

CONTENIDO TEMÁTICO

1. Conceptos básicos de modelación
2. Introducción al análisis y construcción de modelos
3. Elementos de teoría de matrices
4. Teorema fundamental de la programación lineal
5. Algoritmo simplex
6. Problema dual y análisis de sensibilidad
7. Análisis de sensibilidad
8. Rango de optimalidad del análisis de sensibilidad
9. Método de mínimos cuadrados (Gauss)
10. Bibliografía básica

Part I

ELEMENTOS DE MODELACIÓN Y ALGEBRA LINEAL

1 INTRODUCCIÓN

El concepto de optimización surge como un principio subyacente al análisis de muchos problemas complejos de decisión o distribución (asignación). Se caracteriza por tener cierta elegancia filosófica que difícilmente se le disputa y algunas veces ofrece cierto grado de simplicidad operativa indispensable en el área. De acuerdo con esa filosofía de optimización se aproximan los complejos problemas de decisión que involucran la selección de valores para un número de variables interrelacionadas, enfocándose en un solo objetivo diseñado para cuantificar el desempeño y la medida de la calidad de la decisión. Frecuentemente, este objetivo se suele plantear como minimización o maximización sujeta a restricciones que limitan la selección de las variables de decisión. Si un aspecto del problema en estudio se puede aislar y caracterizar mediante un objetivo, como el beneficio o el costo en una situación de negocios, la velocidad o la distancia en un problema físico, la tasa de retorno esperada en un ambiente de inversiones bajo riesgo, o el beneficio social en el contexto de una planeación gubernamental, la optimización podría decirse que ofrece un marco de trabajo conveniente para el análisis.

Por supuesto que es raro que en una situación se representen completamente todas las complejas interrelaciones entre las variables, las restricciones y los objetivos apropiados cuando se aborda un problema complejo de decisión. De ahí que, como sucede con todas las técnicas cuantitativas de análisis, una formulación particular de optimización deberá considerársele como una aproximación. La destreza en la modelación, para capturar los elementos esenciales del problema, y un buen juicio en la interpretación de los resultados se requieren para obtener conclusiones significativas. Entonces, la optimización, debe considerarse como una herramienta para conceptualizar y analizar, más que un principio que subyace a la solución correcta desde el punto filosófico.

Habilidades o destrezas y un buen juicio, con respecto a la formulación del problema y la interpretación de resultados, se mejoran mediante la práctica concreta y a través de un completo entendimiento de la teoría relevante. La formulación en sí misma considera una negociación entre los objetivos conflictivos de construir un modelo matemático suficientemente complejo y preciso para capturar los descriptores del problema y la construcción de un modelo que sea manejable. Los expertos se enfrentan con ambos aspectos de esa dicotomía. Por lo que, si se aspira en convertirse en uno de esos expertos, se debe aprender a identificar y captar los aspectos importantes de un problema por medio de los ejemplos y la experiencia. Es necesario distinguir entre los problemas manejables de los no manejables, conforme se estudian las técnicas y teorías disponibles y mediante el cuidado de fomentar la capacidad para extender la teoría existente a nuevas situaciones, como se pretende en los programas de posgrado.

1.1 Historia y aplicaciones

1.2 Historia y aplicaciones

La Optimización puede definirse como el campo de conocimiento que busca la determinación de las mejores soluciones a ciertos problemas definidos matemáticamente, los cuales a menudo son modelos de la realidad física.

Este campo incluye el estudio de los criterios de optimalidad para los problemas, la determinación de los métodos algorítmicos de solución, el estudio de la estructura de esos métodos y la experimentación computacional con los métodos bajo condiciones de ensayo y con problemas de la vida real. El estudio puede hacerse desde un punto de vista matemático puramente, pero no es el enfoque del curso.

Antes de 1940 el campo de estudio en cuestión se había desarrollado muy poco, se habían realizado algunos cálculos por mínimos cuadrados y los métodos de descenso se habían aplicado en algunos problemas de física. Se conocía el método de Newton con varias variables y se empezaron a desarrollar métodos más sofisticados como el del campo autoconsistente para los problemas de variación en la química teórica.

El surgimiento de las computadoras fue trascendental para el desarrollo de los métodos de optimización y de todo el análisis numérico. En las décadas de los cuarentas y cincuentas se introdujo y desarrolló la programación lineal. De hecho, el término de programación, como sinónimo de optimización, se usó originalmente en el sentido de planeación óptima. Inicialmente, los métodos de optimización tuvieron un rango restringido de aplicación. Después, surgieron los métodos a gran escala, de gran aplicabilidad pero que no se basaron en alguna estructura particular del problema. Los métodos posteriores, al principio, fueron muy crudos e ineficientes. En 1959, se dió un revolucionario auge en el área, con la publicación del reporte de W.C. Davidon sobre la introducción de los métodos de métricas variables. En una reunión por el año de 1961 se discutieron las dificultades

para minimizar funciones con diez variables, pero M.J.D. Powell programó un método basado en las ideas de Davidon, con el que resolvió problemas con 100 variables en un corto tiempo.

La solución de un determinado problema no es lo único que pudiera interesarle al lector, un análisis de sensibilidad de los parámetros del modelo, también es importante, particularmente cuando el modelo matemático no es muy próximo a la realidad o cuando su diseño no es tan preciso como su solución. El lector puede estar interesado en la variación de la solución obtenida, mediante el estudio los parámetros, en rangos bastante amplios, de ser posible sin resolver numerosas veces el modelo.

El libro si se refiere a varios problemas estándares que pueden surgir. De hecho, el material se divide en dos partes, la primera dedicada a la optimización no restringida, en la cual el valor óptimo se encuentra para una función objetivo de muchas variables, no restringidas. Este tema por si mismo es importante, pero además es una herramienta importante para resolver algunos problemas restringidos.

El caso especial de la suma de funciones cuadradas, que provienen de los problemas de ajuste de datos, también se le considera. Lo cual incluye la solución simultánea de conjuntos de ecuaciones no lineales, que de hecho es un tema importante, resuelto a menudo por métodos de optimización.

La segunda parte se refiere a la optimización restringida en la que surgen complicaciones adicionales de varios tipos de restricciones. Un panorama general se trata en la sección 7.

El interés del autor está en presentar métodos prácticos, asociados con la teoría, que se han implementado y del cual se tienen experiencias numéricas satisfactorias. También, es de interés para el autor, la confiabilidad de los algoritmos y una demostración o buenas razones sobre una razonable rapidez de la convergencia de una solución.

No es fácil recomendar un algoritmo particular que mejor resuelve un problema específico, en especial cuando la decisión no es tan clara como pareciera. Hay muchos casos especiales que deben tomarse en cuenta, por ejemplo la relativa facilidad para calcular la función y sus derivadas. Similarmente, consideraciones sobre cómo plantear mejor el problema, en una primera instancia, son relevantes para la selección del método. Finalmente, de mayor importancia, la decisión está sujeta a la disponibilidad de subrutinas o paquetes computacionales para implementar los métodos. En algunos programas, las librerías proveen árboles de decisión para ayudar a los usuarios a seleccionar el método. Aunque son de utilidad, se les debe de usar sólo como una guía, y nunca como un sustituto del sentido común o las recomendaciones de los especialistas en técnicas de optimización.

2 Introducción al análisis y construcción de modelos¹

A continuación se resume el proceso general de solución que es común a todos los tipos de situaciones de decisión.

1. Establecer el criterio a usarse. Por ejemplo. En una situación simple el criterio puede ser seleccionar la acción que maximiza la utilidad.
2. Seleccionar un conjunto de alternativas para consideración.
3. Determinar el modelo a usarse y los valores de los parámetros del proceso. Por ejemplo, podemos decidir que una expresión adecuada para el total de gastos es:

$$\text{Total de gastos} = a + b(\text{unidades vendidas})$$

¹Bonini, C.P., Asuman, W. H., y Bierman, H. Jr. Quantitative Analysis for Management. Chicago: McGraw Hill. IRWIN Series, 1997(9), capítulo 1.

Los parámetros a y b , deben determinarse para utilizar el modelo.

a = costo fijo durante el periodo del proyecto

b = costo variable por unidad

4. Determinar qué alternativa optimiza (es decir, produce el mejor valor), según el criterio establecido en el punto 1.

Example 1 *Tenemos la posibilidad de vender 1,000 unidades de un producto al gobierno a un precio de 50 unidades monetarias por unidad. ¿Se debería aceptar el pedido? Se asume que la compañía tiene capacidad excesiva. Debemos usar el criterio de maximización de la utilidad. Las alternativas son: (a) aceptar el pedido, o (b) rechazarlo. De acuerdo con el criterio de utilidad, aceptaremos el pedido si se incrementa la utilidad o lo rechazamos si no se incrementa la utilidad.*

Solution. *Necesitamos conocer los costos crecientes por producir las 1,000 unidades. El modelo relevante de costos es:*

$$C = a + 1,000b$$

Suponemos que se tienen costos fijos por 5,000 unidades monetarias (a es igual a 5,000) y que los costos variables al producir una unidad son 30 unidades monetarias (b es igual a 30). Entonces el total de costos relevantes si se acepta el pedido son de 35,000 unidades monetarias (5,000 más 30,000).

Una comparación de los ingresos crecientes, 50,000 unidades monetarias, y los costos crecientes, 35,000 unidades monetarias, nos indica que debemos aceptar el pedido. La utilidad sería mayor por 15,000 si “se acepta”, en comparación con la alternativa “se rechaza” el pedido.

2.1 Abstracción y simplificación

Dado que la mente humana no puede considerar cada aspecto de un problema empírico, algunos atributos del problema deben ignorarse si una decisión va a ser tomada. Abstracción y simplificación son pasos necesarios en la solución de cualquier problema humano.

2.2 Construcción del modelo

Después de que el tomador de decisiones ha seleccionado los factores críticos, o variables, de la situación empírica, se les combina de alguna manera lógica de modo que formen un modelo del problema. Un modelo es una representación simplificada de una situación empírica. Idealmente, le quita a un fenómeno natural su confusa complejidad y duplica la conducta esencial del fenómeno natural con algunas variables que están simplemente relacionadas. Entre más simple sea el modelo obtenido, lo mejor para el que toma decisiones, el modelo sirve como una razonable y confiable contraparte del problema empírico.

2.3 Ventajas de un modelo simple

1. Es económico por cuanto al tiempo y pensamiento.
2. Puede ser entendido fácilmente por el que toma decisiones.
3. En caso necesario, el modelo puede modificarse rápida y efectivamente.

2.4 Soluciones

Después que el modelo se ha construido, se pueden derivar conclusiones acerca de su comportamiento por medio del análisis lógico. El que toma decisiones basa entonces sus acciones o decisiones en estas conclusiones.

2.5 Errores

Dos fuentes importantes de error al usar modelos para la toma de decisiones son la exclusión de variables importantes y los errores al definir las relaciones entre las variables. Por ejemplo, en el ejemplo anterior, supongamos que puede esperarse un 40% de pérdida en los rendimientos durante la producción debido a especificaciones restringidas inusuales. Si este factor estuviera presente pero fuese omitido del análisis, el modelo resultante no representaría la situación adecuadamente para los propósitos de decisión (el resultado podría ser una decisión equivocada).

2.6 Técnicas de construcción de modelos

La técnica apropiada para describir y relacionar las variables seleccionadas depende en gran medida de la naturaleza de las variables. Si las variables son susceptibles de alguna forma de medición, y particularmente si pueden dárseles una representación cuantitativa, entonces hay fuertes razones para seleccionar una representación matemática del modelo. Primero, porque hay una disciplina inherente rigurosa en las matemáticas que asegura un procedimiento metódico por parte del investigador: se debe ser específico acerca de qué variables se han seleccionado y qué relaciones se asume que existen entre ellas. Segundo, la matemática es una poderosa técnica para relacionar variables y derivar conclusiones lógicas a partir de premisas dadas. Las matemáticas, combinadas con las modernas computadoras, hacen posible el manejo de los problemas que requieren modelos de gran complejidad y facilita el proceso de toma de decisiones, donde el análisis cuantitativo es aplicable.

El análisis cuantitativo se ha extendido a muchas áreas de las operaciones de negocios de las empresas y se ha vuelto un modo efectivo de enfocar ciertos problemas de decisión.

