

Aprender Estadística con proyectos

**Memoria
de una experiencia replicable**

Mario Miguel Ojeda Ramírez

Coordinador

UNIVERSIDAD VERACRUZANA

Raúl Arias Lovillo

Rector

Porfirio Carrillo Castilla

Secretario Académico

Agustín del Moral Tejeda

Director General Editorial



Director del Instituto de Ciencias Básicas
Dr. Eryck Romeo Silva Hernández

Coordinador de la Maestría en Ciencias Alimentarias
Dr. Iñigo Verdalet Guzmán

Instructor de la Materia de Métodos Estadísticos
Dr. Mario Miguel Ojeda Ramírez

Instructores adjuntos de la materia de Métodos Estadísticos
M. en C. Jesús Hernández Suárez
M. en C. Fernando Velasco Luna

Colaboradores
C. Ma. Nayeli Hernández Carmona
Lic. Guillermo Cruz González

Estudiantes, generación 2007
Araceli Reyes Telléz
Estela Fernández Borboa
Evangelina Ruiz Figueroa
Evelín Martínez Benavidez
Gema Morales Olán
Ingrid Karina López Guzmán
Irit Roxana Meneses Ocampo
Juan Manuel Martínez Alejo
Lilia Ortíz Rodríguez
Leticia Hernández Galán
Marbell Monserrat Álvaro Moreno
Maria Luisa Yolanda Girón R.
Mayra Jannet Cobaxin Márquez
Rosa Leticia López Quiroz
Silvia del Carmen Pereyra Cástro
Viridiana Guzmán Márquez

Edición y Formación
M. en C. Cecilia Cruz López

Video

Coordinación de Medios Audiovisuales
Juan Rivero Valls

Coordinación de Producción
Arely Ruiz Armas

Productor
Pedro Valencia Sánchez

Edición y Post-Producción
Octavio Pérez Peláez

Camarógrafo
Bismarck Andrade Muller

Desarrollo Multimedia

Director General de Tecnología de Información
Dr. Ramón Parra Loera

Director de Desarrollo Informático de Apoyo Académico
Mtro. Juan Carlos Jiménez Márquez

Jefe del Departamento de Desarrollo de Multimedia
Mtro. Sajid Demian Lonngi Reyna

Diseño y Programación
D.G. Yasmán Lince de la Peña
D.G. María Reneé Galindo Mora
D.G. Miriam Santamaría Pérez
L.I. Gilbert Jiménez Quezada
L.I. Raúl Falfán Medina
L.C.T.C. Luis Alberto García Santana
L.C.T.C. Miguel Galero Carretero

Primera edición, mayo 2011
© Universidad Veracruzana
Dirección General Editorial
Hidalgo 9, Centro
Xalapa, Veracruz
Apartado postal 97, C.P. 91000
diredit@uv.mx
Tel/Fax (228) 818-5980, 818-1388

ISBN: 978-607-502-077-8
Impreso en México
Printed in Mexico

Se permite la reproducción total o parcial de este libro siempre y cuando se cite la fuente.

ÍNDICE

Introducción	9
I. Educación estadística, contexto y plan del curso	17
1.1 Enseñanza y aprendizaje	17
1.2 El enfoque basado en proyectos	20
1.3 El contexto del curso	32
1.4 El programa de Métodos Estadísticos	33
1.5 Actividades realizadas y criterios de evaluación	37
II. La metodología estadística y la investigación	39
2.1 Aspectos generales de la metodología de diseño estadístico	41
2.1.1 Diseños experimentales	41
2.1.2 Estudios observacionales	45
2.1.3 Estudios de muestreo	48
2.1.4 Análisis estadístico	58
2.1.5 Análisis exploratorio de datos	62
2.1.6 Gráficas y distribuciones	63
2.1.7 Análisis comparativo	65
2.1.8 Análisis bivariado y multivariado	67
III. Inferencia básica y modelación	70
3.1 Modelos paramétricos determinísticos	71
3.2 El modelo estadístico	72
3.3 El modelo probabilístico normal	74
3.4 Modelación de una sola muestra	77
3.5 Inferencia en la distribución normal	81
3.6 Prueba de hipótesis	83
3.7 Muestras aparejadas	85
3.8 Muestras independientes de dos poblaciones normales	85
3.9 Análisis de la varianza	88
IV. Modelación de regresión lineal	91
4.1 Análisis de regresión lineal simple	92
4.2 Análisis de regresión lineal múltiple	96
V. Elaboración de un reporte y una presentación	99
Anexo 1. Plan de actividades del curso	103
Bibliografía del curso y Referencias	106



Agradecimiento

Esta memoria no podría haberse realizado sin la participación entusiasta y activa de los estudiantes de ésta y anteriores generaciones de la Maestría en Ciencias Alimentarias. Ellos, con los estudiantes de la Facultad de Estadística e Informática, desarrollaron la mayoría de las actividades creativas que dan sustento a este documento. Quiero reconocer la contribución de los estudiantes: Laura Alicia Cuevas Viveros y Jeny Betsabé Vázquez Aguirre, quienes bajo mi dirección elaboraron sus trabajos recepcionales, -de maestría, el primero, y de licenciatura, el segundo- sobre el tema de educación estadística. Dichos trabajos han sido utilizados, en alguna medida, para diseñar y configurar esta memoria.

Agradecimiento especial les debo a María Nayeli Hernández Carmona, Jesús Hernández Suárez, María Luisa Hernández y Cecilia Cruz López, quienes han cargado con gran parte del trabajo de confección y mejora del documento. El CD que acompaña el escrito contiene un video, diseñado y elaborado por Pedro Valencia, con las ideas y soporte del equipo de trabajo completo, y la complicidad de Arely Ruiz Armas, productora de Televisión Universitaria. El resto de los contenidos en este medio son presentaciones y materiales que se deben a los estudiantes y al equipo. Agradezco mucho este material adicional que fue configurado por Juan Carlos Jiménez Márquez, quien tiene a su cargo la Dirección de Desarrollo Informático de Apoyo Académico de la Universidad Veracruzana.



Introducción

Al iniciar el siglo XXI muchos procesos de cambio social y económico se han profundizado por los diversos impactos y fenómenos derivados de la globalización. Varios de estos procesos afectan frontalmente lo educativo y propician la adopción de un nuevo paradigma, basado en el aprendizaje, orientado al estudiante y hacia la educación integral, sustentado en valores sociales y humanistas, y soportado en la tecnología. La práctica educativa se está transformando en todos los países, y México no es la excepción.

Los cambios en la educación superior suceden al realizar modificaciones y adecuaciones del currículum; es así que se promueven modelos de créditos que garanticen flexibilidad e integralidad en la formación. Ahora se asume que es urgente utilizar y adoptar estrategias centradas en el aprendizaje. En todos los ámbitos de la educación superior existe una urgente necesidad de diseñar y aplicar procesos de innovación educativa (Ribas-Navarro, 2000), que aseguren el aprendizaje significativo de por vida, para que produzca habilidades y actitudes en el estudiante con el fin de que siga aprendiendo después de haber concluido su programa de estudios.

Sin embargo, los cambios profundos podrán ser evidentes y palpables para los estudiantes cuando sus cursos sean distintos, cuando las actividades que realicen con los maestros en clase integren las nuevas intenciones y cuando los contenidos y las prácticas que realizan garanticen la formación de las competencias requeridas para su desempeño profesional; es decir, cuando los cambios sean reales y significativos para el estudiante y se manifiesten cotidianamente; esto es, que los cambios verdaderamente se palpen en los hechos. Estos cambios suceden ahora gradualmente y de forma diferenciada para cada disciplina o materia. Por ello se

pueden identificar movimientos de profesores, psicólogos educativos y gestores de educación, que trabajan caracterizando los problemas y las dificultades para el aprendizaje de cada disciplina. En este sentido, podemos encontrar líneas de investigación y desarrollo educativo como educación química, educación en geografía, matemática educativa, enseñanza de la economía, etc. Estos ámbitos de investigación y desarrollo se definen para estudiar y desarrollar estrategias, enfoques, instrumentos, etc., que permitan elevar los niveles de aprendizaje de las diferentes materias o disciplinas y así garantizar en el estudiante un aprendizaje significativo.

Unas de las materias que enfrentan serias dificultades por los bajos niveles de aprendizaje que de ellas resulta, son las relacionadas con los métodos cuantitativos; entre ellas destaca la estadística. Con el paso de los años se ha visto que el aprendizaje de esta materia para estudiantes de otras profesiones, ha sido muy pobre; también se ha concluido que la forma de impartirla ha carecido de elementos que logren en los estudiantes una actitud positiva hacia esta importante disciplina. Para resolver esta situación, los especialistas en educación estadística recomiendan, de manera repetida, los siguientes señalamientos:

- 1) Garantizar que la estadística se valore por su utilidad para resolver problemas reales.
- 2) Orientar las actividades hacia el aprendizaje de conceptos clave y principios estadísticos más que enfatizar en fórmulas y procedimientos.
- 3) Promover el apropiado uso de los métodos y técnicas de la estadística en una amplia variedad de actividades científicas y profesionales, así como en la vida cotidiana.

- 4) Incorporar los adelantos tecnológicos e instruccionales para garantizar el desarrollo de competencias en el uso de la metodología estadística.

El aprendizaje de la estadística es un tema que está adquiriendo gran importancia dentro de la educación internacional, principalmente por el papel que esta metodología desempeña en la investigación cuantitativa. En este sentido, hay ya un acuerdo generalizado entre investigadores educativos y profesores de estadística de asociaciones internacionales dedicadas a la educación estadística, respecto a la necesidad de diseñar programas y estrategias para mejorar la comprensión y aplicación de los métodos y técnicas estadísticas, así como la correcta interpretación de los resultados que se obtienen al aplicar dicha metodología. Existen también claras evidencias de que los educadores y profesores de estadística en muchos países están cerrando filas para garantizar mejoras en los contenidos y en los enfoques, así como en las estrategias que utilizan para hacer los cursos más activos e interesantes para los estudiantes y, sobre todo, para ofrecer cursos actualizados y adecuados a lo que las aplicaciones de la metodología estadística requieren.

La estadística puede ser vista como una disciplina científica, como una profesión o también como una materia que se dedica a sustentar y a promover una metodología para la investigación científica. En este texto nos interesa resaltar la tercera forma de ver la estadística; es decir, la que interesa a profesionales, técnicos e investigadores de otras disciplinas, quienes la usan como metodología para el diseño y desarrollo de sus estudios e investigaciones, bajo el paradigma cuantitativo; a ellos se llama usuarios de la metodología estadística.

Bajo este contexto, cuando un usuario de la estadística sabe valorar la importancia de los principios y

técnicas que la metodología establece para el diseño de una investigación, cuando es capaz de diseñar la obtención de sus datos de la manera más eficaz y eficiente, cuando tiene las habilidades para procesarlos usando, desde luego, programas computacionales estadísticos, y cuando, una vez que ha obtenido los resultados, los interpreta y concluye correctamente sobre las preguntas que dieron origen a la investigación, entonces se dice que es un buen usuario de la estadística, o bien, que posee el pensamiento estadístico.

Entre las innovaciones que la educación estadística plantea se encuentra el desarrollo de enfoques y estrategias que garanticen la formación de las competencias de un buen usuario de la estadística; es decir, que garanticen, que a través de un curso se adquiera el pensamiento estadístico. Así, un curso que persiga estos objetivos deberá replantear sus contenidos, su enfoque, su organización y su desarrollo.

El enfoque basado en proyectos se ha utilizado para mejorar los niveles de aprendizaje de la estadística. Un curso con este enfoque, consiste en que todas las actividades y contenidos se diseñen de manera que los participantes elaboren un proyecto; asimismo, utiliza el esquema de conferencias que se articula con las actividades prácticas desarrolladas por equipos de hasta tres participantes, con el fin de propiciar el aprendizaje cooperativo.

El principal objetivo del enfoque basado en proyectos es crear una actitud positiva para que los estudiantes tengan un marco y un esquema de trabajo que les permita vivir la experiencia en un curso activo, y que a través del diseñar, desarrollar y reportar una pequeña investigación, puedan vivir la aplicación auténtica de la estadística. Esto, desde luego, hará que los estudiantes puedan cambiar su visión y su actitud hacia esta materia. Para los usuarios de la estadística, tal

experiencia es fundamental ya que es justo en ese momento cuando reconocen que la estadística es una herramienta muy útil para su vida profesional.

También se pretende que los estudiantes logren llevar a la práctica lo visto en el curso; tal fin sólo se puede lograr si las actividades del curso se organizan alrededor del estudio de problemas reales. Bajo esta perspectiva, la estadística está fundamentalmente pensada como una metodología para el diseño y la realización de investigaciones factuales; es considerada como parte integral del proceso de investigación, tanto en la realización del protocolo como en la obtención y en el análisis de los datos, así como en la elaboración del informe o reporte. Cabe decir que es esencialmente un enfoque que permite aprender la estadística en el contexto del proceso de obtención de conocimiento, es decir, se aprende estadística haciendo investigación.

La idea al aplicar el enfoque basado en proyectos es que los estudiantes puedan diseñar y conducir una investigación acorde a sus intereses, motivaciones y conocimientos. Los proyectos que en este contexto se diseñan y desarrollan son llamados proyectos estudiantiles y permiten ejercitarse las técnicas de obtención de datos, la aplicación de los métodos de análisis y, finalmente, la elaboración de un reporte escrito y una presentación. En este sentido, los contenidos y actividades del curso se programan y organizan para dar sustento y soporte al diseño y desarrollo del proyecto.

En el curso de Métodos Estadísticos, de la Maestría en Ciencias Alimentarias de la Universidad Veracruzana, se ha aplicado el enfoque basado en proyectos, con la finalidad de lograr que los estudiantes del curso tengan un acercamiento diferente a la estadística y puedan emplearla como una herramienta indispensable en su vida como futuros investigadores y conductores de estudios en tecnología de los alimentos.

El objetivo general de esta memoria es, entonces, presentar los fundamentos de la metodología estadística en el proceso de la investigación, así como reseñar las actividades de diseño instruccional, el diseño del curso y los resultados obtenidos al aplicar este enfoque, enfatizando su aplicación en estudios experimentales, observacionales y muestrales. Se presenta de manera sucinta una serie de temáticas del análisis de datos y otros relativos a la modelación estadística; finalmente se dan una serie de recomendaciones para elaborar un reporte y una presentación tipo congreso.

Se adjuntan ilustraciones de los proyectos estudiantiles que fueron presentados en el marco del curso realizado entre agosto y diciembre 2008; tales presentaciones se incluyen en el material que va en el CD, un video que narra la experiencia vivida por una generación de estudiantes y algunas presentaciones adicionales que esperamos sean de utilidad.

Cabe destacar que esta memoria podrá ser provechosa para los estudiantes de las nuevas generaciones de la Maestría y de otros cursos que persigan objetivos similares, incluso de nivel de licenciatura; también para los profesores que tengan pensado innovar sus cursos y, por supuesto, para todos los interesados en la educación estadística. En cualquier caso, los materiales deben utilizarse de manera conjunta y necesariamente deben pensarse como un vínculo con los libros de la materia, con las propias lecciones del profesor y con las asesorías. Esto quiere decir que estos materiales buscan promover y provocar el estudio, la reflexión y el planteamiento de preguntas para el instructor, los auxiliares y asesores, más que ser una presentación acabada de estos temas. La forma en que sea utilizada esta propuesta dependerá, en todo caso, de la creatividad del instructor del curso en cuestión.

Como coordinador de este esfuerzo tengo la esperanza de que esta memoria sea una contribución para formar usuarios de la metodología estadística, -técnicos y científicos- con la adopción del pensamiento estadístico y que, por tanto, promueva la innovación educativa en muy diversos cursos de estadística aplicada.

Mario Miguel Ojeda Ramírez



I. Educación estadística, contexto y plan del curso

1.1 Enseñanza y aprendizaje

Una enseñanza que aspire a ser exitosa debe garantizar la formación de personas integralmente desarrolladas que apliquen en la práctica los conocimientos asimilados, que sepan solucionar problemas, que actúen creativamente y que posean una actitud a la vez que proactiva, crítica. La formación de una persona está integrada por el conjunto de aprendizajes significativos que ha adquirido, ya sea dentro de procesos escolarizados o fuera de ellos, en experiencias de la vida diaria, personal, familiar, social o profesional. Si la enseñanza considera este marco, seguramente quien la imparta buscará enfoques que consideren las condiciones del individuo en el momento de exponerse a la experiencia educativa.

La esencia de la enseñanza se halla en la transmisión de información mediante la comunicación; por tal motivo, la acción de comunicar representa un elemento de suma importancia para todo profesor. Un buen profesor debe comunicarse con un discurso claro, así como estimulante y motivador; por otra parte, la enseñanza tiene como objetivo lograr que el estudiante reconozca que teniendo habilidades y capacidades que lo faculten y, por tanto, le permitan enfrentar situaciones nuevas, será más apto, más capaz; es decir, mejor profesional.

El proceso de enseñanza consiste, fundamentalmente, en desarrollar una serie de mecanismos que propician en el individuo cambios graduales cuyas etapas se producen y suceden de manera ascendente; es por ello que a dicho proceso se le debe considerar como una progresión de eventos en constante movimiento. Dicho de otro modo, el proceso

de enseñanza produce cambios sucesivos e ininterrumpidos en la actividad cognoscitiva del estudiante; el profesor, en su labor conductora u orientadora, lleva al estudiante hacia el dominio de los conocimientos, hacia la adquisición de las habilidades, los hábitos y las conductas, acorde con su concepción y entorno; le da las herramientas que lo llevarán en su práctica a un enfoque para abordar la realidad, de una forma integral (Gagné y Briggs, 1997).

En la enseñanza se resumen conocimientos; se va desde el no conocer hasta el saber; desde el saber imperfecto, inacabado e insuficiente hasta el saber perfeccionado, suficiente -que sin llegar a ser del todo perfecto-, se acerca bastante a la aspiración que el estudiante pueda tener, basado en sus conocimientos previos y en sus necesidades, aspiraciones y motivaciones. A la enseñanza se le considera parte fundamental de la educación y, por lo tanto, un instrumento para la formación de una concepción determinada del mundo y sobre todo, de la vida profesional. En este sentido los contenidos de la enseñanza determinan, en gran medida, su efecto educativo. La enseñanza debe situar a los estudiantes en una posición que les permita construir un conocimiento bien estructurado a partir de lo que conocen, en su contexto y considerando los objetivos que se puedan plantear a la luz de sus necesidades concretas.

La enseñanza existe, entonces, para el aprendizaje; es decir, la enseñanza tiene que ser vista como un instrumento para el aprendizaje. Mediante la enseñanza, el aprendizaje se estimula, lo que hace posible a su vez que estos dos aspectos conserven -cada uno por separado- sus peculiaridades, y juntos conformen una unidad entre el papel de guía del instructor o profesor y la actividad del estudiante, la cual es fundamental para garantizar el aprendizaje (Nérici, 1980).

Al aprendizaje se le puede considerar como un proceso de naturaleza compleja en el que aparece un nuevo conocimiento, habilidad o capacidad; para que tal proceso pueda ser considerado realmente como aprendizaje, en lugar de una simple retención pasajera, debe ser susceptible de manifestarse en un tiempo futuro y contribuir, además, a la solución de situaciones concretas, incluso diferentes en su esencia a las que motivaron inicialmente el desarrollo del conocimiento, habilidad o capacidad. El aprendizaje se constata en cambios evidentes en la práctica, desde cómo concebirla, comunicarla y, sobre todo, resolverla. Es decir el aprendizaje, también resulta ser un producto, ya que atestigua, de manera concreta, las competencias adquiridas. Aprender es más que concretar un cambio conductual que se lleva a cabo en el sujeto que aprende. No debe olvidarse que la mente del estudiante no se comporta como un sistema de fotocopiado; es mucho más que eso: transforma integralmente los aspectos de la realidad objetiva y permite que el individuo transforme su práctica y a sí mismo. En este sentido, a los componentes asociados al proceso de enseñanza se les debe concebir como un sistema estrechamente vinculado con la actividad práctica que condiciona las posibilidades de conocer, de comprender y asimilar plenamente el conocimiento. Este sistema se perfecciona constantemente, dando mayores garantías a la actividad dirigida al proceso de obtención de los conocimientos, su aplicación creadora en la práctica profesional y en la vida cotidiana (Ausubel, 1978).

