

SIMULACION DIGITAL COMO VEHICULO DE INSTRUCCION
EN CURSOS DE ESTADISTICA PARA INGENIEROS:
UNA EXPERIENCIA PRACTICA.

Jorge Luis Romeu, FIS.

Departamento de Ingenieria Industrial e Investigacion
Operacional, Syracuse University
y IIT Research Institute
Rome, N.Y. 13440/USA.

Summary.

Motivating engineering students in statistics courses is often difficult. The present paper discusses our classroom experience teaching Regression and Data Analysis for Engineers, a senior level course. The organization and support material of this course is overviewed. The simulation experiments used and their engineering interpretations is described. Practical problems encountered and the adaptive solutions given to them is discussed. Finally, the method of evaluating the students and some objective performance measures of the course results are presented.

1.0 INTRODUCCION.

En la presentacion de cursos de estadistica para estudiantes de ingenieria con frecuencia encuentra el profesor dos grandes problemas: falta de tiempo para cubrir el material requerido y falta de motivacion por parte de los alumnos. Muchos estudiantes ven en los cursos de estadistica tan solo un ejercicio teorico mas con pocas aplicaciones practicas en su area de interes profesional. Lo toman solo porque es un requisito del programa y lo colocan bien abajo en la lista de preferencias y prioridades. El resultado es que, el ya reducido tiempo con que cuenta el

instructor para impartir la extensa materia se reduce aun mas por la falta de interes y motivacion de los alumnos.

El presente trabajo se basa en nuestra experiencia practica, impartiendo un curso de regresion y analisis de datos para ingenieros industriales. Este curso, sin ser un requisito de la carrera, se recomienda a aquellos alumnos que se inclinan hacia las actividades de control de calidad y estudios de movimiento y tiempo. Los prerrequisitos para el curso consisten en un semestre en probabilidad (IOR325) y otro de estadistica (IOR326) al nivel del texto de Walpole-Myers (7). Ademas, es conveniente algun conocimiento basico de la simulacion digital, necesario para los ejercicios practicos y experimentos de clases.

El contenido de este curso es algo ambicioso. Incluye, en 13 semanas de clases con una frecuencia de dos sesiones de hora y media a la semana, los topicos de regresion simple y multivariante, metodos de seleccion de variables en regresion, analisis de varianza y covarianza y algunos rudimentos de diseno de experimentos y analisis de superficies de respuesta. Los dos libros de texto son Chatterjee-Price (2) y Box-Hunter-Hunter (1). El de Draper-Smith (3) sin ser texto del curso, fue recomendado como material de lectura y referencia. Es necesario aclarar que las conferencias estan basadas en los textos mencionados, sin embargo tuvimos completa libertad para seleccionar, modificar y presentar el material de curso segun nuestro criterio, basados en nuestra experiencia previa, articulos recientes, etc. (5), (6). El diseno del curso puede encontrarse en el Apendice.

Este articulo resena la organizacion y el material de apoyo de dicho curso. Se describen los experimentos de simulacion digital utilizados para motivar a los alumnos y fijar los conceptos estadisticos, y se discuten los problemas enfrentados asi como las soluciones dadas a los mismos. Por ultimo, se analizan detalladamente el metodo de evaluacion de los alumnos y algunas de las ventajas e inconvenientes de este metodo de presentacion de la asignatura.

2.0 DESCRIPCION DEL CURSO.

Nuestro objetivo fundamental en todo momento es lograr la motivacion del alumno. Este objetivo se persigue a traves de cinco 'experimentos' utilizados para introducir y aplicar, de forma practica y en el campo de la ingenieria industrial, los conceptos estadisticos del curso.

Contenido del Curso.

- 1.0 Modelacion Estadistica
- 2.0 Regresion Lineal Simple
- 3.0 Analisis de los Residuos
- 4.0 Transformaciones
- 5.0 Analisis de Una y Dos Muestras
- 6.0 Comparaciones Multiples
- 7.0 Regresion Multivariante
- 8.0 Seleccion de Variables
- 9.0 Analisis de Varianza y Covarianza
- 10.0 Disenos Factoriales

Los cinco experimentos presentan aplicaciones de técnicas estadísticas a problemas de ingeniería industrial o de la investigación de operaciones tomados de la vida real. La ventaja de este enfoque sobre un ejercicio numérico ordinario estriba en el factor "real" introducido por el mecanismo de simulación, ya que los resultados de los ejercicios son completamente desconocidos para todos, incluido el instructor.