2.7 Factores cualitativos

Por ejemplo: la moral y el liderazgo en una organización, las restricciones de empleo, las acciones afirmativas, la contaminación y otras áreas de responsabilidad social.

2.8 Resumen

En la toma de decisiones, uno debería establecer el criterio para la toma de una decisión, seleccionar alternativas, determinar un modelo y evaluar las alternativas, usando el modelo y seleccionando la mejor alternativa.

Un modelo es una abstracción y simplificación de un problema real, idealmente, incorpora los elementos y relaciones esenciales del mundo real. Resolver un modelo significa obtener conclusiones lógicas que se derivan del mismo, estas conclusiones deben ser una guía efectiva para la toma de decisiones si el modelo está diseñado y resuelto adecuadamente. La toma de decisiones involucra la integración de información cuantitativa, obtenida del modelo, con el juicio intuitivo acerca de los factores cualitativos.

2.9 Clasificación de modelos

Variables principales en un problema de decisión		
Problema de decisión	Certidumbre	Incertidumbre
Simple	Modelos de caso	Análisis de decisión (árboles de decisión)
Complejo	Modelos de caso Programación lineal y entera	Simulación
Dinámico	Modelos de inventario Modelos PERT (trayectorias críticas)	Simulación Modelos de inventario Modelos de líneas de espera

2.9.1 Problemas simples

Todos los problemas son simplificados al construir un modelo para cualquier análisis. Si de esto resulta sólo un número pequeño de factores o variables y relativamente pocas alternativas, entonces el modelo se denomina simple.

Un modelo de caso o escenario es un modelo de un problema de decisión que se analiza ensayando una serie de casos (posibles resultados o escenarios) usando diferentes alternativas o supuestos. El modelo no está programado para encontrar directamente “la mejor solución”. En lugar de eso, el administrador usa el modelo en un proceso de ensayo y error.

Los modelos de optimización usan procedimientos matemáticos para encontrar la solución óptima e incorporan el uso de probabilidades en la toma de decisiones bajo incertidumbre.

2.9.2 Problemas complejos

Muchos problemas de decisión involucran una gran cantidad de factores o variables importantes, o pueden tener varias alternativas para considerar. Por ejemplo, el problema de decisión de programar el suministro de fábricas a consumidores, a fin de minimizar el costo, involucra cientos de variables y restricciones que pueden tener millones de soluciones.

Los modelos de programación lineal y entera son las técnicas más ampliamente utilizadas para resolver complejos problemas de negocios de este tipo. Usan las técnicas matemáticas para encontrar el máximo (o el mínimo) valor de un objetivo (función), sujeto a un conjunto de restricciones.

Simulación es una técnica para modelar problemas complejos que involucran situaciones con incertidumbre. Se diseña un modelo para reproducir el comportamiento de un sistema. Los modelos de simulación usualmente se analizan con el enfoque de estudio de caso por caso (en contraposición de la optimización).

2.9.3 Problemas dinámicos

Problemas de decisión dinámicos consideran un tipo particular de complejidad, cuando hay una sucesión de decisiones interrelacionadas a lo largo de varios periodos de tiempo. Algunos tipos son los modelos de inventario para determinar cuando solicitar un inventario y qué tantas existencias se deben tener; PERT o modelos de rutas críticas para la programación de proyectos y los modelos de líneas de espera para problemas que involucran tráfico o acumulación.

2.10 Sistemas de soporte para decisión

Un sistema de soporte para decisión (DSS por sus siglas en inglés), es un sistema computarizado integrado diseñado para ayudar en la toma de decisiones. Un DSS incorpora generalmente un modelo (alguno de los señalados arriba), y el sistema computarizado desempeña los cálculos necesarios para resolver el modelo. Generalmente, es más que un modelo, incluye además una base de datos que puede emplearse para proporcionar directamente información al administrador (o al modelo), mediante gráficas o diversos reportes que son fáciles de entender por el usuario. Desde luego que incorpora también tecnología computacional para facilitar el análisis que se requiere en el problema de decisión o para indagar en la base de datos la información requerida.

2.11 Resumen

Las decisiones pueden caracterizarse conforme se toman bajo certidumbre o incertidumbre, dependiendo de si o no los factores principales se asumen como conocidos. La toma de decisiones bajo incertidumbre involucra el uso de probabilidades para expresar la probabilidad de eventos inciertos.

Los problemas de decisión pueden clasificarse como simples (si hay pocas variables importantes), complejos (si hay muchas), o dinámicos (si las decisiones se interrelacionan con el transcurso del tiempo).

3 Conceptos básicos de modelación

Como se explicó anteriormente, un modelo es una simplificación de un problema de decisión. La simplificación se realiza al incluir sólo los elementos importantes y omitir consideraciones no esenciales.

3.1 Modelación matemática

Proceso mediante el cual un sistema físico se traduce en un modelo matemático. Esta modelación considera la naturaleza de la modelación matemática y el enfoque del proceso de modelación.

3.2 Enfoque sistémico

Ofrece un marco teórico que permite ver el problema inmerso en un sistema. Se identifican claramente las características del sistema que son fundamentales para el problema. El proceso de construcción del modelo matemático puede verse de manera simple como un proceso iterativo de múltiples etapas. Las etapas principales en la traducción de un problema del mundo real a la descripción matemática según Murthy and Page (1990, 20) son las siguientes.

1. Formulación del problema
2. Descripción matemática
3. Análisis matemático
4. Interpretación del análisis para obtener una solución

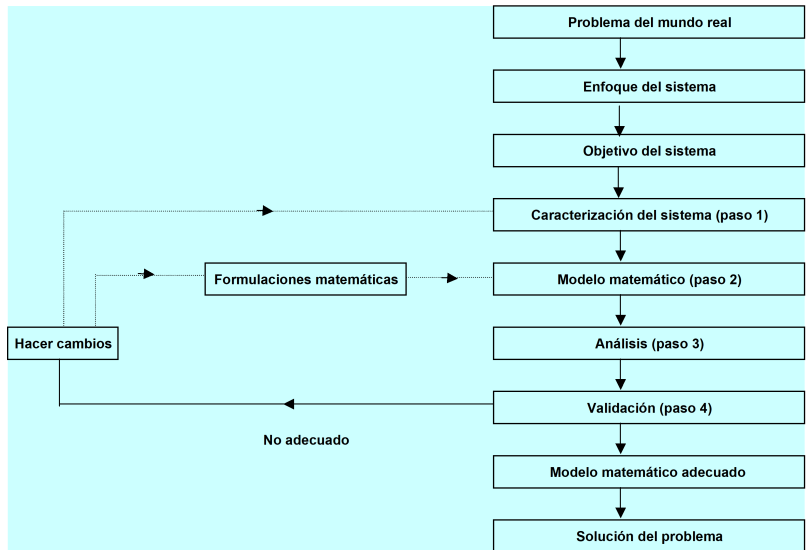


Figure 2

Definition 2 *Un modelo es un conjunto de relaciones entre las variables que en él participan. Visto como una caja negra, las entradas son las variables de decisión que se transforman en medidas de desempeño (las salidas), para un conjunto específico de variables exógenas, políticas y restricciones. El modelo y las relaciones entre variables se ilustran en el siguiente diagrama.*

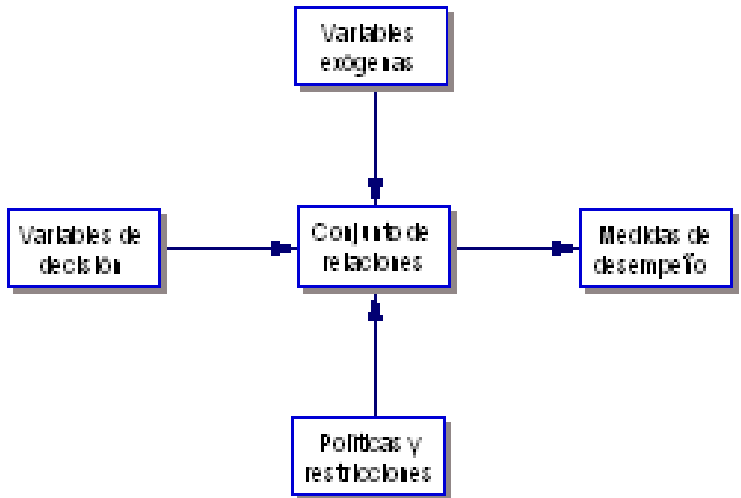


Figure 3: Tomado de Murthy and Page (1990, 20)

Example 3 *Un modelo para la producción y venta de madera contrachapada². La construcción de un modelo que represente las variables y relaciones, más relevantes, en la producción y venta de madera contrachapada (madera, para abreviar), se simplifica para ilustrar el concepto de modelación. Bajo el supuesto de que la administración de una empresa dedicada a la producción y venta de madera contrachapada, hace planes para el próximo año. Los administradores se centran en dos variables de decisión: una sobre la capacidad de la planta y la otra sobre el salario. Respecto a la capacidad de la planta, se plantean las alternativas de expandir o no, el molino, qué tanto y cuándo. En el caso de hacer la expansión ahora, se obtendría capacidad adicional en cualquier trimestre del año en cuestión.*

La segunda decisión tiene que ver con negociaciones laborales que estarían por iniciarse con el sindicato. La empresa y el sindicato tienen que acordar el salario para el próximo año. Esta es una decisión conjunta, resultado del proceso de negociación.

La empresa ha preparado pronósticos de los precios de la madera, que se venderá el año próximo, así como proyecciones sobre cuánto se venderá, es decir, una estimación de la demanda. La empresa sigue una política de producir lo que se ordene, de modo que no se mantenga inventario del producto. Esto significa que la empresa no debe producir más de lo que puede vender, en cualquier periodo.

También, han elaborado pronósticos sobre el precio del tablón, que es la materia prima con la que se elabora la madera.

Para la producción de la madera, la empresa tiene gastos para los salarios, suministros y desde luego la materia prima (tablones). Otros gastos se relacionan con las ventas. La empresa tiene gastos por la renta del equipo para producir. También, hay costos fijos cada periodo (costos hundidos). La siguiente tabla lista los principales factores a considerar en el modelo de producción y venta de madera. La abreviatura MSF, miles de pies cuadrados de superficie, es la unidad de medida para la madera. MBF indica miles de pies de tablón.

²Traducido y adaptado de Bonini (1997,11-14)

	Nombre	Descripción (unidades)
<i>Variables de decisión</i>	<i>Salario</i>	<i>Tasa promedio por miles de empleados (dolares por hora)</i>
	<i>Capacidad adicional</i>	<i>Agregada en cada trimestre (MSF)</i>
<i>Medida de desempeño</i>	<i>Ganancia</i>	<i>Neta trimestral, por operación (miles de dólares)</i>
<i>Variables exógenas</i>	<i>Precio de la madera</i>	<i>Precio de venta trimestral (dólares por MSF)</i>
	<i>Demanda</i>	<i>Demanda trimestral de la madera (MSF)</i>
	<i>Costo del tablón</i>	<i>Costo de compra del tablón (dólares por MBF)</i>
	<i>Productividad</i>	<i>Producción (MSF) por hora de trabajo</i>
<i>Restricciones y políticas</i>	<i>No inventario de madera</i>	
	<i>Producción planeada a coincidir con ventas</i>	
	<i>Análisis trimestral</i>	
<i>Variables intermedias</i>	<i>Rendimiento</i>	<i>Por venta de la madera (miles de dólares por trimestre)</i>
	<i>Gastos de operación</i>	<i>Asociados directamente con la producción (suministro, materia prima, salarios, miles de dólares por trimestre)</i>
	<i>Gastos de envío</i>	<i>(miles de dólares por trimestre)</i>
	<i>Gastos mano de obra</i>	<i>(miles de dólares por trimestre)</i>
	<i>Tablones requeridos</i>	<i>Para la producción (MBF por trimestre)</i>
	<i>Producción de madera</i>	<i>Cantidad de madera producida (MSF por trimestre)</i>
	<i>Capacidad</i>	<i>De producción actual del molino (MSF por trimestre)</i>
	<i>Otros gastos</i>	<i>Por comercialización de la madera (miles de dólares por trimestre)</i>
	<i>Gastos fijos</i>	<i>(miles de dólares por trimestre)</i>
	<i>Gastos por equipo</i>	<i>Renta del equipo (miles de dólares por trimestre)</i>

Diagrama de influencia

3.3 Nota técnica. Flujos de efectivo y valor presente³

De importancia particular para el curso son los problemas que involucran flujos de efectivo en varios años, como por ejemplo inversiones de capital que generan flujos en el tiempo. El periodo de tiempo en que se recibe el dinero es un aspecto importante de su valor. No sería lo mismo recibir 1,000 U.M. ahora que dentro de cinco años. Si no se les necesita inmediatamente se podrían invertir y tener algo más que 1,000 U.M. dentro de cinco años.

