

COMPRESIÓN DE LAS DISTRIBUCIONES MUESTRALES EN UN CURSO DE ESTADÍSTICA PARA INGENIEROS

UNDERSTANDING OF SAMPLE DISTRIBUTIONS FOR A COURSE ON STATISTICS FOR ENGINEERS

Lidia Retamal P.¹ Hugo Alvarado M.¹ Rodrigo Rebolledo V.²

Recibido 10 de abril de 2006, aceptado 2 de marzo de 2007

Received: April 10, 2006 Accepted: March 2, 2007

RESUMEN

Se describe una propuesta didáctica de enseñanza contextualizada y con uso del programa @risk de las distribuciones muestrales en el aula de estadística para ingenieros de segundo año universitario. Apropiándonos de la teoría de las funciones semióticas, desarrollada en la Universidad de Granada-España, caracterizamos los elementos de significados de las propiedades importantes de las distribuciones muestrales y evaluamos, mediante campos de problemas algebraicos y de simulación, los errores o dificultades que los alumnos ponen de manifiesto en las aplicaciones de simulación de procesos en las ciencias de la ingeniería. Como consecuencia, al considerar los elementos de significados adquiridos en las respuestas de los estudiantes, proponemos la simulación para muestras pequeñas y grandes de forma intuitiva, como primer acercamiento del alumno hacia la construcción del significado de las distribuciones muestrales, usando el lenguaje gráfico con apoyo del computador, para posteriormente analizar con los estudiantes su forma algebraica según la naturaleza de las variables aleatorias.

Palabras clave: Enseñanza y aprendizaje de la estadística, distribuciones muestrales, significado y comprensión, simulación.

ABSTRACT

In this work we describe a contextual didactic approach using the software @risk, for teaching "sample distributions" in a Statistics course for engineers. Using the theory of semiotic functions, developed by Universidad de Granada in Spain, we characterize the meaning elements of the main properties of sample distributions. Then, using algebraic and simulation problems, we determine and evaluate the different errors and difficulties that emerge when the students simulate processes in Engineering. After considering the meaning elements obtained from the students' answers, we propose simulation as a first approach towards the construction of the meaning for sample distributions for small and big samples, using intuitive forms by means of a graphic language via computer support, so as to analyze with the students the algebraic form, depending on the nature of the random variables.

Keywords: Teaching and learning of Statistics, sample distributions, meaning and understanding, simulation.

INTRODUCCIÓN

Una problemática generalizada en las aulas universitarias es determinar y secuenciar los elementos prioritarios que debe considerar el profesor para una buena enseñanza de conceptos y proposiciones matemáticas. Esto nos conduce a reflexionar acerca de si en la clase de estadística se favorece la comprensión de conceptos y si buenas presentaciones en el aula consideran más de una forma de acercarnos a los enunciados estadísticos o simplemente llegan a

los alumnos como un producto dogmático y acabado [1]. En la enseñanza actual de la estadística aumenta el interés didáctico sobre la enseñanza de las Distribuciones Muestrales, en adelante DM, donde enunciados como el teorema central del límite se considera frecuentemente en situaciones profesionales a la ingeniería y como herramienta de investigación.

Entendiendo la Estadística como la ciencia que comprende la recopilación, tabulación, análisis e interpretación de

¹ Área de Ciencias Básicas, Universidad Católica de la Santísima Concepción. Concepción, Chile. lretamal@ucsc.cl, alvaradomartinez@ucsc.cl

² Facultad de Ingeniería, Universidad Católica de la Santísima Concepción. Concepción, Chile. rrebolledo@ucsc.cl

los datos cuantitativos y cualitativos, se pretende que los estudiantes de ingeniería adquieran, mediante la correcta aplicación de propiedades importantes de las DM a situaciones de la ingeniería, razonar a partir de los datos empíricos, considerar las variaciones de los tamaños muestrales y evaluar el riesgo al tomar decisiones en ambiente de incertidumbre. Este trabajo de diseño y aplicación de las DM conjuga el conocimiento matemático en sí mismo con la teoría didáctica y los antecedentes de investigaciones sobre el tema. Se destaca que, en la línea actual de la docencia matemática universitaria de pregrado, se incorporan recursos computacionales a la acción didáctica. Elegimos el programa de simulación estadística @risk para Microsoft Excel, debido a que el uso de los computadores debe reforzar el aprendizaje de los conceptos a tratar, al proporcionar varias representaciones aportando al significado del tema.

En esta propuesta se analizan algunas propiedades fundamentales de las DM en una muestra de libros de estadística aplicada, usados en la enseñanza a ingenieros en Chile, que varían desde su enunciado más simple a la presentación formal avanzada. Nuestra metodología se inspira en el marco teórico de las funciones semióticas, desarrollado en la Universidad de Granada, donde se consideran los objetos matemáticos como emergentes de la actividad de resolución de campos de problemas y se diferencian diversos elementos en su significado. Durante el primer semestre académico 2005 se analizó el diseño de un proceso de estudio en el que pretendemos en el aula de estadística la apropiación de conceptos principales de las DM por parte de los alumnos, plasmado en una correcta aplicación en diversos elementos de significados de las DM en la resolución de problemas propios de la ingeniería. Una vez establecida la metodología de enseñanza de las DM se llevó a cabo su aplicación en el segundo semestre y evaluación de los errores y dificultades en las respuestas de los estudiantes a los diversos problemas algebraicos y de simulación de las DM.

Considerando las investigaciones relacionada con su comprensión, se presentan algunos campos de problemas contextualizados que requieren de las DM como herramienta de análisis de datos en ingeniería e intentamos responder algunas cuestiones de investigación que orientan y fundamentan el diseño del proceso de estudio de las distribuciones muestrales, tales como: ¿es posible introducir la simulación en conceptos de las distribuciones muestrales, debido a la gran complejidad de su aparato analítico?, ¿los estudiantes aplican correctamente los diversos elementos de significados de propiedades en la resolución de problemas?, ¿adquieren destrezas en plantear un diseño muestral?