La enseñanza se diseña y orienta por los objetivos de aprendizaje que desempeñan la función de determinar los contenidos, los métodos y las formas de organizar su desarrollo; en consecuencia, los objetivos de aprendizaje presuponen entonces las transformaciones que se desean alcanzar en el estudiante al cual se enseña. Tales objetivos sirven además para orientar el trabajo tanto de los instructores como de los estudiantes

en el proceso de aprendizaje; también dan la pauta para seleccionar el mejor enfoque que se debe dar a los contenidos; es decir, cómo se deben enseñar. No es posible, por tal motivo, que se conciba a la enseñanza fuera de la perspectiva sistémica; la enseñanza carece de sentido si no se piensa en función de objetivos de aprendizaje.

1.2 El enfoque basado en proyectos

Un aspecto importante en el sistema educativo es el método de enseñanza; en este sentido debemos partir de que la didáctica orienta el proceso de enseñanza-aprendizaje. La característica principal del método de enseñanza consiste en que tiene un objetivo de aprendizaje e incluye las operaciones y acciones para este fin, como son la planificación y la sistematización adecuada. Nérici (1980) afirma que el método de enseñanza es el conjunto de movimientos y técnicas lógicamente coordinadas para dirigir el aprendizaje del estudiante hacia determinados objetivos. Para Dewey (1989) el método no es exterior al material, sino simplemente su tratamiento. Para él, material significa contenido y el método, el camino para aprenderlo.

Siempre que se habla de enseñanza se piensa en el maestro, pero cuando se habla de aprendizaje quien pasa al centro es el estudiante. El aprendizaje cambia al individuo y lo pone en una perspectiva mejor ante la vida y los retos de su ejercicio profesional futuro. Se puede decir entonces es ése el fin último de la educación. El aprendizaje es una actividad que debe asumirse con suma responsabilidad y conciencia -lo cual implica un convencimiento, un gusto y una motivación especial-. Asimismo, puede ser pensado como un reto individual específico, en el que gran parte del logro que se tiene es del ámbito personal. La motivación para el aprendizaje es imprescindible, se convierte en condición

necesaria y el factor decisivo en la realización del aprendizaje, que se logra con la actividad constructiva del estudiante; es él quien en último término modifica y reelabora sus esquemas de conocimiento, construyendo su propio aprendizaje. El profesor actúa como guía y facilitador en la construcción de aprendizajes significativos, los que se establecen a partir de relaciones entre los conocimientos nuevos, y las experiencias y los conocimientos anteriores (Estepa-Murillo, 2000).

En este contexto, en un curso de estadística se debería pensar en cambiar actitudes, deseos, habilidades y conocimientos de los estudiantes, con la finalidad de que ellos aprovechen los conocimientos y puedan a su vez utilizarlos en los problemas que se les presentarán en su vida futura como profesionales. Aunque esta pretensión parecería natural, se presenta con poca frecuencia entre quienes tienen la responsabilidad de hacerlo: los profesores e instructores.

Para el logro del aprendizaje, tanto alumnos como profesores deben tener interés en sus tareas, es decir, poseer una motivación positiva hacia el aprendizaje los unos, hacia la enseñanza, los otros. Si ello se cumple, y después se cuenta con un marco conceptual, información suficiente, y contenidos temáticos bien planteados se tiene entonces la base para desarrollar el aprendizaje. Sin embargo, el conocimiento no es la información, ni los contenidos temáticos, ni las lecciones impartidas en sí; pensado así se diría entonces que conocer sería igual a acumular información, lo cual si bien es importante, requiere de otros procesos del pensamiento. Hasta aquí el proceso está incompleto; el siguiente nivel de aprender se da con la experiencia, con la vivencia, con la puesta en práctica. Entonces es cuando el conocimiento, las habilidades y las actitudes se conjugan para dar lugar a las competencias, fundamentales para garantizar el aprendizaje significativo; para lograr tal fin es necesario que el

estudiante realice una fase de experiencia personal “aplicando el conocimiento” (Snee, 1993; Weldon, 1994).

Es importante que los profesores consideren al diseñar y conducir su enseñanza las actuales transformaciones de la educación y para ello deben tener presente una serie de preguntas que guiarán este proceso: ¿Qué estudiantes son? ¿Cómo es que aprenden? ¿Cómo debe enseñárseles para facilitar un aprendizaje adecuado? ¿Qué les motiva? ¿Qué les disgusta de los contenidos? ¿Qué métodos y enfoques motivan el aprendizaje?

Los cambios que se han producido afectan todas las áreas de la enseñanza y del aprendizaje; por ejemplo, ahora, en el trabajo en clase se utilizan diferentes ejercicios que propician la colaboración en parejas y pequeños grupos; se ha demostrado que este tipo de actividades resultan muy recomendables. Se dice que los cursos tienen que ser activos, y los materiales de estudio y de trabajo extraclase bien diseñados, para revalorar el proceso de aprendizaje. Por ello, la planeación y programación del curso son fundamentales.

El proceso de aprendizaje en todos los casos: el aúlico y el que sucede en los distintos ámbitos (familiar, laboral, comunitario, etc.) es producto de la interacción racional y emocional en cada persona. Por esta razón, aquellos profesores que además de enseñarnos, nos motivaron al aprendizaje al transmitirnos el amor y compromiso que sentían por su materia, es a quienes recordamos mejor y son aquellos precisamente quienes influyeron en nuestras posteriores decisiones. Incluir la perspectiva emocional en el tratamiento de los contenidos, permite, la mayoría de las veces, un trabajo académico más significativo entre los profesores y sus estudiantes. Para nosotros (Ver figura 1.2.1), hacerlo, fue muy importante porque (1) pusimos más atención a sus presentaciones, objetivos y teorías, (2) a su estudio y trabajo; asimismo (3) ello nos permitió tener

otras habilidades basadas en esos conocimientos; y (4) el ciclo del aprendizaje consistió en conocer las exigencias de los problemas de investigación en situaciones reales para revalorar y reformular los conocimientos y habilidades obtenidas.

Este ciclo, que aquí presentamos en la figura 1.2.1, sucede muy pocas veces en el contexto de la educación institucional; sin embargo, si se considera la manera en que se forman los médicos, por decir una de las profesiones que utilizan obligadamente la enseñanza teórico práctica, nos parece que, naturalmente debiera usarse la misma metodología cuando se plantea y desarrolla un curso destinado a usuarios de metodología estadística. Con los programas educativos habituales, comúnmente se logra arribar al punto (3) del diagrama. Entonces, se generan habilidades y hábitos que se encuentran soportados en conocimientos teóricos, pero que requieren cerrar el proceso de aprendizaje con la práctica. Por eso, el punto central de nuestro planteamiento será el valor de la motivación para llegar al final, que es de aprendizaje significativo. Abundaremos al respecto.

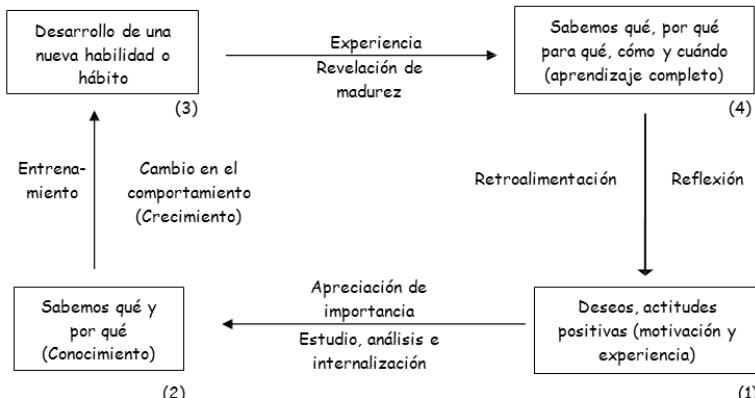


Figura 1.2.1. Ciclo del crecimiento, el aprendizaje completo y el cambio conductual general.

¿Por qué alguien decide estudiar algo? A esta pregunta se puede responder simplemente: tiene una motivación para ello. El motivo que impulsa a estudiar una materia en particular puede ser de diferente naturaleza y puede ocupar diferentes escalas o grados respecto del aprendizaje. Revisando algunas razones que los estudiantes enfrentan, un motivo para estudiar la materia estadística podría ser porque están obligados, al formar ésta parte del plan de estudios, por lo que les es absolutamente indispensable obtener esos créditos. Tal podría ser el caso de un estudiante de ingeniería, biología, administración, economía o sociología. Puesto así, el estudiante no tiene como propósito aprender, sino aprobar: el estudiante no tiene ningún motivo personal para desarrollar una actitud positiva hacia el aprendizaje de la estadística, sino por el contrario, el estudiante se predispone y no realiza su mejor esfuerzo en las tareas y actividades que el profesor plantea, argumenta una antipatía por los métodos cuantitativos y por los números. Cuando un estudiante está huyendo de las matemáticas y se encuentra que tiene más de un curso de estadística en su currícula, indudablemente muestra una predisposición negativa (Behar y Ojeda, 1995). Lo anterior es algo que el profesor no puede ignorar cuando diseña las actividades y lecciones del curso.

Una motivación más significativa para el estudiante podría darse cuando al cursar estadística, se percata de que en su futuro papel como ingeniero, biólogo, economista o sociólogo no podrá dar lugar a la lógica determinista y que, de pretender usarla, no le permitiría resolver satisfactoriamente una gran variedad de problemas. Si el estudiante se percata de que la estadística le será útil en su vida profesional de manera determinante, entonces tendrá una actitud distinta hacia su aprendizaje. A este respecto Slotnick (1989) menciona que los estudiantes de otras disciplinas que llevan un curso de estadística sólo aprenderán si se les muestra por qué les conviene aprender y para qué les va a servir

concretamente en su vida profesional lo que aprenderán. Él recomienda que a estos estudiantes se les ilustre cómo la estadística sirve de herramienta, tan valiosa como los conocimientos de la disciplina que estudian, ya que enseña el razonamiento científico. Se debe recordar que la estadística es una metodología que nos dota de principios, procedimientos y técnicas para la investigación cuantitativa, y su adecuado aprendizaje propicia las competencias para entender la dinámica y evolución de fenómenos colectivos, al tiempo que dota de habilidades para colectar, analizar e interpretar datos. Estas habilidades mejoran los niveles de aprendizaje de la estadística. Tal estrategia contribuye, dice este autor, a que los cursos de esta disciplina dejen de recordarse como los peores, los más tortuosos, los más frustrantes, y ya no se caracterice a la estadística como árida, con excesivos cálculos numéricos y conceptos difíciles de comprender.

El aprendizaje, la motivación y el rendimiento son tres conceptos básicos en la psicología, en la pedagogía; que un profesor debe manejarlos. Para saber de manera cercana las causas de satisfacción de la actividad docente, bajo la escala planteada por Zubieta y Susinos (1992) debemos destacar que la mayoría de los profesores se sienten más satisfechos por la materia que explican, que por las relaciones que establecen con sus estudiantes. Por otro lado, en el marco de la innovación educativa, se dice que el profesor en la actualidad debe seleccionar y organizar los contenidos y las actividades del curso, guiar a los estudiantes en el registro e integración de la información, ser competente en procedimientos y métodos de su disciplina, mantener la curiosidad intelectual de sus estudiantes y promover el aprendizaje independiente. Un buen profesor debe ser ante todo sincero y sencillo, ya que ambas características propiciarán un clima agradable para el intercambio y el aprendizaje, entre, y con los estudiantes.

Lamentablemente, la estadística es un buen ejemplo del fracaso de la educación tradicional. La gran mayoría de los profesores que la imparten perciben que los alumnos presentan una actitud poco entusiasta cuando platican sus experiencias. Kempthorne (1980) dice: "Ha habido una gran falla en la enseñanza de la estadística, originada por los propios profesores de estadística. Uno toma la ruta fácil de enseñar una especie de matemáticas. Uno puede tener la justificación parcial de que esa especie de matemáticas es una parte del área completa. Lo que debería ocurrir es que las ideas y metas estadísticas deberían determinar las matemáticas que deben ser enseñadas para la estadística y no al revés". Para agravar la situación, es que abundan los profesores carentes de competencias para propiciar la innovación educativa, así como el reiterado uso de textos obsoletos y de sistemas educativos anquilosados. Se puede decir que la mayoría de los cursos de estadística están mal concebidos y mal conducidos. Un curso tradicional de estadística mantiene las siguientes características:

- El profesor quiere enseñar.
- El alumno participa muy poco.
- Los contenidos están sobrecargados de formalismo.
- Se hace énfasis en algoritmos.
- El curso se imparte fuera del contexto de los estudiantes.
- Se hace uso de textos obsoletos.

Aun con buena intención y considerando que en el mejor de los casos se muestran ilustraciones, que no siempre cumplen con los requisitos necesarios para no inducir a

distorsiones del pensamiento estadístico, entonces se producen lastres en los contenidos y se afectan los objetivos de aprendizaje. La realidad es que los estudiantes están ante un escenario muy poco alentador. Sin embargo, hay ciertos avances que documentan optimismo.

Frecuentemente, a los cursos tradicionales se les critica por una o varias de las razones siguientes:

- Hacen mayor énfasis en probabilidad y matemáticas sin dar suficiente claridad a los conceptos estadísticos clave.
- Utilizan una amplia variedad de métodos presentados para aplicaciones que proporcionan poca o nula utilidad.
- Sobreestiman la importancia de cálculo y otorgan poco énfasis al razonamiento estadístico.
- Atienden insuficientemente la conducción de inferencias estadísticas a partir de datos y problemas reales.
- Estimulan la capacidad intelectual de los estudiantes, de manera escasa o nula.
- Plantean actividades que producen una excesiva angustia en los estudiantes y en algunos hasta un trauma, lo que evidentemente bloquea el aprendizaje.

Entonces, se puede decir que los cursos tradicionales de estadística no representan una buena motivación para incrementar el interés y aprecio por las técnicas, los modelos y la inferencia estadística; es decir, no producen el cambio de actitud necesario sobre la valoración de las herramientas estadísticas.

Al definir qué libro utilizar para un curso, se debe de tomar en cuenta que no reproduzca la metodología de los cursos tradicionales. Un buen libro es de gran ayuda, especialmente cuando se tiene poca experiencia como instructor y cuando se conoce poco de la disciplina en la cual se aplicará dicha enseñanza; sin embargo, debe tomarse en cuenta que el contenido no lo es todo, sino sólo uno de los componentes que integran el trabajo de enseñanza de la estadística aplicada.

Se sugiere que por todos los medios el profesor intente que durante la enseñanza de la estadística se presenten los vínculos que esta materia tiene entre un tema y otro, para despejar del estudiante la falsa idea de que la estadística es un conjunto de técnicas aisladas que sirven para propósitos difíciles de identificar. A este respecto Weinberg (1993) dice: "Otras conexiones presentadas con claridad ayudarán a reducir el sentimiento entre los estudiantes de que la estadística es una colección amorfa de tópicos..."

Podemos establecer que la problemática de la educación estadística para no estadísticos es compleja, pero es posible tener un acercamiento distinto si la abordamos desde una perspectiva que resuelva en la práctica la dificultad frecuente en el aprendizaje de los conceptos, principios y procedimientos de esta metodología. Como aspecto principal es posible considerar la falta de motivación y una actitud positiva hacia el aprendizaje, que ya mencionamos. El razonamiento estadístico se puede entender al contextualizar la metodología estadística en un proceso de investigación, particularizando en la disciplina para la que se enseña. Eso, y señalar al estudiante en términos de los problemas a los que se enfrentará cuando sea un profesional, le permitirá que resignifique su experiencia y se desprejuicie sobre la materia, lo que se traducirá en una valoración de la estadística como una herramienta de investigación muy útil.

Entonces, en muchos ámbitos se ha señalado ampliamente que los cursos de estadística para estudiantes de otras disciplinas deben -bajo un contexto de aprendizaje participativo- brindar oportunidades de trabajar con problemas reales (Snee, 1993; Weldon, 1994). La psicología del aprendizaje en el marco de una nueva pedagogía (Mayor et al., 1993; Beltrán, 1993) se distingue por buscar esquemas, enfoques y modelos que garanticen el aprendizaje significativo y que produzcan habilidades y actitudes para que el estudiante continúe aprendiendo (aprender a aprender) después de sus estudios. Para tal fin, se han promovido estrategias que reorientan su rol para que sea más activo, participativo, y que trabaje en el contexto de problemas reales, proponiendo soluciones en equipo (aprendizaje cooperativo) (Ojeda, 2001); esto es, propiciar que los estudiantes se expongan a la experiencia de aplicar la estadística, aún cuando fuera sólo con datos de problemas sencillos; indudablemente un mejor resultado se logra cuando los estudiantes realizan un proyecto de aplicación completo. A este enfoque se le llama precisamente aprendizaje basado en proyectos.

La idea de hacer un proyecto no implica un curso de investigación; si no más bien educar a los estudiantes para que aprendan a plantear problemas simples de índole estadística, de la vida cotidiana, de los problemas de su profesión y dentro del ámbito de sus intereses. Asimismo, que sepan manejar los conceptos clave del diseño estadístico: cómo diseñar una pequeña encuesta, un estudio observacional sencillo o un experimento simple. Los estudiantes deben ser capaces de plantear los objetivos y las preguntas de investigación de manera específica y clara. Si lo anterior se logra, ellos dimensionan la enorme utilidad de la estadística, pero enmarcada en el proceso de conducir un proyecto. Para entonces, el estudiante vive ya una experiencia de la estadística dentro del proceso de investigación.

Para que los estudiantes aprendan esta materia, sobre todo aquellos de otras disciplinas diferentes a la estadística, se hace necesario que se expongan a problemas reales de aplicación, y esto implica participar en la secuencia desde el primero hasta el último punto (Ver figura 1.2.2). Es claro que la mayor dificultad para este requerimiento será cómo lograrlo. La táctica a usar para cada curso corresponde decidirla a cada maestro, lo que implica diseñar e instrumentar cada curso particular. Un profesor no debe descuidar el dar buenas lecciones y conferencias, debe además incluir actividades como la presentación y discusión de proyectos ya realizados en los que se puedan observar los pasos de la secuencia señalados aquí en la figura 1.2.2; también puede utilizar seminarios estudiantiles, prácticas con datos reales, ejercicios realizados, presentados y discutidos en grupos, etcétera.

La idea central al diseñar y conducir un curso en este enfoque debe estar orientada a la mayor participación de los estudiantes en todas las actividades programadas; es decir, menos lecciones y conferencias del profesor, más trabajo en el aula, más trabajo con datos y sobre todo mayor involucramiento de los estudiantes en la solución de problemas reales; de esta manera, la estadística muestra su ayuda estratégica para plantear y resolver problemas. Debe quedar claro que la estadística no es una rama de las matemáticas, que tampoco es solamente procesamiento de datos y sobre todo que no es una ciencia teórica. En la figura 1.2.2 podemos ver las dos grandes fases de la investigación y la serie de actividades que se desarrollan en ese marco; en cada una de estas actividades los estudiantes encontrarán una conexión muy clara entre investigación y técnicas y métodos estadísticos; incluso hasta en la elaboración del reporte de los resultados y su comunicación.

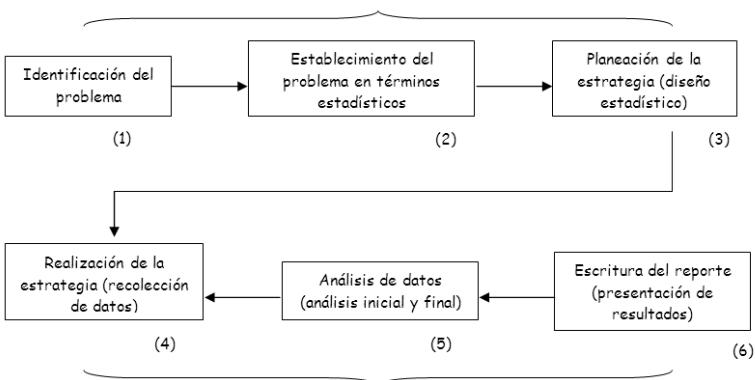


Figura 1.2.2. Esquema del protocolo de investigación que considera aplicar estadística.