Un enfoque general de los problemas relativos a flujos de efectivo en el tiempo es convertirlos a sus equivalentes en valor presente usando un descuento o tasa de interés y calcular el interés compuesto. El valor presente de 1,000 U.M. a cinco años, recibidos hoy usando una tasa de descuento del 6 por ciento es

$$\frac{1,000}{(1 + 0.06)^5}$$

Lo anterior significa que si se invierten 747.26 U.M. en una cuenta bancaria a una tasa de interés del 6 % (compuesta anualmente), se tendrán 1,000 U.M. al final de los cinco años.

En general, el valor presente de una cantidad A , con una tasa de descuento r , recibido en n años es

$$\frac{A}{(1 + r)^n}$$

Algunas ocasiones se tienen varios flujos a lo largo de cierto número de años, con la misma tasa. Por ejemplo, el valor presente de A recibidos al final del año 1, más B recibido al final del año 2 es

$$\frac{A}{(1 + r)^1} + \frac{B}{(1 + r)^2}$$

4 ELEMENTOS DE TEORÍA DE MATRICES

4.1 Conceptos básicos de álgebra lineal

Definition 4 *Un conjunto no vacío V es un espacio vectorial sobre un campo K si en V están definidas dos operaciones: una suma de elementos de V , denominados vectores, y una multiplicación de elementos de V por elementos de K , llamados escalares, como sigue.*

Axiom 5 *Cerradura de la suma o adición vectorial, la cual asocia a cada par de elementos u, v de V , un único elemento de V , denotado $u + v$, que satisface los siguientes axiomas de la adición.*

A.1. Conmutatividad. Para dos elementos cualesquiera u, v de V se cumple

$$u + v = v + u$$

A.2. Asociatividad. Para tres elementos cualesquiera u, v, w de V se cumple

$$(u + v) + w = u + (v + w)$$

A.3. Existencia del neutro aditivo. Existe un elemento único en V , denominado neutro y denotado por θ tal que

$$u + \theta = u$$

³Adaptado del Apéndice, Capítulo 1, Bonini.

para todo u en V .

A.4. Existencia del inverso aditivo. Para todo u en V existe un único elemento, denominado inverso aditivo y denotado por $-u$, tal que

$$u + (-u) = 0$$

Axiom 6 Cerradura de la multiplicación por escalares, la cual asocia a todo u en V y cada k en K , un único elemento de V , denotado por ku , que se satisface los siguientes axiomas de la multiplicación por escalares.

M.1. Distributividad. Para todos los escalares k y cualesquiera elementos u, v de V

$$k(u + v) = ku + kv$$

M.2. Distributividad. Para todos los escalares k y m y todo u en V ,

$$(k + m)u = ku + mu$$

M.3. Asociatividad. Para todos los escalares k y m y todo u en V

$$k(mu) = (km)u$$

M.4. Para todo u en V , $1u = u$

Corollary 7 Consecuencias inmediatas de los axiomas de espacios vectoriales o lineales son la unicidad del elemento cero y el inverso aditivo.

Definition 8 El producto escalar de dos vectores $x = (x_1, \dots, x_n)$ y $y = (y_1, \dots, y_n)$ se define como

$$x \cdot y = \sum_{i=1}^n x_i y_i$$

Proposition 9 Desigualdad de Cauchy-Schwarz

$$|x^T \cdot y| \leq \|x\| \|y\|$$

Example 10 Los siguientes conjuntos son ejemplos de espacios vectoriales

1. \mathbb{R} con la adición y multiplicación de números reales usuales.
2. \mathbb{C} con la adición y multiplicación de complejos usuales.
3. \mathbb{R}^n con la adición de vectores y multiplicación por reales usual del espacio euclideo de dimensión n .
4. El conjunto V de todos los vectores de \mathbb{R}^n que son ortogonales a un vector fijo a , no nulo. En particular, cuando $n = 2$, V es una recta que pasa por el origen con a como vector ortogonal. Si $n = 3$, V es un plano que pasa por el origen y tiene como vector normal a .

Definition 11 Un subconjunto no vacío S de un espacio lineal V que es a su vez un espacio lineal con la adición y multiplicación definidas en V se denomina subespacio.

Criterion 12 *Un criterio simple para verificar si un subconjunto S de V es un subespacio de V consiste en comprobar solamente que se cumplen los axiomas de cerradura.*

Definition 13 *Sean u_1, \dots, u_n vectores de V y k_1, \dots, k_n escalares, la expresión*

$$k_1 u_1 + \dots + k_n u_n$$

se denomina combinación lineal finita de los vectores u_1, \dots, u_n .

Corollary 14 *De las definiciones de espacio vectorial y combinación lineal se desprende fácilmente que para un determinado conjunto de vectores u_1, \dots, u_n de V , el conjunto S de todas las combinaciones lineales de esos vectores, es un espacio vectorial, denominado el espacio generado por el conjunto de vectores u_1, \dots, u_n .*

Definition 15 *Sea S un subconjunto no vacío del espacio lineal V . Un elemento x de V de la forma*

$$x = \sum_{i=1}^n c_i x_i$$

donde x_1, \dots, x_n son todos elementos de S y c_1, \dots, c_n son escalares, es una combinación lineal finita de elementos de S .

Proposition 16 *El conjunto de todas las combinaciones lineales de elementos de S satisface los axiomas de cerradura y de ahí que es un subespacio de V .*

Notation 17 *Al subespacio generado por S , o la extensión lineal de S , lo denotamos por $L(S)$.*

Notation 18 *Si S es vacío, definimos a $L(S)$ por $\{0\}$, el conjunto que sólo contiene al cero.*

Definition 19 *La dimensión de un subespacio S es el número máximo de vectores linealmente independientes en S .*

4.2 Espacios Euclidianos

Definition 20 *E^n es el espacio lineal n -dimensional sobre el campo de los números reales \mathbb{R} o bien, el espacio euclideo de dimensión n , cuyos elementos son vectores $x = (x_1, x_2, \dots, x_n)$, con n componentes.*

Notation 21 *Si cada componente del vector es no-negativo, se anota*

$$x \geq 0$$

Corollary 22 *Cualquier subespacio M de E^n es un subconjunto cerrado en las operaciones de suma vectorial y multiplicación escalar.*

Definition 23 *El segmento de recta que une a los vectores x , y es el conjunto*

$$[x, y] = \alpha x + (1 - \alpha) y$$

con

$$0 \leq \alpha \leq 1$$

Como consecuencia de que un mismo subespacio puede ser generado por diferentes conjuntos, véase por ejemplo E^2 , el cual es generado por

$$\{i, j\}, \{i, j, i + j\}$$

ó

$$\{0, i, -i, i - j, i + j\}$$

surgen preguntas como las siguientes:

1. ¿Cuáles espacios pueden generarse con un conjunto finito de elementos?
2. Si un espacio se genera con un conjunto finito de elementos, ¿cuál es el menor número de elementos requerido para generarlo?

Problem 24 *De respuesta a las dos cuestiones anteriores.*

Definition 25 *El complemento ortogonal del subespacio M de E^n , denotado por M^\perp , es el conjunto de todos los vectores ortogonales a todos los vectores de M .*

Proposition 26 *M^\perp es un subespacio, porque*

1. Si x y $y \in M^\perp$ y $c \in M$, entonces

$$(x + y) \cdot c = 0$$

luego la suma es cerrada.

2. La multiplicación por escalares es cerrada, porque, si λ es un escalar, entonces $\lambda x \in M^\perp$ ya que

$$\lambda x \cdot c = 0$$

para todo $c \in M$.

Proposition 27 *M y M^\perp generan a E^n porque, para todo $x \in E^n$*

$$x = \text{proy}_M(x) + \text{proy}_{M^\perp}(x)$$

4.3 Ortogonalidad e independencia lineal

Definition 28 *Dos vectores $x = (x_1, \dots, x_n)$ y $y = (y_1, \dots, y_n)$ son ortogonales si y solo si*

$$x \cdot y = 0$$

Definition 29 *Si además, $x \cdot x = y \cdot y = 1$ se dice que los vectores son ortonormales.*

Definition 30 *La norma o magnitud de un vector x se denota y define como*

$$\|x\| = (x \cdot x)^{1/2}$$

Definition 31 *Generalizando, los vectores u_1, \dots, u_n son ortogonales, si $u_i \cdot u_j = 0$ para cualesquiera $i \neq j$ y ortonormales si, además, $u_i \cdot u_i = 1$ para $i = 1, \dots, n$.*

Example 32 *Cualquier conjunto, no vacío, de vectores ortogonales, no nulos, siempre puede convertirse en un conjunto de vectores ortonormales, cuando se multiplica cada vector por el recíproco de su norma. Es decir, si u_1, \dots, u_n son ortogonales, entonces los vectores*

$$\frac{u_1}{(u_1 \cdot u_1)^{1/2}}, \dots, \frac{u_n}{(u_n \cdot u_n)^{1/2}}$$

son vectores ortonormales.

Definition 33 *El conjunto de vectores u_1, \dots, u_n es linealmente dependiente si existen escalares $\lambda_1, \lambda_2, \dots, \lambda_n$, no todas cero, tales que*

$$\sum_{i=1}^n \lambda_i u_i = 0$$

Definition 34 *Si no existe dicho conjunto, esto es, si la única solución de la ecuación anterior es la solución trivial 0, se dice que los vectores son linealmente independientes.*

Definition 35 *Un conjunto de vectores linealmente independientes u_1, \dots, u_n que generan al espacio E^n es una base de ese espacio.*

Corollary 36 *Toda base de E^n tiene exactamente n vectores linealmente independientes.*

4.4 Transformaciones lineales¹

A continuación consideramos funciones cuyo dominio y rango son subespacios de espacios lineales, que cumplen las propiedades siguientes. Estas funciones las denominamos transformaciones, mapeos u operadores lineales.

Definition 37 *Si V y W son dos espacios lineales sobre un mismo campo K , una función se llama transformación lineal de V en W si tiene las siguientes dos propiedades.*

1. $T(x + y) = T(x) + T(y)$ para todo $x, y \in V$
2. $T(kx) = kT(x)$ para todo $x \in V$ y $k \in K$

Las propiedades anteriores significan que T preserva la adición y la multiplicación por escalares. Combinando las dos propiedades se obtiene la expresión

$$T(a_1x + a_2y) = a_1T(x) + a_2T(y)$$

donde $a_1, a_2 \in K$ y $x, y \in V$.

Generalizamos, por inducción, la propiedad combinada en

$$T\left(\sum_{i=1}^n a_i x_i\right) = \sum_{i=1}^n a_i T(x_i)$$

donde $a_i \in K$ y $x_i \in V$.

Definition 38 La norma de una transformación lineal $T : V \rightarrow W$ se define como

$$\|T\| = \max_{\|x\| \leq 1} \|T(x)\|$$

Proposition 39 La norma anteriormente definida tiene la propiedad siguiente

$$\|T(x)\| \leq \|T\| \cdot \|x\|$$

para todo $x \in V$.

Example 40 La transformación identidad $T(x) = x$. usualmente denotada por I .

Example 41 La transformación cero. $T(x) = 0$ denotada por O .

Example 42 La multiplicación por un escalar. $T(x) = kx$

Example 43 Un sistema de ecuaciones lineales. Consideremos los espacios euclidianos $V = E^n$ y $W = E^m$, y $m \times n$ números reales, a_{ij} , donde $i = 1, 2, \dots, m$ y $j = 1, 2, \dots, n$. Definimos la transformación lineal como el mapeo que manda cada vector $x = (x_1, \dots, x_n)$ de V sobre un vector de W , mediante la regla de correspondencia siguiente

$$y_i = \sum_{j=1}^n a_{ij} x_j$$

para cada $i = 1, 2, \dots, m$.