El primer programa de simulación se escribió en lenguaje FORTRAN. El programa modela un simple sistema de una cola, con una estación de servicio según la distribución exponencial y la regulación del servicio por estricto orden de llegada (FIFO). Las llegadas siguen una distribución de Poisson. El problema se describió a los alumnos como la modelación de una máquina-herramienta (torno). La estación de servicio es el torno, el operador consume un tiempo (distribuido exponencialmente) en cada trabajo y la cola son las piezas a trabajar. Utilizando este modelo se introdujo la técnica de regresión lineal simple, el análisis de residuos y la comparación de dos muestras apareadas.

Para introducir la regresión, se pidió a los alumnos que evaluaran la capacidad de producción de distintas máquinas-herramientas (distintos tiempos de servicio) respecto a distintos tipos de piezas (distintos tiempos de llegada al sistema). Se hizo hincapié en la validación de los supuestos de la regresión como primicia para la validez de los resultados estadísticos.

Table 2: Batch Size Trade-off.

Responses	B A T C H S I Z E S		
	200	100	50
No. of Batches	5	10	20
Differences Detected Between the Strategies	7.64	8.49	8.71
Std. Deviation	2.62	1.37	0.45
T Test	2.62	6.18	19.28
T _{n-1, 0.95}	2.77	2.26	2.09

Part 2-A: Descriptive Statistics of Pairwise T Tests.

Legend: Pairwise Comparison of the Simulated Service Time for Two Queuing Strategies:

- using 5 batches of 200 customers
- using 10 batches of 100 customers
- using 20 batches of 50 customers

Para introducir el problema de comparacion de dos muestras apareadas, presentamos el concepto de prioridad en Teoria de Colas. Se presento a los alumnos el problema de comparar dos politicas de produccion: una con estricto orden de llegada y otra en la cual los dos tipos de piezas procesadas en el torno tenian prioridades diferentes. Se les pidio que analizaran el mismo flujo de llegada a la estacion de servicio siguiendo estas dos politicas de produccion, utilizando la prueba T para datos apareados sobre la variable "productividad" (medida en terminos de piezas producidas por hora de trabajo).

Los demas programas utilizados en los 'experimentos' fueron escritos en lenguaje GPSS (4) de simulacion. El segundo programa modela un sistema de dos estaciones de servicio, en serie, con un tampon de capacidad limitada entre ambos. Las distribuciones de llegada y de servicio seguian siendo Poisson y Exponencial, respectivamente. La introduccion de un sistema mas complicado permitio la inclusion de un mayor numero de variables independientes y dependientes asi como variables cualitativas y cuantitativas.

Las variables independientes fueron:

- disciplina de la cola (cualitativa)
- flujo de llegada (cuantitativa)
- tiempo de servicio (cuantitativa)
- Capacidad del tampon (cuantitativa)

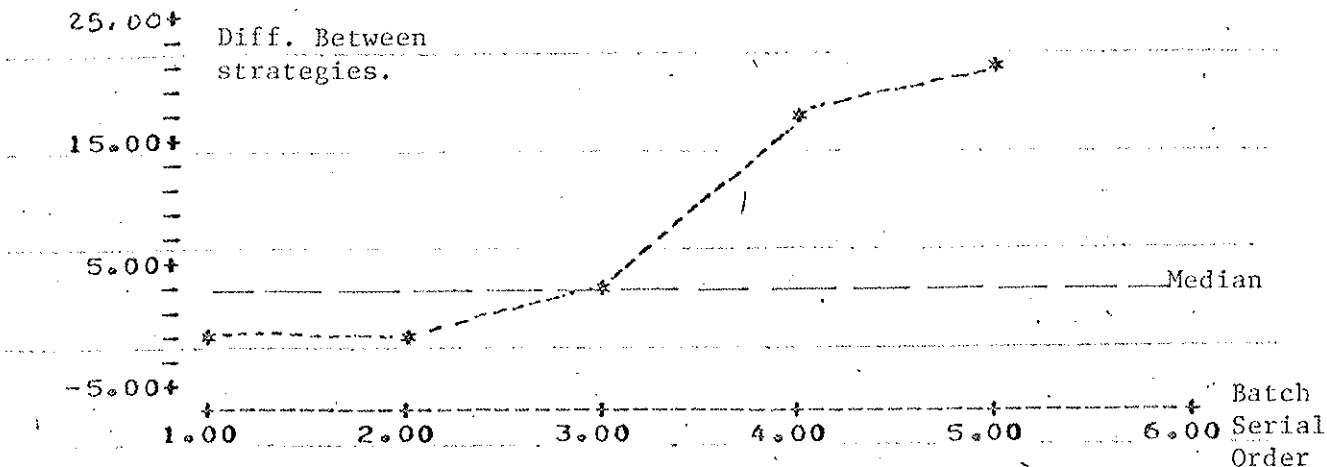
Se analizaron las siguientes variables dependientes:

- tiempo total en el sistema (por categoria)
- porcentaje de utilizacion de los recursos
- flujo (productividad)

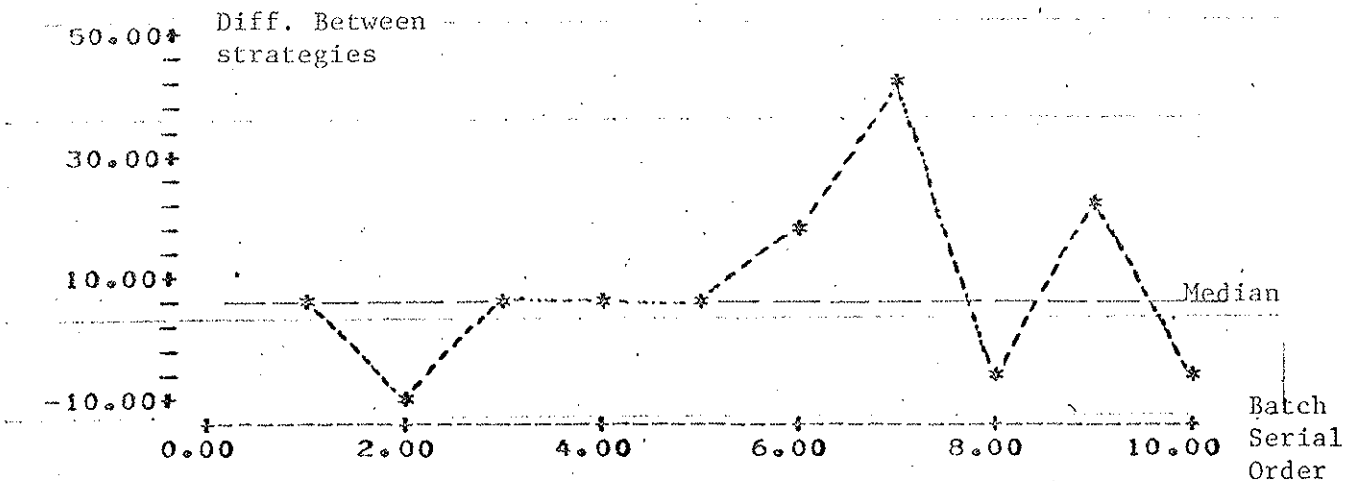
Table 2: Batch Size Trade-Off (continued).

Part 2-B: Serial Plots of Strategy Differences by Batch Sizes.

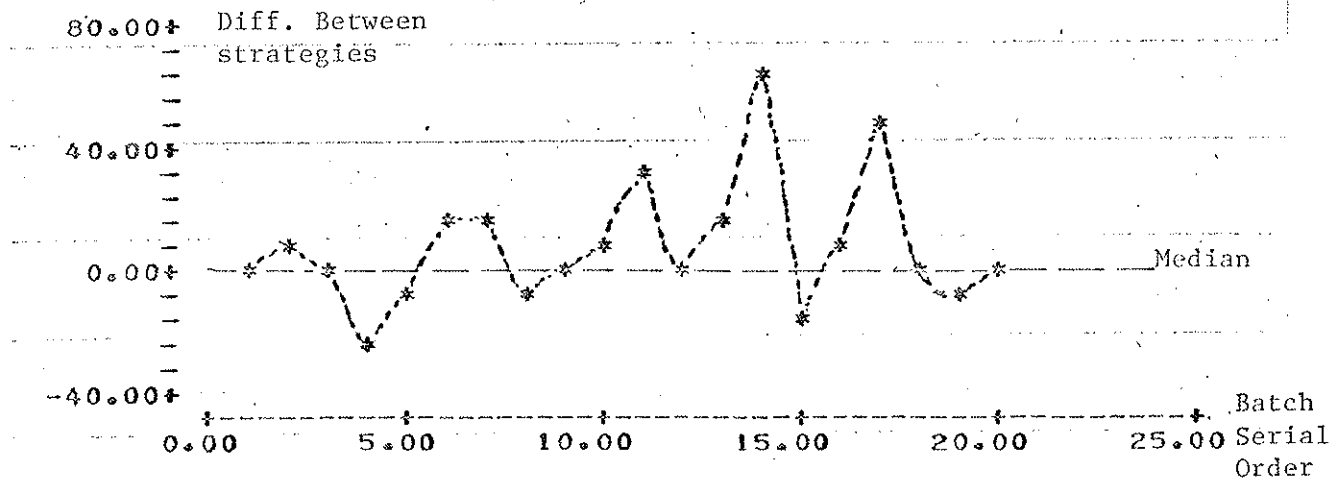
a) Batch Size 200.