La recomendación general que se pueda dar respecto a los contenidos, es que los cursos deben estar orientados sobre conceptos que son clave en la aplicación. Se escuchan insistenteamente las sugerencias de enseñar menos rudimentos matemáticos de la probabilidad y estadística matemática –combinatoria, deducción de fórmulas, demostraciones, etc.– menos recetas, menos trabajo mecánico irrelevante y más aspectos esenciales para la valoración crítica y la adecuada aplicación de los métodos estadísticos (Weldon, 1994; Weinberg, 1993). Una recomendación general para cualquier curso de estadística aplicada es realizar una caracterización de las necesidades de cada profesión para que se consideren los conceptos clave.

Es importante señalar que la estadística se aprende haciendo estadística, aplicándola a la solución de problemas reales. Por tal motivo se debe considerar el enfoque basado en proyectos estudiantiles partiendo de un buen programa, acorde a las necesidades de la profesión que los estudiantes ejercerán en el futuro y a la buena elección de un libro de texto. Sin embargo,

también es necesario reconocer la importancia que recae en la estrategia específica del profesor y su táctica para lograr los objetivos del curso, y que el logro del aprendizaje de sus estudiantes puede convertirse en un interesante reto para él.

Para la adquisición del aprendizaje significativo, es preciso que se hallen presentes -de manera permanente- la motivación del estudiante, su interés y su compromiso, así como también la comprensión del material estudiado en clase. Como se dijo, es de considerable importancia que el estudiante participe de manera activa en el proceso de asimilación y elaboración de la información y, sobre todo, en la aplicación de esa información en la solución de problemas de la vida real. Por ello se hace imprescindible dar las herramientas necesarias para que ellos posteriormente puedan transferir el uso de la estadística a un ámbito ajeno al curso. En este sentido podemos reiterar que el enfoque basado en proyectos se usa ampliamente en diversos contextos disciplinarios y hoy en día se reconoce su utilidad para elevar los niveles de aprendizaje y el aprecio por la estadística.

1.3 El contexto del curso

Los estudiantes de posgrado son, en general, estudiantes maduros, saben que lo que han decidido estudiar les implica un mayor compromiso y que los niveles de exigencia son mayores que en licenciatura. Otra característica que los distingue es que ya han probado la investigación, aun cuando haya sido de manera superficial. Fue ése el caso frecuente que encontramos con los estudiantes que ya realizaron su trabajo recepcional de licenciatura. Algunos ya han utilizado la metodología estadística en investigación y han puesto en práctica lo que aprendieron en cursos previos. Se puede decir que son estudiantes iniciados, la mayoría con ciertas ideas erróneas sobre la estadística, otros ya han tenido

experiencia laboral y son personas responsables y, en general, todos ellos están bastante dispuestos a asumir el control de su proceso formativo.

Los estudiantes de la Maestría de Ciencias Alimentarias son egresados del área de ciencias aplicadas como Ingeniería, Química, Nutrición, etc. y tienen facilidad para entender e involucrarse en los procesos de investigación. De hecho, varias de estas profesiones se ejercitan así; es decir, realizan experimentos y estudios sobre muestras, o utilizan expedientes y bases de datos ya existentes (estudios observacionales), por lo que los métodos y técnicas de la estadística pueden no ser desconocidas para ellos. En la medida que encontrábamos mayor o menor experiencia por parte de cada estudiante, era que mostraban una mayor motivación por comprender y capacitarse para ser un mejor investigador o tecnólogo en alimentos.

La metodología estadística que se estudia en las ingenierías se encuentra estandarizada por lo que los estudiantes tienen la experiencia de uno o varios cursos previos de carácter tradicional: con énfasis en procedimientos, fórmulas y con muchos ejercicios, razón por la que encontramos también algunos prejuicios y confusiones respecto a la dificultad de la estadística.

1.4 El programa de Métodos Estadísticos

El programa de la Maestría en Ciencias Alimentarias describe los contenidos y actividades del curso Métodos Estadísticos, que deben desarrollar los estudiantes bajo la coordinación del profesor, con el apoyo de los auxiliares y de estudiantes asesores de los últimos años de la carrera de Licenciado en Estadística, que trabajan aprendiendo cómo ejercer la consultoría estadística.

El curso de Métodos Estadísticos se encuentra ubicado en el área de formación básica en el primer semestre; cuenta con una duración de 45 horas y 6 créditos, aunque es necesario destacar que las actividades extraclase son de una intensidad y duración mayor al doble del tiempo considerado.

La estadística es una metodología fundamental para la investigación en ciencias alimentarias, tanto en la fase de diseño como para el análisis de datos y presentación de resultados a través de tablas, cuadros y gráficas. Los tecnólogos de alimentos usan intensivamente esta metodología, a tal grado que para un maestro en ciencias alimentarias es fundamental el uso de la metodología estadística en todo el proceso de investigación, incluso desde el diseño del protocolo, donde se requieren principios y conocimientos de diseño y análisis de este tipo. Es menester mencionar que para el diseño y desarrollo de nuevos propuestas, cuando se hacen valoraciones de los alimentos ya existentes y de los recién creados, así como diferentes estudios de impacto en la salud humana, el uso de metodología estadística es de suma importancia.

Bajo esta perspectiva, a continuación se presentan el objetivo general y los particulares, asociados a los contenidos y actividades que los estudiantes deben desarrollar para lograr un aprendizaje correcto de esta metodología.

El objetivo general es, entonces, dotar al alumno de los conocimientos básicos y de las habilidades fundamentales para constituirse en un buen usuario de la metodología estadística en el proceso de la investigación, bajo un enfoque orientado al desarrollo de proyectos y teniendo como condición el adecuado aprovechamiento de la tecnología computacional; es decir, el uso de paquetes de programas especializados para el análisis estadístico.

A continuación se presentan las unidades de estudio, los objetivos por unidad y los contenidos temáticos a desarrollar.

Unidad 1: La metodología estadística y la investigación

Objetivo: dotar a los participantes del marco conceptual de la metodología estadística en el contexto de la investigación, para proponer la elaboración de un proyecto.

- 1a. Caracterización del proceso de investigación.
- 1b. Diseño estadístico.
- 1c. Análisis estadístico.
- 1d. Elaboración del reporte.

Unidad 2: Aspectos generales de la metodología de diseño estadístico

Objetivo: dotar de conocimientos y desarrollar las habilidades que les permita distinguir las áreas de diseño estadístico, así como realizar el diseño estadístico asociado a un proyecto particular.

- 2a. Diseños experimentales.
- 2b. Muestreo estadístico.
- 2c. Estudios observacionales.
- 2d. Estudios mixtos.

Unidad 3: Análisis exploratorio de datos

Objetivo: desarrollar las habilidades para realizar adecuados análisis descriptivos y exploratorios usando el paquete Statística, y para diseñar la estrategia de análisis inicial en el marco del proyecto de trabajo en el curso.

- 3a. Tipos de datos.
- 3b. Gráficas y distribuciones.
- 3c. Análisis comparativos.
- 3d. Análisis bivariados y multivariados.
- 3f. Herramientas básicas no convencionales.

Unidad 4: Inferencia estadística básica

Objetivo: dotar de conocimientos y habilidades para realizar los procesos de inferencia estadística utilizando el enfoque de la modelación a través del paquete *Statistica*, así como para el diseño de estrategias de modelación estadística del proyecto.

- 4a. El modelo estadístico.
- 4b. Modelo de medias.
- 4c. Estimación y prueba de hipótesis.
- 4d. Análisis de varianza y ordenación de medias.

Unidad 5: Modelación de regresión lineal

Objetivo: desarrollar habilidades para los modelos de regresión lineal a través del paquete *Statistica*, así como para el diseño de estrategias de modelación estadística del trabajo en el curso.

- 5a. Regresión lineal simple.
- 5b. Regresión lineal múltiple.
- 5c. El proceso de modelación.
- 5d. Aplicaciones de la regresión.

Unidad 6: Estadística no paramétrica

Objetivo: enseñar los conocimientos y habilidades requeridas para realizar análisis básico de datos discretos y aplicar procedimientos no paramétricos de prueba de hipótesis a través del paquete *Statistica*, acorde a las necesidades concretas de los proyectos.

- 6a. Datos categóricos.
- 6b. Pruebas no paramétricas.
- 6c. Análisis de varianza no paramétrico.

Unidad 7: Análisis multivariado

Objetivo: garantizar el buen uso de las técnicas multivariadas usadas en los objetivos y procedimientos generales, al reconocer cuándo usar una u otra técnica.

cómo usarla con paquetes de programas disponibles y cómo interpretar y reportar los resultados.

- 7a. Análisis multivariado descriptivo.
- 7b. Análisis de componentes principales.
- 7c. Análisis cluster.
- 7d. Análisis de correlación canónica.
- 7e. Análisis de correspondencia.

1.5 Actividades realizadas y criterios de evaluación

Se prepararon conferencias para dar de manera rápida los contenidos; es decir, se hizo la transmisión de un marco conceptual básico. Estas conferencias estuvieron soportadas en los materiales necesarios para realizar un estudio y el análisis detallado de los temas. Para el aprendizaje fue necesaria la realización de prácticas individuales y por equipos, usando un software estadístico, por lo cual es muy importante que se siga un acucioso plan de actividades. A su vez, para integrar los contenidos se desarrollaron los proyectos, que fueron presentados inmediatamente después de la tercera unidad. El informe de cada proyecto se presentó con una exposición al finalizar el curso, utilizando un formato de artículo científico en un evento tipo congreso.

La evaluación se consideró de la siguiente manera: participaciones en las conferencias, 20%; forma y contenido de los reportes de las prácticas individuales y por equipo, 30%; proyecto y presentación, 20%; borrador del artículo y su presentación en el congreso, 30%. La dinámica del curso implicó seguimiento semanal de las actividades. Cada estudiante tuvo un expediente en el que se registraba cada producto y su correspondiente evaluación. Cada mes se realizaban sesiones de retroalimentación que permitieron ajustar actividades y dar soporte a los estudiantes con algún rezago.

Para las presentaciones públicas se utilizó un esquema de atributos predefinidos como referentes; se estableció una presentación tipo, así como sucesivas revisiones antes de la presentación final.

En el Anexo 1 se presenta una descripción detallada de las actividades que se desarrollaron en el marco del curso, tanto por parte del profesor como de los estudiantes, destacando los productos que éstos últimos debieron elaborar y entregar.

II. La metodología estadística y la investigación

La metodología estadística es un conjunto de principios, técnicas y métodos para diseñar y realizar una estrategia en la obtención correcta de conocimientos; está organizada en procedimientos para obtener y analizar datos e interpretar los resultados obtenidos; su fundamento son los principios, su lógica y su forma de pensar y actuar es consistente con el método científico. Se dice que más importante que conocer los métodos estadísticos es tener el pensamiento estadístico, traducido como una forma de pensar y actuar ante los problemas reales que requieran de metodología estadística para resolverlos. En este capítulo se da una visión panorámica de dicha metodología.

La estadística comprende tres aspectos básicos en el desarrollo de una investigación: (1) El diseño adecuado para la obtención de datos, (2) su análisis; y (3) la interpretación y presentación de los resultados en forma apropiada.

El diseño es la guía que conduce todo el proceso. Desempeña el mismo papel que el itinerario en un viaje, es lo que lleva al estudiante de un punto inicial u origen, al sitio final o resultados. Conduce a la formulación de la metodología que se utilizará para obtener los datos de acuerdo con las necesidades de información.

Entre los criterios que se emplean para formular la metodología del trabajo está el que los datos se colecten de la manera más rápida, económica y sencilla. El análisis de datos se da a partir de métodos y procedimientos existentes para explotarlos de manera que sea posible extraer de ellos lo relevante y resolver las preguntas que dieron origen al estudio o investigación. Finalmente, en la interpretación y presentación de los resultados, una serie de principios y procedimientos de la estadística proporcionan los lineamientos generales para elaborar

formatos de presentación y graficación, además de los elementos para construir juicios de valor de los resultados del análisis estadístico. Asimismo, la estadística da elementos para la adecuada construcción de conclusiones, recomendaciones y toma de decisiones en contextos de incertidumbre.

En la figura 2.1.1 se muestra que la metodología estadística se adapta al proceso de diseño y realización de la investigación, por lo que su particularidad en términos del qué y el cómo en mucho está definida por el problema objeto de estudio.

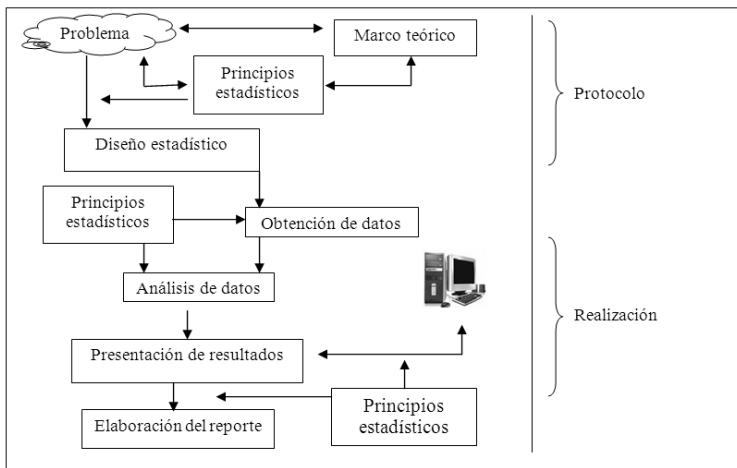


Figura 2.1.1. Proceso de aplicación de la metodología estadística en el contexto de la investigación en ciencias y disciplinas factuales.

El proceso general de la aplicación de la metodología estadística requiere necesariamente de un buen nivel de involucramiento en el problema en cuestión, en términos del manejo cabal de los conceptos clave de la estadística. Esto es fundamental para el profesor ya que le permitirá una buena comunicación con sus alumnos y la rápida comprensión del problema para contribuir en el diseño de la estrategia para resolverlo.

2.1 Aspectos generales de la metodología de diseño estadístico

La parte fundamental para ligar la metodología estadística con los pasos de una investigación fáctica se denomina diseño estadístico. En esta parte se definen los objetivos -el general y los particulares-, y se determinan las necesidades de información; asimismo, se determina el colectivo y la unidad de estudio, el esquema de obtención de datos y, lo más importante, qué se va a medir (variables) y cómo se va a medir (escalas y métodos de medición). Hay tres tipos de diseños estadísticos que tienen que ver con la forma de obtener el colectivo de estudio y la forma de utilizar la aleatorización. A continuación se presenta una descripción muy sucinta de estos tres tipos de grandes metodologías del diseño estadístico.

2.1.1 Diseños experimentales

Un experimento es una investigación donde ciertos factores se manipulan, definiendo los niveles y conjugaciones para constituir lo que se llama tratamiento; cada tratamiento se aplica a la unidad de estudio, que, en el caso de los estudios experimentales, se llama unidad experimental, a la cual se le miden una o varias respuestas. Las variables independientes son los factores y las dependientes, la o las respuestas. Es decir, una combinación de los niveles de los factores constituyen el tratamiento, y lo que se observa como resultado de la aplicación del tratamiento es la respuesta. De manera explícita, un experimento es el estudio de una relación causa-efecto que es controlada por el investigador, en la medida en que otros factores de influencia en la respuesta se aíslan o se reduce su influencia.

El colectivo de estudio en el caso de una investigación experimental es el conjunto de unidades

experimentales; la aleatorización y el control local constituyen dos ejes en la planeación del experimento e indican las actividades y procedimientos a seguir para desarrollarlo. A continuación revisamos algunos diseños básicos, donde los tratamientos están definidos por el investigador en el marco del estudio.

El diseño experimental más simple es el Diseño Completamente al Azar, en el que se cuenta con T tratamientos, donde T es un número mayor o igual que dos. Estos tratamientos se aplican a grupos que se forman aleatoriamente con las unidades experimentales. Una vez que se determinan los T tratamientos con precisión, se consiguen e identifican unidades experimentales, que serían $n \times T$, donde cada grupo tendría n unidades. Acto seguido, se realiza una selección aleatoria en las primeras unidades para formar el grupo 1, luego en las segundas para formar el grupo 2, y así sucesivamente; después se aplican los tratamientos a estos grupos.

Otro diseño es el llamado Diseño en Bloques Completos al Azar en donde hay una operación de control local sobre las unidades experimentales que se llama bloqueo; las unidades experimentales se dividen en b bloques; estas unidades se agrupan de acuerdo con diferentes niveles de un factor, tratando de que las unidades entre sí sean homogéneas y que los grupos sean diferentes. Se supone que el factor de bloqueo induce variabilidad en la respuesta, y que se pretende controlar el efecto de este factor; por tal motivo siempre hay que definir cuál es el factor que permite hacer el bloqueo. En cada bloque se tienen tantas unidades experimentales como tratamientos; es decir, si hay cinco tratamientos, cada bloque tendrá que tener cinco unidades experimentales; en cada bloque se realiza una selección aleatoria de esas cinco unidades para definir qué tratamiento recibirá cada una.

El Diseño en Cuadrado Latino, o Cuadro Latino como también se le llama, permite asignar T tratamientos con T repeticiones; esto determina el aspecto de "cuadro latino", el número de tratamientos es igual al número de repeticiones, entonces lo que se tiene es un conjunto de unidades experimentales bloqueadas por dos criterios. Los pasos son: 1) determinar los T tratamientos; 2) tener el cuadrado con las unidades experimentales bloqueadas en las dos direcciones, un factor fila y un factor columna; 3) asignar un "cuadrado latino", que es un arreglo de letras latinas a estas unidades de estudio. Los cuadros latinos aparecen tabulados en libros de diseños experimentales (Mead, 1983); hay una serie de arreglos de letras latinas que se llaman los cuadrados básicos, que se utilizan para realizar una permutación aleatoria de filas y columnas, y de esta manera: 4) determinar el tratamiento que se le asigna a cada una de las unidades en cada renglón y en cada columna.

Por ejemplo, un cuadrado 3×3 donde cada una de las letras latinas se repiten una vez en cada uno de los renglones y una vez en cada una de las columnas; una permutación aleatoria sería hacer una selección aleatoria de los números 1, 2 y 3; es decir, obtener al azar el primero, después el segundo y después el tercero, y esto determinaría una permutación del orden; supóngase que sale 3, 2, 1; eso querría decir, que las filas quedarían en ese orden; lo mismo se hace para las columnas, y estas permutaciones aleatorias mantienen las características del arreglo: cada letra aparece una sola vez en cada renglón y en cada columna. Entonces se puede decir que la aleatorización se hace sobre la base de las permutaciones aleatorias de los cuadrados.

Hasta aquí se ha revisado cómo asignar los tratamientos a las unidades experimentales, ya sean homogéneas o ya sean bloqueadas en una dirección o en dos direcciones. Sin embargo, es necesario establecer cómo conformar los tratamientos. Existe un tema en

diseños experimentales llamado diseño de tratamientos; y también, un esquema denominado esquema factorial completo, el cual implica que se ensayan tantos tratamientos como combinaciones de todos los niveles de todos los factores a estudiar; por ejemplo, si el factor *A* tiene tres niveles (a_0, a_1 y a_2) y el factor *B* tiene dos niveles (b_0 y b_1), el *C* tienen cuatro niveles (c_0, c_1, c_2 y c_3), ¿Cuántos diferentes tratamientos se pueden formar con estos niveles de estos factores? Se pueden formar $3 \times 2 \times 4$, o sea, 24 tratamientos, que serían todas las combinaciones de los niveles de los diferentes factores. Al experimento que considera todos los tratamientos se le llama un experimento factorial completo.

Dentro de los experimentos factoriales completos destacan las series 2^k , donde se ensayan dos niveles, que se identifican genéricamente como el nivel bajo y el nivel alto, de cada uno de *k* factores; por ejemplo, 2^3 a la 3 sería un factorial donde se tiene del factor *A* dos niveles, del factor *B* dos niveles y del factor *C* dos niveles; los tratamientos que produciría serían ocho.