Example 44 Producto interior con un elemento fijo. Sea $V = E^n$, y z un elemento fijo en V , definimos

$$T : V \rightarrow W$$

como

$$T(x) = x \cdot z$$

para todo $x \in V$.

Example 45 Proyección en un subespacio. Sea V un espacio lineal y S un subespacio de dimensión finita de V , definimos a

$$T : V \rightarrow W$$

como

$$T(x) = s$$

donde

$$s = \sum_{i=1}^n (x e_i) e_i$$

si

$$\{e_i\}$$

es una base ortonormal de S , con $i = 1, 2, \dots, n$. La s así definida es la proyección de x en el subespacio S .

Example 46 *El operador derivación.* Sea V el espacio lineal de todas las funciones reales derivables definidas en el intervalo (a, b) . Sea W el espacio de todas las funciones derivadas de elementos de V . La transformación

$$D : V \rightarrow W$$

asocia cada función f de V con su derivada f' , elemento de W , es decir

$$D(f) = f'$$

Example 47 *El operador integración.* Sea V el espacio lineal de todas las funciones reales continuas definidas en el intervalo $[a, b]$. El operador

$$\int : V \rightarrow W$$

tiene por regla de correspondencia

$$\int (f) = g$$

donde g es la función en V definida por

$$g(x) = \int_a^x f(t) dt$$

si

$$a \leq x \leq b$$

4.5 Tópicos sobre matrices

Notation 48 *A continuación se representan un vector columna, un vector fila y una matriz*

$$a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

$$b = (b_{11} \quad b_{12} \quad \cdots \quad b_{1n})$$

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}$$

Definition 49 *El rango de una matriz A es el número máximo de columnas linealmente independientes de A , y es igual al número de vectores fila linealmente independientes.*

Definition 50 *Una matriz $A_{m \times n}$ se dice que es de rango completo si su rango r es igual al mínimo entre m y n .*

Definition 51 Dos matrices A y B cuadradas, son similares si existe una matriz no singular S tal que

$$B = S^{-1}AS$$

Notation 52 Utilizando vectores fila y submatrices, a continuación, se hace una descomposición de la matriz A

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix}$$

en un arreglo de 2×2

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

donde

$$A_{11} = [a_{11} \quad a_{12}]$$

$$A_{12} = [a_{13} \quad a_{14}]$$

$$A_{21} = \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

y

$$A_{22} = \begin{bmatrix} a_{23} & a_{24} \\ a_{33} & a_{34} \end{bmatrix}$$

Otra descomposición, que frecuentemente se utiliza, es

$$A = [B|C]$$

donde

$$B = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

y

$$C = \begin{bmatrix} a_{14} \\ a_{24} \\ a_{34} \end{bmatrix}$$

Definition 53 Una matriz de dimensión $n \times n$ es ortogonal si se cumple la ecuación matricial siguiente

$$A^T A = A A^T = I$$

donde I es la matriz identidad

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

La ecuación anterior puede interpretarse diciendo que A es ortogonal si sus columnas y renglones, vistos como vectores, son ortonormales.

4.6 Ecuaciones lineales y matrices inversas

Una de las situaciones más comunes en las que se presentan matrices es en los sistemas de ecuaciones lineales

$$\sum_{j=1}^n a_{ij}x_j = b_i, \quad i = 1, \dots, n \quad (1.3.1)$$

Se asume que a_{ij} y b_i son coeficientes conocidos, y se desea encontrar los valores de x_1, \dots, x_n para los cuales la ecuación (1.3.1) se satisface. La ecuación (1.3.1) generalmente se escribe en su forma matriz-vector como:

$$Ax = b \quad (1.3.2)$$

El sistema (1.3.2) tiene solución única x si el $\det A \neq 0$. Un método para obtener esta solución es el proceso de Eliminación Gaussiana.

Suponga que $a_{11} \neq 0$. Se multiplica la primer ecuación de (1.3.1) por $a_{11}^{-1}a_{21}$ y se resta de la segunda ecuación. Después se multiplica la primer ecuación por $a_{11}^{-1}a_{31}$ y se resta de la tercera ecuación. Se continua este proceso hasta que el múltiplo correspondiente del primer renglón ha sido restado de cada uno de los otros renglones. El sistema se reduce a la forma:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{22}^1x_2 + \dots + a_{2n}^1x_n &= b_2^1 \\ &\vdots \\ a_{n2}^1x_2 + \dots + a_{nn}^1x_n &= b_n^1 \end{aligned} \quad (1.3.3)$$

donde los 1's en los coeficientes indican que en general, han sido cambiados de sus valores originales. El efecto de estas modificaciones, es eliminar el x_1 de las últimas $n-1$ ecuaciones.

El sistema modificado (1.3.3) tiene las mismas soluciones que el sistema original (1.3.1), ya que si x_1, \dots, x_n son una solución de (1.3.1) también satisfacen (1.3.3). Inversamente, haciendo las modificaciones análogas de (1.3.3) (i.e. multiplicar $a_{11}^{-1}a_{21}$ por el primer renglón y sumarlo al segundo renglón, etc.) para regresar a el sistema original (1.3.1), luego si x_1, \dots, x_n son una solución de (1.3.3), también son una solución de (1.3.1).

Después, se supone que $a_{22}^1 \neq 0$ y se repite el proceso para las últimas $n-1$ ecuaciones para eliminar los x_2 desconocidos de las últimas $n-2$ ecuaciones. Continuando con este proceso, se llega a el sistema triangular de ecuaciones:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{22}^1x_2 + \dots + a_{2n}^1x_n &= b_2^1 \\ &\vdots \\ a_{n-1,n-1}^{(n-2)}x_{n-1} + a_{n-1,n}^{(n-2)}x_n &= b_{n-1}^{(n-2)} \\ a_{nn}^{(n-1)}x_n &= b_n^{(n-1)} \end{aligned} \quad (1.3.4)$$

el superíndice indica el número de veces que el coeficiente ha sido cambiado. Este sistema de ecuaciones tiene exactamente las mismas soluciones que el sistema original.

Luego, supongamos que $a_{nn}^{(n-1)} \neq 0$. Se obtiene x_n utilizando únicamente la última ecuación. Conociendo x_n se obtiene x_{n-1} sustituyendo en la penúltima ecuación, trabajando de esta forma, el sistema triangular permite encontrar las soluciones x_{n-2}, \dots, x_1 . Por lo tanto, bajo las suposiciones anteriores se demuestra que el sistema (1.3.4), y por ende el sistema (1.3.1), tiene solución única.

1.3.1 Solución de Sistemas Lineales. Sea A una matriz real ó compleja $n \times n$ con $\det A \neq 0$ y sea b un n -vector dado. Entonces el sistema $Ax = b$ tiene solución única.

4.7 Matriz Inversa

En ocasiones es útil escribir la solución de $Ax = b$ en términos de otra matriz, conocida como la inversa de A y denotada por A^{-1} . Supongamos que $\det A \neq 0$, y sea e_i el vector con un 1 en la i -ésima posición y 0's en las restantes. Por (1.3.1) tenemos que cada uno de los sistemas

$$Ax_i = e_i, \quad i = 1, \dots, n \quad (1.3.5)$$

tiene una solución única x_i . Sea $X = (x_1, \dots, x_n)$ el vector cuyas componentes son dichas soluciones, entonces (1.3.5) es equivalente a

$$AX = I \quad (1.3.6)$$

donde, I es la matriz identidad.

Podemos encontrar una matriz Y tal que

$$XY = I$$

Multiplicando por A y usando (1.3.6), se concluye que $Y = A^{-1}$, tal que $XA = I$. Ahora se define $A^{-1} = X$. Entonces las relaciones $XA = I$ y $AX = I$ son equivalentes a

$$A^{-1}A = I = AA^{-1} \quad (1.3.7)$$

que son las relaciones básicas que la inversa A^{-1} debe satisfacer. Se concluye también que:

$$\det A \det A^{-1} = 1 \quad \text{ó} \quad \det A^{-1} = (\det A)^{-1} \quad (1.3.8)$$

Se puede multiplicar la ecuación $Ax = b$ por A^{-1} resultando

$$x = A^{-1}b \quad (1.3.9)$$

En conclusión, si $\det A \neq 0$, entonces una matriz A^{-1} que satisface (1.3.7) existe. Inversamente, si esta matriz existe, entonces (1.3.8) demuestra que $\det A \neq 0$. Se dice que A es no singular o invertible si $\det A \neq 0$ o si A^{-1} existe, de otra forma A es singular.

1.3.2 Si A y B son matrices $n \times n$, entonces AB es no singular si y sólo si ambas A y B son no singulares. Además, $(AB)^{-1} = B^{-1}A^{-1}$.

4.8 Formulación matricial de la Eliminación Gaussiana.

Supongamos que todos los divisores en el proceso son distintos de cero, así que el intercambio de renglones no es necesario. Los cálculos que conducen al sistema reducido (1.3.3) puede representarse como la multiplicación de matrices $L_1Ax = L_1b$ donde

$$L_1 = \begin{bmatrix} 1 & & & & & & \\ -l_{21} & 1 & & & & & \\ & 0 & 1 & & & & \\ & & \vdots & \ddots & & & \\ -l_{n-1} & 0 & \cdots & 0 & 1 & & \end{bmatrix}, \quad l_{i1} = \frac{a_{i1}}{a_{11}}, i = 2, \dots, n \quad (1.3.10)$$

Similarmente, el k -ésimo paso se realiza por la multiplicación de la matriz

$$L_k = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & \cdot & & 1 & \\ & & \cdot & & -l_{nk+1k} & \\ & & \cdot & & \vdots & \\ & & \cdot & & \cdot & \\ 0 & & 0 & & -l_{nk} & & 0 & 0 & 1 \end{bmatrix}, \quad l_{ik} = \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}, i = k+1, \dots, n \quad (1.3.11)$$

Por lo tanto el sistema triangular (1.3.4) se obtiene por

$$\widehat{L}Ax = \widehat{L}b, \quad \widehat{L} = L_{n-1} \cdots L_2 L_1 \quad (1.3.12)$$

La matriz \widehat{L} es una matriz triangular inferior con 1's en la diagonal.

Se denota por U a la matriz triangular superior del sistema (1.3.4) y tenemos que

$$\widehat{L}A = U \quad (1.3.13)$$

Como \widehat{L} es una matriz triangular inferior con 1's en la diagonal principal el $\det \widehat{L} = 1$, y la inversa de \widehat{L} existe. Multiplicando (1.3.13) por \widehat{L}^{-1} y haciendo $\widehat{L}^{-1} = L$ tenemos que:

$$A = LU \quad (1.3.14)$$

Notar que L es también una matriz triangular inferior con 1's en la diagonal principal. De aquí, podemos ver que $A = LU$ es la factorización de A en el producto de dos matrices triangulares superior e inferior, lo anterior se conoce como factorización LU ó descomposición LU de A .

Una matriz de permutación es una matriz $n \times n$ con exactamente un 1 en cada renglón y en cada columna y 0's como valores restantes. Por ejemplo

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

La matriz identidad es el ejemplo trivial de una matriz de permutación, ya que cualquier matriz de permutación puede obtenerse de la matriz identidad reacomodando renglones o columnas. Si P es una matriz de permutación, el efecto de las multiplicaciones PA y AP es reacomodar renglones o columnas de A , respectivamente, en exactamente la misma forma en como P fue obtenida de la matriz identidad.

1.3.3 Si A es una matriz $n \times n$ no singular, entonces existe una matriz de permutación P tal que

$$PA = LU \quad (1.3.17)$$

donde L es una matriz triangular inferior con 1's en la diagonal principal, y U es una matriz triangular superior.

4.9 Otro criterio para la No Singularidad.

Un caso especial del Teorema 1.3.1 es cuando $b = 0$; en este caso, si A es no singular, $x = 0$ es la única solución de $Ax = 0$. Si representamos a Ax como la combinación lineal

$$Ax = \sum_{i=1}^n x_i a_i$$

de las columnas a_1, \dots, a_n de A , decir que $Ax = 0$ tiene solamente la solución $x = 0$ es equivalente a decir que las columnas de A son vectores linealmente independientes. Por lo tanto, A es no singular si y solo si sus columnas son linealmente independientes.