b) Batch Size 100



c) Batch Size 50



Este modelo permitio introducir los conceptos de regresion multivariante, seleccion de variables y de docimasia de hipotesis en regresion multivariante. Estos conceptos fueron introducidos en el tercer experimento e incluyeron, ademas, validacion de los supuestos del modelo, problemas de correlacion entre las variables independientes, transformaciones, etc.

El cuarto experimento utilizo el mismo modelo de simulacion descrito el en parrafo anterior. Incluyendo, convenientemente, distintas disciplinas de cola (por estricto orden de llegada y seleccion priorizada) fueron introducidos los conceptos del Modelo Lineal General y el analisis de covarianza.

El tercer y ultimo programa de simulacion escrito para este curso modelaba varias estaciones de servicio en paralelo (especificamente cinco) con iguales tiempos de servicio, de llegada y una sola cola comun para todas. Las distribuciones siguieron siendo Exponenciales y de Poisson. Este modelo fue utilizado para el ultimo experimento y para el proyecto de fin de curso.

Este ultimo (quinto) experimento sirvic para introducir los conceptos de disenos factoriales y algunos elementos del analisis de superficies de respuesta. El proyecto de fin de curso, totalmente desarrollado por pequenos grupos de alumnos, resumia los elementos fundamentales de entre los estudiados en este curso.

3.0 OPERACION DEL CURSO.

El curso duraba de 13 a 14 semanas, con una frecuencia de dos clases de hora y media a la semana. Teniendo en cuenta el examen intra-semestral y el proyecto final, este programa permitia alternar una tarea y un experimento cada fin de semana. Las tareas consistian en el desarrollo de un problema tomado de los textos. Los experimentos y el proyecto final se desarrollaban de la manera que pasamos a describir.

El objetivo fundamental de los experimentos era proporcionar una "experiencia real" en el analisis de datos y motivar a estos alumnos de ingenieria. Esto se lograba, fundamentalmente, a traves de la interpretacion dada a los modelos y al caracter aleatorio de los resultados obtenidos a partir de ellos.

Al primer experimento se le asigno la interpretacion de una maquina-herramienta, asi como al segundo. El tercer experimento se introdujo como la macro-modelacion de la operacion de una planta de ensamblaje. La primera estacion de servicio constituia toda la operacion de ensamblaje de este proceso. La segunda estacion de servicio representaba la funcion de control de calidad, mientras que el tampon representaba las restricciones de almacenamiento. En el cuarto experimento, este proceso de ensamblaje se complico con la introduccion de mas de una linea de produccion, por ejemplo, ensamblaje de autos y camiones y se discutio algo tan real como la definicion de una politica de produccion optima bajo estas condiciones. Por ultimo, el quinto modelo de simulacion se introdujo como la representacion de la operacion de una oficina de banco, aunque tambien podia

representar la operacion de una centralita de telefonos, de un supermercado, etc.

Cada experimento se presentaba conceptualmente, o sea, se discutia la problematica del modelo. A continuacion, se introducía el modelo de simulacion que representaba la situacion real y seguidamente la utilidad practica obtenida al resolverlo. Con esto se lograba espolear el interes de los alumnos en buscar una solucion practica al problema. Solo entonces se introducían las herramientas estadísticas que podrian servirle para encontrar dicha solucion. A cada estudiante se le entregaba una copia del programa de simulacion y un valor DIFERENTE para generar los flujos de llegada y tiempos de servicio en el programa de simulacion. De esta forma se aseguraba que los resultados individuales fueran todos distintos. Con esto, podian los alumnos cooperar libremente en la resolucion de sus problemas. No solamente no habia dos resultados numericos iguales sino que hasta se obtenian soluciones diferentes, debido al caracter aleatorio de los experimentos y a los tamanos de muestra tomados. Este ultimo factor permitia desarrollar amplias y fructíferas discusiones en clase sobre problemas fundamentales como son el tamaño de la muestra, el nivel de la docima y la funcion de potencia.