FUNDAMENTOS DEL ESTUDIO

Se describe el marco teórico empleado, que ha situado los elementos de significado institucional y personal de la distribución muestral y a continuación algunas investigaciones previas relacionadas con sus propiedades y la simulación en ingeniería.

Significado y comprensión de la distribución muestral

En el marco teórico considerado [9-11] las matemáticas se asumen como una actividad humana implicada en la solución de cierta clase de situaciones problemáticas de la cual emergen y evolucionan progresivamente los objetos matemáticos. Se pretende elaborar un modelo de los procesos de comprensión de las matemáticas que tenga en cuenta los factores institucionales y socioculturales implicados en los mismos. El autor considera diferentes entidades primarias como constituyentes del significado de un objeto matemático (por ejemplo, distribución muestral), que son las que se analizan en este trabajo:

- *Problemas y situaciones* que inducen actividades matemáticas y definen el campo de problemas asociado al objeto.
- *Procedimientos, algoritmos, operaciones*. Cuando un sujeto se enfrenta a un problema y trata de resolverlo, realiza distintos tipos de *prácticas*, que llegan a convertirse con el tiempo en objeto de enseñanza.
- *Representaciones* materiales utilizadas en la actividad de resolución de problemas (términos, expresiones, símbolos, tablas, gráficos).
- *Abstracciones* (conceptos, proposiciones). Las definiciones y propiedades características del objeto y sus relaciones con otros conceptos.
- *Demostraciones* que empleamos para probar sus propiedades y que llegan a formar parte de su significado.

Importancia de las distribuciones muestrales en la ingeniería

La importancia de la estadística en la ingeniería ha sido subrayada por la participación de la industria en el aumento de la calidad, ya que las técnicas estadísticas pueden emplearse para describir y comprender la variabilidad, que es el resultado de cambios en las condiciones bajo las que se hacen las observaciones [14]. La simulación en la industria es apropiada en situaciones complejas donde el tamaño del problema dificulta el uso de técnicas analíticas [6]. La simulación es el proceso de diseñar un modelo de un sistema real y llevar a cabo experiencias

con él a través del computador, con la finalidad de estudiar el comportamiento de sistemas y evaluar las estrategias para organizar el funcionamiento del modelo que representa [4]. En este trabajo consideramos una simulación simplemente como un experimento estadístico de muestreo [15], que plantea un problema del análisis estadístico de las muestras. Según Vallecillos [18] muchos alumnos tienen dificultad con el cálculo de probabilidades y, sin embargo, es necesario que adquieran comprensión y competencia en las diferentes distribuciones de probabilidad y en las nociones básicas de muestreo. Estas dificultades de comprensión de las mismas podrían influir en los errores de aplicación a la ingeniería en tópicos de estimación por intervalos o los contrastes de hipótesis.

Partiendo de la base que las distribuciones muestrales son el puente entre las distribuciones de probabilidades y la inferencia estadística, colegas de ingeniería nos indicaron posibles áreas específicas de aplicación de procesos de simulación mediante la estimación de distribuciones de probabilidades de la media muestral. Algunas situaciones problemas a la ingeniería de donde surge la distribución muestral son:

- a. Procesos de llegada: clientes a un servicio, piezas a una máquina, vehículos a un semáforo, clientes a un hospital, buses a un paradero, etc. En este caso la variable aleatoria representa el tiempo entre llegadas y las distribuciones más utilizadas son: uniforme y de la familia gamma la exponencial y erlang.
- b. Datos de desembarque: simulación de los desembarques y sus utilidades, de días con pesca, de los tiempos de proceso de fileteado, congelado, etc. (generalmente hay colas o tiempos de espera en cada etapa del proceso).
- c. Diseño de edificios: las cargas máximas a la que están sometidos no se conoce de manera exacta, por lo que se puede considerar una variable aleatoria. De igual forma, la resistencia de los materiales tampoco se conoce de manera exacta, ya que en el proceso de fabricación, transporte y montaje pueden ocurrir solicitaciones no calculadas.
- d. Duración de servicios: tiempo en procesar una pieza, tiempo de atención en un hospital o un cajero automático. Distribuciones continuas son usadas, como gamma, weibull, lognormal.
- e. Número de piezas o fracción de unidades falladas en una muestra (muestreo de calidad utilizando la distribución binomial o beta, respectivamente).

Investigaciones previas sobre las distribuciones muestrales

Las investigaciones sobre el tema buscan medios didácticos que hagan asequibles estos conceptos a los alumnos, en que muchos métodos estadísticos requieren la condición de normalidad para su correcta aplicación, puesto que la omnipresencia de la distribución normal ha hecho que se hayan desarrollado muchos más métodos basados en dicha distribución. Sin embargo, los estudios sobre la comprensión de las distribuciones muestrales no han sido tratados con estudiantes de ingeniería.

Well, Pollatsek y Boyce [19] realizaron tres experimentos con estudiantes de psicología sobre la comprensión de las distribuciones de la media en el muestreo, cuyo fin era identificar las razones que conducen a utilizar la heurística de la representatividad y los diversos juicios que realizan los sujetos al justificar sus respuestas. En general, los sujetos parecen comprender que los promedios de muestras más grandes se acercan más a la media de la población, pero no comprenden las implicaciones de esto sobre la variabilidad de la media muestral. La variable más importante que influye en el éxito de la tarea es la similitud entre la media muestral y poblacional. Méndez [13], al comparar las creencias sobre el teorema central del límite en estudiantes de doctorado y alumnos noveles, representó el conjunto de conocimientos implícito en el teorema por medio de un mapa conceptual, utilizando cuatro propiedades básicas que deben entenderse para poder lograr una comprensión sólida del teorema: a) La esperanza de la distribución muestral de la media muestral es igual a la media de la población, a medida que el tamaño de la muestra tiende al infinito; b) La varianza de la distribución muestral es menor que la de la población (cuando $n > 1$); c) La forma de la distribución muestral tiende a ser acampanada a medida que se incrementa el tamaño muestral, y aproximadamente normal, independientemente de la forma de la distribución en la población; d) La forma de la distribución muestral crece en altura y decrece en dispersión a medida que el tamaño muestral crece.

