

LA ENSEÑANZA Y EVALUACIÓN DEL ANÁLISIS DE DATOS EN LA ASIGNATURA ESTADÍSTICA DE LOS ESTUDIOS DE GRADO EN CIENCIAS AMBIENTALES DE LA UNIVERSIDAD POLITÉCNICA DE VALENCIA

Carmen CAPILLA¹

Departamento de Estadística e Investigación Operativa Aplicadas y Calidad¹

Universidad Politécnica de Valencia¹

Resumen

Las competencias específicas de los estudios de grado en Ciencias Ambientales incluyen la capacidad de interpretación cualitativa y cuantitativa de datos. La aplicación de métodos de análisis de datos es necesaria en los procesos de toma de decisiones y resolución de problemas medioambientales. Por estos motivos el plan de estudios contempla la asignatura Estadística en el bloque de materias instrumentales. Esta comunicación recoge la experiencia en la enseñanza y evaluación de esta competencia en dicha asignatura en la Universidad Politécnica de Valencia. Las actividades que integra tienen como objetivo que el alumno adquiera entre otras, habilidades como tratamiento de datos, interpretación de resultados y manejo de programas estadísticos. Entre estas actividades se incluye la realización de prácticas con los programas SPSS y Statgraphics. Las prácticas con software permiten hacer un mayor énfasis en los conceptos y en la interpretación de resultados, dejando en un segundo plano de importancia la realización manual de cálculos. Los resultados obtenidos con estos dos programas indican que los alumnos tienen menos dificultades con SPSS cuando analizan tablas de frecuencias bidimensionales y cuando aplican métodos de análisis descriptivo univariante. En el análisis descriptivo de datos multivariantes los resultados son mejores con Statgraphics.

Palabras Clave: *competencias, análisis de datos, ciencias ambientales, evaluación, software estadístico.*

1. Introducción

En el proyecto para los estudios de grado de Ciencias Ambientales adaptados al espacio europeo de educación superior, se establecen los objetivos formativos, una vez analizadas y definidas las competencias genéricas o transversales y las específicas del perfil profesional de dicha titulación. Entre las primeras se indican: comunicación oral y escrita, conocimientos de informática relativos al ámbito de estudio, trabajo en equipo y razonamiento crítico. Entre las competencias específicas (saber) se incluye la capacidad de interpretación cualitativa y cuantitativa de datos. La aproximación científica a cualquier cuestión medioambiental y ecológica, requiere la aplicación de metodología científica apropiada para obtener los datos necesarios, así como para asegurar los análisis e interpretaciones correctos [1]. Todo ello hace que en la propuesta del plan de estudios se incluya la estadística aplicada al medio ambiente en el bloque de materias instrumentales. Se señalan también como habilidades o destrezas a adquirir, además de la competencia específica mencionada:

1. Tratamiento de datos.
2. Interpretación de resultados estadísticos.
3. Manejo de programas estadísticos.

El objetivo del presente trabajo es analizar la enseñanza del análisis de datos medioambientales mediante la utilización de software estadístico. La discusión utiliza como caso práctico el manejo de dos programas de software comercial en un curso de estadística aplicada de los estudios de grado en Ciencias Ambientales de la Universidad Politécnica de Valencia (España). El enfoque de

dicha asignatura es expuesto en las siguientes secciones de esta comunicación y trata de formar al alumno en las competencias y habilidades mencionadas en el párrafo anterior.

2. Enseñanza del análisis descriptivo de datos

2.1 Descripción del contexto

La titulación de Ciencias Ambientales de la Universidad Politécnica de Valencia se implantó en el año 1997 como una enseñanza de segundo ciclo. Está diseñada con un enfoque interdisciplinar. A ella pueden acceder alumnos que hayan cursado un primer ciclo de ingeniería o estudios de grado en Biología, Química, etc. Actualmente la titulación está adscrita a la Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos. Uno de los objetivos principales de dicha titulación es preparar a los estudiantes para sus futuras carreras profesionales en la gestión y toma de decisiones en medio ambiente. Por tanto, el plan de estudios incluye la asignatura estadística, con el fin de formar al alumno en el método científico para obtener y tratar la información necesaria para resolver cuestiones medioambientales.

La formación previa del estudiante es uno de los factores que más pueden influir en el aprendizaje de la estadística [2]. Todos los alumnos que se matriculan en el curso de estadística de Ciencias Ambientales han estudiado cálculo y álgebra en sus primeros años en la universidad. Un porcentaje elevado (cerca al 70%) ha tenido también un curso de introducción a la estadística centrado sobre todo en cálculo de probabilidades. La mayoría no tiene experiencia previa en el análisis estadístico de datos. Teniendo en cuenta estas características y la heterogeneidad de los alumnos, la asignatura estadística de Ciencias Ambientales se inicia con la enseñanza de técnicas para análisis descriptivo de datos. En las primeras clases se introducen los conceptos básicos de este tipo de análisis en los casos univariante y multivariante.