Debe ser claro que un tratamiento define las combinaciones de las diferentes dosis de cada factor que lo conforma; así entonces, las series 3^k son utilizadas con bastante frecuencia; aquí cada factor tiene identificados tres niveles genéricos: el nivel bajo, el nivel medio y el nivel alto; por ejemplo, un 3^2 , producirá nueve tratamientos.

A veces *k* es grande; es decir, el número de factores es un número grande, digamos 6, o mayor. Cuando se tienen series 2^k ó 3^k con $k > 5$ entonces se cuenta con un número de tratamientos muy difícil de trabajar en la práctica; por ejemplo, si se tienen 64 tratamientos, ensayar esto en un cuadrado latino implicaría 64×64 unidades de estudio, que serían demasiadas; en esas circunstancias se puede ensayar sólo una fracción de los tratamientos; es decir, no todos

los tratamientos, sino por ejemplo la mitad o la cuarta parte para los casos de los factoriales 2^k , o bien terceras partes o novenas partes para los factoriales 3^k ; obviamente cuando sólo se ensaya una fracción de tratamientos hay una serie de efectos que no se pueden estimar. Para una introducción a esta temática el lector puede ver el capítulo XVI de Mead (1983).

En un plan experimental, antes que otra cosa, hay que hacer la determinación de objetivos; después la definición de factores, niveles, factores de bloqueo y variables; sigue la determinación de instrumentos y procesos de medición, hasta que se llega a la determinación del diseño experimental, que fundamentalmente es la metodología de diseño de un estudio experimental; una vez que se realiza el experimento se obtienen los datos; después vienen dos actividades finales que son el análisis de datos y la elaboración del reporte, donde se confinan las conclusiones y recomendaciones. Estas serían las fases de una investigación experimental.

2.1.2 Estudios observacionales

En los experimentos el investigador construye una situación a la que son expuestas las unidades experimentales, la cual consiste en aplicar un tratamiento, que es una condición o estímulo bajo determinadas circunstancias, para después evaluar los efectos de la aplicación. Por decirlo de una manera, en un experimento "se construye una realidad"; en un estudio observacional, sólo se observan las unidades de estudio en el contexto, que se delimita; por eso a veces a los estudios observacionales se les llama pseudoexperimentos. En este tipo de estudios las variables independientes ya han ocurrido por lo que no es posible manipularlas; es decir, el investigador no tiene control directo sobre las variables, sólo las observa y mide sus

efectos. Los diseños de estudios observacionales pueden clasificarse en transversales y longitudinales, y pueden también ser descriptivos, exploratorios, o correlacionales/causales. Cuando un estudio es longitudinal, interesa el estudio de la evolución de un colectivo en el tiempo e identificar la tendencia y observar los cambios.

Los estudios transversales son aquellos en los cuales se recolectan datos en un sólo momento; es decir, en un tiempo único, con el propósito de describir variables y analizar su incidencia e interrelación; por ejemplo, determinar el nivel de escolaridad de los empleados de un sindicato o investigar el estado de salud física y emocional de las personas que ingresan a un hospital. En estos ejemplos hay un conjunto de casos que se someten a medición.

Los diseños transversales pueden ser exploratorios, que se utilizan para una investigación inicial. Por lo general, se aplican a problemas nuevos o poco conocidos, y pueden describir a uno o varios grupos caracterizados por niveles de factores independientes. Muchos estudios en ciencias sociales son transversales exploratorios.

Por otro lado, los estudios transversales descriptivos tienen como objetivo indagar la incidencia y los valores en los que se manifiesta una o más variables (dentro del enfoque cuantitativo). Su procedimiento consiste en medir o ubicar un grupo de personas u objetos, en una variable o concepto y proporcionar su descripción; por lo tanto, son estudios puramente descriptivos y cuando se establecen hipótesis, éstas son también descriptivas. Estos estudios nos presentan un panorama del estado de uno o más grupos de personas, objetos o unidades en determinado momento. Por ejemplo, un investigador que desea describir el nivel de empleo en tres ciudades (Mérida, Mexicali y Sinaloa) o describir la vida de los niños

huérfanos que viven en la calle de tres barrios de la ciudad de México.

Los estudios transversales correlacionales/causales sirven para describir relación entre dos o más categorías, conceptos o variables en un momento determinado. A veces únicamente en términos correlacionales, otras en términos de relación causa-efecto (razones por las que se manifiesta una categoría, una variable, un suceso o un concepto) pero siempre en un momento específico. Por lo tanto, los diseños correlacionales/causales pueden limitarse a establecer relaciones entre variables sin precisar sentido de causalidad ni pretender analizar relaciones de causalidad. Cuando se limitan a relaciones no causales, se fundamentan en ideas o hipótesis correlacionales; y cuando buscan evaluar relaciones causales, se basan en ideas o hipótesis causales.

Los estudios observacionales longitudinales permiten obtener datos a través del tiempo en puntos o períodos, para hacer inferencia respecto al cambio, sus determinantes y consecuencias. Tales puntos o períodos se especifican por lo común al diseñar el estudio, y entonces las variables respuesta se van midiendo conforme avanza el estudio, siempre que el estudio sea prospectivo. Por ejemplo un investigador que desea analizar cómo evolucionan los niveles de empleo durante 5 años en una ciudad, o uno que busca observar cómo se desarrolla una comunidad indígena en varios años, con la llegada de la computadora e Internet a sus vidas.

Los diseños de investigación longitudinales se dividen en longitudinales de tendencia y los de evolución de grupo; los primeros son para analizar cambios a través del tiempo (en categorías, variables, conceptos o sus relaciones) dentro de alguna población en general. Por ejemplo, una investigación para analizar cambios en la actitud hacia el aborto en una comunidad. Dicha actitud se mide o se recolectan datos en varios puntos en el

tiempo (digamos, anualmente o en períodos no establecidos durante 10 años) y se examina su evolución a lo largo de este periodo. Por otro lado, los estudios longitudinales de evolución de grupo examinan cambios en la evolución del tiempo en subpoblaciones o en grupos específicos. Su atención son las cohortes o grupos de individuos vinculados de alguna manera.

Finalmente, los estudios longitudinales de panel son parecidos al de evolución de grupo, sólo que el mismo grupo de sujetos es medido u observado en todos los tiempos o momentos. Por ejemplo, una investigación que observa anualmente, durante cinco años, los cambios en actitudes de un grupo de ejecutivos en relación con un programa para elevar la productividad.

Las principales áreas de aplicación de los estudios observacionales son las áreas de ciencias médicas, principalmente la epidemiología, las ciencias sociales, la ecología y las ciencias de la conducta; pero también la economía, la administración y las áreas de los negocios, donde se tienen sistemas de información que cada vez es más urgente explotarlos para la toma de decisiones.

2.1.3 Estudios de muestreo

Con mucha frecuencia, en el contexto de las tareas de diagnósticos organizacionales o en la definición de los criterios que permiten establecer las especificaciones de un proceso de producción o de servicios, se recomienda la realización de una encuesta. Realizar un trabajo estadístico de este tipo no se reduce a elaborar una serie de preguntas y aplicar los cuestionarios para después obtener algunas gráficas y cuadros que una vez interpretados den los elementos para la toma de decisiones. Las investigaciones deben cumplir con criterios de validez y seguir una metodología diseñada y

conducida de manera escrupulosa, todo para aportar resultados confiables.

Cuando se va a tomar una muestra debemos preocuparnos por dos cosas: 1) cuántos elementos muestrear y 2) cómo seleccionarlos. El segundo problema se resuelve con un esquema de muestreo, es decir, con un método que nos lleve a obtener un subconjunto de los elementos de la población de muestreo. El primer punto nos indica cuál es el número mínimo de elementos que requerimos para lograr un propósito específico de inferencia con una precisión y confiabilidad establecida.

En general los esquemas de muestreo se dividen en probabilísticos o aleatorios y no aleatorios. En el caso de los estudios enumerativos se recomienda usar los probabilísticos. Recuérdese que en este caso se conoce N , el tamaño de la población de muestreo. En este capítulo se describen los principales esquemas de muestreo aleatorio y se discuten algunos procedimientos prácticos y rápidos para calcular el tamaño de muestra. También se incluyen algunos comentarios sobre muestreo no probabilístico, y al final hay una sección con recomendaciones para el diseño del cuestionario, aspecto que es fundamental para las encuestas de opinión y estudios de mercado. A continuación revisamos algunos esquemas de muestreo.

Muestreo Aleatorio Simple. Es el esquema probabilístico más simple y se constituye en la base de muchos de los otros procedimientos de esta familia. En él se supone que se tiene una población de tamaño N . La idea es darle a cada elemento la misma probabilidad de salir electo en la muestra, lo que se garantiza a través de una selección aleatoria simple de n números entre 1 a N , a partir de una tabla o con la ayuda de una calculadora. Estos números nos indicarán cuáles elementos se deben elegir y observar. Para hacer esta selección aleatoria se debe tener un listado de los elementos de la población

de muestreo que se denomina marco muestral. El esquema que considera este aspecto se trata a continuación en la figura 2.1.3.1 donde se representa gráficamente el muestreo aleatorio simple.

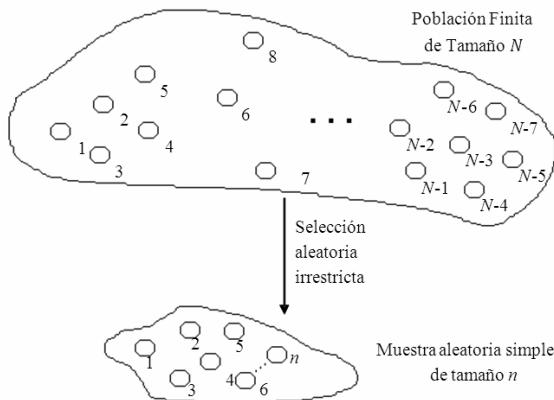


Figura 2.1.3.1. Esquema de un muestreo aleatorio simple

La principal desventaja de este muestreo es que la variabilidad en la muestra es mayor que la que resultaría al usar otros tipos de muestreo, y el riesgo de obtener una muestra poco representativa es más alta. También debe destacarse que con frecuencia quien realiza el estudio puede obtener información que le permita construir estratos (grupos más homogéneos), lo que da una mayor garantía de representatividad.

En los problemas reales es más frecuente tener una población estratificada, es decir, una población de muestreo compuesta por varios grupos bien identificados a los que se llama estratos. Los individuos pertenecen solamente a uno de los estratos. Para seleccionar una muestra estratificada de tamaño n , se procede de la siguiente manera. Sea $N = N_1 + N_2 + \dots + N_L$, donde N_L es el número de elementos o unidades en el L -ésimo estrato. Se determina n y se distribuye en los N estratos, por ejemplo usando asignación proporcional, es decir:

$$n_L = \frac{N_L}{N} n$$

Cabe hacer notar que hay otras formas de hacer la asignación, una de las cuáles se trata en la penúltima sección de este capítulo que es dedicada a la determinación del tamaño de muestra.

Entonces una vez determinado n_L se procede a realizar un muestreo aleatorio simple en cada estrato.

Debe destacarse que a partir de una muestra estratificada se ha aumentado la representatividad de la muestra con lo que la precisión de las inferencias será mayor. Además, ahora es posible realizar comparaciones entre los estratos. En muchas ocasiones el muestreo estratificado eleva sustancialmente las posibilidades de obtener un conocimiento más preciso sobre la situación de estudio, es quizás por eso uno de los esquemas más utilizados.

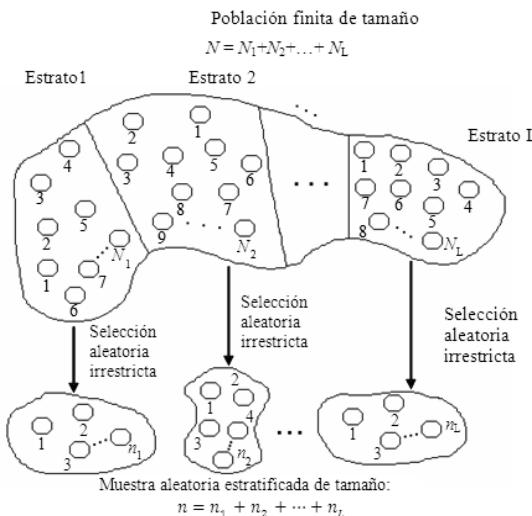


Figura 2.1.3.2. Esquema de un muestreo aleatorio estratificado

En los esquemas de muestreo anteriores es necesario conocer el marco, un listado de las unidades de muestreo de las que sea posible hacer la selección aleatoria y la ubicación de las unidades a muestrear, lo que en muchas situaciones prácticas es muy costoso. Sin embargo, es posible obtener un marco de grupos de unidades elementales, cuando éstas aparecen naturalmente agrupadas en pequeños grupos llamados conglomerados. Tal sería el caso de un estudio sobre ambiente organizacional en una cadena de tiendas. Aquí las unidades de estudio son los empleados, los cuales están agrupados por sucursal, y aunque tal vez podríamos obtener un listado de todos los empleados de la cadena resulta más práctico trabajar sólo con un listado de las tiendas. Si podemos obtener un listado de conglomerados es posible, a partir de éste, obtener una muestra aleatoria de unidades básicas.

Sea $1, 2, \dots, M$ el marco de conglomerados. Nosotros podemos seleccionar una muestra de unidades básicas en tres etapas:

1. Se selecciona una muestra de m conglomerados usando muestreo aleatorio simple. De cada conglomerado seleccionado obtenemos un marco de las N_i unidades $i = 1, 2, \dots, m$.
2. Seleccionamos una muestra aleatoria de tamaño n_i , $i = 1, 2, \dots, m$, de cada uno de los conglomerados.
3. Así la muestra total será de tamaño $n = n_1 + n_2 + \dots + n_m$. La forma de determinar n_i para cada conglomerado puede ser por separado o bien determinar n y después distribuirla sobre los m conglomerados. Ver figura 2.1.3.3.

En muchas situaciones, como en el caso del estudio de ambiente organizacional sobre la cadena de tiendas, los

conglomerados son en algún sentido también unidades de estudio, ya que hay un efecto contextual que hace que las unidades elementales de un conglomerado sean más homogéneas entre ellas que cuando son comparadas con las unidades de otros conglomerados.

El uso de los esquemas de muestreo en varias etapas, del cual el muestreo por conglomerados es un caso particular, enfrenta una serie de dificultades de tipo teórico, pero su valor práctico lo hace muy popular entre los investigadores.

Observe que el muestreo aleatorio y el estratificado no son alternativas excluyentes con el muestreo por conglomerados o con el sistemático. Pueden construirse estratos de conglomerados y en cada estrato realizar muestreo aleatorio de conglomerados y en cada conglomerado realizar muestreo sistemático. Evidentemente las expresiones para el cálculo de los estimadores se vuelven un poco complejas, pero es perfectamente factible diseñar e implementar estrategias combinadas.

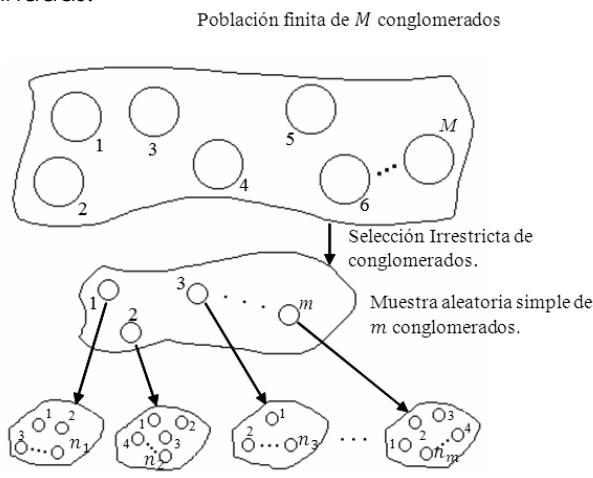


Figura 2.1.3.3. Esquema de un muestreo por conglomerados con selección de unidades en dos etapas.

Algunas poblaciones aparecen ordenadas físicamente en filas, gavetas, etc., o bien en el tiempo. Por ejemplo, podemos mencionar los expedientes de crédito, las cuentas, los clientes en el orden en el que llegan, etc. Una manera de aprovechar el orden para elegir una muestra es haciendo una elección sistemática. Para esto el total N de la población debe dividirse en n grupos, cada uno de tamaño k , así $N = nk$. Entonces de los primeros k elementos se selecciona uno aleatoriamente. El resto de los elementos de la muestra se obtiene sistemáticamente tomando siempre el elemento $j + ik$, $i = 1, 2, \dots, (n - 1)$. (Ver figura 2.1.3.4)

Si la población tiene un comportamiento cíclico la muestra puede ser poco representativa, por lo que se recomienda tener cierta información para garantizar la efectividad de este esquema cuando sea utilizado.

Con mucha frecuencia en la práctica resulta necesario combinar varios esquemas. Así se podría tener un muestreo estratificado y por conglomerados que en la segunda etapa utilice el procedimiento sistemático. Por ejemplo, para el caso de la cadena de tiendas, éstas pueden aparecer agrupadas a lo largo y ancho del país en regiones geográficas, lo cual daría un criterio para estratificar. Otro ejemplo es el caso de un estudio para determinar factores que definen la calidad de las tortillas elaboradas por medio de máquinas, que en México se denominan tortillerías. Para esto se obtuvo un listado de las tortillerías de la ciudad de Xalapa, Veracruz, las que fueron clasificadas de acuerdo a una categorización que de ellas ha hecho la Secretaría de Comercio. Dado que las unidades de muestreo son los clientes, se seleccionó una muestra estratificada de tortillerías y en cada una se tomó una muestra sistemática de clientes. Las características de la población y el problema bajo estudio en general serán los elementos que permitan confeccionar el plan de la forma más adecuada.

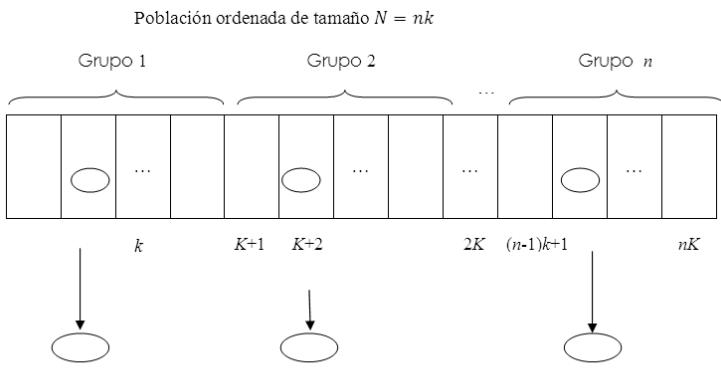


Figura 2.1.3.4. Esquema de una selección sistemática de tamaño n de una población ordenada de tamaño $N = nk$.

En algunas ocasiones se dificulta la selección al azar de la muestra, en estos casos, para realizar el estudio se puede tomar la muestra con base en el juicio. En la medida en que tengamos conocimiento de la población que deseamos muestrear, con métodos no aleatorios, podemos seleccionar aquellas muestras que pueden ser representativas. Si hay un experto de la población bajo estudio podría proponer una muestra "a juicio" que intente garantizar representatividad. Dado que esa muestra proviene del buen juicio de un experto, estrictamente hablando, las expresiones deducidas para muestreo probabilístico no resultarían adecuadas a esta situación, sin embargo, cuando no es posible realizar muestreo probabilístico, el buen juicio de un experto se convierte en la única alternativa. Cuando esta tarea se realiza con base en un equipo interdisciplinario, la muestra lograda puede arrojar una buena idea acerca de las características de la población.

Otra forma de construir una muestra es a través de cuotas establecidas en términos de algunas variables que

definen representatividad. Estas cuotas se les establecen a los trabajadores de campo y encuestadores, y ellos eligen las unidades específicas que pertenecerán a la muestra con las características que ya se han definido para su cuota. Aquí se deben establecer criterios específicos para que el sesgo de la elección se disminuya. En la medida que se garantice esto, y se consideren más variables para definir la representatividad, mayor calidad tendrá la muestra.

Hay que decir muy claramente que siempre que se pueda obtener una muestra probabilística, ésta se debe preferir, ya que con ella se garantizará la disminución de sesgos de elección, y en el sentido del proceso, se tendrá garantía de validez externa.