1.3.4 Criterios para la No Singularidad. Sea A una matriz $n \times n$ en los Reales o los Complejos. Entonces es equivalente

- a) $\det A \neq 0$;
- b) Existe A^{-1} matriz $n \times n$ tal que $AA^{-1} = A^{-1}A = I$;
- c) $Ax = b$ tiene solución única para b ;
- d) $Ax = 0$ tiene solamente la solución $x = 0$;
- e) Las columnas (o renglones) de A son linealmente independientes.

4.10 Forma Echelon y Rango

Supongamos que la matriz A es singular o, una matriz $m \times n$. Podemos aplicar el proceso de eliminación Gaussiana como antes, utilizando intercambio de renglones si es necesario; pero puede ocurrir que cuando se encuentre un elemento pivote igual a cero, cada uno de los elementos por debajo de este pivote en esta columna sea también cero. En este caso simplemente nos movemos a la siguiente columna, en el mismo renglón y continuamos el proceso. En general, la forma reducida puede tener una apariencia como

$$\begin{bmatrix} * & * & * & \cdot & \cdot & \cdot & \cdots & * \\ 0 & 0 & 0 & * & \cdot & \cdot & \cdots & * \\ \cdot & & & 0 & * & \cdot & \cdots & * \\ \cdot & & & & 0 & * & \cdots & * \\ \cdot & & & & & & \cdots & 0 \\ \cdot & & & & & & & \vdots \\ 0 & \cdot & \cdot & \cdot & & & & 0 \end{bmatrix} \quad (1.3.21)$$

donde los asteriscos indican elementos distintos de cero. Esta es la forma echelon or row echelon de la matriz. En general, el primer elemento distinto de cero en cada renglón aparece por lo menos una posición a la derecha del primer elemento distinto de cero en el renglón previo.

Por la eliminación Gaussiana para matrices no singulares, se puede expresar este proceso como

$$PA = LU$$

donde P es una matriz de permutación $m \times m$, L es una matriz triangular inferior no singular $m \times m$ y U es una matriz $m \times n$ de la forma (1.3.21). Si A es $n \times n$ y no singular, entonces, por todo lo anterior, U es una matriz $n \times n$ con elementos distintos de cero en la diagonal, y por lo tanto la forma row echelon de una matriz no singular es una matriz triangular superior no singular.

1.3.5 El rango renglón y el rango columna de cualquier matriz A $m \times n$ son iguales.

Se define el rango de una matriz A $m \times n$, denotado por $\text{rank}(A)$, como el número de renglones linealmente independientes de A o, equivalentemente, el número de columnas linealmente independientes.

Part II

ELEMENTOS DE CONVEXIDAD⁴

4.11 Conceptos Topológicos

Definition 54 Una sucesión de vectores $\{x_k\}_{k \in \mathbb{N}}$ en E^n converge al límite x si

$$|x_k - x| \rightarrow 0$$

cuando

$$k \rightarrow \infty$$

Es decir, si dado $\varepsilon > 0$ existe $N > 0$ tal que

$$|x_k - x| < \varepsilon$$

cuando

$$k \geq N$$

Cuando se sobrentiende que el conjunto de índices varía en \mathbb{N} , se escribe simplemente que $\{x_k\}$ converge a x , o bien

$$\{x_k\} \rightarrow x$$

Definition 55 Un punto $x \in E^n$ es un punto límite de la sucesión $\{x_k\}$ si existe una subsucesión convergente a x . Esto es, x es un punto límite de $\{x_k\}$ si existe un subconjunto K de \mathbb{N} tal que la subsucesión $\{x_k\}_{k \in K}$ converge a x .

Definition 56 Un entorno o vecindad de $x \in E^n$, con radio ε , es un conjunto de la forma

$$V = \{y \in E^n : |y - x| < \varepsilon\}$$

para algún $\varepsilon > 0$. También se suele llamar esfera con centro en x y radio ε .

Definition 57 Un subconjunto S de E^n , es abierto si para cada punto de S existe un entorno del mismo punto contenido en S . De manera equivalente, se dice que S es abierto si dado $x \in S$ existe $\varepsilon > 0$, tal que

$$|y - x| < \varepsilon \Rightarrow y \in S$$

Example 58 La esfera sólida con centro en el origen y radio 1 (sin incluir la superficie o cáscara)

$$\{x \in E^n : \|x\| < 1\}$$

es un subconjunto abierto de E^n .

⁴Lang, Serge (1966) Linear algebra. MASS: Addison-Wesley Publishing Company, Appendix 1, pp. 279 – 286.

Definition 59 El interior de cualquier conjunto S de E^n es el conjunto de puntos $x \in S$, que son el centro de alguna vecindad contenida en S .

Corollary 60 De las dos definiciones anteriores se tiene que el interior de un conjunto siempre es un conjunto abierto. De hecho, es el mayor conjunto abierto contenido en el conjunto.

Example 61 El interior del conjunto $\{x \in E^n : \|x\| < 1\}$ es el mismo conjunto $\{x \in E^n : \|x\| < 1\}$, el cual es un conjunto abierto.

Definition 62 Un conjunto P es cerrado si todo punto que está arbitrariamente cerca del conjunto P , es un miembro de P . De manera equivalente, si contiene a todos sus puntos límite, es decir, P es cerrado si $x \in P$ para todo x para el cual se cumpla $\{x_k\} \rightarrow x$ con $x_k \in P$.

Example 63 La esfera con centro en el origen y radio 1 (que incluye la superficie o cáscara)

$$\{x \in E^n : \|x\| \leq 1\}$$

es cerrado.

Definition 64 El cierre o cerradura de cualquier conjunto P en E^n , es el menor conjunto cerrado que contiene a P .

Definition 65 Un punto x pertenece a la frontera de un conjunto S si y solo si toda vecindad del punto x contiene puntos de S y de su complemento.

Proposition 66 La frontera de un conjunto es aquella parte del cierre que no está en el interior del conjunto.

Example 67 El conjunto

$$\{x \in E^n : \|x\| = 1\}$$

es la frontera de los conjuntos $\{x \in E^n : \|x\| < 1\}$ y $\{x \in E^n : \|x\| \leq 1\}$.

Definition 68 Un conjunto es compacto si es cerrado y acotado⁵

Theorem 69 Sea S un conjunto compacto y $\{x_k\}$ una sucesión en S , entonces $\{x_k\}$ tiene un punto límite en S , es decir, existe alguna subsucesión de $\{x_k\}$ que converge a un punto de S .

⁵El término acotado significa que el conjunto está contenido dentro de alguna esfera de radio finito.

Definition 70 Un subconjunto S de E^n , se llama convexo si S contiene al segmento que une cualquier par de puntos de S . Es decir, para todo par de puntos x_1, x_2 de S , el segmento

$$x_1 + t(x_2 - x_1) = tx_2 + (1 - t)x_1$$

con

$$0 \leq t \leq 1$$

está contenido en S .

Notation 71 Los puntos en E^n , pueden darse como n -eadas o n -uplas (x_1, x_2, \dots, x_m) , o mediante letras mayúsculas.

Remark 72 Note que el miembro izquierdo de la igualdad es la ecuación del segmento dirigido de x_1 a x_2 ; mientras que el miembro derecho se conoce como una combinación lineal de los vectores de posición x_1 y x_2 , en la que los coeficientes son números reales t entre 0 y 1 inclusive, cuya suma es igual a 1.

Definition 73 Dado un conjunto de puntos $S = \{x_1, x_2, \dots, x_m\}$ en E^n , se dice que el punto

$$x = \sum_{i=1}^m a_i x_i$$

es

1. Combinación lineal de los elementos de S si $a_i \in \mathbb{R}$ para toda $i = 1, 2, \dots, m$.
2. Combinación lineal no negativa de S si $a_i \geq 0$, para toda $i = 1, 2, \dots, m$.
3. Combinación lineal convexa de los elementos de S si $a_i \in [0, 1]$, para toda $i = 1, 2, \dots, m$ y

$$\sum_{i=1}^m a_i = 1$$

Theorem 74 Sean x_1, x_2, \dots, x_m puntos en E^n , cualquier conjunto convexo que los contenga contendrá también todas las combinaciones lineales

$$\sum_{i=1}^m a_i x_i$$

con $0 \leq a_i \leq 1$ y

$$\sum_{i=1}^m a_i = 1$$

La demostración se hace por inducción sobre m y se factoriza $(1 - a_m)$ de los $m - 1$ primeros términos, para aplicar la hipótesis de inducción.

Theorem 75 Sean x_1, x_2, \dots, x_m puntos en E^n . El conjunto de todas las combinaciones lineales convexas

$$\sum_{i=1}^m a_i x_i$$

con $0 \leq a_i \leq 1$ y

$$\sum_{i=1}^m a_i = 1$$

es un conjunto convexo.

Corollary 76 En virtud de los dos teoremas anteriores se concluye que el conjunto de todas las combinaciones lineales descritas en ambos teoremas, es el menor conjunto convexo que contiene a todos los puntos x_1, x_2, \dots, x_m .

Definition 77 Sea S un subconjunto de E^n , el conjunto $C(S)$ formado por todas las combinaciones lineales convexas de elementos de S se denomina cierre convexo de S .

Proposition 78 Propiedades para conjuntos convexas

1. Si S y S' son conjuntos convexas en E^n , entonces $S \cap S'$ también es un conjunto convexo.
2. Si S y S' son conjuntos convexas en E^n , entonces $S + S'$ también es un conjunto convexo.
3. Sea $T : E^n \rightarrow E^m$ una transformación lineal. Si S es convexo en E^n , entonces la imagen directa

$$T(S)$$

es convexa en E^m .

4. Sea $T : E^n \rightarrow E^m$ una transformación lineal y $S' \subset E^m$, un convexo. Si $S = T^{-1}(S')$, la imagen inversa de S' , entonces S es convexo.

Example 79 Los siguientes conjuntos son convexas³

1. El semiplano inferior, limitado por la recta $x_1 = x_2$:

$$S_1 = \{(x_1, x_2) \in E^2 | x_1 \leq x_2\}$$

en E^2 , es convexo porque todo segmento, que une cualquier par de puntos de S_1 , está contenido en S_1 . Es decir, si (x_1, x_2)

$x(x_1, x_2)$ y $y(y_1, y_2)$ son elementos de S_1 , entonces el segmento \overline{xy} que los une está contenido en S_1 . Para demostrar esto último, tomemos cualquier punto $z(z_1, z_2)$ del segmento que une a los puntos anteriores, el cual se expresa como

$$\begin{aligned} z &= ty + (1-t)x \\ (z_1, z_2) &= t(y_1, y_2) + (1-t)(x_1, x_2) \end{aligned}$$

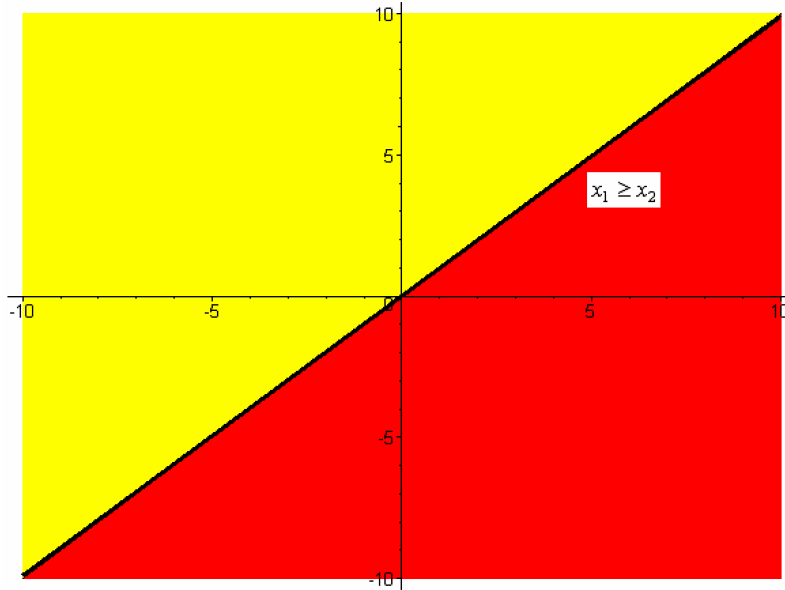


Figure 5

con

$$0 \leq t \leq 1$$

De la igualdad se desprenden las expresiones

$$z_1 = (1 - t) x_1 + t y_1$$

$$z_2 = (1 - t) x_2 + t y_2$$

Como $x = (x_1, x_2)$ y $y = (y_1, y_2)$ son elementos de S_1 se cumplen las desigualdades

$$x_1 \geq x_2$$

$$y_1 \geq y_2$$

Si multiplicamos por los reales no negativos $(1 - t)$ y t , a la primera y segunda desigualdades, respectivamente, obtenemos

$$(1 - t) x_1 \geq (1 - t) x_2$$

$$t y_1 \geq t y_2$$

Sumando ambas desigualdades

$$(1-t)x_1 + ty_1 \geq (1-t)x_2 + ty_2$$

$$z_1 \geq z_2$$

Por lo tanto, concluimos que el punto $z(z_1, z_2)$ pertenece al conjunto S_1 , es decir, que el segmento está contenido en S_1 .