La asignatura dispone de 60 horas presenciales. Estas horas se desglosan de la manera siguiente: 30 horas son para la exposición en el aula de los conceptos teóricos apoyándose en casos prácticos medioambientales. Diez horas se dedican a la resolución de ejercicios y actividades en el aula. La 20 horas restantes son para la realización de prácticas en aula informática. Estas 20 horas se reparten en diez sesiones de 2 horas cada una a lo largo del cuatrimestre.

2.2 Organización de las prácticas

La asignatura incorpora como actividades importantes las prácticas en aula informática. La clase se divide en cuatro subgrupos. En cada uno los estudiantes trabajan en equipos de 2 a 3 personas, que se constituyen al inicio del cuatrimestre y que son fijos durante el mismo.

Cada clase en el aula informática consta de dos partes. En la primera el profesor explica el manejo de las opciones del software que permiten aplicar las técnicas de análisis vistas en las clases de teoría. Se utiliza un cañón de video para proyectar las pantallas de resultados del software. Los alumnos siguen la explicación realizando al mismo tiempo los análisis en el PC que tienen asignado. En esta explicación inicial habitualmente se resuelve algún ejemplo con datos medioambientales reales que ya se han utilizado en las clases de teoría. Entre estos ejemplos se encuentran algunos en los que se analiza la base de datos de contaminantes atmosféricos de Valencia [3], disponible a través de internet. En otros casos se analizan datos climatológicos proporcionados por el Instituto Nacional de Meteorología y extraídos de las referencias [4], [5] y [6].

En la segunda parte de la práctica los equipos de estudiantes resuelven un problema similar con cuestiones que requieren el análisis de datos reales utilizando software estadístico. Cada grupo debe presentar en un plazo máximo de quince días desde la realización de la práctica, un informe escrito con las conclusiones más relevantes de los análisis realizados en esta segunda parte. Se hace especial hincapié en la interpretación de los resultados que proporcionan los programas estadísticos. Los informes de las prácticas son un feedback importante sobre las dificultades y los avances de los alumnos. Se utilizan como evaluación continua. La calificación media obtenida en las diez sesiones de aula informática tiene un peso del 20% en la nota final. La asistencia a las prácticas cuenta un 10% adicional en dicha nota.

2.3 Prácticas de análisis descriptivo con Statgraphics y SPSS

Los programas utilizados actualmente en las prácticas son Statgraphics y SPSS, que se encuentran disponibles con licencia campus en la Universidad Politécnica de Valencia. Hasta el año 2004 las prácticas se realizaban sólo con Statgraphics, según hace la mayor parte de las materias del departamento al que está adscrita la asignatura que describimos en esta comunicación. Statgraphics es un programa de manejo sencillo, que funciona en el entorno windows, y que es apropiado para alumnos en los primeros cursos en la universidad. Teniendo en cuenta que un elevado porcentaje de los alumnos de la titulación en Ciencias Ambientales ya ha cursado previamente una estadística con manejo de Statgraphics, se decidió complementar su formación con el manejo de SPSS. SPSS es un programa potente, utilizado en el mundo profesional.

2.3.1 Ejemplos de ejercicios de las prácticas

El primer curso en el que se empezó a trabajar con los dos programas se realizaron sólo dos prácticas con SPSS. Concretamente las que correspondían a la primera parte del programa de la asignatura: análisis descriptivo univariante (tablas de frecuencias, gráficos, cálculo de parámetros muestrales) y multivariante (diagramas de dispersión, covarianza y correlación). Se hicieron un total de cuatro sesiones en aula informática dedicadas a esta materia. En las dos primeras sesiones los alumnos trabajaron con Statgraphics: la práctica 1 requería la aplicación de técnicas de análisis univariante con este programa, y la práctica 2 la aplicación de métodos descriptivos multivariantes con el mismo. En las dos sesiones siguientes tuvieron que utilizar SPSS para aplicar métodos descriptivos univariantes (práctica 3) y multivariantes (práctica 4). El nivel de dificultad de los ejercicios realizados con ambos programas fue similar. A continuación se recogen dos ejemplos de los análisis que tuvieron que hacer en esas prácticas.

Ejemplo 1: Comparación a nivel descriptivo de la concentración de ozono en tres zonas de la ciudad de Valencia.

Se dispone de observaciones de la concentración de ozono (ppb) registrada los viernes en tres estaciones de la red automática de medición de la contaminación atmosférica en Valencia [3].