La determinación del tamaño de muestra es un tema que ha dado origen a mucho trabajo de investigación por parte de los estadísticos. Dado que éste depende de muchos factores, varias soluciones teóricas enfrentan dificultades prácticas o restricciones de costo. Los factores más importantes que deben tomarse en cuenta para la determinación del tamaño de muestra son:

- La(s) variable(s) de interés.
- La variabilidad en la población.
- El tamaño de la población.
- Los objetivos de la inferencia.
- La precisión y la confiabilidad.

A continuación se hará una justificación de los conceptos involucrados en el último inciso, ya que son de suma importancia.

La precisión de una estimación puede expresarse generalmente a través de dos elementos: El error tolerable (δ) y la confianza (γ) o confiabilidad. El error tolerable es la diferencia que estamos dispuestos a aceptar entre el verdadero valor poblacional (θ) y el calculado con la muestra $\hat{\theta}$. Por otro lado, la confianza es justamente la probabilidad de que el error tolerable no sea sobrepasado.

La relación entre el tamaño n de la muestra y el tamaño N de la población, para una precisión constante especificada, es tal que el tamaño de muestra crece muy lento, aun con grandes incrementos del tamaño de la población. Así, por ejemplo, para un tamaño de población $N = 300$ resulta una muestra de $n = 120$. Sin embargo, si el tamaño de la población se duplicara a 600, la muestra sería de 150. Nótese que no se duplica. Es más, si $N = 900$, el tamaño de muestra será de $n = 164$. Si la población fuese muy grande, digamos $N = 1,000,000$ el tamaño de muestra sería $n = 200$, el cual es el valor límite (tope), manteniendo en todos los casos el mismo nivel de precisión requerido.

Una manera muy simple de determinar el tamaño de una muestra es cuando es posible fijar el interés en una sola variable y el objetivo de inferencia es estimar por intervalo a la media. Así, el tamaño de muestra se obtendría usando la fórmula:

$$n = \frac{\sigma^2 Z_{(\alpha/2)}^2}{\varepsilon^2},$$

donde $Z_{(\alpha/2)}$ es un valor de tablas que determina un nivel de significancia o confianza (90%, $Z = 1.65$; 95%, $Z = 1.95$; 99%, $Z = 2.34$); σ^2 es la varianza de la población, que se puede sustituir por una estimación adecuada; y ε es la precisión con la que se desea obtener la estimación.

Para calcular el tamaño de muestra para una población estratificada se puede proceder calculando un tamaño de muestra para cada estrato, lo que implicaría contar con información sobre la varianza de cada estrato, o bien calculando un tamaño de muestra global y después haciendo una distribución sobre los estratos. El segundo procedimiento resulta más práctico y barato, pero la asignación del tamaño de muestra para cada estrato puede hacerse sólo en dependencia del tamaño del estrato. Si se cuenta con información sobre la variabilidad en cada estrato es posible utilizar una asignación o distribución del tamaño de muestra utilizando la ecuación:

$$n_L = \frac{nN_h\sigma_h}{\sum_{h=1}^L N_h\sigma_h},$$

donde N_h y σ_h son el tamaño del estrato de la formación de la muestra y la desviación estándar, correspondientes al estrato L - ésmo. Con esta expresión se toman más muestras de un estrato más grande y de los más variables o heterogéneos.

2.1.4 Análisis estadístico

Una vez que concluye la realización o implementación del diseño estadístico se tienen los datos, los cuales son revisados y capturados en un arreglo de renglones (casos o unidades) y columnas (variables). El proceso que se aplica a la matriz de datos con el propósito de obtener respuestas a las preguntas de investigación se denomina análisis estadístico, el cual tiene varias fases: la primera es de naturaleza descriptiva y exploratoria, esto implica la aplicación de procedimientos de conteo y la obtención de tablas de frecuencias y porcentajes con el propósito de tener la primera información sobre los patrones y la variabilidad.

Al análisis que se realiza de cada columna de la tabla de datos por separado se le llama análisis marginal, éste proporciona la base para las primeras conclusiones sobre el estudio. Al análisis marginal le sigue el análisis bivariado o cruzado, esto implica la selección de una serie de preguntas de interés, las cuales nos permiten identificar las variables a cruzar. La elaboración de conteos cruzados y la construcción de tablas de resumen permiten tener soporte para contestar las preguntas de investigación. Cuando hay una cantidad grande de variables explicatorias y de variables respuesta, entonces se requiere de esfuerzos técnicos especiales para el tratamiento de estos datos; se dice que entonces deben usarse técnicas multivariadas, las cuales son parte de las técnicas avanzadas del análisis estadístico. En realidad cada objetivo del estudio sugiere una estrategia de análisis, la cual a su vez puede disponer de una o más herramientas estadísticas.

Para hacer un análisis estadístico hay que entender la estructura de la matriz de datos y los datos mismos. Se deben realizar análisis marginales e ir construyendo poco a poco juicios sobre el colectivo. En este sentido el análisis estadístico de los datos es un proceso iterativo de descubrimiento, lo cual quiere decir que los procedimientos de cálculo aritmético y la realización manual o con calculadoras ha dejado de ser práctico y eficiente, ahora adquiere importancia el saber cómo hacerlo y cómo entender los resultados o salidas, pero a través de paquetes estadísticos.

Ante esta perspectiva, y por la situación del desarrollo de la computación estadística, se debe poner mayor atención a la forma correcta de explotar las facilidades que brindan los paquetes estadísticos, porque es muy probable, y se da de manera frecuente, el hecho de pensar que por aprender a operar un paquete estadístico es posible hacer cualquier análisis estadístico olvidándose de la validez externa o interna del estudio,

tal actitud ha multiplicado los malos usos de la metodología estadística. A este respecto, (Chatfield, 1988) plantea una serie de reglas para analizar datos en el contexto de un estudio o investigación en general. Aquí presentamos las llamadas “Seis Reglas Básicas”.

1. No intentar analizar los datos antes de tener un entendimiento claro de qué es lo que se está midiendo y por qué, tratando además de encontrar si existe información anterior o primaria acerca de los posibles efectos que pueda introducir cada variable en el comportamiento general del problema o fenómeno. En este orden de ideas, el analista de los datos deberá hacerse muchas preguntas con la finalidad de: clarificar los objetivos del estudio o análisis del problema; conocer el significado de cada variable y las unidades en que se están midiendo; conocer el significado de los símbolos especiales que se estén utilizando (si los hay); y si existen experiencias similares que aporten información complementaria sobre el problema o fenómeno en cuestión que apoye los análisis, entonces deberá acceder a la revisión de ésta.
2. Una vez realizado lo anterior es necesario conocer cómo fueron recolectados los datos. Aquí se destaca básicamente la necesidad de conocer si hubo un proceso de aleatorización u otra forma de garantizar la validez externa que sea apropiada y que garantice la confiabilidad de las mediciones. Si los datos provienen de un proceso no aleatorizado propiamente, posiblemente sólo sea justificado realizar un análisis descriptivo simple, lo cual tendrá que ser explícitamente indicado. Hay muchas técnicas estadísticas que se soportan sobre supuestos restrictivos, que de no cumplirse le restan validez a los resultados.

3. Especifique cuál es la estructura de los datos. Siendo importante aquí contestar las siguientes preguntas: ¿Son suficientes las observaciones para explicar el problema o fenómeno? ¿Son muchas o pocas las variables explicativas? Aquí es necesario distinguir los diferentes tipos de variables que se vayan a estudiar, definiendo si son controlables, explicatorias o variables respuesta, etc. Además debe hacerse una clasificación de variables por tipo de medida o escala, y por la naturaleza: continuas o discretas, cualitativas, binarias, etc. Todo ello porque los análisis resultantes dependen críticamente de la estructura que guarden los datos.
4. Posteriormente los datos deben ser examinados en una forma exploratoria antes de tratar de intentar un análisis más sofisticado. Para llevar a efecto este análisis es necesario el cálculo de estadísticas básicas y el ajustar gráficas de funciones a los datos en cualquier forma que aparezca apropiada, haciendo esto separadamente para cada variable (y en algunos casos para pares de ellas). Se recomienda el uso de histogramas, diagramas de cajas y alambres, así como diagramas de dispersión, de tallos y hojas, para hacerse una idea de la distribución que pueda suponerse para los datos, así como también para tratar de observar los efectos de los valores faltantes o valores extremos y que puedan, o no, afectar los posibles análisis.
5. Utilizar el sentido común todo el tiempo. Aunque los resultados de aplicar la estadística pueden contradecir algunas ideas preliminares, la explicación que se logra de los resultados es consistente con la lógica del sentido común.

6. Reportar los resultados en una forma clara y explicativa por sí mismos. Debemos dejar que las evidencias hablen, que sean el eje del discurso de las recomendaciones y conclusiones.

2.1.5 Análisis exploratorio de datos

Las características o conceptos que se miden en las unidades de estudio se denominan variables, y se clasifican, por su naturaleza, en continuas y discretas. Las variables discretas son aquellas características en las que las categorías que puede tomar la variable son un número finito. Las continuas son variables que, en principio, su medición puede resultar cualquier valor en un continuo. En el proceso de medición todas las variables son discretas debido a la precisión de los instrumentos que restringe el conjunto de valores posibles siempre a un conjunto finito. También pueden encontrarse variables que combinan las características de las continuas y discretas; son llamadas variables mixtas. Aunque son raras, también es posible encontrar variables de esta naturaleza.

Usualmente los datos se obtienen a partir de motivaciones específicas y es posible establecer una serie de preguntas que permiten clasificar a las variables como independientes o explicativas y como variables respuesta. Esto proporciona los elementos para especificar con precisión los objetivos del análisis estadístico, sobre todo cuando se establece como objetivo el estudio de una relación causa-efecto.

El producto de la medición son los datos, los cuales se organizan en una matriz o tabla de doble entrada en la que los individuos o unidades de estudio son los renglones y las mediciones en las diferentes variables son las columnas, algunas de las cuales sirven para definir la

estructura del colectivo; ésta es, junto con las preguntas de investigación, la materia prima del análisis estadístico.

La matriz de datos usualmente se constituye en un reto para la creatividad del analista. Al analizar esta matriz hay que tener siempre en mente las preguntas que originaron el estudio, el objetivo preciso y las ideas clave de la estadística. Con los medios modernos de la computación, hacer un análisis estadístico requiere del diseño de estrategias muy precisas.

El análisis exploratorio es un conjunto de técnicas numéricas y gráficas que permiten realizar un análisis de los datos sin tener que considerar modelos probabilísticos; es decir, el análisis exploratorio nos permite ver y evidenciar patrones, tendencias y grado de dispersión de los datos. Muchas veces la aplicación apropiada de estas técnicas nos permite obtener respuestas para las preguntas de investigación, y muchas veces esto es suficiente para dar conclusiones y recomendaciones.

2.1.6 Gráficas y distribuciones

Debido a que un conjunto grande de datos no nos dice gran cosa, algunas veces se desean condensar los datos de una manera más manejable, esto se hace fácil utilizando una distribución de frecuencias.

Los datos se denotan mediante una variable X y se le asignan sus respectivas frecuencias (f), esto no es otra cosa más que el número de veces que ocurre el valor x en la muestra.

Las distribuciones de frecuencias pueden estar agrupadas y no agrupadas. La primera significa que los valores de X no se combinan para formar grupos, sino que cada x es un grupo en sí. En la segunda, los datos se agrupan en intervalos de clase; esto es recomendable,

cuando en un grupo grande de datos en un continuo existen diferentes valores de X en lugar de unos cuantos repetidos.

Todas las clases que se utilicen deben de ser de la misma amplitud y éstas se deben de establecer de manera que a cada dato pertenezca una sola clase. Es recomendable que la amplitud de la clase sea un número impar. Para formar las clases se busca el dato mínimo y el máximo y se obtiene la amplitud, se selecciona el número de clases deseado, se divide la amplitud entre el número de clases y se obtiene el ancho de la clase, se elige un valor inicial, este valor debe de ser un poco más pequeño que el valor mínimo. Los algoritmos computacionales de los paquetes estadísticos hacen estas operaciones.

Los histogramas son un tipo particular de gráfica de barras que representa a un conjunto de datos en una escala de razón o de intervalo, está compuesto de:

- Un título que identifica las frecuencias de las clases.
- Una escala vertical que identifica las frecuencias de las clases.
- Una escala horizontal que identifica a la variable X .

Cuando los valores de la escala vertical se cambian a fracciones

$$\left(\frac{i}{\text{num. de observaciones}} \right)$$

donde i es el número de veces que aparece un valor en el rango; el histograma se convierte en un histograma de frecuencias relativas.

Los histogramas son de gran utilidad, ya que nos ayudan a ver la forma de la distribución de los datos. El histograma de una muestra debe de tener una distribución cuya forma sea similar a la de la población de la cual se tomó la muestra. Si el que analiza el histograma está familiarizado con la variable considerada, podrá interpretar varias características importantes. Los histogramas pueden ser de diferentes formas: normal, uniforme, sesgado hacia la derecha o izquierda, bimodal (las dos clases con mayor frecuencia están separadas por una o más clases, a menudo esta situación indica que se han muestreado dos poblaciones), la normal, etc.

2.1.7 Análisis comparativo

Una distribución de frecuencias puede convertirse en una distribución de frecuencias acumuladas muy fácilmente. Esto se realiza colocando un subtotal de las frecuencias junto a cada clase. Una ojiva es una distribución de frecuencias (o frecuencias relativas) acumuladas, igual que el histograma ésta se compone por un título; una escala vertical, que identifica las frecuencias acumuladas o las frecuencias relativas acumuladas; y una escala horizontal que identifica las fronteras superiores de cada clase. Es útil cuando se quiere representar el rango porcentual de cada valor en una distribución de frecuencias.

Muchas veces en la práctica la muestra o conjunto de casos se presenta con una estructura de grupo. Existen variables como género, edad, etc. que nos pueden explicar la variabilidad que estamos estudiando. Estas variables llamadas “indicadoras” nos permiten dividir un conjunto de datos en varios grupos; otras veces se aplican diferentes tratamientos a las unidades de estudio, por lo que se tienen varias muestras. Esto es lo más

frecuente en las situaciones reales, y en estos casos se requiere un análisis comparativo.

Las tablas de contingencia nos permiten tabular los datos nominales en varios grupos (muestras estratificadas); en la representación gráfica de este tipo de tablas se pueden utilizar, barras comparativas o pasteles comparativos. Para ilustrar el concepto suponga que se aplicó una encuesta por colonias sobre una muestra de viviendas estratificadas, de acuerdo a la apariencia de la vivienda:

- Viviendas de clase media alta y alta.
- Viviendas de clase media.
- Viviendas de clase baja.

Suponga que se le preguntó si consumían carne regularmente al menos 3 veces por semana. La respuesta fue dicotómica (Sí, No). La tabla 2.1.7 es una presentación de los datos.

Tabla 2.1.7. Datos hipotéticos de una encuesta sobre un marco estratificado.

Respuestas	Estratos			Totales
	A	B	C	
Si	75	58	27	160
No	13	69	86	168
Totales	88	127	113	328

Para graficar los datos de una tabla de contingencia se puede utilizar el gráfico de barras comparativo (se usa cuando las observaciones que se están estudiando se subdividen en grupos); sirve para visualizar cada grupo de estudio y compararlos.

También se puede utilizar un gráfico de pastel, en el cual se puede examinar la distribución de frecuencias para cada categoría. Aquí, dado que los datos están divididos, se elabora un gráfico para cada grupo.

Para cuando se tiene un conjunto de datos cuantitativos y éstos son pocos, se puede utilizar un diagrama de tallos y hojas, también se puede utilizar un histograma, en el que se puede examinar la distribución de frecuencias para los valores de una variable. En el caso de que los datos están divididos, se elabora un gráfico para cada grupo. Otro gráfico que se puede utilizar es el de tallos y hojas comparativo.

2.1.8 Análisis bivariado y multivariado

El análisis bivariado es aquel que nos permiten explorar la asociación entre dos variables. El diagrama de dispersión se utiliza para visualizar relaciones entre dos variables X y Y ; los datos individuales representados por puntos etiquetados en un espacio bidimensional. Las dos coordenadas que determinan la posición de cada punto corresponden a los valores específicos de las dos variables X , Y . Si las dos variables están relacionadas fuertemente, entonces los puntos forman una figura sistemática (por ejemplo, una línea recta o una curva). En dichos gráficos también se pueden identificar datos “outliers” (datos atípicos), los cuales se perciben fuera del patrón que siguen los datos.

El coeficiente de correlación de Pearson es un valor numérico que nos dice el grado de asociación lineal entre dos variables X y Y . Dicho coeficiente de correlación se denota con la letra r , y lo calculamos con la siguiente expresión:

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}},$$

donde s_x y s_y son las sumas de cuadrados de cada variable y s_{xy} es la suma de productos cruzados; es decir:

$$s_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$
$$s_x^2 = \sum_{i=1}^n (x_i - \bar{x})^2$$
$$s_y^2 = \sum_{i=1}^n (y_i - \bar{y})^2$$

El diagrama de dispersión con histogramas es una gráfica en la que se observa la asociación entre sus variables y la distribución de cada variable; para el diagrama de dispersión se especifican las dos variables (X, Y) y para el histograma las variables son desplegadas a lo largo de los ejes (X, Y) del respectivo diagrama de dispersión. Este tipo de análisis permite tener una visión univariada para cada variable así como la visión de la asociación entre las dos variables. Es una herramienta que enfoca cualquier complejidad que pueda ser intrínseca a los datos; ya que éstos pueden ser no normales, e incluso tener una asociación no lineal. El rol del análisis de datos es extraer la máxima información de los datos, y presentar un producto concreto y tan útil como sea posible.

El análisis multivariado generalmente es aplicado a situaciones donde:

- a) se quieren encontrar patrones sistemáticos o grupos de observaciones en problemas que implican la consideración de varias variables respuesta,
- b) se quieren explorar posibles relaciones complejas entre varias variables.

Una de las técnicas estadísticas en el análisis multivariado es la de Componentes Principales, que se utiliza para detectar cuáles variables tienden a mostrar niveles de asociación lineal. Esta técnica es muy popular y permite construir índices y estudiar agrupaciones, tanto de individuos como de variables.

El gráfico de las caritas de Chernoff se basa en la representación de valores de variables en las características de una cara. Se pueden representar hasta 20 variables, las cuales deberán ser asociadas con una de las partes de la cara. Esto permite desplegar cada caso o entidad (que puede ser un grupo de casos) y así estudiar agrupaciones.

Gráfica de polígonos o estrellas, en este diagrama se representa el valor de cada variable como una cantidad sobre una línea de un conjunto de k – líneas igualmente espaciadas que surgen del centro de un círculo. Para formar la estrella se unen los puntos de las k – líneas formando un polígono.

La matriz de correlación resume la relación lineal entre diversas variables en una matriz de tantas entradas como variables se tengan en estudio. El gráfico matricial de correlación presenta diagramas de dispersión por pares de variables y también se conoce como diagramas de escalera, permite estudiar la asociación de un grupo de variables.

III. Inferencia básica y modelación

El término modelo fue empleado originalmente por Niels Bohr en la descripción del átomo de hidrógeno publicada en 1913, y a partir de esa fecha, el uso y la generalización del concepto de modelo ha sido tal que se le ha llegado a considerar como uno de los máximos logros del desarrollo humano en el siglo XX.

Existen diferentes definiciones de modelo, que es un concepto versátil. Aquí se parte del concepto de modelo matemático, el cual se puede caracterizar simplemente diciendo que es una explicación teórica del fenómeno bajo estudio. Esta explicación primero se bosqueja en forma verbal y luego se formaliza mediante una o más ecuaciones. En un modelo matemático se debe identificar una o varias variables dependientes y una o varias variables independientes; el modelo, entonces, plantea una relación causa-efecto. Ahora bien, es posible utilizar diferentes tipos de funciones matemáticas para proponer un modelo que representa a una variable respuesta como función de una o más variables independientes, pero todos estos tipos de funciones matemáticas las podemos identificar como modelos determinísticos, que son ampliamente utilizados en los libros de cálculo y de física. Por ejemplo, la primera Ley de Newton $F = ma$, es un modelo determinístico, porque en situaciones cotidianas establece predicciones con un margen de error pequeño, que el modelo no considera.