2. La región limitada inferiormente por la parábola vertical $x_2 = x_1^2$:

$$S_2 = \{(x_1, x_2) \in E^2 \mid x_2 \geq x_1^2\}$$

en E^2

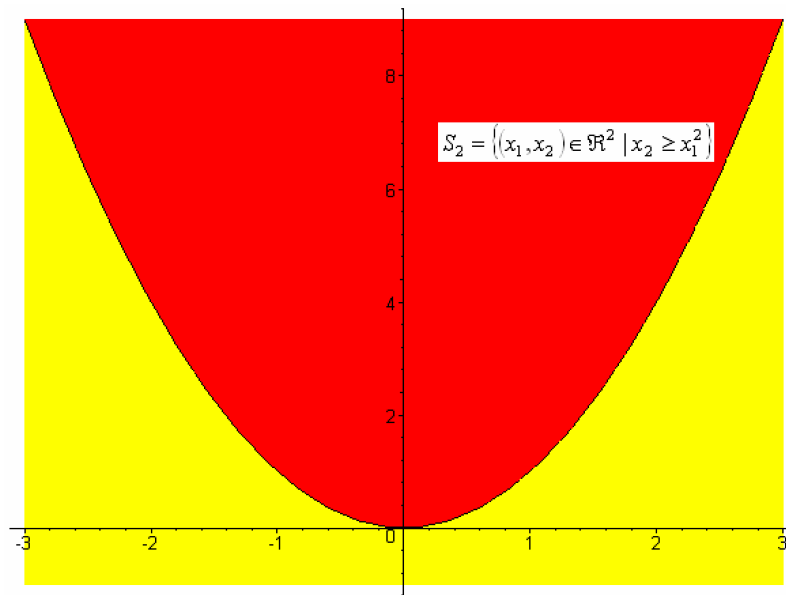


Figure 6

Un argumento similar al utilizado en el ejemplo anterior, explica la convexidad del conjunto S_2 . Consideremos un par de puntos x y y de S_2 , entonces se cumplen las desigualdades

$$x_2 \geq x_1^2$$

$$y_2 \geq y_1^2$$

y

$$(1-t)x_2 \geq (1-t)x_1^2$$

$$ty_2 \geq ty_1^2$$

que resultan de multiplicar por reales no negativos: $(1 - t)$ y t . Consideramos de nuevo un punto arbitrario z del segmento que une a los puntos x y y ,

$$z = ty + (1 - t)x$$

con

$$0 \leq t \leq 1$$

Por lo que, sus coordenadas son

$$z_1 = (1 - t)x_1 + ty_1$$

$$z_2 = (1 - t)x_2 + ty_2$$

Estas coordenadas satisfacen la relación

$$z_2 \geq z_1^2$$

que define al conjunto S_2 , porque se obtiene de las siguientes dos desigualdades

$$\begin{aligned} z_2 &= (1 - t)x_2 + ty_2 \geq (1 - t)x_1^2 + ty_1^2 \\ (1 - t)x_1^2 + ty_1^2 &\geq [(1 - t)x_1 + ty_1]^2 = z_1^2 \end{aligned}$$

La segunda desigualdad es consecuencia de que la diferencia siguiente es no negativa

$$\begin{aligned} (1 - t)x_1^2 + ty_1^2 - [(1 - t)x_1 + ty_1]^2 &= \\ t(1 - t)x_1^2 + t(1 - t)y_1^2 - 2t(1 - t)x_1y_1 &= \\ t(1 - t)[x_1^2 - 2x_1y_1 + y_1^2] &= \\ t(1 - t)[x_1 - y_1]^2 &\geq 0 \end{aligned}$$

3. La recta con pendiente a y ordenada al origen b :

$$S_3 = \{(x_1, x_2) \in E^2 \mid x_2 = ax + b\}$$

en E^2 .

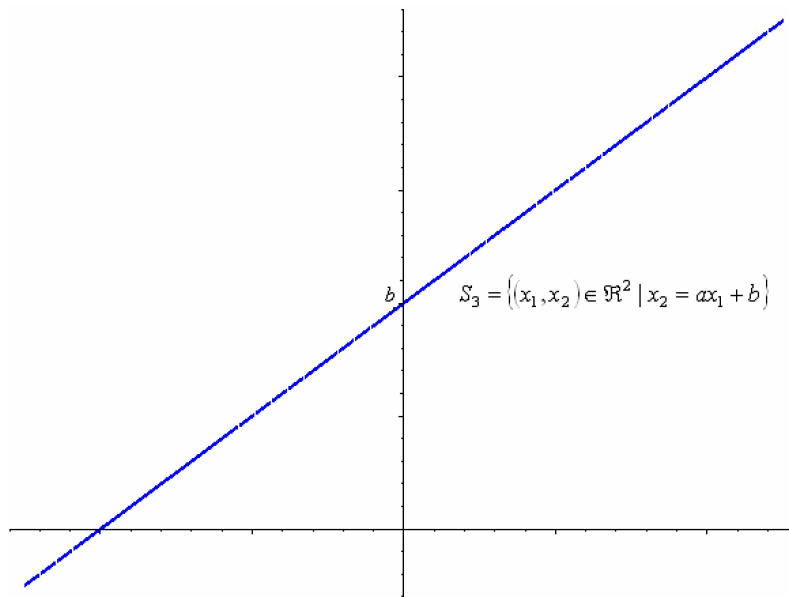


Figure 7

4. El círculo con centro en el origen y radio r :

$$S_4 = \{(x_1, x_2) \in E^2 \mid x_1^2 + x_2^2 = r^2\}$$

en E^2 .

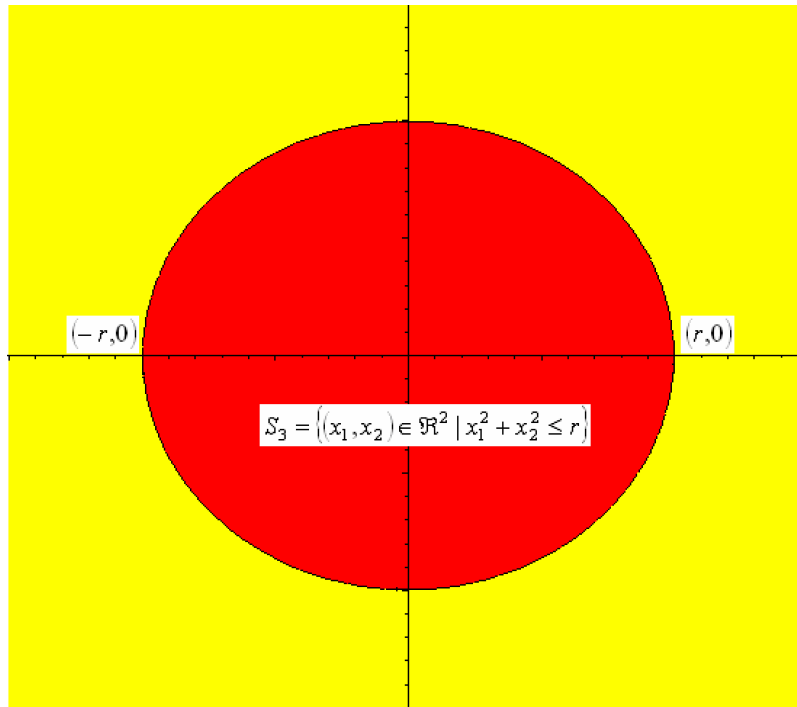


Figure 8

5. El hiperplano H consistente de todos los puntos x de E^2 tales que

$$A \cdot x = c$$

con $A \in E^n$ y $c \in \mathbb{R}$, es un conjunto convexo por ser imagen directa de la transformación lineal

$$T(x) = A \cdot x$$

6. Los semiespacios abiertos

$$S = \{x \in \mathbb{R} \mid x > s\}$$

y

$$S = \{x \in E^n \mid A \cdot x > c\}$$

son convexos.

7. También lo son los semiespacios cerrados

$$S = \{x \in E^n \mid A \cdot x \geq c\}$$

8. En E^2 , un ejemplo de hiperplano es el definido por la ecuación

$$3x - 2y = -1$$

y de semiespacio cerrado, el que está determinado por ese hiperplano.

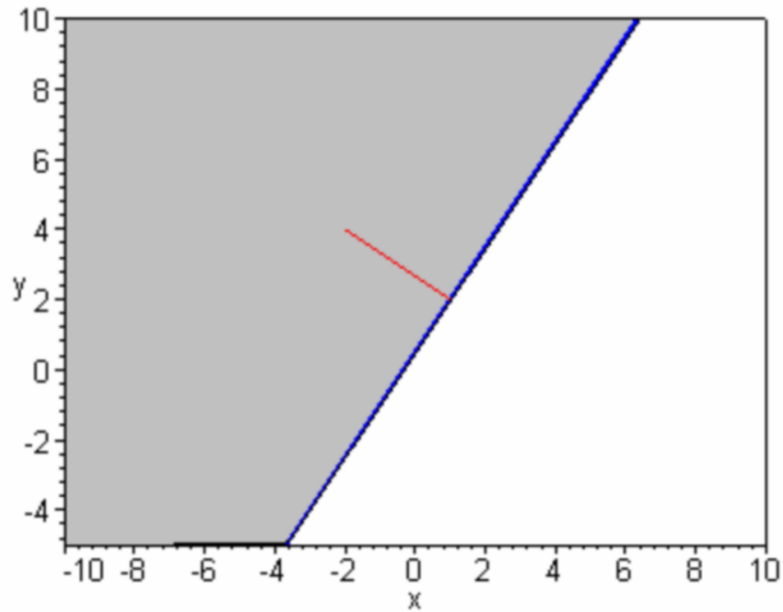


Figure 9

El hiperplano (la recta) pasa por el punto $(1, 2)$ y el vector $n = (3, -2)$ es ortogonal al hiperplano. El semiespacio iluminado en gris está definido por

$$x \cdot n \leq -1$$

9. En general, el hiperplano

$$x \cdot n = c$$

determina dos semiespacios cerrados

$$x \cdot n \geq c$$

y

$$x \cdot n \leq c$$

o dos semiespacios abiertos

$$x \cdot n > c$$

y

$$x \cdot n < c$$

Observemos que la intersección de un número finito de semiespacios es convexa conforme a la propiedad 1 de los conjuntos convexos.

10. Además, en E^2 , la intersección de un número finito de semiplanos puede ser o no acotada, en el sentido de que un conjunto S es acotado si existe un número real $c > 0$ tal que

$$\|x\| \leq c$$

para todo $x \in S$.

Theorem 80 *Separación por hiperplanos.* Sea S un conjunto cerrado convexo de E^n , $p \in E^n$. Entonces p pertenece a S , o existe un hiperplano H que contiene a p , tal que S está contenido en uno de los semiespacios determinados por H .

Proof. Si el punto p no pertenece a S , la idea de la demostración es construir el hiperplano que contiene al punto p con la propiedad de separación enunciada.

Sea $f(x) = \|x - p\|$, una función definida en S , la cual es claramente continua y acotada inferiormente. De ahí que, f alcanza su mínimo⁶ en al menos un punto q de S , es decir,

$$\|q - p\| \leq \|x - p\|$$

para todo $x \in S$.

