da indicios de una posible solución al problema. Es importante aclarar que en ambos diagramas de cajas no se están graficando los mismos datos, sino que si observamos la tabla se puede notar que a cada semana se le asocia una precipitación, una temperatura mínima y una máxima (filas), por ende la amplitud o variación de temperatura que es calculada por la estudiante, que consiste en la diferencia entre estas dos temperaturas, también se selecciona por formar parte de la tabla. Así que los puntos comunes que vemos en ambos diagramas de caja son respectivamente la precipitación y amplitud de temperatura asociadas a una misma semana.

Seguidamente se muestra lo que ocurre con la tabla de datos al seleccionar esos puntos:

Semana	Temper...	Temper...	Precipitaci...	variación	
1	1-7Enero	26,26	12,47	0,716259	13,79
2	8-14Enero	25,91	12,08	0,2461	13,83
3	15-21En...	26,29	12,07	0,4645	14,22
4	22-28En...	26,63	12,16	0,693859	14,47
5	29-4Enero...	26,96	13,29	0,797659	13,67
6	5-11Febr...	27,52	13,11	0,2441	14,41
7	12-18Fe...	28,49	13,69	0,3181	14,8
8	19-25Fe...	28,66	14,48	0,694559	14,18
9	26-4Feb...	30,15	14,8798	0,3292	15,4702
10	5-11Marzo	30,36	15,2198	0,3976	15,1926
11	12-18Ma...	31,71	15,9698	0,369	15,7402
12	19-25Ma...	32,49	16,4698	0,1841	16,0202
13	26-1Mar...	32,78	17,5598	0,4317	15,2202
14	2-8Abril	34,2493	17,8198	0,725659	16,4295
15	9-15Abril	33,9693	18,6098	0,838759	15,3695
16	16-22Abril	34,4293	19,38	2,0464	15,0493
17	23-29Abril	35,51	20,92	1,8825	14,59
18	30-5Abril	35,81	21,21	1,64196	14,2
19	7-13Mayo	34,04	21,64	2,2253	13,3

Figura 3: Temperaturas y precipitación promedio de la región del Mante en 40 años, por semana

Lorena y Alejandra proponen una solución también correcta, pero, usando una estrategia de análisis diferente a la empleada por Marcela, donde extraen de la tabla de datos original, los datos correspondientes a cada uno de los meses (subtablas). La siguiente figura muestra la estrategia de transnumeración hecha por Lorena:

Enero				
	Semana	Tem_Max	Tem_Min	Precipit...
1	1-7Enero	26,26	12,47	0,6755
2	8-14Enero	25,91	12,08	0,2461
3	15-21En...	26,29	12,07	0,4645
4	22-28En...	26,63	12,16	0,6531

Figura 4: Temperaturas y precipitación registradas en la región del Mante en el mes de enero

Alejandra además de lo anterior, construyó una tabla adicional que muestra los promedios de las variaciones de la temperatura y la precipitación por mes como se muestra a continuación:

	MES	Promedio_d...	Promedio_de...
1	ENERO	14,0775	0,509800
2	FEBRERO	14,2650	0,493225
3	MARZO	15,6560	0,342320
4	ABRIL	15,3300	1,352950
5	MAYO	13,6625	2,647450
6	JUNIO	12,5620	6,405480
7	JULIO	11,3600	6,371550
8	AGOSTO	12,9340	5,382680
9	SEPTIEM...	11,7450	5,904300
10	OCTUBRE	12,7950	2,649880
11	NOVIEM...	13,3320	0,908940
12	NOVIEMBRE	14,0050	0,581225

Figura 5: Temperaturas y precipitación promedios registradas en la región del Mante por mes

Esta tabla le permitió a Alejandra identificar los meses en los que se presentaba la mayor variación de la temperatura y las precipitaciones mínimas. Lorena implementó una tabla similar a la de Alejandra, pero, que a diferencia de ella utiliza las medias de las temperaturas mínimas y las medias de las temperaturas máximas, las medias de las precipitaciones y finalmente la diferencia entre las medias de las temperaturas mínimas y máximas por mes.

Esteban y Sonia construyeron un gráfico de línea para la precipitación pluvial y otro para la variación de la temperatura, como se muestra a continuación:

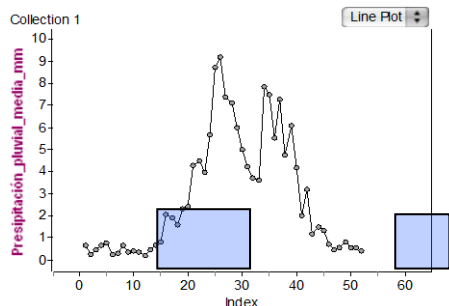


Figura 6: Precipitación pluvial en la región del Mante por semana

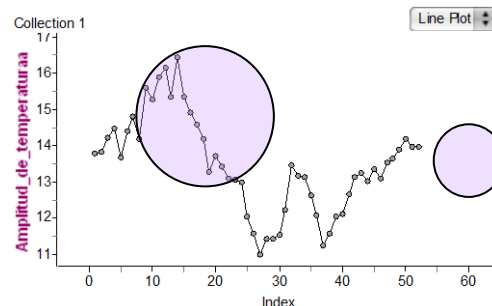


Figura 7: Variación de temperatura en la región del Mante por semana

Este tipo de gráfico mediante una superposición muestra claramente los intervalos donde las precipitaciones mínimas y las variaciones de temperatura máximas convergen, lo que conduce a la propuesta de solución correcta para Esteban y Sonia.