En general, denotemos por un modelo determinístico a la relación funcional:

$$y = f(x)$$

donde y es un escalar o vector y x es también un escalar o vector.

De acuerdo a la forma de la función $f(x)$ es posible clasificar a estos modelos determinísticos. A continuación hacemos una presentación de los más importantes para la estadística.

3.1 Modelos paramétricos determinísticos

En esta clase de modelos la función $f(x)$ se hace depender de un vector $\boldsymbol{\theta}$ de parámetros, y entonces se denota a la relación funcional como:

$$y = f(x; \boldsymbol{\theta})$$

Algunos ejemplos de modelos paramétricos determinísticos son los siguientes:

1. $y = \beta_0 + \beta_1 x$
2. $y = \beta_0 + \beta_1 x + \beta_2 x^2$
3. $y = \beta_0 + \beta_1 x + \beta_2 x_2 + \cdots + \beta_p x_p$
4. $y = \frac{1}{1+e^{-x\beta}}$
5. $y = e^{-\beta x}$
6. $y = x^\beta e^{-\beta_1 x}$

Los primeros tres ejemplos son de los llamados modelos lineales, ya que la función $f(x, \boldsymbol{\theta})$ se expresa como una combinación lineal del vector $\boldsymbol{\theta}$, que en el segundo caso es $\boldsymbol{\theta}^t = (\beta_0, \beta_1, \beta_2)$, y en el tercero $\boldsymbol{\theta}^t = (\beta_0, \beta_1, \dots, \beta_p)$. El primero es un modelo de línea recta, donde β_0 es la ordenada al origen y β_1 es la pendiente de la recta. Este modelo expresa la relación funcional como una función del valor escalar x , al igual que el segundo

ejemplo; sólo que en éste tenemos una ecuación cuadrática.

Sin embargo, en esencia el segundo modelo sigue siendo una combinación lineal de los parámetros en $\boldsymbol{\theta}^t = (\beta_0, \beta_1, \beta_2)$. El tercer modelo es la generalización de esta familia, por lo que a esta forma funcional se le llama el modelo lineal general.

Los siguientes tres ejemplos son casos de los llamados modelos no lineales, ya que no pueden ser expresados como una función lineal del vector de parámetros $\boldsymbol{\theta}$. Sin embargo, estos modelos comparten una propiedad interesante: si se les aplica una transformación adecuada pueden convertirse en un modelo lineal.

3.2 El modelo estadístico

La inferencia estadística parte de la postulación, explícita o implícita, de un modelo que describe el comportamiento de una o varias variables aleatorias; este modelo está compuesto por dos partes: 1) sistemática (explicada); 2) aleatoria (no explicada). El modelo estadístico será:

$$y = f(A, B, \dots) + e(F, G, \dots)$$

donde A, B, \dots es un conjunto de factores o variables que el investigador conoce (controla), f es una función de esos factores o variables, y e es una parte que describe la contribución de la aleatoriedad, siempre presente en los fenómenos, debida a los factores F, G, \dots no controlados y a veces hasta desconocidos.

La postulación de la parte sistemática depende del conocimiento del investigador sobre el comportamiento del fenómeno bajo estudio. En algunas ocasiones por teoría física el investigador conoce de la forma como deberían los niveles en los factores A, B, \dots afectar a la respuesta y , aunque esta influencia no se dará de manera exacta, ya que en la realización de cada valor de y contribuirán otros factores F, G, \dots e incluso concursarán también factores desconocidos.

En los procesos que observamos en tiempo real esto es intuitivamente claro. El diagrama que aparece en la figura 3.2.1 lo muestra.

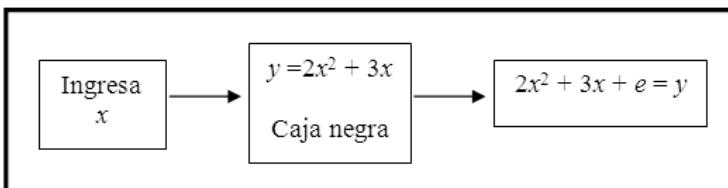


Figura 3.2.1. Esquema general de la idea del modelo estadístico.

La caja negra es el conocimiento del investigador. La calidad de la caja negra depende de la magnitud de e .

La estadística aborda la circunstancia de la siguiente manera: 1) de acuerdo a un diseño estadístico se obtienen los datos (x, y) bajo condiciones controladas y se propone en base a los conocimientos una función $f(x, y)$ que sea razonable. La postulación expresa de esta función nos lleva a una definición del tipo de análisis específico que faremos, a las hipótesis que probaremos o a la estimación de parámetros. Para realizar estimaciones o pronósticos a continuación desarrollaremos formas particulares del modelo estadístico lineal, aunque es posible tener mayores generalizaciones (Draper y Smith, 1981; Mason et al., 1990).

3.3 El modelo probabilístico normal

El modelo más utilizado en los procesos de inferencia estadística es el modelo normal o Gaussiano (en honor a Karl Gauss quien es uno de los que primeramente lo estudiaron). Para que en una situación particular se pueda usar esta distribución teórica (como también se le refiere al modelo) se requiere que la característica de interés y sea una variable de naturaleza continua, por ejemplo un peso o una longitud, el consumo semanal de energía, la concentración de una sustancia, etc. La segunda condición que se debe cumplir es que al graficar los datos del colectivo a través de un histograma, un diagrama de caja, o un diagrama de tallos y hojas, más o menos aparezcan gráficos de la forma como se observa en la figura 3.3.1.

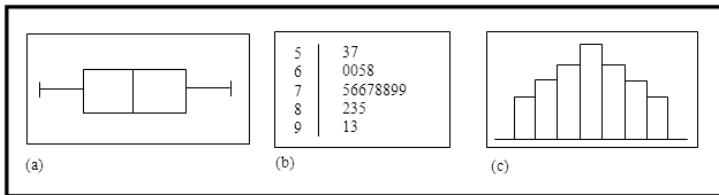


Figura 3.3.1. Tres ejemplos de distribuciones empíricas aproximadamente normales.

El modelo normal es el comportamiento teórico de la variable y , si se hacen infinitamente pequeños los intervalos del histograma y se hace infinitamente grande el tamaño de la población. El modelo normal se caracteriza por una función matemática, que tiene la siguiente forma:

$$f(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(y-\mu)^2}{\sigma^2}}$$

donde y es cualquier número real; μ , es el llamado parámetro de centralidad, que puede ser también cualquier número real; y σ^2 , es el llamado parámetro de dispersión (forma) que es un número positivo mayor que cero. En esta notación $f(y; \mu, \sigma^2)$ nos describe que la distribución de y depende de dos valores específicos μ, σ^2 por lo que se está hablando de una familia de distribuciones.

Dado que para cada μ y cada σ^2 distinto corresponde un modelo normal distinto se dice que al hacer referencia a esta distribución se habla de una familia de modelos parametrizada por μ y σ^2 , a los que desde luego se les llama parámetros. Si se conociera (que nunca sucede) a μ y σ^2 , podríamos saber exactamente a qué miembro de la familia nos referimos.

La figura 3.3.2 presenta diferentes modelos normales, en los cuales varían μ y σ^2 . Nótese que μ indica la ubicación del centro de la campana y σ^2 la esbeltez de la misma.

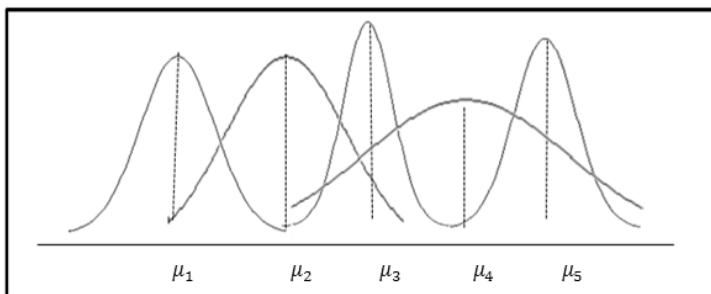


Figura 3.3.2. Distribuciones normales con medias $\mu_1 < \mu_2 < \mu_3 < \mu_4 < \mu_5$ y

$$\sigma_3^2 < \sigma_5^2 < \sigma_1^2 < \sigma_2^2 < \sigma_4^2$$

La principal utilidad del modelo radica en que a partir de él se pueden hacer inferencias. Sin embargo antes de incursionar en este tema, se hablará del modelo normal estándar ya que es el miembro de la familia que siempre

usaremos. Para este miembro se han elaborado tablas y es el que se usa y refiere en los procedimientos inferenciales. Tiene la característica de que su media $\mu = 0$ y su $\sigma^2 = 1$. A veces se abrevia, diciendo que la característica Z se distribuye normal estándar, por $Z \sim N(0,1)$. Para esta distribución se han tabulado los cuantiles teóricos y las áreas bajo la curva hasta el cuantil dado. A esta área se le llama distribución acumulativa, que se denota por $\phi(Z)$, la cual es el área bajo la curva hasta el valor Z ; esto es:

$$\phi(Z) = \int_{-\infty}^Z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz; \quad |-\infty < z < +\infty|$$

que se corresponde al área especificada en la figura 3.3.3.

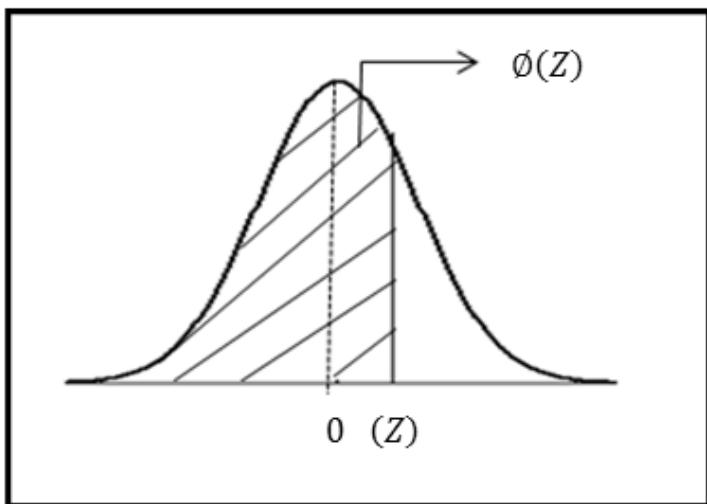


Figura 3.3.3. Valor de $P(Y \leq y) = F(x) = \phi(Z)$ para la distribución normal estándar.

Cualquier variable y que se suponga tiene una distribución normal con media μ y varianza σ^2 se puede llevar al modelo normal a través de una transformación

que se llama estandarización, que consiste en aplicar la fórmula:

$$Z = \frac{Y - \mu}{\sigma}$$

Mediante esta transformación los nuevos valores seguirán una distribución normal estándar, de esta forma podemos evaluar probabilidades de eventos en Y .

3.4 Modelación de una sola muestra

Uno de los problemas que se presentan con relativa frecuencia en las aplicaciones es el de modelar datos obtenidos a partir de una muestra de una sola población considerando la medición en una variable Y de naturaleza continua. Interesa esencialmente describir o inferir sobre la población de la que los datos provienen. Es decir, a partir de una muestra aleatoria Y_1, Y_2, \dots, Y_n , que produce datos y_1, y_2, \dots, y_n , se desea inferir sobre la población de donde las variables Y_1, Y_2, \dots, Y_n , fueron muestreadas. Para este caso no tenemos variables explicatorias del tipo X , por lo que el modelo estadístico apropiado deberá expresar la parte sistemática en términos de una constante solamente; es decir, el modelo razonable para este caso sería:

$$y_i = \mu + e_i;$$

$$i = 1, 2, \dots, n$$

Este modelo plantea que la medición y_i depende sólo de una constante general o promedio μ y de una contribución de otros factores no controlados que pueden ser incluso errores de medición, que se agrupan en la cantidad e_i . Se supone que los e_i son variables aleatorias independientes, con distribución normal, de media cero y varianza σ^2 .

De esto se sigue que:

1. $E(y_i) = E(\mu + e_i) = E(\mu) + E(e_i) = \mu + 0 = \mu$
2. $V(y_i) = V(\mu + e_i) = V(\mu) + V(e_i) = 0 + \sigma^2 = \sigma^2$

Nótese que entonces es equivalente plantear que $y_i \sim NID(\mu, \sigma^2)$; $i = 1, 2, \dots, n$. En esta situación los parámetros a estimar son μ y σ^2 , que caracterizan a la distribución normal específica de la que provienen los datos. Para estimar estos parámetros hay varios métodos, como máxima verosimilitud o momentos, y se declara una serie de propiedades deseables que estos estimadores deben cumplir, como insesgamiento, mínima varianza y consistencia, entre otros.

Los estimadores de μ y σ^2 que más se usan por sus propiedades son respectivamente:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n y_i$$
$$\hat{\sigma}^2 = \frac{1}{(n-1)} \sum_{i=1}^n (y_i - \bar{y})^2$$

De esta manera, a partir de datos obtenidos como realización de una muestra aleatoria y_1, y_2, \dots, y_n , se evalúan estos estimadores y se caracteriza el modelo estimado.

Existen dos gráficos adicionales que se proponen para explorar normalidad en los residuos o en los datos crudos, son el P-Plot y el Q-plot. El P-Plot, también llamado gráfico de probabilidad normal, simplemente gráfica los datos ordenados, digamos $e_{(1)}, e_{(2)}, \dots, e_{(n)}$, contra los cuantiles de la distribución normal estimada usando la media y la varianza muestrales, digamos que estos valores

son $z_{(1)}, z_{(2)}, \dots, z_{(n)}$. Si los datos provienen de la distribución normal entonces las parejas $e_{(i)}, z_{(i)}$; $i = 1, 2 \dots, n$, deberán ajustarse, en el correlograma, a una línea recta. Desviaciones de este patrón deberán asociarse a falta de cumplimiento en el supuesto.

Nótese que dado que los datos individuales se predicen a partir del valor de la media y entonces, en el eje de la abscisa el valor en el que se acumulan los valores es éste. Además, en este caso no se presentan fuera del conjunto de puntos, los que se corresponderían con outliers o atípicos; es decir, no hay evidencia alguna de violación del supuesto.

A continuación se presentan los gráficos correspondientes a una muestra de 100 observaciones simuladas de una normal estándar, con el propósito de tener una imagen de referencia que se corresponda con datos normales. En la figura 3.4.1 se presenta el histograma con distribución normal ajustada y en la figura 3.4.2 aparece el correspondiente P-Plot.

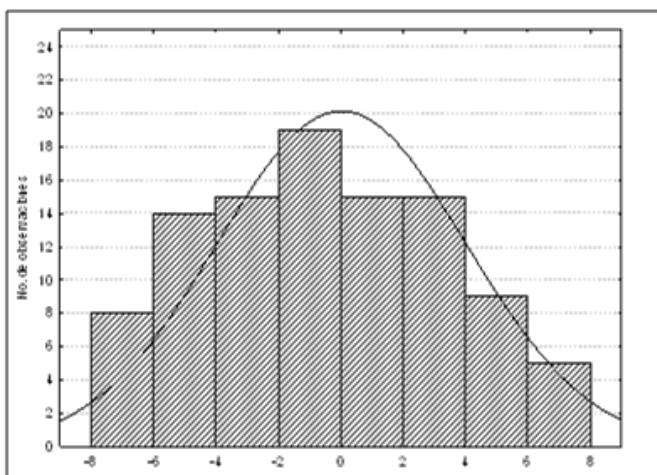


Figura 3.4.1. Histograma con distribución normal ajustada para 100 datos simulados de una distribución normal estándar.

En términos muy generales, según se aprecia en la figura 3.4.1, esta muestra es bastante ajustada al comportamiento teórico de la distribución normal. Tampoco se puede ver apariencia de presencia de puntos atípicos.

En la figura 3.4.2 podemos ver un ajuste muy notorio al patrón que se esperaría bajo normalidad exacta de los datos.

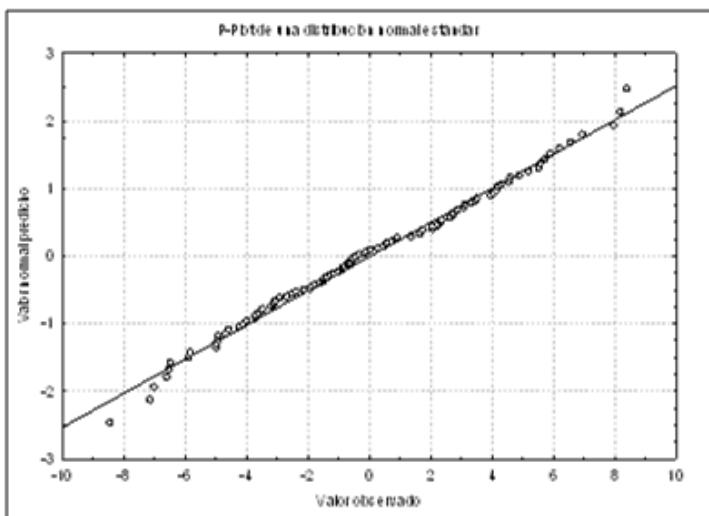


Figura 3.4.2. P-Plot para los 100 datos simulados de una distribución normal estandar.

La etapa de diagnóstico nos permite valorar el ajuste del modelo y determinar su utilidad para considerarlo en la fase de inferencia. Si la presencia de outliers es de influencia para las inferencias se debe decidir qué hacer con estos datos atípicos en la muestra. Si se cuenta con suficientes datos y los puntos se dan en ambos lados de la escala las inferencias no se verán afectadas. Más de 20 datos son suficientes según estudios de robusticidad de las pruebas (Tiku, et al., 1986) que

usaremos en este capítulo, aunque no debemos guiarnos por el único criterio del tamaño de muestra, ya que como sabemos un punto demasiado extremo puede influir notoriamente a la estimación de la tendencia central y a la dispersión. Una forma empírica de investigar la influencia de los puntos atípicos es realizar dos inferencias, una quitando los atípicos y otra sin quitarlos. Si las inferencias coinciden podríamos despreocuparnos de la presencia de estos datos.

3.5 Inferencia en la distribución normal

Uno de los objetivos de la estadística es poder valorar la incertidumbre en torno a una inferencia. En el caso de la estimación de un parámetro en el modelo estadístico interesa saber la precisión con la que se da esta estimación. La precisión se valora en términos de la varianza que esperamos el estimador tenga de muestra a muestra; si la varianza de éste es muy grande entonces podemos tener poca confianza en esa estimación puntual.

En general, si σ es el parámetro de interés y $\hat{\sigma}$ es el correspondiente estimador puntual denotaremos por $V(\hat{\sigma})$ a la varianza de este estimador. Un intervalo de confianza será un intervalo construido con base en estas dos cantidades que nos expresará un conjunto de valores que muy probablemente (con un nivel de confianza especificado) cubrirán al verdadero valor σ , el cual se supone fijo pero desconocido. Hay varios métodos para construir intervalos de confianza; nosotros enunciaremos la forma general de su construcción en cada caso. El que utilizamos a continuación se conoce como el método de pivoteo.

Antes de que se tomen los datos se puede establecer que, por el Teorema de Límite Central, si

tenemos una muestra aleatoria Y_1, Y_2, \dots, Y_n , de una población $N(\mu, \sigma^2)$, entonces:

$$\bar{Y} \sim NI\left(\mu, \frac{\sigma^2}{n}\right)$$

Asimismo, tenemos que:

$$Z = \frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1).$$

Por lo tanto, es válido establecer que:

$$Pr\left\{-Z_{(\alpha/2)} \leq \frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq Z_{(\alpha/2)}\right\} = 1 - \alpha,$$

donde $Z_{(\alpha/2)}$ es el cuantil $\alpha/2$ de la distribución normal estándar.

Si despejamos en la desigualdad obtenemos que:

$$Pr\left\{\bar{Y} - Z_{(\alpha/2)}\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{Y} + Z_{(\alpha/2)}\left(\frac{\sigma}{\sqrt{n}}\right)\right\} = 1 - \alpha.$$

De esta manera al intervalo

$$\bar{Y} - Z_{(\alpha/2)}\left(\frac{\sigma}{\sqrt{n}}\right), \quad \bar{Y} + Z_{(\alpha/2)}\left(\frac{\sigma}{\sqrt{n}}\right)$$

se le llama intervalo al $(1 - \alpha) * 100\%$ de confianza para μ .