Tanto la investigación anterior como la de Vallecillos [17] indican que estas propiedades no son sencillas. En la investigación de esta última autora se observa falta de comprensión del efecto del tamaño de la muestra sobre la variabilidad de la distribución muestral y confusión entre la media de la población (parámetro) y la media muestral. Estas mismas dificultades son observadas por delMas, Garfield, y Chance [7] tras una enseñanza basada en la simulación, utilizando un programa de elaboración propia. Los autores sugieren que la tecnología por sí sola no es suficiente para la comprensión, sino que las

actividades y la enseñanza de tipo constructiva tienen un rol importante. Otros trabajos recientes de propiedades importantes relacionados con las DM son el de Tauber [16], quien describe los múltiples elementos de significado relacionados con la distribución normal y Alvarado [2] sobre el significado del teorema central del límite.

METODOLOGÍA

Esta es una investigación que analiza principalmente variables cualitativas, como los elementos de significados institucional y personal de las distribuciones muestrales y los errores presentados en un curso de estadística a nivel universitario. Nuestro trabajo contiene componentes teóricos y aplicados [5]. Consta de un estudio cuasi-experimental. Por su objetivo sería una investigación evaluativa, según las fuentes de datos, empírica y por su enfoque, descriptiva y exploratoria [12].

Población de estudio

La población objetivo fueron 134 estudiantes de ingeniería de cinco especialidades (Acuicultura y Pesca, Portuario, Informático, Industrial y Civil) que cursan Estadística. La asignatura impartida en la Facultad de Ingeniería de la Universidad Católica de la Santísima Concepción es de carácter mínimo del área de Ciencias Básicas y tiene prerrequisito el curso de Probabilidades. Por tanto, los alumnos tienen conocimiento de estadística descriptiva, cálculo de probabilidades, variables aleatorias, distribuciones de probabilidades clásicas y distribuciones bivariadas. Además, han trabajado con los procesadores de planilla Excel y de texto Word.

La inscripción de los alumnos en el curso la realizaron por Internet (requisito es haber aprobado el curso de probabilidades), de manera que una vez que se completa una sección debe incorporarse a la que está con cupos disponibles. Por lo tanto, los alumnos participantes se dividieron de forma aleatoria en dos secciones de 55 y 79 alumnos.

Desarrollo de la enseñanza

La planificación de la secuencia de enseñanza de la distribución muestral se concluyó al final del primer semestre (agosto); considerando, por ejemplo, el tiempo de ejecución, las condiciones de recurso tecnológico disponible en el aula y el número de alumnos inscritos, y se finalizó con la aplicación y evaluaciones del proceso de estudio en diciembre. Las actividades desarrolladas fueron:

a. Se establecieron los contenidos propios que abarca el tema de las distribuciones muestrales; revisando 10 libros de textos de estadística matemática y aplicada a la ingeniería que tratan el tema analizando la estructura de cómo presentan el tema, su secuencia y contenidos previos. Hemos determinado que un 80% de los textos conducen el proceso de estudio de DM y luego la estimación puntual, lo cual discutimos y modificamos debido a nuestra experiencia docente, favoreciendo la base de conceptos previos. Usualmente vemos las distribuciones muestrales, estimación puntual de parámetros y luego intervalos de confianza, lo que hace verse de forma un tanto discontinuada. La propuesta es comenzar con las propiedades deseables de estimadores y criterios para encontrar buenos estimadores (media muestral), seguido de encontrar la función de distribución de los estimadores, que nos permitirá conocer mejor las características de los estimadores utilizados.

b. Se seleccionaron las categorías de los elementos de significados de las DM; determinando los campos de problemas específicos de las DM que vamos a desarrollar, apoyado en las investigaciones [16, 2, 3].

c. La lección (desarrollada en cada Unidad con teoría, propiedades y actividades) de la DM para estudiantes de ingeniería considera la siguiente secuencia:

Unidad 0. Contenidos previos. Introducción y conceptos básicos: definición de estadística, muestreo estadístico, población, muestra.

Estimación puntual de parámetros: Muestra aleatoria, estimadores, propiedades de los estimadores, métodos de estimación.

Unidad 1. Distribución de los estimadores de poblaciones con distribución Normal en una y dos muestras. Distribución muestral de la media muestral y sus aplicaciones de proximidad al parámetro que estima calcular el tamaño de muestra necesario dada una probabilidad de cercanía. Obtener un intervalo de confianza para el parámetro. Calcular probabilidades de estimadores de la media y varianzas muestrales, usando tablas estadísticas de las distribuciones Chi-cuadrado, t-student y de Fisher.

Unidad 2. Distribución de los estimadores de variables discretas de poblaciones no Normales en una y dos muestras

Unidad 3. Distribución de los estimadores de variables continuas de poblaciones no Normales en una y dos muestras

d. Se establecieron las funciones del docente y del alumno; evaluando la adecuación del tiempo disponible de la ejecución de la lección y de los conocimientos previos. En el proceso de enseñanza el profesor guía a los estudiantes a definir un problema, llevar a cabo un plan de diseño muestral y su administración de la información, recolección de datos y su exploración, plantear hipótesis, y finalmente resultados, conclusiones y comunicación. Se destacan conceptos importantes mostrando su necesidad en la presentación de algunos problemas con base de datos usando el computador (comparando de forma visual los datos empíricos con la distribución normal), seguido de su formalización y aplicación de tipo algebraica. El profesor conduce a los estudiantes en la aplicación de resolución de un problema ingenieril. Los alumnos intentaron resolver, individualmente y en grupos de dos, los problemas dirigidos a descubrir la necesidad de uso de algunas propiedades de la DM y luego el profesor guía al grupo a la solución correcta, destacando las actividades de cálculo e interpretación de aplicaciones a la ingeniería.

e. La experiencia de aula se desarrolló en la asignatura de Estadística, con inscripción inicial de 134 alumnos de ingeniería y divididos en dos secciones. Las dos cátedras fueron realizadas por profesores de jornada completa y las prácticas por profesores con algunos años de experiencia práctica en la docencia de estadística. Las sesiones de clase en el aula estuvieron en función de resolución de campos de problemas, partiendo de situaciones concretas y luego a situaciones de la ingeniería. Previo a la evaluación en el laboratorio de una situación problema, estaba disponible en la plataforma un demo del programa @risk con los pasos técnicos de construcción de un histograma y la generación de números aleatorios en forma de matriz; se destinó una tarde de asesoría por parte de un profesor a consulta de práctica en el laboratorio y una semana a consulta vía correo electrónico o en oficina.