Los datos están en el fichero SPSS OZONO.SAV. Para abrir el fichero puedes utilizar desde la barra de menú las instrucciones SPSS:

Archivo....Abrir....Datos

Nombre: OZONO

La variable OZONO de este fichero (disponible en el editor de datos) contiene 277 valores de la concentración de este contaminante en tres estaciones distintas. La variable ESTACIÓN da los códigos de la zona de muestreo (1, 2 ó 3) correspondiente a cada valor de ozono.

Compara descriptivamente la posición, dispersión y forma de la distribución de la concentración de ozono entre las tres estaciones. Para ello utiliza el cálculo de parámetros muestrales y las representaciones gráficas de los histogramas y el diagrama box-whisker múltiple. En la respuesta aporta los resultados numéricos e interpretaciones de los mismos y los gráficos que justifiquen los resultados.

Para hacer el análisis que se pide en el párrafo anterior puedes utilizar la opción:

Analizar....Estadísticos descriptivos....Explorar

Dependientes: OZONO

Factores: ESTACION

Ejemplo 2: Análisis descriptivo de la relación existente entre parámetros climáticos.

[6] recoge observaciones de temperatura en °C (T), precipitación en mm (Pcion) y presión en mbar (P), del mes de junio en una zona de la costa del Pacífico. Los años de muestreo son de 1951 a 1970. En 1951, 1953, 1957, 1965 y 1969 se presentó el fenómeno de El Niño.

Los datos están en el fichero SPSS GUAYAQUIL.SAV. Se puede acceder a ellos desde este programa con:

Archivo....Abrir....Datos

Nombre: GUAYAQUIL

Estudia mediante la matriz de diagramas de dispersión las pautas y relaciones entre las tres características.

Esta representación gráfica la puedes obtener con:

Gráficos....Dispersión....Matricial

Variables en la matriz: Temperatura en °C [T]

Precipitación en mm [Pcion]

Presión en mb [P]

Calcula las matrices de varianzas-covarianzas y de correlaciones. Relaciona los valores de estos parámetros con las pautas observadas en la matriz de diagramas de dispersión.

El cálculo de dichas matrices se puede realizar con:

Analizar....Correlaciones....Bivariadas

Variables: Temperatura en °C [T]

Precipitación en mm [Pcion]

Presión en mbar [P]

Los dos ejemplos anteriores se han extraído de los guiones de prácticas 3 y 4 que se hicieron el primer año en que se empezó a utilizar también el programa SPSS.

2.3.2 Resultados de las prácticas

La Tabla 1 recoge los valores medios y las desviaciones típicas de las calificaciones obtenidas en las cuatro primeras prácticas, el primer año en que se empezaron a utilizar los dos programas Statgraphics y SPSS.

Práctica	Media	Desviación típica
1: Análisis descriptivo univariante con Statgraphics	6.36	1.61
2: Análisis descriptivo multivariante con Statgraphics	8	1.05
3: Análisis descriptivo univariante con SPSS	7.34	1.79
4: Análisis descriptivo multivariante con SPSS	7.19	1.36

Tabla 1 Resultados de los alumnos (42 grupos) en las cuatro primeras prácticas

Un análisis detallado de las calificaciones obtenidas en las cuatro primeras prácticas aportó la siguiente información. La única correlación significativa fue la estimada entre las notas de las prácticas 1 y 2 (valor 0.51). La comparación de las calificaciones medias de las prácticas con el análisis de datos apareados arrojó los resultados:

Comparación práctica 2-práctica 1: Diferencia significativa. El intervalo de confianza para la diferencia de notas medias al 95% es [1.03, 2.03]. Esto indica que los alumnos obtuvieron unas

calificaciones medias significativamente superiores cuando aplicaron con Statgraphics métodos descriptivos multivariantes respecto a los métodos univariantes.

Comparación práctica 4-práctica 3: No hay diferencia significativa en las notas medias. Trabajando con SPSS la aplicación de métodos univariantes y multivariantes presentó el mismo nivel de dificultad.

Comparación práctica 3-práctica 1: Diferencia significativa. El intervalo de confianza para la diferencia de notas medias al 95% resulta [0.18, 1.77]. Refleja que cuando los alumnos aplicaron métodos de análisis univariante tuvieron menos dificultades con el programa SPSS respecto a Statgraphics.

Comparación práctica 2-práctica 4: Diferencia significativa. El intervalo de confianza para la diferencia de medias entre ambas prácticas, al 95% es [0.41, 1.35]. Con Statgraphics cuando aplicaron método descriptivos multivariantes los alumnos cometieron menos errores y presentaron mejores informes que con SPSS.