En muchas aplicaciones, sin embargo, σ^2 es también desconocida, y entonces el intervalo se tiene que construir en base a una estimación de este parámetro; se usaría s^2 , y entonces tenemos que:

$$\frac{\bar{Y} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{(n-1)}$$

donde $t_{(n-1)}$ representa la distribución t de Student con $(n - 1)$ grados de libertad. Procediendo de manera equivalente a cuando se conoce σ^2 se obtiene que:

$$Pr \left\{ \bar{Y} - t_{(\alpha/2), (n-1)} \left(\frac{s}{\sqrt{n}} \right) \leq \mu \leq \bar{Y} + t_{(\alpha/2), (n-1)} \left(\frac{s}{\sqrt{n}} \right) \right\} = 1 - \alpha$$

y entonces el intervalo al $(1 - \alpha) * 100\%$ de confianza para μ cuando σ^2 es desconocida sería:

$$\left(\bar{Y} - t_{(\alpha/2), (n-1)} \left(\frac{s}{\sqrt{n}} \right), \quad \bar{Y} + t_{(\alpha/2), (n-1)} \left(\frac{s}{\sqrt{n}} \right) \right)$$

En este intervalo $t_{(\alpha/2), (n-1)}$ es el cuantil $\alpha/2$ de la distribución t de Student con $(n - 1)$ grados de libertad; es el valor análogo a $Z_{(\alpha/2)}$ para el caso en que se conoce σ^2 .

Estos intervalos tienen una interpretación frecuencial una vez que se obtienen los datos y_1, y_2, \dots, y_n y se calcula \bar{Y} ; es decir, se dice que, de 100 muestras que se tomen de la población muestreada, en 95 de ellas esperaríamos cubrir al verdadero valor del parámetro, que es considerada una cantidad desconocida pero fija.

3.6 Prueba de hipótesis

Como establecimos anteriormente, en muchas ocasiones el investigador puede plantear de antemano un supuesto acerca del valor de algún o algunos parámetros en el modelo estadístico. Por ejemplo, para el caso del modelo

que estamos estudiando podría establecer que la media toma un valor específico, digamos μ_0 . Esto se podría plantear así en términos de hipótesis estadísticas.

$$H_0: \mu = \mu_0$$

vs

$$H_1: \mu \neq \mu_0$$

A la hipótesis H_0 se le llama hipótesis nula y a la H_1 hipótesis alterna. En este caso, la alterna es bilateral porque establece la diferencia en una dirección u otra. Se podrían también plantear hipótesis unilaterales, hacia la derecha $H_1: \mu > \mu_0$ o hacia la izquierda $H_1: \mu < \mu_0$. El procedimiento de prueba es equivalente pero se especifica el nivel de significancia acumulado de un sólo lado.

Cuando se conoce σ^2 el procedimiento de prueba es el siguiente:

1. Se calcula $Z_c = \frac{\bar{Y} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$, al que se le llama el estadístico de prueba.
2. Se calcula el nivel de probabilidad empírico (p – value)
3. Si este valor es más pequeño que el α especificado ($\alpha = 0.01, \alpha = 0.05, \alpha = 0.1$), entonces se declara que hay suficiente evidencia para rechazar H_0 .

$$\begin{aligned} p &= p_1 + p_2 \\ &= Pr(Z \geq Z_c) + Pr(Z \leq Z_c) \end{aligned}$$

En el caso de que la prueba fuese unidireccional entonces se tomaría sólo $p = p_1$ ó $p = p_2$, según sea la dirección.

Si la varianza σ^2 es desconocida entonces el estadístico de prueba es:

$$t_c = \frac{\bar{Y} - \mu_0}{\frac{s}{\sqrt{n}}}$$

el cual bajo H_0 cierta, sigue una distribución t de Student.

3.7 Muestras aparejadas

En algunas ocasiones el estudio considera la realización de dos mediciones repetidas, digamos Y_1 y Y_2 sobre la misma unidad o individuo. A los datos así obtenidos se les llama datos aparejados o de dos muestras aparejadas. En este caso los datos para una muestra aleatoria de tamaño n serían: $(y_{11}, y_{21}), (y_{12}, y_{22}), \dots, (y_{1n}, y_{2n})$.

En este tipo de situaciones realmente lo que interesa es estudiar la diferencia $D = Y_2 - Y_1$, ya que las observaciones al pertenecer al mismo individuo están correlacionadas; es decir, los datos que interesarían son: $d_1 = y_{21} - y_{11}$, $d_2 = y_{22} - y_{12}$, ..., $d_n = y_{2n} - y_{1n}$. Si suponemos que estas diferencias se distribuyen $N(\delta, \sigma_d^2)$, entonces podemos usar el modelo estadístico:

$$d_i = \delta + e_i; \quad i = 1, 2, \dots, n$$

3.8 Muestras independientes de dos poblaciones normales

En la práctica, con mucha frecuencia, interesa comparar dos entidades, como por ejemplo, dos turnos cuando se estudia la productividad de trabajadores, dos métodos cuando se desea tomar una decisión sobre cuál resulta ser mejor o comparar dos tratamientos para realizar una

tarea o para atacar un padecimiento uno tradicional y otro que es una nueva propuesta.

En este tipo de casos es necesario realizar monitoreos comparativos y experimentos que tienen como propósito producir datos para establecer una comparación formal entre las entidades o grupos bajo estudio. Si los grupos son independientes en el sentido de que están formados por individuos distintos, o bien es válido y razonable suponer que las mediciones son independientes, entonces el modelo estadístico propuesto será aquel para dos muestras independientes. Este modelo se presenta de la siguiente manera: en cada uno de dos grupos independientes se observa una variable respuesta Y . Sean los datos y_{ij} ; $i = 1, 2$; $j = 1, 2, \dots, n_i$, donde n_i es el tamaño del i -ésimo grupo. Los grupos son balanceados si $n_1 = n_2 = n$. El modelo estadístico adecuado para este caso es:

$$y_{ij} = \mu_i + e_{ij}; \quad i = 1, 2; \quad j = 1, 2, \dots, n_i$$

En este modelo, otra vez se supone que $e_{ij} \sim N(0, \sigma^2)$; nótese que entonces:

1. $E(y_{ij}) = E(\mu_i + e_{ij}) = E(\mu_i) + E(e_{ij}) = \mu_i + 0 = \mu_i; \quad i = 1, 2.$
2. $V(y_{ij}) = V(\mu_i + e_{ij}) = V(\mu_i) + V(e_{ij}) = 0 + \sigma^2 = \sigma^2.$

Así tenemos que los parámetros de interés son μ_i ; $i = 1, 2$ y σ^2 . Aquí estaríamos interesados en probar la hipótesis de que los dos procedimientos son equivalentes en promedio; es decir:

$$\begin{aligned} H_0: \mu_1 &= \mu_2 \\ &\text{vs} \\ H_1: \mu_1 &\neq \mu_2 \end{aligned}$$

Para poder probar esta hipótesis se utiliza el método de análisis de varianza, que constituye un procedimiento aritmético que produce la llamada tabla de ANOVA.

En la tabla 3.8.1 se presenta el formato general del ANOVA, donde ordenamos una serie de cálculos para realizar el análisis, que consiste en implementar la prueba de la hipótesis enunciada.

Tabla 3.8.1. Tabla general de análisis de varianza para el caso de dos muestras independientes.

F. V.	G. L.	S. C.	C. M.	F _c
Grupos	1	$G = \sum_{i=1}^2 n_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$	$\frac{G}{1} = CMG$	$F_c = \frac{CMG}{CME}$
Error	(n - 2)	$E = \sum_{i=1}^2 \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot})^2$	$\frac{E}{n-2} = CME$	
Total	(n - 1)	$T = \sum_{i=1}^2 \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{\cdot\cdot})^2$		

En la tabla 3.8.1 tenemos que $\bar{y}_{i\cdot}$ y $\bar{y}_{\cdot\cdot}$ representan las medias de grupo y la media general, respectivamente;

$$\bar{y}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}; \quad i = 1, 2$$

$$\bar{y}_{\cdot\cdot} = \frac{1}{\sum_{i=1}^2 n_i} \sum_{i=1}^2 \sum_{j=1}^{n_i} y_{ij}$$

En esta tabla F. V., se lee como Fuentes de Variación, G. L., Grados de Libertad, S.C., Suma de

Cuadrados, C. M. Cuadrados Medios y F_c como relación F o F calculada.

Se puede verificar que las cantidades T, G y E cumplen $T = G + E$, a la cual se le llama identidad básica de la suma de cuadrados. A T se le llama suma de cuadrados totales, a G se le llama suma de cuadrados debida a grupos, y a E se le llama suma de cuadrados debida al error.

Un estimador adecuado de $\hat{\sigma}^2$ en este caso del modelo estadístico es CME ; es decir, $\hat{\sigma}^2 = CME$.

Asimismo, se puede demostrar que, bajo las condiciones generales, los estimadores óptimos de μ_1 y μ_2 resultan ser: $\hat{\mu}_1 = \bar{y}_1$ y $\hat{\mu}_2 = \bar{y}_2$.

Una forma general de probar la hipótesis de igualdad de medias es haciendo referencia al hecho de que F_c sigue una distribución F de Fisher con 1 grado de libertad en el numerador y $(n - 2)$ grados de libertad en el denominador. De esta manera se calcula el p -value asociado a $P(F \geq F_c)$; si este valor resulta ser muy pequeño (menor que 0.1, 0.05, 0.01 ó 0.001) se concluye que existe suficiente evidencia en contra de H_0 .

3.9 Análisis de la varianza

En diversas situaciones interesa comparar varias poblaciones, con el propósito de probar la hipótesis nula de que no son diferentes en su promedio. Esto puede ser de interés en muchos casos que son generalizaciones de los tratados en la sección anterior. Por ejemplo, en lugar de comparar dos métodos o tratamientos, estaríamos interesados en comparar g tratamientos donde $g > 2$.

Veamos que entonces estamos en una simple generalización del caso de dos muestras. Ahora tenemos datos y_{ij} ; $i = 1, 2, \dots, g$; $j = 1, 2, \dots, n_i$, donde g es el número

de grupos o muestras, las cuales deben ser independientes.

El modelo estadístico para esta situación es:

$$y_{ij} = \mu_i + e_{ij}; \quad i = 1, 2, \dots, g; \quad j = 1, 2, \dots, n_i,$$

donde nuevamente $e_{ij} \sim NI(0, \sigma^2)$. La hipótesis de interés sería:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_g$$

vs

$$H_1: \mu_i \neq \mu_j \text{ para al menos un } i \neq j$$

La tabla de análisis de la varianza para este caso tendría la forma general de la tabla 3.9.1.

Tabla 3.9.1. Tabla general de análisis de varianza para el caso de g grupos.

F. V.	G. L.	S. C.	C. M.	F _c
Grupos	$(g - 1)$	$G = \sum_{i=1}^2 n_i (\bar{y}_{i.} - \bar{y}_{..})^2$	$\frac{G}{g - 1} = CMG$	$F_c = \frac{CMG}{CME}$
Error	$(n - g)$	$E = \sum_{i=1}^2 \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2$	$\frac{E}{n - g} = CME$	
Total	$(n - 1)$	$T = \sum_{i=1}^2 \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2$		

Al ajustar el modelo a los datos obtenemos que los estimadores son:

$$\hat{\mu}_i = \bar{y}_i; \quad i = 1, 2, \dots, g \quad \text{y} \quad \hat{\sigma}^2 = CME$$

De esta manera tenemos que los valores predichos serían $\hat{y}_{ij} = \hat{\mu}_i = \bar{y}_i$, por lo que los residuos serían:

$$\hat{e}_{ij} = y_{ij} - \hat{\mu}_i = y_{ij} - \bar{y}_i$$

El procedimiento de análisis para esta generalización se sigue de manera inmediata del caso de dos grupos. Es decir, se observa el *p*-value en la tabla de análisis de varianza, y se realiza el diagnóstico usando residuos. Si el modelo da un ajuste razonable; se dice que hay evidencia para rechazar H_0 y no se tiene razón para pensar que alguno o algunos supuestos se estén violando; entonces se procede a obtener conclusiones.

IV. Modelación de regresión lineal

El término regresión, fue usado por primera vez por el científico inglés Francis Galton, quien observó que aunque existe una tendencia de los padres altos a tener hijos altos y de los padres bajitos a tener hijos de baja estatura, la distribución de las estaturas de una población no cambia substancialmente de generación a generación. Su explicación mostraba que había una tendencia de la estatura promedio de los hijos con padres de determinada estatura a moverse o regresar a la estatura promedio de toda población. A esta regularidad la llamó "Ley de la Regresión Filial". Esta ley fue confirmada por el matemático, también inglés Karl Pearson, quien recolectó más de mil datos de estaturas de miembros de grupos familiares. Pearson encontró que la estatura promedio de los hijos de padres altos era menor que la estatura promedio de sus padres, mientras que la estatura promedio de los hijos de padres bajitos era mayor que el promedio de la de sus padres.

El análisis de regresión es una técnica estadística que se utiliza para investigar y modelar la relación entre variables. En una definición moderna del análisis de regresión, en términos generales, trata de la dependencia de una variable (dependiente o respuesta), con una o más variables (independientes o explicatorias), con el objetivo de estimar o predecir la media o valor promedio de la variable dependiente con base en los valores conocidos o fijados (en muestras repetidas) de las variables independientes. Las aplicaciones de la regresión son numerosas y ocurren en casi todos los campos, incluyendo la ingeniería, las ciencias físicas, las ciencias biológicas, las ciencias sociales y la economía, entre otras. De hecho, el análisis de regresión puede ser considerado como la técnica estadística más ampliamente utilizada.

4.1 Análisis de regresión lineal simple

El análisis de correlación sirve únicamente para saber si existe relación lineal entre dos variables: cuando dos variables tienen un coeficiente de correlación muy alto, los investigadores sabrán que tienen datos que establecen una fuerte relación lineal entre las variables. El caso más sencillo en regresión se presenta cuando se tiene la variable respuesta Y con una sola variable explicatoria X y se espera relacionarlas por medio de un modelo lineal, al que se le conoce como modelo de regresión lineal simple. Por lo tanto, como sólo se tienen dos variables, es fácil graficarlas y analizarlas.

El análisis de regresión permite predecir el cambio de la variable dependiente si la variable independiente se cambia en una unidad, esto se realiza por medio de una ecuación lineal.

El análisis de regresión es una técnica con la que se adecua una ecuación matemática a un conjunto de datos (Boyd, 1987). Los datos deben de consistir en mediciones de dos o más variables continuas y además el tamaño de la muestra debe de ser cuando menos el doble o el triple del número de variables medidas.

La ecuación matemática es:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

donde y es una variable dependiente y x una variable independiente o explicatoria, y β_1 es un coeficiente que indica el efecto que tiene en y un cambio en una unidad en la variable x , β_0 es la ordenada al origen o el valor promedio cuando $x = 0$. Debido a que esta ecuación tiene una variable independiente, se le conoce como ecuación de regresión simple.

El hacer una regresión lineal a partir de una población es casi imposible, se recurre a la obtención de muestras de la población y con éstas se realiza el análisis de regresión.

Por medio de las técnicas de regresión y en base a una muestra es posible estimar una ecuación que se asemeje lo más posible a la ecuación poblacional, llamada Función de Regresión Poblacional (FRP). A esta función de estimación se le llama Función de Regresión Muestral (FRM). La FRM hará lo posible para estimar los valores medios poblacionales de y dado un valor de x , con cierto error. Para cada muestra existe una FRM, pero no se podrá saber cuál es la que se asemeje más a la FRP ya que ésta es desconocida. Pero, lo que si se asegura es que las técnicas de regresión en base a los datos de la muestra harán posible la obtención de la FRM que más se ajuste a ellos.

La FRM se puede expresar de la siguiente manera:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

donde:

\hat{y}_i = es el estimador del valor $E(y/x_i)$

$\hat{\beta}_0$ = es el estimador de β_0

$\hat{\beta}_1$ = es el estimador de β_1

El procedimiento utilizado comúnmente para calcular la línea de regresión que más se ajuste a los datos, es conocido como método de Mínimos Cuadrados; el cual reduce al mínimo la suma de cuadrados de los residuos, es decir, este método permite obtener los estimadores que reducen al mínimo la sumatoria

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2$$

donde $y_i - \hat{y}_i$ es el residuo.

La estimación del coeficiente β_1 se obtiene mediante la siguiente fórmula:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

La estimación del coeficiente β_0 se calcula de la siguiente manera:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

donde \bar{x} y \bar{y} son medias muestrales de x y y respectivamente.

Siempre que se lleve a cabo un análisis de regresión, los investigadores tendrán la necesidad de saber qué tan bueno es, es decir, si la ecuación de regresión explica adecuadamente la variación observada de la variable dependiente (y). Para esto, se calcula el coeficiente de determinación (R^2). Este coeficiente puede tomar valores entre 0 y 1; si $R^2 = 1$ indica que la ecuación de la regresión explica al 100% la variación que se da en la variable dependiente con relación a su media, y por lo tanto todos los residuos deben ser iguales a cero. Cuando $R^2 = 0$ la ecuación de la regresión no se ajusta a los datos, y algunos residuos serán mayores de cero.

$$R^2 = \frac{\left[\begin{array}{l} \text{Varianza total en la} \\ \text{variable dependiente} \end{array} \right] - \left[\begin{array}{l} \text{Varianza "no explicada" por} \\ \text{la ecuación de regresión} \end{array} \right]}{\text{Varianza total en la variable dependiente}}$$

Si la recta de regresión explica totalmente la variación de y , todos los residuos y la varianza no explicada por la ecuación serán iguales a cero, y el coeficiente de determinación será:

$$R^2 = \frac{\left[\begin{array}{l} \text{Varianza total en la} \\ \text{variable dependiente} \end{array} \right] - 0}{\text{Varianza total en la variable dependiente}} = 1$$

Cuando el valor de R^2 se encuentra entre 0.50 y 1, se considera que la ecuación de regresión explica satisfactoriamente la variación en y .

La fórmula para obtener el coeficiente de determinación es:

$$R^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Para saber si la variable x explica significativamente el comportamiento de la variable y , se utilizan hipótesis empleando análisis de varianza.

La hipótesis que se plantea es la siguiente:

$$H_0: \beta_1 = 0$$

vs

$$H_1: \beta_1 \neq 0$$

Tabla 4.1.1. Cuadro de análisis de varianza para la regresión lineal simple.

F.V.	G.L.	S.C.	C.M.
Debido a la regresión	SCR	1	$\frac{SCR}{1}$
Debido a los residuos	SCE	$n-2$	$\frac{SCE}{(n-2)}$
Total	SCT	$n-1$	

donde:

$$SCR = \beta_1^2 \sum_{i=1}^n (x_i - \bar{x})^2$$

$$SCE = \sum_{i=1}^n \hat{e}_i^2$$

$$SCT = \sum_{i=1}^n y_i^2$$

El estadístico de prueba es el siguiente:

$$F_c = \frac{\frac{\beta_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n \hat{e}_i^2}}{(n-2)}$$

La variable F_c sigue una distribución F de Fisher con $(1, n-2)$ grados de libertad.

4.2 Análisis de regresión lineal múltiple

En la sección anterior, se revisó el análisis de regresión lineal simple en donde se utiliza sólo una

variable independiente. Sin embargo, es muy común que los datos recopilados abarquen varias variables independientes. En este caso, a este tipo de estudio se le denomina análisis de regresión lineal múltiple, éste permite realizar un estudio de regresión más realista debido a que una variable dependiente tiende a recibir la influencia de los valores de diferentes variables independientes.

La regresión lineal múltiple se utiliza para describir la naturaleza de una relación lineal entre una variable dependiente y varias independientes, y se utiliza también para predecir el valor de la variable dependiente a partir de los valores conocidos de las variables independientes.

El tamaño de la muestra que se utilice en este tipo de análisis debe de ser cuando menos el doble o el triple de las variables medidas.

En el análisis de regresión lineal múltiple tratamos de encontrar la recta que mejor se ajuste a un conjunto de datos. En este caso se trata de tres dimensiones y la expresión que mejor se ajuste será un plano bidimensional.