f. El tiempo dedicado en el aula al estudio de la DM fue de cuatro semanas (16 módulos de 80 minutos) y cuatro módulos en el laboratorio de computación. Se llevaron a cabo cuatro sesiones por semana, 3 de cátedra y una de práctica. En dos sesiones paralelas de cátedra se trabajó con apoyo de notebook y cañón multimedia y la otra sesión con el pizarrón.

g. Se utilizó el paquete estadístico @risk del programa Excel para el trabajo final de aplicación a la ingeniería.

h. Se evaluó al final de cada semana y al final del semestre académico, aplicando una prueba escrita a ambas secciones.

i. Se utilizó como apoyo a la gestión administrativa del curso, la plataforma de entorno virtual de aprendizaje EV@ de la dirección de Docencia y en el laboratorio la planilla Excel y procesador de texto Word para los informes de las tareas. El material escrito estuvo disponible a los estudiantes en la plataforma EV@ previo a cada unidad.

j. Cada tres clases se recogían las respuestas escritas de los alumnos a los controles propuestos. Por la limitación de espacio describimos sólo dos evaluaciones aplicadas.

Instrumentos de recolección de la información

Se utilizaron dos instrumentos de recolección de datos para la evaluación en estadística, que recogen la lectura directa de datos y reconocimiento de propiedades clasificando las respuestas como correctas o incorrectas, y la capacidad de análisis de datos reales, las estrategias utilizadas [8]. Para medir la eficacia docente del profesor o *qué comprenden* los estudiantes de la DM, consideramos la evaluación en el estudio como la correspondencia entre el *significado institucional* presentado en la enseñanza y el *significado personal* efectivamente construido por los alumnos [9], identificando sus dificultades y errores.

Diseño de prueba conceptual y de desarrollo algebraico.

Presentamos el instrumento en la Tabla 1, con ítems referidos a escribir una síntesis de enlaces de contenido, ítems de tipo verdadero/falso con justificación en caso de responder falso, de selección múltiple y finalmente de desarrollo algebraico de cálculo de probabilidades de estimadores. El alumno debe relacionar varios conceptos (ítem 1) y expresarse en términos de lenguaje simbólico (ítem 2a). El objetivo de los ítems 3 y 4 es determinar la comprensión de conceptos y aplicar propiedades de las DM. También se ha medido en problemas de desarrollo (ítems 5 y 6), definir la variable aleatoria, calcular probabilidades de estimadores y su interpretación (ítem 2b), mediante el comportamiento de la distribución de probabilidades de la media muestral y varianza muestral. Analizar patrones de variabilidad de algunas distribuciones clásicas de probabilidades (normal, uniforme) de valores observados para distintos tamaños muestrales. Contestaron 134 alumnos los ítems 1, 2 y 5; 106 los ítems 3 y 4; 129 el ítem 6.

Diseño de prueba de aplicación contextualizada con programa estadístico.

Otro instrumento utilizado, que evalúa la comprensión de información de datos y capacidad de análisis y síntesis, fue las respuestas escritas de los estudiantes a una actividad práctica individual propuesta al final del semestre, de desarrollo de un informe

investigativo de un problema a la especialidad aplicando la DM mediante la simulación con apoyo del programa @risk de Excel. En un plazo de dos semanas el alumno debería utilizar las propiedades de la DM a un problema de ingeniería, definiendo la variable de estudio sea discreta o continua, objetivos y referencias (búsqueda de problemas en las tesis de ingenieros egresados, libros de ciencias de la ingeniería, o recolección personal de bases de datos), incorporar tablas de datos y gráficas como el histograma, obtener la distribución de probabilidad clásica asociada, estimar la solución mediante la simulación para valores distintos de muestras y las conclusiones.

ANÁLISIS DE LOS RESULTADOS

De acuerdo al marco teórico del significado institucional y personal de un objeto matemático [10] se ha establecido el *significado institucional local*, que corresponde a lo planificado por el profesor cuando diseña la secuencia didáctica de la Distribución Muestral.

A continuación, pasamos a caracterizar los principales elementos de significados de las distribuciones muestrales adquiridos por los estudiantes y los errores manifestados en las respuestas a los dos diseños de pruebas. Ver tabla 1.

Prueba conceptual y de desarrollo algebraico

Tauber [16] y Alvarado [2] han determinado una variedad de elementos de significados relacionados con la distribución normal y el teorema central del límite; conceptos importantes que nos permiten acercarnos a la comprensión de la Distribución Muestral. A continuación se describen, tabla 2, los elementos de significados considerados en la enseñanza durante el segundo semestre académico y los porcentajes de respuestas contestadas correctamente.

Campos de problemas. Los alumnos respondieron correctamente en reconocer y escribir las diferentes situaciones-problemas de las DM, considerando la población inicial normal o no. La dificultad se mostró en las aplicaciones de las propiedades de las DM, donde sólo un 47% respondió correctamente un problema de obtener tamaño de muestra adecuado y un 58% logró resolver la situación de intervalo de confianza.

Lenguaje. Un 63% de los alumnos han escrito correctamente los términos verbales y simbólicos y se han apropiado de

las notaciones en el caso de variables continuas. Cabe señalar que un 35% de los alumnos presentan errores de lectura de las tablas estadística de la distribución normal y chi-cuadrado.

Procedimientos. Se observa que los estudiantes han realizado de forma correcta el procedimiento de cálculo de probabilidades del estadístico de varianza muestral y también para el caso de la suma de variables aleatorias continuas. Continúan presentando un 51% de errores en llegar a la solución correcta de la probabilidad usando la media de la distribución muestral, y sólo un 34% logró resolver el caso de aplicación de obtener el tamaño de muestra, proceso inverso a la tipificación.