Los experimentos fueron evaluados sobre la base de enfoque, comprobacion de los supuestos del modelo e interpretacion de los resultados. Nunca se tomaron en cuenta determinados valores numericos. Los estudiantes tenian completa libertad para escoger su enfoque, siempre y cuando pudieran sustentarlos bajo un

criterio racional. Por ejemplo, en el problema de la seleccion de variables, en regresion multiple, se introdujeron los metodos de seleccion ascendente, descendente y paso a paso. Sin embargo, se dio a los alumnos completa libertad para escoger el metodo que estimasen conveniente, siempre y cuando los utilizaran correctamente. En algunos casos, a motu proprio, varios alumnos utilizaron mas de un metodo y compararon los resultados, yendo mas alla de lo requerido en la tarea asignada.

Tambien, fueron asignadas por los alumnos distintas variables dependientes, lo que permitio la libre interaccion entre los estudiantes. Asi podian ganar mas experiencia al enfocar distintos problemas desde distintos angulos, y comprobar como la solucion para una variable no necesariamente era la misma solucion para otra. Por ejemplo, en el experimento del analisis de dos politicas distintas de prioridad, pudieron los estudiantes verificar como la optimizacion de una estacion especifica de servicio ofrecia una solucion distinta de la optimizacion del flujo total del sistema.

El proyecto final era asignado dos semanas antes del fin de curso y substituia el examen final. Se daba a los alumnos dos politicas de produccion a evaluar y escoger la politica optima (sobre la base de una pre-determinada variable de respuesta). Debian disenar un experimento, llevarlo a cabo, analizar sus resultados y escribir un informe final dirigido al comite de direccion de la empresa con las conclusiones de su trabajo. Debian utilizar el Modelo Lineal General, validar los supuestos del modelo, realizar las pruebas de hipotesis necesarias y sacar

conclusiones adecuadas a SUS resultados numericos. El espiritu de este ejercicio era prepararlos para comenzar a trabajar en el mundo real.

El metodo de evaluacion del curso consistia en dar igual peso a cuatro componentes: las tareas de clase, los experimentos, el examen intra-semesteral y el proyecto final. Con este sistema era muy dificil suspender o descollar, a menos que se fuera realmente bueno o malo. Por otra parte, este sistema obligaba al alumno a trabajar de manera constante a traves de todo el curso. Los resultados obtenidos en los dos anos en que fue llevado a la practica, fueron altamente encomiables para los alumnos que pasaron por el mismo.

4.0 PROBLEMAS CONFRONTADOS DURANTE EL CURSO.

El desarrollo del presente curso presupone la existencia de una computadora grande, del calibre de una IBM-370, que permita la insercion de un compilador GPSS, la inclusion de la biblioteca estadistica SAS o IMSL y capacidad para producir graficos. Estas herramientas fueron, precisamente, las utilizadas en nuestra implementacion.

Por otra parte, sabemos que no todos los centros universitarios en iberoamerica poseen estos recursos ni se encuentran en capacidad de adquirirlos. Afortunadamente, el advenimiento de la Computadora Personal (PC), bien asequible hace posible que este tipo de curso pueda impartirse en cualquier parte.

La mayor dificultad para impartir un curso de este tipo se encuentra en el desarrollo de los necesarios programas de simulación. Lenguajes especializados del tipo GPSS están completamente descartados, debido al tamaño de computadora que requieren. Sin embargo, nuestro primer programa de simulación, utilizado en el experimento número uno, fue enteramente escrito en FORTRAN, y nada impide que pueda ser escrito en BASIC o cualquier otro lenguaje utilizado por los actuales PC's. La generación de números pseudoaleatorios requiere una computadora de 32 bits o más, para neutralizar los problemas de ciclo. Sin embargo, existen métodos de generación de números pseudoaleatorios que, aunque defectuosos, pueden ser fácilmente implementados en una PC. El objetivo fundamental del presente curso es llegar al alumno y proporcionarle una rica experiencia educativa. Utilizar un generador defectuoso, explicando al detalle por qué es así, tan solo puede enriquecer su experiencia haciéndolo participe de los riesgos de una inconsciente o defectuosa aplicación de un método de trabajo.