2.4 Evaluación del nivel de comprensión de conceptos de análisis descriptivo de datos

Con el fin de evaluar el nivel de comprensión de los conceptos básicos de análisis descriptivo de datos, se realizó en el curso 2006 un estudio utilizando algunos items del cuestionario descrito en [7]. Este cuestionario es conocido como "Statistical Reasoning Assessment Test" (SRAT) y se ha aplicado en diversos contextos para evaluar el nivel de razonamiento estadístico de estudiantes tras un curso introductorio en estadística. El estudio consistió en pasar a principio de curso, antes de ver la materia de análisis descriptivo de datos, un test con los items 1, 4, 5, 14, 15, 16 y 17 de SRAT. Estos items están relacionados con el tema de análisis descriptivo de datos. El mismo test se pasó después de explicar en teoría y de realizar las prácticas de aula informática de dicho tema.

La Tabla 2 muestra los resultados obtenidos con los items seleccionados, en relación con el razonamiento sobre diversos conceptos, antes y después de la realización de todas las actividades del curso sobre análisis descriptivo. Se observa que el razonamiento sobre parámetros de posición como la media mejoró después de dichas actividades. También mejoró la interpretación de tablas de frecuencias cruzadas.

	Antes (n=34)	Después (n=22)
Interpreta correctamente tablas de frecuencias cruzadas	1.09 (0.17)	1.39 (0.19)
Error: la media es el número más frecuente	17.1% (0.06)	13% (0.07)
Error: no tiene en cuenta la presencia de "outliers" cuando calcula la media	5.7% (0.04)	0%
Error: confunde media con mediana	2.9% (0.03)	0%
Error: conclusiones e inferencia sobre la población basándose en muestras pequeñas	1.43 (0.15)	1.22 (0.21)

Tabla 2 Puntuaciones y porcentajes obtenidos al razonar sobre diversos conceptos de análisis descriptivo de datos.

3. Conclusiones

Entre las competencias específicas del grado en Ciencias Ambientales se incluye la capacidad de interpretación cualitativa y cuantitativa de datos. Por dicho motivo el plan de estudios correspondiente incluye la asignatura estadística, con el fin de formar a los futuros profesionales

en técnicas que permitan obtener y analizar de forma científica la información necesaria para la toma de decisiones en cuestiones medioambientales. El presente trabajo analiza concretamente la formación en métodos estadísticos para análisis descriptivo de datos medioambientales, mediante la realización de prácticas con software. El objetivo de dichas prácticas es que el alumno adquiera habilidades en el tratamiento de datos, en la interpretación de resultados de análisis estadísticos y en el manejo de programas estadísticos.

Es muy frecuente en sistemas medioambientales la disponibilidad de técnicas de medición que proporcionan muestras de gran tamaño, lo que hace prácticamente imposible la realización manual de cálculos. Por otro lado, los métodos gráficos para visualizar la información son cada vez más complejos. Por todo ello, hoy en día, la formación de especialistas en medio ambiente requiere la incorporación de software de análisis de datos. En esta comunicación se presentan algunos resultados de la utilización de los programas estadísticos Statgraphics y SPSS., y se compara la evaluación de los alumnos con ambos programas al aplicar métodos de análisis descriptivo univariante y multivariante. La heterogeneidad de conocimientos previos y las características de los alumnos influyen en las calificaciones obtenidas en las actividades que se realizan con los dos programas. La comparación de los mismos indica que los alumnos tuvieron menos dificultades con SPSS cuando aplicaron técnicas univariantes. En el análisis descriptivo de datos multivariantes los resultados fueron mejores con Statgraphics. La realización de una prueba con preguntas de respuesta múltiple antes y después de las actividades relacionadas con la enseñanza del análisis descriptivo de datos, indicó que mejoró el razonamiento sobre parámetros de posición como la media y la interpretación de tablas de frecuencias cruzadas. Los resultados que se recogen en este trabajo aportan una información útil como punto de partida para los cambios que requiere la adaptación al espacio europeo de educación superior.

4. Referencias

- [1] BARNETT, V. *Environmental Statistics*. New York: Wiley, 2004.
- [2] BATANERO, C.; GODINO, J.D.; VALLECILLOS, A.; GREEN, D.R.; HOLMES, P. "Errors and difficulties in understanding elementary statistical concepts". *International Journal of Mathematical Education in Science and Technology*. Vol. 25, No. 4, 1994, pp. 527-547.
- [3] Laboratorio de Medio Ambiente. *Base de datos de contaminantes atmosféricos de Valencia*. Ayuntamiento de Valencia, 2008.
- [4] BERTHOUEX, P.M.; BROWN, L.C. *Statistics for Environmental Engineers*. Boca Ratón, FL: Lewis Publishers, 2002.
- [5] PIEGORSCH, W.W.; BAILER, A.J. *Statistics for Environmental Biology and Toxicology*. Boca Ratón, FL: Chapman and Hall, 1997.
- [6] WILKS, D.S. *Statistical Methods in the Atmospheric Sciences*. New York: Academic Press, 2006.
- [7] GARFIELD, J.B. "Assessing statistical reasoning". *Statistical Education Research Journal*. Vol.2, No. 1, 2003, pp.22-32.