Conceptos. Este elemento de significado es el más débil, donde el grupo curso carece de sintetizar conceptos relacionados, sólo son capaces de reproducir definiciones de forma aislada.

Propiedades. En promedio más de un 70% de las respuestas de tipo conceptual de las propiedades señaladas han sido reconocidas e incorporadas al conocimiento del tema por los alumnos. Cabe señalar que sólo un 22% contestó correctamente la propiedad que relaciona los conceptos de media muestral y media poblacional; esto puede justificarse por trabajar con la simbología más que con términos verbales o la rápida lectura del ítem.

Argumentos. Este elemento de significado no ha sido comprendido por los alumnos; en general utilizan bien en su algoritmo algunas propiedades, definiciones y procedimiento algebraicos, pero no interpretan ni hacen síntesis de los resultados; un 30% aplica el teorema central del límite pero no lo fundamenta de acuerdo al tamaño muestral, la naturaleza de las variables y los parámetros de la distribución que consideran. Ver tabla 2.

Prueba de aplicación contextualizada de las DM con el programa @risk

El trabajo final de aplicación fue formulado en el procesador Word, analizada la experimentación con el programa @risk, y enviado a la plataforma virtual EV@. Previo a la ejecución de este trabajo, se levantó en la plataforma un demo de una actividad de comandos y gráfica de simulación con @risk de la distribución de weibull (ver figura 1).

Tabla 1. Frecuencias de respuestas contestadas correctamente en los ítems de las evaluaciones.

Ítems	Contestadas correctamente	
	Frec.	%
[1] Enuncie una definición que involucre los conceptos de Distribución Muestral, estadístico y muestra aleatoria.	4	3
[2] Hemos dado el siguiente teorema en clase; <u>Teorema 1</u> : Considere una muestra aleatoria extraída de una población normal de media μ y varianza σ^2 . Entonces, la variable aleatoria de la media muestral tiene distribución normal con media μ y varianza σ^2/n . a) Enuncie en lenguaje simbólico el teorema 1. b) Explique su importancia y alcance de aplicación.	a) 85 b) 42	63 31
[3] Responda Verdadero (V) si la proposición es siempre verdadera. En caso contrario justifique sustituyendo las palabras en el casillero Falso para las que hagan siempre verdadera la proposición: a) Una distribución muestral describe los valores que tomaría un estadístico en muchas repeticiones de una muestra o de un experimento en las mismas condiciones. (V) b) La desviación estándar de la media muestral aumenta a medida que crece el tamaño de la muestra. (F, disminuye) c) Los histogramas de todas las distribuciones muestrales tienen forma simétrica. (F, algunas) d) Un parámetro es la medida de alguna característica de una muestra. (F, estadístico) e) Cuando el estadístico de prueba es t y el número de grados de libertad se hace muy grande, el valor crítico de t está bastante próximo al normal estándar z . (V) f) El valor esperado de la distribución muestral de las \bar{X} es igual a la media de la muestra. (F, población)	a) 74 b) 70 c) 79 d) 15 e) 80 f) 23	70 66 75 14 75 22
[4] Marque con una X su respuesta. El peso promedio de alumnas de primer año universitario es 52 k con desviación estándar 12 k. Se toma una muestra aleatoria de 36 alumnas de ingreso a la UCSC. Con probabilidad del 68% podemos decir que la media de la muestra estará comprendida: (a) Entre 40 y 64 kilos (b) Entre 50 y 54 kilos (c) Entre 48 y 56 kilos (d) Entre 51 y 53 kilos	a) 11 b) 62 c) 07 d) 26	10 58 07 25
[5] Algebraico. Un inspector federal de pesos y medidas visita una planta de empaçado para verificar que el peso neto de las cajas sea el indicado en éstas. Considere la v.a. peso neto de las cajas con una distribución normal. El gerente de la planta asegura al inspector que el peso promedio de cada caja es de 750 g con una desviación estándar de 5 g. El inspector selecciona, al azar, 100 cajas y encuentra que el peso promedio es de 748 g. a) Bajo estas condiciones, ¿qué tan probable es tener un peso medio de 748 g o menos? ¿Qué actitud debe tomar el inspector? b) ¿Cuál es la probabilidad de que la diferencia entre la media de la muestra y la media verdadera no exceda de 1 gramo? c) Calcular la probabilidad de que el peso de una caja, seleccionada al azar de un lote, supere los 754 gramos. d) Si el inspector selecciona, al azar, 81 cajas, encuentre la probabilidad de que la varianza muestral sea mayor que 31,84 gramos ² .	a) 66 b) 83 c) 91 d) 98	49 62 68 73
[6] Algebraico. Al sumar números, un computador aproxima cada número al entero más próximo. Supongamos que todos los errores de aproximación son independientes y distribuidos uniformemente entre (-0.5, 0.5). a) Si se suman 1500 números, ¿cuál es la probabilidad de que la magnitud del error total exceda de 15? b) ¿Cuántos números pueden sumarse juntos a fin de que la magnitud del error total sea menor que 10, con probabilidad 0,90?	a) 108 b) 44	84 34

Tabla 2. Elementos de significados de las DM considerados en el cuestionario.

Tipo de elementos	Elementos de significados	%
Campos de problemas	Búsqueda de distribución de la media muestral de población normal	80
	Obtener la distribución del error muestral de población normal	67
	Distribución de una variable aleatoria normal	81
	Distribución de la varianza muestral	73
	Estimación por intervalos de confianza de la media para muestras grandes	58
	Distribución de la suma de variables aleatorias continuas	69
	Obtener el tamaño de muestra de variables aleatorias continuas	47
Lenguaje	Distribución de la media muestral de población normal (simbólico)	63
	Tabla de la distribución normal, chi-cuadrado	65
	Notación de suma de variables aleatorias continuas	90
Procedimientos	Cálculo de probabilidad de media muestral de población normal	49
	Cálculo y transformación algebraica de una variable aleatoria normal	68
	Cálculo de probabilidad de la varianza muestral	83
	Cálculo de probabilidad de suma de variables aleatorias continuas	84
	Operación inversa a la tipificación en variables continuas	34
Conceptos	Distribución Muestral	70
	Relacionar Distribución Muestral, estadístico y muestra aleatoria	3
	Estadístico	14
Propiedades	La desviación estándar de la media muestral aumenta para n grande	66
	Estandarizar la media muestral de población normal	66
	Simetría de las distribuciones muestrales	75
	Teorema central del límite para la distribución t-student	75
	Valor esperado de la media muestral coincide con la media poblacional	22
	Media de la distribución de una suma de variables aleatorias continuas	95
	Varianza de la distribución de la suma de variables aleatorias continuas	95
Argumentos	Explicativo en palabras de la distribución de media muestral	31
	Teorema central del límite para variables continuas	30