La ecuación de regresión múltiple que se utiliza es:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

donde y es la variable dependiente y x_1, x_2, \dots, x_p , son las variables independientes, β_0 es un coeficiente estructural necesario para complementar la ecuación, generalmente tiene poca importancia; los coeficientes $\beta_1, \beta_2, \dots, \beta_p$, indican como y tiende a cambiar con cada cambio de x , a condición de que las otras x permanezcan constantes. En ocasiones a estos coeficientes se les llama coeficientes de regresión parcial y se calculan de la misma manera que en análisis de regresión simple.

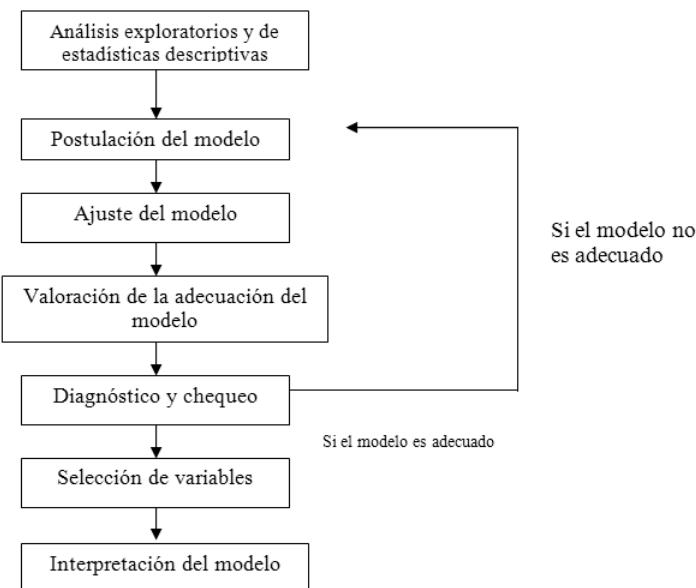


Figura 4.2.1. Fases de modelación en la regresión lineal múltiple.

En el diagrama que se presenta en la figura 4.2.1 se ilustra el proceso iterativo de la modelación usando el modelo de regresión lineal múltiple. Nótese que el proceso de verificación y diagnóstico nos puede llevar a replantear el modelo. Estas fases, en las que se define un ciclo, terminan cuando el modelo que tenemos es el más “aceptable”, en términos de la mayor R^2 y el menor, $\hat{\sigma}^2 = CME$ y sin problemas de multicolinealidad o influencia de observaciones.

Una vez que se ha obtenido un modelo aceptable en los términos establecidos viene una fase final, que consiste en elegir el modelo más simple.

V. Elaboración de un reporte y una presentación

La última tarea, y una de las más importantes en la aplicación de la estadística, consiste en diseñar y elaborar un reporte y una presentación de los resultados. En este documento, que a veces se hace en forma de un artículo, se debe presentar la descripción del problema, la metodología de obtención y análisis de datos, así como los resultados y las conclusiones. La tarea de diseñar y elaborar un reporte y una presentación requiere una clara comprensión tanto del problema como del proceso que se siguió para su solución. Cuando esto se hace en forma de artículo se deben tomar en cuenta las recomendaciones que en la guía para autores da la revista a la que se desea someter el trabajo para su publicación.

En general, se debe escribir un marco teórico y una revisión del estado del arte respecto al problema abordado, revisando trabajos similares y conexos. En este marco se presenta la descripción del problema y las preguntas de investigación, así como los propósitos del trabajo, la estrategia seguida y sobre todo los resultados.

En un apartado fundamental, que se denomina materiales y métodos, del reporte se debe dar cuenta de lo que se hizo y cómo se hizo; qué materiales y métodos se utilizaron, para qué y cómo. A veces se le llama también sección de Metodología. En este sentido, el manejo de los conceptos clave de la estadística es fundamental, sobre todo en lo que se refiere a la descripción de la metodología de obtención y análisis de datos.

El análisis de los resultados, una vez realizado, debe ser plenamente justificado y los resultados deben presentarse de manera apropiada. Usualmente se recomienda diseñar y elaborar sólo unos cuantos

cuadros, tablas y gráficos, no hace falta presentar todo lo que se hizo; hay que seleccionar lo más relevante, lo directamente relacionado con las preguntas que dieron origen a la investigación o estudio. Por otro lado, la discusión de los resultados debe incluir los aspectos estadísticos pero también aspectos del marco de referencia del problema, debe tratar de dar una explicación de los resultados y no solamente hacer una recapitulación de los mismos.

Para escribir reportes de análisis estadísticos no hay en la literatura más que recomendaciones generales. En este sentido una práctica consciente propiciará, sin duda, la adquisición de las habilidades que se requieren para hacerlo de manera adecuada. Hay en la literatura manuales y guías para escribir artículos científicos, como el de Day y Gastel (2008), que se recomienda a los estudiantes del curso.

Con seguridad es posible decir que una investigación o estudio no se ha terminado hasta que se reporte y difunda. El reporte de la investigación es la prueba de fuego para el que está encargado de la investigación o estudio, ya que es allí donde concurren muchos problemas que son producidos o propiciados por la mala conducción de algunos pasos en el desarrollo de la misma. La buena organización y administración de los pasos y fases del estudio producirán materiales fundamentales para elaborar un buen reporte; sin embargo, también se requiere un esfuerzo de síntesis y un manejo de la comunicación escrita en el estilo técnico. La síntesis de los resultados y un buen esfuerzo de diseño, revisión repetida y un buen manejo del estilo producen, en general, buenos reportes.

Después de un adecuado reporte de investigación o estudio técnico, es inmediata y relativamente sencilla la preparación de una ponencia para un público específico, ya que la mayoría de las veces los resultados

de los estudios se deben presentar ante equipos de trabajo de la misma organización.

A continuación se enlistan una serie de recomendaciones para preparar una presentación final en Power Point.

- La presentación se diseña. Recomiendo hacerla primero con papel y lápiz, redactando cuidadosamente cada lámina. Eso dará la posibilidad de concebirla antes de realizarla.
- Seleccionar apropiadamente los colores de fondo y letra, garantizando alto contraste.
- Los tamaños de letra deben ser al menos 28 y debe evitarse más de 10 líneas por lámina (de preferencia 8 ó menos).
- Utilizar esquemas, diagramas y figuras de manera equilibrada y armonizada con el texto.
- No utilizar textos largos, preferir el formato de viñetas.
- El número de láminas no deberá exceder 25 (se recomienda alrededor de 20).
- La primera lámina deberá llevar el título, el nombre de los autores y un resumen de no más de 50 palabras.

Las secciones a desarrollar deberán ser:

- Marco conceptual y antecedentes;
- Problemática;

- Justificación;
- Objetivos;
- Diseño estadístico;
- Análisis estadístico;
- Resultados;
- Discusión;
- Referencias.

Hay que considerar que el diseño y el análisis estadístico reportan la estrategia metodológica que se siguió; el análisis estadístico reporta las técnicas y métodos que se aplicaron y para qué se aplicaron; aquí se declara el modelo que se ajustó, en el caso de que se usara. En la sección de resultados se presenta sólo lo relevante y lo significativo (evitar presentar salidas de la computadora); hay que diseñar tablas y figuras que permitan comunicar los resultados del estudio. La Discusión debe retomar el Marco conceptual y antecedentes; evitar redundar en los resultados. Las referencias se deben preparar siguiendo el “modelo Harvard”.

Una vez editada la presentación revisar tipografía, ortografía, contrastes, tamaño de letra y elementos de la guía.

Preparar un guión del discurso a desarrollar y repasar y ensayar la presentación.

Anexo 1: Plan de actividades del curso

SEMANAS	ACTIVIDADES DEL CURSO	ACTIVIDADES EXTRACLASE Y PRODUCTOS
1	Conferencia: Introducción al curso y presentación sobre la caracterización del proceso de investigación y el diseño estadístico	Contestar preguntas de repaso.
2	Repaso sobre la caracterización del proceso de investigación y el diseño estadístico. Conferencia de las unidades (análisis estadístico y elaboración del reporte)	Contestar preguntas de repaso. Tarea 1.
3	Conferencia sobre diseños experimentales, diseños de muestreo y protocolo.	Contestar preguntas de repaso. Tarea 2.
4	Repaso de los estudios de muestreo y experimentales. Conferencia sobre los estudios observacionales.	Contestar preguntas de repaso y entrega del anteproyecto.
5	Conferencia sobre los tipos de datos y variables, gráficas y distribuciones.	Contestar preguntas de repaso.
6	Repaso sobre: Tipos de datos y variables, gráficas y distribuciones. Conferencia sobre los análisis comparativos, análisis bivariados y multivariados. Sesión de presentación de protocolos.	Contestar preguntas de repaso.
7	Conferencia sobre los modelos estadísticos y el modelo de medias.	Contestar preguntas de repaso. Entrega de práctica 1.
8	Exposición de protocolos.	Exposición y entrega de protocolos.
9	Repaso sobre los modelos	Contestar preguntas

	estadísticos y el modelo de medias. Conferencia sobre estimación y prueba de hipótesis, análisis de varianza y ordenamiento de medias.	de repaso.
10	Conferencia sobre regresión lineal simple y una introducción a regresión lineal múltiple.	Contestar preguntas de repaso. Entrega de práctica 2.
11	Conferencia sobre regresión lineal múltiple y el proceso de modelación. Sesión de presentación de avances.	Contestar preguntas de repaso y seminario.
12	Conferencia sobre los datos categóricos y el análisis de varianza no paramétrico.	Contestar preguntas de repaso y seminario.
13	Conferencia sobre el análisis multivariado descriptivo y el análisis de componentes principales. Presentación de avances.	Contestar preguntas de repaso y seminario.
14	Conferencia sobre el análisis multivariado descriptivo y el análisis de componentes principales. Presentación de avances.	Contestar preguntas de repaso y seminario.
15	Congreso final	Presentación y entrega de artículos.



Bibliografía del curso

y

Referencias

Bibliografía del curso

- Arroyo, V. y Vázquez, J. B. (2005). *Herramientas Estadísticas con STATISTICA para Estudios en Ciencias Alimentarias*. LINAЕ, Facultad de Estadística e Informática, Universidad Veracruzana. Xalapa, Veracruz, México.
- Batanero, C. (2001). *Didáctica de la Estadística*. Universidad de Granada, Dpto. de la Didáctica de la Matemática. Granada, España.
- Bishop, G. and Talbot, M. (2001). Statistical thinking for novice researchers in the biological sciences. International Association for Statistical Education and International Statistical Institute. In *Training researches in the Use of Statistics* Batanero C. (Ed.), pp. 215-226.
- Chatfield, C. (1995). *Problem Solving: A statistician's guide*. Second Edition. Chapman & may. Londres.
- Day, R. A. y Gastell, B. (2008). *Cómo Escribir y Publicar Trabajos Científico*. Publicación Científica y Técnica No. 621, Organización Panamericana de la Salud. Washington, D.C. EEUU. 333 pp.
- Devore, J. L. (2001). *Probabilidad y Estadística para Ingeniería y Ciencias*. 5^a Edición. Internacional Thomson Editores. México.
- Hair, J., Anderson, E., Tatham, L. and Black, C. (1999). *Ánálisis Multivariante*, Prentice-Hall Inc. New York.
- Montgomery, D. C. and Runger, G. C. (1996). *Applied Statistics and Probability for Engineers*. Wiley, New York.

- Moore, D. S. (2000). *The Basic Practice of Statistics*. Second Edition, W. H. Freeman. New York.
- Ojeda, M. M., Díaz, J. E., Apodaca, C. y Trujillo I. (2004). *Metodología de Diseño Estadístico*. Textos Universitarios, Universidad Veracruzana. Xalapa, Veracruz, México.
- _____. (2001). "Aspectos Generales de la Metodología Estadística". LINAЕ, Facultad de Estadística e Informática, Universidad Veracruzana. Xalapa, Veracruz, México.
- _____. (1994). "La importancia de una buena cultura estadística en la investigación" *La Ciencia y el Hombre*, No. 17, 143-156.
- _____. y De León G. (1996). *Metodología Estadística Básica: Principios y Herramientas para Plantear y Resolver Problemas en un Contexto Organizacional*. Facultad de Matemáticas, Universidad Autónoma de Guerrero. Chilpancingo, Guerrero, México.
- Ott, L. (1993). *An Introduction to Statistical Methods and Data Analysis*. Fourth Edition. Duxbury Press. Belmont, California.
- Silva, E. R. (1994). "Estadística Multivariada Aplicada a la Ciencia de los Alimentos". Trabajo Receptacional, Especialización en Métodos Estadísticos, Facultad de Estadística e Informática, Universidad Veracruzana. Xalapa, Veracruz, México.
- Schaeffer, R. L., Gnanadesikan, M., Watkins, A. & Witmer, J. (1996). *Activity-Based Statistics*. Springer Verlag. New York.

Starkings, S. (1997). Assessing students projects. In: *The Assessment Change in Statistics Education*. (Eds.) Gal and Garfield. IOS Press. Amsterdam.

Sylwester, D. L. and Mee, R. W. (1992). Students' projects: "An important element in the beginning statistics course". Proceeding of the ASA Section on Statistical Education, pp. 137-141. American Statistician Association. Alexandria, Virginia, USA.

Vázquez, J. B. (2005). "Enfoque Basado en Proyectos para el Curso de Métodos Estadísticos de la Maestría en Ciencias Alimentarias, de la Universidad Veracruzana". Tesis de Licenciatura en Estadística, Facultad de Estadística e Informática, Universidad Veracruzana, Xalapa, Veracruz, México.

Velasco, M. L. (1999). "Introducción a la Metodología Estadística en la Investigación". Tesis de Licenciatura en Estadística, Facultad de Estadística e Informática, Universidad Veracruzana, Xalapa, Veracruz, México.

Watt, T. A. (1998). *Introductory Statistics for Biology Students*, Second Edition. Chapman & Hall/CRC. Boca Raton, Florida.

Weldon, L. K. (1994). "On teaching the big ideas of statistics: A Project-based approach". Research Report No. 94-05, Simon Fraser University. Victoria, British Columbia, Canadá.

Referencias

- Ausubel, D. (1978). *Adquisición y retención del conocimiento. Una perspectiva cognitiva.* Barcelona, Paidós.
- Behar, R. y Ojeda, M. M. (1995). "La problemática de la enseñanza y el aprendizaje de la estadística en la educación superior". *La Ciencia y el Hombre.* Pp. 19-29.
- Beltrán, J. (1993). *Procesos, Estrategias y Técnicas de Aprendizaje.* Madrid. 1993. 380 p.
- Chalfield, C. (1988). *Problem Solving; A Statistician Guide.* Chapman and may. London, UK.
- _____. (1995). *Problem Solving; A Statistician Guide.* 2da. Ed. Chapman and may. London, UK.
- Day, R. A. y Gastell, B. (2008). *Cómo Escribir y Publicar Trabajos Científico.* Publicación Científica y Técnica No. 621, Organización Panamericana de la Salud. Washington, D.C. EEUU. 333 pp.
- Dewey, J (1989). *Cómo Pensamos: Nueva Exposición de la Relación entre Pensamiento y Proceso Educativo.* Pról. a la ed. española: Antonio Caparros; tr., Marco Aurelio Galmarini. Barcela Paidós, 1989.
- Draper, N. R., y Smith, H (1981). *Applied Regression Analysis.* 2a. ed. New York: Wiley.
- Estepa, M. (2000). "Los procesos y las preocupaciones del profesorado: Análisis de un caso". *Revista de la Facultad de Ciencias de la Educación.* vol 2. 2000 pags 41-64 (Ejemplar dedicado a: El profesor ante los nuevos retos de la Enseñanza Universitaria)

- Gagné, R. y Briggs, L. (1997). *La planificación de la enseñanza*. México, Trillas.
- Kempthorne, O. (1980). "The Teaching of Statistics; Content Versus Form". *The American Statistician*. February, Vol. 34 No. 1, pp. 17-21.
- MacVean, M.K., and P.J. Mason, (1990). "Cloud-Top Entrainment Instability through Small-Scale Mixing and Its Parameterization in Numerical Models". *J. Atmos. Sci.*, 47, 1012-1030.
- Mayor, J., Suengas, A. y González M., (1993). *Estrategias metacognitivas. Aprender a aprender y aprender a pensar*. Editorial Síntesis, S.A. Madrid, España.
- Mead, R (1983). *Statistical Methods in Agriculture and Experimental Biology*. London, New York; Chapman and hall.
- Montgomery, D. C. and Runger, G. C. (1996). *Applied Statistics and Probability for Engineers*. Wiley, New York.
- Nérici, I. (1980). *Metodología de la enseñanza*. Kapelusz Mexicana. México.
- _____. (1985). *Metodología de la Enseñanza*, 4a. edición. Kapelusz Mexicana. México.
- Ojeda, M. M. (1994). "La importancia de una buena cultura estadística en la investigación". *La ciencia y el Hombre*. Pp. 17, 143-152.
- _____. (1998). *Aspectos Básicos del diseño Estadístico de Experimentos*. Editorial de la Universidad Veracruzana. Xalapa, Veracruz, México.

- _____. (2001). *Los retos de la educación en el siglo XXI. Expresiones de Veracruz*, Gobierno del Estado de Veracruz.
- _____. Díaz, J. E., Apodaca, V. C. y Trujillo L. I. (2004). *Metodología de Diseños Estadístico. Textos Universitarios*, Universidad Veracruzana. Xalapa, Veracruz, México.
- Rao, P. S. R. S. (2000). *Sampling Methodologies whit Applications*. Chapman and Hall/CRC, London.
- Ribas, M. (2000). *Innovación Educativa: Teoría, Procesos y Estrategias*. Editorial Síntesis. Madrid.
- Silva, E. (1994). "Estadística Multivariada Aplicada a la Ciencia de los Alimentos". Trabajo Receptacional, Especialización en Métodos Estadísticos, Facultad de Estadística e Informática, Universidad Veracruzana.
- Slotnick, H. B. (1989). "The timing and teaching of statistics in medical school." *Bulletin of teaching of statistics in Health Science (News Letter of the Subsection of statistical Education Section)*. December 1989. American Statistical Association, Alexandria, Virginia. pp. 1-3.
- Snee, R. D. (1993). "What's missing in statistical education?" *The American Statistician*. Pp. 47, 149-154.
- Thomas, E. y David, K. H. (2002). "Statistics education and the making statistics more effective in schools of business conferences". *The American Statistician*. Vol. 56, No. 2. pp. 107-111.
- Tikul, M.L., Tan, W.Y., and Balakrishnan, N. (1986). *Robust inference*. Marcel Dekker. New York.

Vázquez, J. B. (2005). "Enfoque Basado en Proyectos para el Curso de Métodos Estadísticos de la Maestría en Ciencias Alimentarias, de la Universidad Veracruzana". Tesis de Licenciatura en Estadística, Facultad de Estadística e Informática, Universidad Veracruzana. Xalapa, Veracruz, México.

Velasco, M. L. (1999). "Introducción a la Metodología Estadística en la Investigación". Tesis de Licenciatura en Estadística, Facultad de Estadística e Informática, Universidad Veracruzana. Xalapa, Veracruz, México.

Watt, T. A. (1998). *Introductory Statistics for Biology Students*. Second Edition. Chapman & Hall/CRC. Boca Raton, Florida.

Weinberg, S. L. (1993). "Conceptualizing applied statistics: A current need Proceeding of the ASA" Section on Statistical Education. Pp. 233-238.

Weldon, L. K. (1994). "On Teaching the Big Ideas of Statistics: A project-based Approach". *Research Report* No. 94-05, Simon Fraser University. Victoria, British Columbia.

Zubieta, J. C. y Susinos, T. (1992). "Las satisfacciones e insatisfacciones de los enseñantes". Madrid: CIDE.

ADCB 16 Running

Run Status

Run time: 19.00 min

Stop time: 50.00 min

 Running No Fault

Channel A

Level: 0.460 mV

Range: 10 Volts

Sampling Frequency

Run: 4 (10 Hz)

Channel B

Level: 463.760 mV

Range: 10 Volts

Sync Status

Enabled

View B Only

20 Channel B