Finalmente, el autor de este trabajo lidio durante más de 10 años con todas las limitaciones en equipo y tecnología propias de un país pobre y en vías de desarrollo y sabe muy bien del esfuerzo adicional que tiene que poner de su parte un instructor para compensarlos. También conoce el carácter y la dedicación de nuestros educadores universitarios y no tiene la menor duda de que este inconveniente tan solo servirá de acicate a nuestros profesores.

5.0 CONCLUSIONES.

El objetivo fundamental del curso aquí esbozado, así como del presente trabajo, es el de proveer de un ejemplo de experiencia docente rica y a la vez interesante para el alumno. Una experiencia que lo haga sumergirse en su trabajo estadístico y adquirir algunas técnicas útiles en su ulterior vida profesional.

Personalmente, estos dos años constituyeron una experiencia muy provechosa. Independientemente, algunas variables objetivas que pueden servir de indicadores para evaluar este curso son las siguientes:

Al comienzo del curso, cada alumno fue investido de una cuenta en la computadora, para realizar su trabajo de curso. Los mejores alumnos invariablemente solicitaban más dinero a mediados del curso, debido al uso intensivo que hacían de la misma. De más está decir que nunca faltó dinero adicional para esto.

En varios casos y por propia iniciativa, los alumnos realizaron diferentes variantes del mismo experimento y compararon los resultados. En otros casos, hicieron los experimentos con las variables dependientes que les fueron asignadas y con otras, asignadas a otros compañeros, y compararon los resultados.

Algunos alumnos no quedaron satisfechos usando el software estadístico asignado (del sistema APL) y solicitaron herramientas más sofisticadas. Siempre fueron complacidos en el sentido de que tuvieran acceso a software del tipo SAS y IMSL y que se les

intruyo en la forma de utilizarlo.

Por ultimo, este curso se impartia en el semestre final de la carrera, justo antes de ingresar en las filas profesionales. Era altamente placentero verlos partir diciendo que les gustaria encontrar un trabajo en un area donde pudieran aplicar las herramientas recién adquiridas.

6.0 AGRADECIMIENTOS.

Agradecemos al Departamento de Ingenieria Industrial e Investigacion Operacional de Syracuse University la oportunidad y la libertad concedidas para enseñar este curso, y al Profesor J. P. Vilaplana sus sugerencias para mejorar la presentacion de este trabajo.

BIBLIOGRAFIA.

- (1) Box, G. G., W. Hunter and J. Hunter. Statistics for Experimenters. John Wiley, 1978.
- (2) Chatterjee, S. and B. Price. Regression Analysis by Example. John Wiley, 1978.
- (3) Draper, N. and H. Smith. Applied Regression Analysis. John Wiley, 1966.
- (4) IBM Corp. General Purpose Simulation System/360. Introductory User's Manual. IBM Applications Program GH20-0304-4.
- (5) Romeu, J. L. "Algunas Aplicaciones de la Estadística a la Ingeniería Civil". Revista de Ingeniería Civil, No. 2, 1976. La Habana, Cuba.
- (6) Romeu, J.L. "Un Ensayo para el Analisis y la Clasificación de las Fuentes de Información Mediante Métodos de la Estadística Matemática". Revista de Ingeniería Civil, No.4, 1978. La Habana, Cuba.
- (7) Walpole, R. and R. Myers. Probability and Statistics for Engineers and Scientists. MacMillan, 1978.

Apéndice: Diseño del Curso.

1ra. Semana:	Revisión de Conceptos Estadísticos Modelación Estadística
2da. Semana:	Regresión Lineal Simple Ejemplo Numérico
3ra. Semana:	Análisis de los Residuos Transformaciones
4ta. Semana:	Ejemplos/Experimento No. 1 Comparación de Dos Muestras (Prueba T)
5ta. Semana:	Función de Potencia Prueba de Hipótesis
6ta. Semana:	Experimento No. 2 Examen Intra-Semestral
7ma. Semana:	Análisis de Varianza (clasificación simple) Residuos/Comparaciones Múltiples
8va. Semana:	Regresión Multivariada Regresión (cont.)
9na. Semana:	Experimento No. 3 Selección de Variables
10ma. Semana:	Variables Cualitativas (COVARIANZA) Experimento No. 4
11va. Semana:	Autocorrelación Multicolinealidad
12va. Semana:	Experimentos Factoriales Experimento No. 5
13ava./14ava. Semanas:	Factoriales Fraccionados Superficies de Respuesta