Amelia prefiere trabajar con histogramas, como se muestra seguidamente:

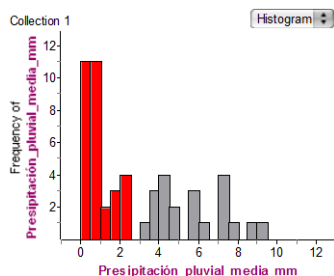


Figura 8: Precipitación pluvial en la región del Mante por semana

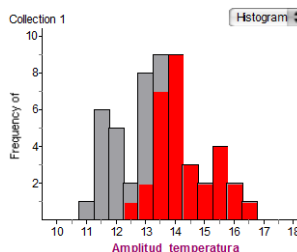


Figura 9: Variación de temperatura en la región del Mante por semana

En este caso se seleccionaron las barras, que representan las semanas donde la precipitación es menor y automáticamente se resaltan las barras en el otro histograma de amplitud de temperatura, donde se aprecia coincidencia con las semanas donde la variación de la temperatura es mayor, esto lleva a Amelia a proponer una solución válida al problema.

Finalmente, queda por analizar el caso de Roger. Este estudiante intentó hacer un análisis por bloques, que consistía en seleccionar la información respectiva de cada mes y luego construyó un diagrama de caja para cada uno de los meses, tanto de temperatura como de precipitación, pero, esta estrategia, según su criterio, no le proporcionó ninguna información útil, por lo que confeccionó su propia tabla de manera manual en papel.

Note que en la tabla él estima la diferencia entre las temperaturas para saber las variaciones entre las medias, de las temperaturas mínimas y máximas de cada mes. Además Determina los meses de mayor precipitación pluvial para descartar esos meses, que en este caso son junio, julio, agosto, septiembre, pero descarta agosto, septiembre, octubre y noviembre, no así junio y julio porque un mes de mucha lluvia afecta hasta dos meses después y en este caso no se descarta porque mayo y abril serían los mese que definen la situación de estos.

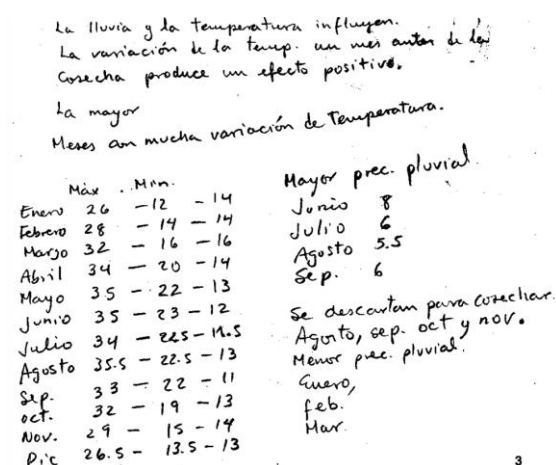


Figura 10: Estrategia propuesta por Roger para resolver el problema

Además percibe que los meses de menor precipitación son enero, febrero y marzo. Finalmente con esta información concluye que: “De acuerdo con los datos obtenidos se concluye que se debe cosechar del 26 de febrero al 3 de mayo”, una buena propuesta.

Tabla 1

Recursos usados por los estudiantes dentro de la estrategia de resolución del problema, según tipo de pensamiento estadístico

Pensamiento Estadístico	Descripción de recursos
Necesidad de los datos	Los estudiantes consultan la tabla de datos y la estudian
Transnumeración	Segmentación de tablas. La tabla disponía de una organización de la información de previo.
Consideración de la Variación	Amplitud de temperatura, identificación de variación de temperatura y precipitación en los modelos.
Razonamiento con modelos	Gráficos de línea, diagramas de caja, histograma, promedios,
Análisis del contexto	Se consideran los supuestos básicos para la zafra al buscar los datos relevantes, se proponen soluciones prácticas en períodos de tiempo.

Conclusiones

La resolución de problemas estadísticos en contextos reales y mediante planteamientos no guiados, muestra la dificultad en los estudiantes para tomar la iniciativa a pesar de tener los conocimientos necesarios, tanto en Estadística como a nivel del instrumento computacional usado. Llama la atención que, si esta situación es evidente en estudiantes que cuentan con niveles de educación formal superior, entonces, se debe revisar lo que está ocurriendo en niveles más bajos de educación formal.

Los resultados finales muestran un nivel de pensamiento Estadístico avanzado y diverso de los sujetos del estudio, caracterizado por los diferentes modelos estadísticos creados por autoría propia y sin variar las condiciones del caso, llegando hasta el nivel más alto de pensamiento integrando el conocimiento del contexto con el conocimiento estadístico. Hay conjeturas hechas que se dan gracias a las bondades del programa Fathom 2.0, no obstante, se rescata el caso de un

estudiante que a pesar de conocer de forma competente este instrumento, persiste en trabajar con sus estrategias formación en estadística tradicional, haciendo uso de lápiz y papel. El sistema educativo en todos los niveles debería considerar dentro de su currículo la enseñanza de la estadística a partir de la incorporación de problemas similares, donde el estudiante disponga de experiencias cercanas al contexto que implica resolver problemas cotidianos con el uso de la estadística y comprender así, su aplicabilidad.

Referencias

- Arteaga, P. (2011). Evaluación de conocimientos sobre gráficos estadísticos y conocimientos didácticos de futuros profesores. *Tesis Doctoral*. Universidad de Granada. Directora: C. Batanero
- Batanero, C. (2001). *Didáctica de la Estadística*. Grupo de Educación Estadística, Universidad de Granada, España.
- Moore, D.S. (1997a). New pedagogy and new content: The case of statistics. *International Statistical Review*, 65(2), 123-165.
- Snee, R. (1990). Statistical thinking and its contribution to quality. *The American Statistician*, 44, 116-121.
- Wild, C & Pfannkuch, M. (1999). Statistical thinking and empirical enquiry. *International Statistical Review*, 67.