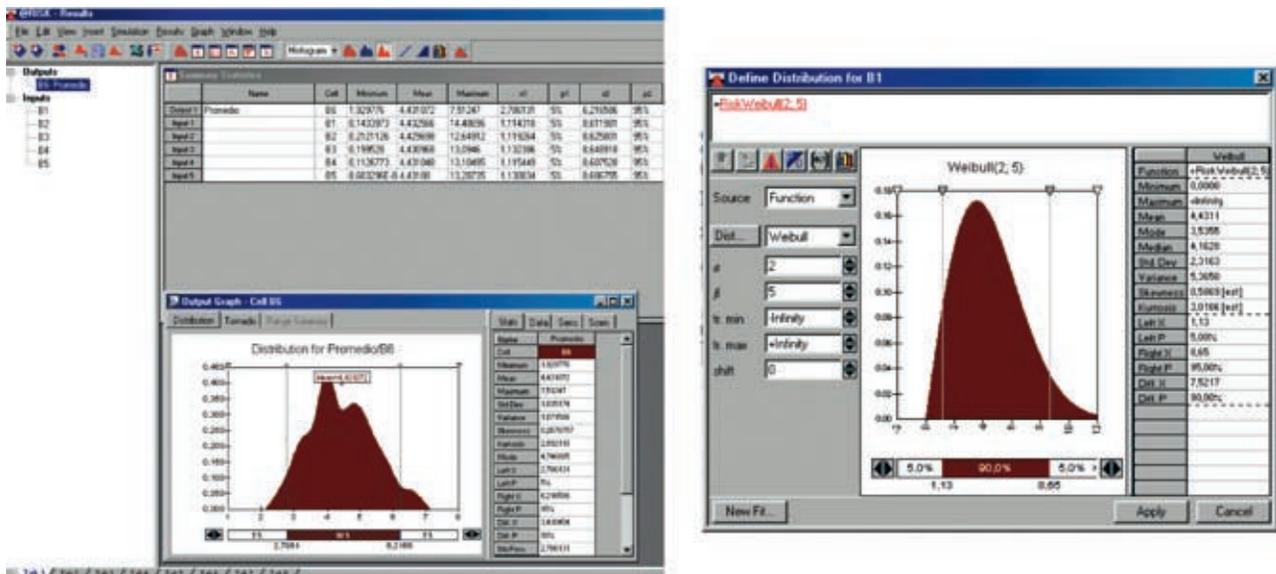


Figura 1. Distribución de muestreo de \bar{X} cuando la distribución de la población es Weibull, con $\alpha = 2$ y $\beta = 5$ utilizando el programa @risk.

A continuación se presentan los principales elementos de significados de las DM explicitados por los alumnos en su informe de aplicación con apoyo del computador.

Situaciones. Los alumnos han empleado variados temas de su interés; seleccionaron 42 variables aleatorias discretas y 22 variables aleatorias continuas, definiendo diversas variables a la ingeniería tales como: Número de salmónidos contagiados por el parásito nematodo *Hysterothycacium* en cultivos marinos del sur de Chile (binomial), Número de camiones cargados por el cargador frontal en un período de una hora (Poisson), Tiempo en horas que demora una máquina de pino en obtener un metro cúbico de producción (exponencial), Precio de cada paquete comprado en la transacción en la bolsa de comercio (uniforme).

En general, los problemas planteados no fundamentaron su importancia y muy pocos definieron correctamente la variable aleatoria con alguna distribución clásica. Además, notamos una confusión en definir una variable aleatoria con distribución Poisson con la exponencial.

En la tabla 3 se muestra un resumen de las distribuciones de probabilidad consideradas y el área de especialidad analizada a la ingeniería.

Lenguaje. Los términos y expresiones verbales más utilizados fueron: Variable aleatoria, distribución de probabilidad, distribución normal, muestra aleatoria, estadístico, distribución muestral, teorema central del límite, suma de variables aleatorias independientes e idénticamente distribuida, simulación, histograma, intervalo de confianza, dispersión, población. Las notaciones simbólicas permiten realizar operaciones con los conceptos, trabajando a un alto nivel de complejidad.

En los trabajos, las expresiones simbólicas de la distribución muestral más recurridas son: X_1, X_2, \dots, X_n para referirse

a una muestra aleatoria, $S_n = \sum_{i=1}^n X_i$ al enunciar la suma de n variables aleatorias, $\bar{X} \approx N(\mu, \sigma^2/n)$ al describir la distribución muestral de la media de la muestra para n suficientemente grande, la fórmula de estandarización de la media muestral $\frac{\bar{x} - \mu}{s/\sqrt{n}}$. En cuanto a las *Representaciones gráficas*, las consideradas fueron: Gráfico de barras, Histograma, Polígono de frecuencia, Curva densidad normal y Distribuciones clásicas de probabilidad. En la *simulación* se trabajó con la planilla Excel y el programa estadístico @risk.

Procedimientos. El algoritmo más reiterativo de resolución de la situación-problema fue el siguiente:

Definir la variable aleatoria a la ingeniería; generar una matriz de datos, tablas de datos; representar por un histograma la distribución muestral de la media muestral o suma de variables aleatorias; interpretar gráficas, por ejemplo, el histograma; obtener distribución probabilidad clásica asociada, cálculo de parámetros de la distribución de probabilidad escogida; obtener la distribución de muestreo de la media muestral cuando la población no es normal; comparar valores teóricos esperados con su correspondiente valor empírico, cálculo de probabilidad en intervalos de la distribución normal.