Construimos el hiperplano que pasa por p y es perpendicular al vector $n = q - p$, el cual no es nulo porque $p \notin S$. Este hiperplano cumple con el requerimiento de separación de que S está contenido en alguno de los semiespacios definidos por él. Digamos que S está contenido en el semiespacio abierto

$$x \cdot n > c$$

con

$$c = p \cdot n$$

porque si consideramos a cualquier q' de S distinto de q , entonces para todo t tal que

$$0 < t \leq 1$$

se cumplen las siguientes implicaciones

$$\begin{aligned} \|q - p\| &\leq \|q + t(q' - q) - p\| = \|(q - p) + t(q' - q)\| \\ &\Rightarrow \\ (q - p)^2 &\leq (q - p)^2 + 2t(q - p) \cdot (q' - q) + t^2(q' - q)^2 \\ &\Rightarrow \\ 0 &\leq 2(q - p) \cdot (q' - q) + t(q' - q)^2 \\ &\Rightarrow \\ 0 &\leq (q - p) \cdot (q' - q) \end{aligned}$$

cuando $t \rightarrow 0$. Esto último, implica que

$$0 \leq n \cdot (q' - q)$$

⁶De acuerdo con el teorema de Weierstrass, para un conjunto acotado inferiormente.

Sumando y restando P en el lado derecho de la desigualdad y asociando convenientemente se obtiene

$$\begin{aligned} 0 &\leq n \cdot (q' - p) + n \cdot (p - q) \\ &\Rightarrow \\ 0 &\leq n \cdot (q' - p) - n \cdot n \\ 0 &\leq n \cdot (q' - p) \end{aligned}$$

porque

$$n \cdot n > 0$$

Finalmente, como q' es un elemento arbitrario de S y se cumple la relación

$$q' \cdot n \geq p \cdot n$$

con esto, se concluye que q' pertenece al semiespacio definido por

$$x \cdot n > p \cdot n$$

es decir, S está contenido en dicho semiespacio. ■

Theorem 81 La cerradura⁴ de un conjunto convexo S , denotada por \bar{S} , es convexa.

Proof. La demostración del enunciado anterior se sigue de que es posible encontrar dos sucesiones $\{p_n\}$ y $\{q_n\}$ de puntos de S , que tengan por límite a cualesquier par de puntos p y q de \bar{S} , respectivamente; de que el límite de

$$tp_n + (1 - t)q_n$$

es

$$tp + (1 - t)q$$

para toda t , y del hecho de que S sea un conjunto convexo. ■

Remark 82 Un punto p es un punto frontera de S si toda vecindad abierta

$$V(p, \varepsilon)$$

de p es tal que

$$\begin{aligned} V \cap S &\neq \phi \\ V \cap (E^n - S) &\neq \phi \end{aligned}$$

Definition 83 Un semiplano H se dice que es un soporte de S , un conjunto convexo, en p , un punto frontera de S , si p pertenece a H y S está contenido en uno de los dos semiespacios cerrados determinados por H .

Theorem 84 Sea S un conjunto convexo en E^n y p un punto frontera de S . Entonces existe un hiperplano H soporte de S en p .

Proof. Sin pérdida de generalidad⁵ podemos suponer que S es cerrado. La construcción del hiperplano con la propiedad deseada se hace con ayuda de las propiedades de acumulación de los elementos de \bar{S} . Para cada $k > 2$ podemos encontrar p_k que no está en S , a una distancia de p menor que $1/k$, porque p es un punto frontera de S . Por el teorema de separación obtenemos un q_k en S cuya distancia a p_k es mínima. Sea $n_k = (q_k - p_k)$ y n'_k el vector unitario en la dirección de n_k . La sucesión de vectores n'_k tiene un punto de acumulación n' en la esfera unitaria porque es compacta. Por el mismo teorema de separación se tiene que para toda $x \in S$ y toda k ,

$$x \cdot n_k > p_k \cdot n_k$$

lo que implica que

$$x \cdot n'_k > p_k \cdot n'_k$$

para toda k y por consiguiente la relación se cumple para el punto de acumulación n' de la sucesión y para p , el límite de la sucesión p_k . Esto es,

$$x \cdot n' = p \cdot n'$$

Esto último prueba que el hiperplano $x \cdot n' = p \cdot n'$ soporta a S en p . ■

Remark 85 Sea S un conjunto convexo y H el hiperplano definido por la ecuación

$$x \cdot n = a$$

Supongamos que para todo $x \in S$ se tiene

$$x \cdot n \geq a$$

Si p es un punto de S que está sobre el hiperplano, entonces p es un punto frontera. En caso contrario, para $\varepsilon > 0$ suficientemente pequeño, $p - \varepsilon n$ debería ser un punto de S , con la propiedad contraria a la suposición. Concluimos de ahí que H es un hiperplano de soporte de S en p .

Definition 86 Un punto p es extremo de S , un conjunto convexo, si p pertenece a S y si no existen puntos distintos q_1 y q_2 tales que p puede expresarse en la forma

$$p = tq_1 + (1 - t)q_2$$

con

$$0 < t < 1$$

En otras palabras, que p no esté dentro de un segmento de recta contenido en S , a menos que sea uno de los puntos extremos del segmento.

Definition 87 Un conjunto S está acotado por debajo o inferiormente si existe un vector $b = (b_1, \dots, b_n)$ tal que para todo $x = (x_1, \dots, x_n)$ en S tenemos que $x_i \geq b_i$ para todo $i = 1, \dots, n$.

Theorem 88 Sea S un conjunto convexo cerrado y acotado inferiormente. Entonces cualquier hiperplano de soporte de S contiene al menos un punto extremo.

Proof. La idea de la demostración consiste en probar que la intersección T de S con cualquier hiperplano de soporte de S contiene un punto extremo de T que también es un punto extremo de S porque T está contenido en S . Nótese que T hereda también las propiedades de ser convexo, cerrado y acotado inferiormente. ■

Proposition 89 *El menor de los conjuntos convexos que contiene a un conjunto E arbitrario con al menos un elemento es la intersección de todos los convexos que lo contienen. Describimos dicho conjunto de la manera siguiente. Sea E^c el conjunto de todas las combinaciones lineales convexas de puntos en E , es decir, los puntos de E^c se expresan en la forma*

$$\sum_{i=1}^m t_i p_i$$

con p_1, \dots, p_m en E con coeficientes reales tales que

$$0 \leq t_i \leq 1$$

y

$$\sum_{i=1}^m t_i = 1$$

Claramente, E^c es convexo y cualquier convexo que contenga a E debe contener a E^c , y de ahí que E^c es el conjunto convexo mínimo que contiene a E .

Definition 90 *Nombramos a E^c , la cerradura convexa de E .*

Proposition 91 *Sean S un conjunto convexo y E el conjunto de sus puntos extremos. Entonces E^c está contenida en S .*

Theorem 92 *de Krein-Milman. Sea S un conjunto cerrado, acotado y convexo. Entonces S es la cerradura convexa de sus puntos extremos, es decir,*

$$S = E^c$$

Proof. Sólo resta demostrar que

$$S \subset E^c$$

Lo cual se demuestra por reducción al absurdo. Sea $p \in S$ y supongamos que $p \notin E^c$. Por el teorema de separación existe un hiperplano H que pasa por p definido por una ecuación del tipo

$$x \cdot n = c$$

tal que

$$x \cdot n > c$$

para todo $x \in E^c$.

Sea $L : E^n \rightarrow \mathbb{R}$, una transformación lineal con regla de correspondencia

$$L(x) = x \cdot n$$

entonces

$$L(p) = c$$

porque p pertenece al hiperplano H , y además

$$L(p) \not\subseteq L(E^c)$$

porque si ocurriera lo contrario, entonces P pertenecería a E^c , contrariamente a la suposición.

$L(S)$ es un conjunto cerrado, acotado y convexo, por ser imagen, bajo una transformación lineal, de S , el cual es cerrado, acotado y convexo. De ahí que $L(S)$ es un intervalo cerrado de \mathbb{R} , digamos $[a, b]$, que además contiene a c .

Sea H_a un hiperplano definido por la ecuación

$$x \cdot n = a$$

de la nota anterior se sabe que H_a es un hiperplano de soporte de S , de ahí que contenga un punto extremo de S , es decir un punto de E^c . Obtenemos entonces una contradicción del hecho de que

$$x \cdot n > c \geq a$$

para todo x en E^c . Por tanto se cumple el teorema de Krein-Milman. ■

Exercise 93 *Demostrar que el conjunto solución del sistema de ecuaciones lineales*

$$Ax = b$$

con b un vector columna de E^n y A una matriz real de dimensión $n \times n$, es un conjunto convexo en E^n .

4.12 Funciones cóncavas y convexas

Definition 94 *Una función f definida en un conjunto convexo S del espacio euclideo de dimensión n , se llama convexa si para cualquier par de puntos*

$$\begin{aligned} x &= (x_1, \dots, x_n) \\ y &= (y_1, \dots, y_n) \end{aligned}$$

de S se cumple

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

donde

$$0 \leq t \leq 1$$

Example 95 *La función cuadrática*

$$f(x) = x^2 - 4x + 5$$

es una función convexa porque la ordenada de cualquier punto de la parábola comprendido entre los puntos

$$\begin{aligned} (x, f(x)) \\ (y, f(y)) \end{aligned}$$

es menor o igual que la ordenada del punto del segmento, con esos extremos y con la misma abscisa.

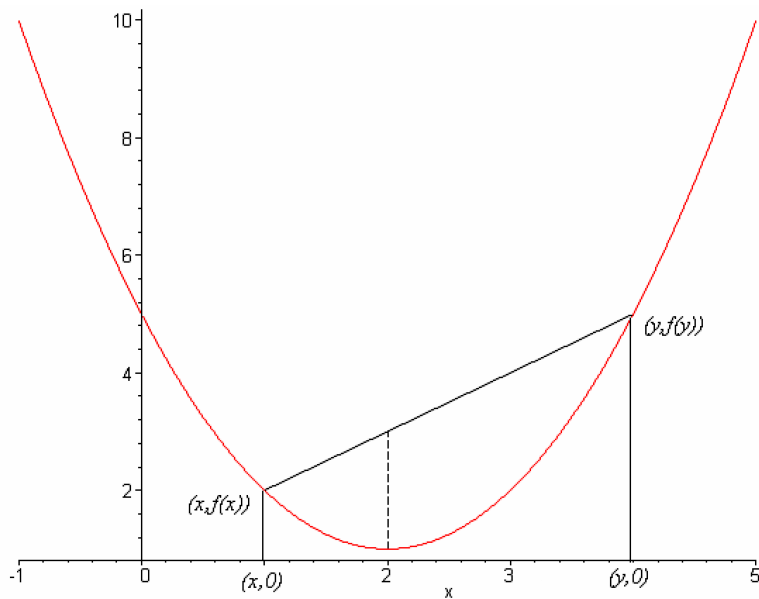


Figure 10

Definition 96 Una función f definida en un conjunto convexo S del espacio euclideo de dimensión n , se llama cóncava si para cualquier par de puntos

$$x = (x_1, \dots, x_n)$$

$$y = (y_1, \dots, y_n)$$

S se cumple

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

donde

$$0 \leq t \leq 1$$

Example 97 La función cuadrática

$$f(x) = -x^2 + 4x - 5$$

es una función cóncava porque la ordenada de cualquier punto de la parábola comprendido entre los puntos

$$(x, f(x))$$

$$(y, f(y))$$

es mayor o igual que la ordenada del punto del segmento, con esos extremos y con la misma abscisa.

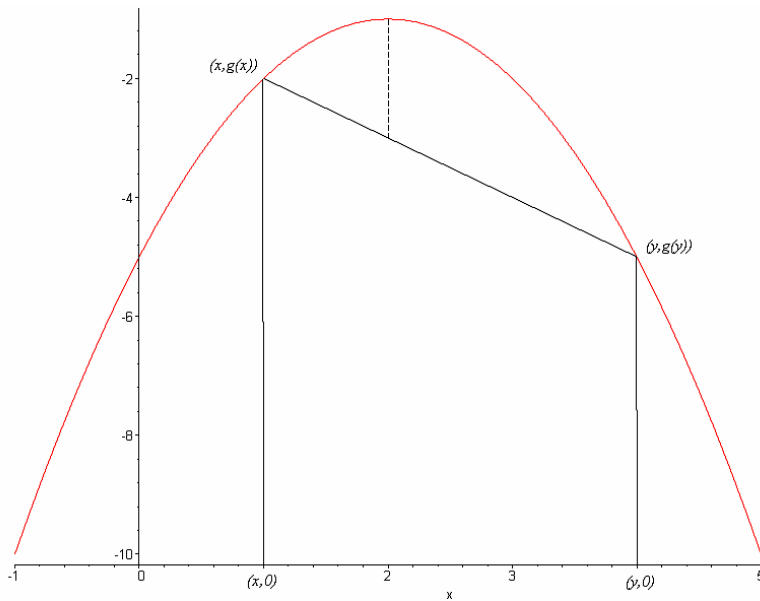


Figure 11

Part III

PROGRAMACIÓN LINEAL⁷

Los administradores usan modelos para resolver y entender problemas de decisión en los negocios. En general, pretenden optimizar algún objetivo, como las utilidades y los costos. En esas situaciones problemáticas, existen factores sujetos a control del administrador denominados variables de decisión. También, existen factores que limitan o restringen las acciones que puede realizar, como las restricciones en la capacidad de la planta y el equipo, o límites en la demanda, el procesamiento o el envío de unidades.