Conceptos. Los más empleados por los alumnos son los siguientes: distribución muestral, enunciado del teorema central del límite, estadístico, media muestral, distribución de probabilidades.

Propiedades. Las que utilizaron los alumnos fueron:

La media muestral sigue una distribución normal cuando la muestra es grande;
La convergencia de la suma de variables aleatorias a la distribución normal mejora conforme aumenta el tamaño de la muestra;

Tabla 3. Distribuciones de probabilidades y temas de ingeniería empleados por los estudiantes

Distribución Teórica	Especialidad de Ingeniería	Área de desarrollo
Binomial 24	Acuicultura y Pesca 11	Apícola, cultivo, industria salmonera, Transporte, flota, pesca, computación, programación, control de calidad, servicio atención clientes, consumo empresas, forestal, central hidroeléctrica, tiempo de falla, seguridad vial, educación, deporte, automotriz, encuesta laboral, bolsa de comercio, ganadería, estructura de construcción, resistencia de materiales
Poisson 18	Marítimo Portuario 7	
Normal 3	Informática 10	
Exponencial 14	Industrial 34	
Uniforme 2	Civil 7	
Gamma 1		
Logística 1		
Weibull 1		
No define distribución 3		

Aproximación de la distribución binomial por la normal como caso particular del teorema central del límite;
 Aproximación de algunas distribuciones clásicas a la distribución normal;
 Aproximación a la normal mejora según la naturaleza de las variables aleatorias;
 La media del promedio de la muestra es igual a la media de la población;

La forma de la distribución muestral tiende a ser acampanada a medida que se incrementa el tamaño muestral, y aproximadamente normal, independiente de la forma de la distribución en la población.

Argumentaciones. Para validar y explicar las proposiciones los estudiantes emplearon la comparación de los valores teóricos y empíricos, representación gráfica y la síntesis.

Muy pocos estudiantes llegan a una conclusión final correcta en el contexto del problema, relacionando gran parte de los elementos de significados, realizando un análisis de cada propiedad y concluyendo con una síntesis. Un ejemplo, es la alumna de ingeniería en Acuicultura y Pesca que inició bien su trabajo, obtuvo una variable aleatoria, X: “número de pérdidas de especies en los sistemas de cultivo por linterna”, a partir de un problema

situado a su área de especialidad, y en su resolución asocia correctamente la variable con la distribución de Poisson. A continuación, toma una muestra aleatoria de linternas y define la suma de variables aleatorias con aproximación a la distribución normal, por medio del teorema central del límite (propiedad). Determina los valores exactos de la distribución teórica Poisson y de la normal y luego los compara. En el cálculo utiliza correctamente la corrección por continuidad (propiedad). También relaciona las propiedades con la forma para varios polígonos de frecuencias (lenguaje) mediante el programa @risk (figura 2). Interpretó las gráficas mediante la curtosis (propiedad) comentando que a medida que se va aproximando a la normal, se vuelve más simétrica. Concluye, para varias muestras aleatorias de tamaño 100, 300, 1000 y distintas iteraciones de 20, 500 y 5000, que a más grande el número de iteraciones, manteniendo el parámetro de Poisson $\mu=5$ y aumentando la cantidad de especies de ostras (*argopecten purpuratus*), la distribución Poisson se aproxima bien a la distribución normal, además la varianza disminuye a medida que aumenta la muestra ya que depende de la cantidad de especies que se utilizaron. Un comentario es que las representaciones gráficas obtenidas del programa estadístico entrega los resúmenes estadísticos con el cálculo de intervalos de confianza clásicos, que en este caso no fue analizado e interpretado.



Sistema de linterna (consta de 10 pisos)



Simulación de 300 especies de ostras con $\mu=5$

Figura 2. Análisis de muestreo en el campo de acuicultura: situación de cultivo de ostras por medio del sistema de linterna en suspensión mediante @risk.

CONSIDERACIONES FINALES

Debido a la amplitud de las propiedades y elementos relacionados de la Distribución Muestral y su apropiación de conceptos por parte de los alumnos, la evaluación de habilidades generales es sólo parcial, ya que los controles fueron varios con objetivos específicos.

La descripción de las características principales en el significado personal que los alumnos construyeron sobre las distribuciones muestrales, en base a los significados institucionales dado en la tabla 1, indican que más de la mitad del grupo de estudiantes respondió correctamente los elementos de significados de las DM en las dos pruebas analizadas en este trabajo, destacando la aplicación de estimadores de la media de variables aleatorias provenientes de muestras de una distribución binomial.

Los alumnos respondieron correctamente la resolución de un problema algebraico del caso de variables de tipo continua a un promedio del 63%, lo que indica que muchos de los elementos considerados en la enseñanza de este tema pasarán a formar parte del significado personal de los alumnos.

Si bien, aplican en un contexto situado las ideas básicas de parámetro, estadístico y distribución de probabilidades, observamos dificultades en interpretar los resultados de aproximaciones a la distribución normal y también ciertos gráficos.

Observamos que los errores más comunes que se deducen de las respuestas, y que denotan desajustes con el significado institucional pretendido en el curso, son: aplican el teorema central del límite pero no lo justifican; no definen correctamente la variable aleatoria de la suma de variables aleatorias en un contexto a la ingeniería; no aplican la corrección por continuidad del caso discreto aproximado por una distribución continua (normal); aún persisten errores en la lectura de la tabla estadística de la distribución normal estándar y chi-cuadrado; y finalmente, los alumnos calculan la probabilidad de estimadores de variables aleatorias pero no interpretan correctamente. También, hubo errores en los estudiantes de ingeniería en relacionar correctamente la media poblacional y la media muestral, lo que coincide con el estudio de Vallecillos [17].

Consideramos positiva la experiencia de trabajar la DM mediante la simulación con el computador, que a pesar de disponer con un 25% de equipos con el programa

utilizado en el laboratorio, los alumnos bajaron por Internet el @risk con vigencia limitada. Un análisis detallado de las respuestas de los estudiantes muestra que ellos aprendieron a utilizar el programa @risk, obteniendo tablas de la distribución de la suma de variables aleatorias con muestras de distintos tamaños muestrales y la gráfica de barras e histograma como un medio de validar la forma simétrica de la distribución de la media muestral a medida que aumenta el tamaño muestral; una de las propiedades propuesta por Méndez [13]. Los estudiantes en general utilizan como criterio de análisis el tipo de variables aleatorias para identificar la distribución clásica de probabilidades y aplicar algunas propiedades de aproximación a la distribución normal. Aunque, observamos en algunos estudiantes que confunden una variable discreta con una continua, como es el caso de la distribución Poisson con la distribución exponencial.

La mitad del grupo curso manifestó adquirir ciertas habilidades, como la de identificar y formular un problema a la ingeniería, trabajar en equipo con sus pares en las actividades planteadas. Sin embargo, es débil la predisposición a realizar experimentos, analizar la información y la correcta interpretación de resultados del análisis de datos. No hacen un trabajo de análisis en las tareas al plantear un problema a la ingeniería, para comprobar las propiedades importantes de la DM y síntesis para obtener una conclusión del conjunto de análisis realizado. Son capaces de manejar la planilla Excel y el software @risk pero no su interpretación de resultados. La mayoría de los alumnos se conforma con aplicar la estimación de la media en distintas muestras pero no justifican si es adecuado o no tal solución.

Debemos continuar este trabajo, identificando las competencias estadísticas prioritarias de un ingeniero y su desarrollo en la asignatura.

La investigación didáctica sobre la enseñanza y aprendizaje de conceptos avanzados de estadística es todavía muy escasa. Este trabajo descrito es sólo una primera fase de investigación que nos ha permitido de forma sistemática caracterizar en base a las respuestas de los alumnos, los errores en la comprensión de las propiedades de las DM y que puede orientar la labor del profesor, tanto para planificar la enseñanza y evaluación del tema, como para fundamentar la enseñanza posterior de la estimación por intervalos y los contrastes de hipótesis. Una segunda etapa será repetir la metodología de enseñanza con otros estudiantes de ingeniería y validar las evaluaciones construidas.

AGRADECIMIENTOS

Este informe ha contado con el apoyo de la Dirección de Investigación y Postgrado de la Universidad Católica de la Santísima Concepción, mediante la adjudicación del proyecto interno de investigación DIN 06/2005.

REFERENCIAS

- [1] H. Alvarado. "Significado y comprensión de un teorema estadístico: elementos básicos en el desarrollo profesional del profesor para una buena enseñanza". *Boletín de Investigación Educativa*. Vol. 19 N° 1, pp. 227-244. 2004.
- [2] H. Alvarado. "Elementos del significado del teorema central del límite en textos de estadística para ingenieros". *Memoria de Tercer Ciclo*. Universidad de Granada. 2004.
- [3] H. Alvarado y C. Batanero. "El significado del teorema central del límite: evolución histórica a partir de sus campos de problemas". *Investigación en Didáctica de las Matemáticas*. Editorial A. Contreras, pp. 13-36. ISBN: 84-8439-268-6. Grupo FQM126. Universidad de Granada. 2005.
- [4] R. Ardanuy y Q. Martín. "Estadística para ingenieros". Salamanca: Hespérides. 1998.
- [5] R. Bisquerra. "Métodos de investigación educativa". Barcelona, España. P.P.U. 1989.
- [6] R. Chase, N. Aquilano y F. Jacob. "Administración de producción y operaciones". McGraw-Hill. Santa Fe, Bogotá. 2000.
- [7] R. C. del Mas, J. B. Garfield y B. L. Chance. "A model of classroom research in action: developing simulation activities to improve students' statistical reasoning". *Journal of Statistic Education*. Vol. 7 N° 3. 1999. Online: <http://www.amstat.org/publications/jse/secure/v7n3/delmas.cfm>.
- [8] I. Gal. "Assessing students' interpretations of data: Conceptual and pragmatic issues". *Papers on Statistical Education presented at ICME-8*. Editorial B. Phillips, pp. 49-58. Swinburne University of Technology. 1997.
- [9] J. D. Godino. "Un enfoque ontológico y semiótico de la cognición matemática". *Recherches en Didactique des Mathématiques*. Vol. 22 N° 2/3, pp. 237-284. 2002.
- [10] J.D. Godino. "Teoría de las funciones semióticas. Un enfoque semiótico de la cognición e instrucción matemática". Granada: El autor. 2003.
- [11] J.D. Godino y C. Batanero. "Clarifying the meaning of mathematical objects as a priority area of research in Mathematics Education". En A. Sierpiska y J. Kilpatrick (Eds.), *Mathematics education as a research domain: A search for identity*, pp. 177-195. Dordrecht: Kluwer. 1998.
- [12] R. Hernández, C. Fernández y C. Baptista. "Metodología de investigación". McGraw-Hill. México. 1998.
- [13] H. Méndez. "Understanding the Central limit Theorem". Tesis para optar al grado de doctor. Universidad de California. UMI 6369. 1991.
- [14] D. Montgomery y G. Runger. "Probabilidad y estadística aplicadas a la ingeniería". México: McGraw-Hill. 1996.
- [15] J. Pazos, A. Suárez y R. Díaz. "Teoría de colas y simulación de eventos discretos". Prentice Hall. Madrid, España. 2003.
- [16] L. Tauber. "Significado y comprensión de la distribución normal a partir de actividades de análisis de datos". Tesis para optar al grado de doctor. Universidad de Sevilla. España. 2001.
- [17] A. Vallecillos. "Estudio teórico-experimental de errores y concepciones sobre el contraste de hipótesis en estudiantes universitarios". Tesis para optar al grado de doctor. Universidad de Granada. 1994.
- [18] A. Vallecillos. "Comprensión de la lógica del contraste de hipótesis en estudiantes universitarios". *Recherches en Didactique des Mathématiques*. Vol. 15 N° 3, pp. 53 - 81. 1996.
- [19] A. Well, A. Pollatsek y S. Boyce. "Understanding the effects of sample size on the variability of the mean". *Organizational Behavior and Human Decision Processes*. Vol. 47, pp. 289-312. 1990.