Un modelo matemático es una representación simplificada del problema de decisión en el cual las variables de decisión, el objetivo y las restricciones se representan por símbolos y ecuaciones. Un modelo de programación lineal (PL) es un tipo particular de modelo matemático en el cual las relaciones entre las variables son lineales y sólo se considera una medida de desempeño u objetivo. La ventaja de este modelo es que existe una técnica matemática denominada programación lineal que determina la mejor u óptima decisión, aunque se tengan miles de variables y relaciones.

Considérense las variables de decisión

$$x_1, x_2, x_3, \dots, x_n$$

El modelo de programación lineal se diseña para maximizar (o minimizar) una función objetivo

⁷Bonini, C.P., Asuman, W. H., y Bierman, H. Jr. (1997) Quantitative Analysis for Management. Chicago: McGraw Hill. IRWIN Series. Novena Edición, Capítulo 2.

de la forma

$$f = \sum_{i=1}^n c_i x_i$$

donde f es algún objetivo económico como las utilidades, producción, costos, semanas laborables o toneladas suministradas. Los criterios de optimización generalmente son escoger los valores de las variables de decisión que produzcan mayores utilidades o menores costos.

Los coeficientes de la función objetivo son constantes y las variables de decisión tienen exponente uno, por lo que la función f es una combinación lineal de las variables de decisión.

En general, el administrador no puede seleccionar los valores de las variables de decisión arbitrariamente, la selección está limitada por un conjunto de relaciones o restricciones en términos de las variables de decisión

$$\sum_{j=1}^n a_{ij} x_j \leq b_i$$

para $i = 1, 2, \dots, m$. Los coeficientes a_{ij} son constantes y la constante b_i restringe el valor de la función objetivo f puesto que restringe los valores de las variables de decisión

$$x_1, x_2, x_3, \dots, x_n$$

mediante desigualdades o igualdades.

La solución obtenida por medio de la programación lineal es el conjunto de valores de las variables de decisión que general el valor óptimo de la función objetivo (máximo o mínimo) dentro de las restricciones.

La programación lineal no permite incertidumbre en ninguna de las relaciones ni variables aleatorias, de ahí que el problema de maximizar la función objetivo sujeta a las restricciones es conceptualmente simple. La programación lineal hace posible manejar problemas con gran cantidad de restricciones, de manera ordenada. Esta técnica es excepcionalmente fuerte y general, es aplicable a gran variedad de problemas en los negocios y puede manejarse de manera rutinaria con la ayuda de computadoras personales.

4.13 Análisis marginal

Un ejemplo simple de un modelo de programación lineal, que puede resolverse por sentido común o por un *análisis marginal* es el siguiente. Supongamos que la utilidad marginal de un producto A es 5 UM por unidad y de un producto B es 2 UM por unidad; que se puede vender todo lo que se produzca de ambos productos y además, que ambos productos se producen con el mismo equipo.

Se quiere determinar qué producto se va a producir una vez que encontremos la capacidad de las instalaciones. Por ejemplo, el equipo puede producir 100 unidades del producto A o 600 unidades del B, por día.

En este ejemplo simple, obviamente el producto B es más deseable que el A, puesto 1,200 UM de utilidades diarias es mejor que 500 UM.

Supongamos ahora que la empresa es capaz de producir 20 distintos productos en 15 departamentos diferentes y que cada producto requiere distinto tiempo de producción en cada departamento. Si la dificultad del problema no es impresionante, supongamos que cada departamento realiza 10 procesos y que cada producto requiere distinto tiempo de producción en cada proceso ¿Cómo determinar la combinación de producción óptima? Un problema de este tipo, se resuelve mejor por

medio de programación lineal. El término lineal es apropiado puesto que todas las utilidades y las relaciones de producción se suponen proporcionales o lineales; es decir, que el grado máximo de cualquier variable es 1 y que ninguna de las variables se multiplica por alguna otra.

El método simplex es el procedimiento más común para resolver problemas de programación lineal y se le incluye en la mayoría de las hojas de cálculo para computadora.

4.14 Formulación de problemas de programación lineal

Antes de resolver problemas de programación lineal es importante aprender cómo definir variables y formular ecuaciones, esto es, cómo plantear un problema de negocios en la forma de maximizar (o minimizar) una función lineal, sujeta a restricciones lineales. Usamos el término formulación para traducir un problema del mundo real en relaciones matemáticas. La formulación es la parte que ofrece mayores retos al analizar problemas de negocios.

Example 98 *Un problema de combinación de productos (plan de producción). Una empresa de manufactura produce dos productos, A y B. Cada uno de estos productos se procesa en dos máquinas diferentes. Una máquina tiene una capacidad disponible de 24 horas, la segunda máquina tiene una capacidad de 16 horas. El producto A requiere de dos horas de tiempo en cada máquina, mientras que el producto B requiere de 3 horas en la primera máquina y una hora en la segunda. La utilidad marginal del producto A es 6 UM por unidad (el incremento en la utilidad por unidad producida) y del producto B es de 7 UM por unidad. La empresa puede vender tantas unidades de ambos productos como produzca.*

El objetivo de la empresa es maximizar las utilidades. El problema es determinar el número de unidades de cada producto, A y B, deben producirse dentro de los límites de capacidad disponible de cada máquina.

Formulación: Definición de variables: sean x_1 = la cantidad de unidades del producto A a producirse, x_2 = la cantidad de unidades del producto B a producirse y P el incremento total de utilidades de la empresa. La función objetivo es

$$P = 6x_1 + 7x_2$$

La empresa desea maximizar el valor de P , cuyo valor se incrementa conforme se incrementa el valor de las variables de decisión x_1 , x_2 (6 UM por el número de unidades vendidas del producto A más 7 UM por el número de unidades vendidas del producto B).

La primera restricción relaciona la disponibilidad de tiempo de la primera máquina, que se expresa como

$$2x_1 + 3x_2 \leq 24$$

es decir, cada unidad del producto A usa dos horas de la primera máquina y 3 horas para el producto B; de ahí que el total de horas utilizadas en la primera máquina se expresa como el lado izquierdo de la desigualdad, mientras que el nivel disponible de horas (la capacidad) es el lado derecho de dicha desigualdad.

Para la segunda máquina, la restricción es similar

$$2x_1 + x_2 \leq 16$$

Finalmente, la formulación de un problema de programación lineal como éste, incluye la restricción de que las variables de decisión x_1 y x_2 toman valores no-negativos. Esto significa, en términos del problema, que la empresa produce cantidades positivas o cero de ambos productos.

En resumen, la formulación completa del modelo de PL para el problema de combinación de productos es

Maximizar:

$$P = 6x_1 + 7x_2$$

Sujeta a:

$$2x_1 + 3x_2 \leq 24$$

$$2x_1 + x_2 \leq 16$$

$$x_1, x_2 \geq 0$$

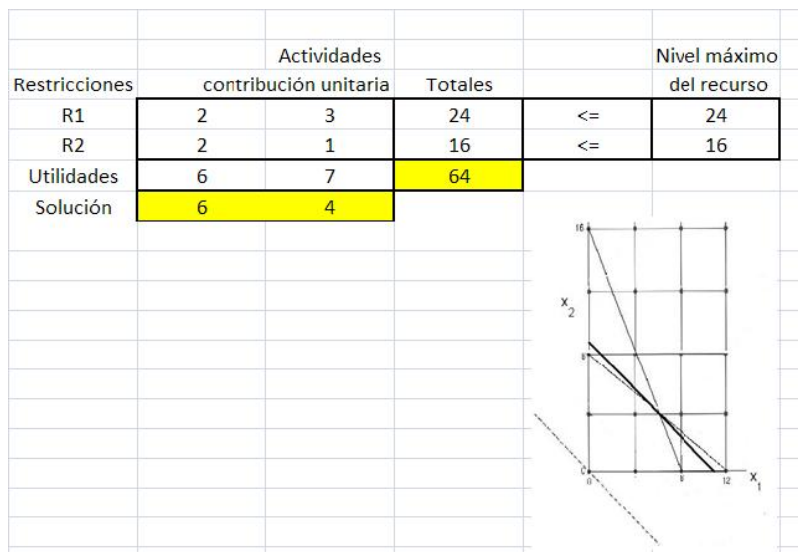


Figure 12

Solution 99

Example 100 *Un problema de transporte. Un productor de jabón y detergentes tiene tres plantas localizadas en las ciudades de Cincinnati, Denver y Atlanta. Los principales almacenes están localizados en Nueva York, Boston, Chicago, Los Ángeles y Dallas. La preocupación de la empresa es saber qué fábrica debe suministrar a qué almacén. Los requerimientos de ventas para el siguiente año de cada almacén están en la tabla 1 y los de los costos de envíos en la tabla 2. siguientes. La capacidad de cada fábrica en cada ubicación es limitada. Cincinnati tiene una capacidad anual de 100,000 piezas, Denver de 60,000 y Atlanta de 50,000.*

La empresa desea determinar un plan de envío (o suministro) que minimice los costos totales

C, de envío.

Tabla 1		Requerimientos
Ubicación del almacén	Ventas anuales (miles de piezas)	
Nueva York	50	
Boston	10	
Chicago	60	
Los Ángeles	30	
Dallas	20	
Total	170	

Tabla 2		Costos de envío (por miles de piezas)				
De	a	Nueva York	Boston	Chicago	Los Ángeles	Dallas
Cincinnati		240	300	160	500	360
Denver		420	440	300	200	220
Atlanta		300	340	300	480	400

Formulación: Definición de variables: denotemos a las variables de decisión con dos subíndices, el primero para indicar el origen (la fábrica) y el segundo para indicar el destino (el almacén). Por ejemplo, x_{11} = el número de piezas enviadas de la primera fábrica (Cincinnati) al primer almacén (Nueva York), etc. De manera general tendremos que x_{ij} es el número de piezas enviadas del origen i (i -ésima planta) al destino j (j -ésimo almacén), donde $i = 1, 2, 3$ y $j = 1, 2, 3, 4, 5$.

Entonces el objetivo es minimizar la función costo total de enviar los productos por cada una de las rutas posibles (de una fábrica a un almacén):

$$C = 240x_{11} + 300x_{12} + 160x_{13} + 500x_{14} + 360x_{15} + 420x_{21} + 440x_{22} + 300x_{23} + 200x_{24} + 220x_{25} + 300x_{31} + 340x_{32} + 300x_{33} + 480x_{34} + 400x_{35}$$

Existen dos conjuntos (tipos) de restricciones, de demanda y capacidad. Las restricciones que indican que la demanda se satisface son:

$$\begin{aligned} \text{Nueva York} & \sum_{i=1}^3 x_{i1} = 50 \\ \text{Boston} & \sum_{i=1}^3 x_{i2} = 10 \\ \text{Chicago} & \sum_{i=1}^3 x_{i3} = 60 \\ \text{Los Ángeles} & \sum_{i=1}^3 x_{i4} = 30 \\ \text{Dallas} & \sum_{i=1}^3 x_{i5} = 20 \end{aligned}$$

Las restricciones que indican el nivel máximo de la capacidad de cada fábrica son:

$$\begin{aligned} \text{Cincinnati} & \sum_{j=1}^5 x_{1j} \leq 100 \\ \text{Denver} & \sum_{j=1}^5 x_{2j} \leq 60 \\ \text{Atlanta} & \sum_{j=1}^5 x_{3j} \leq 50 \end{aligned}$$

Finalmente, cada variable de decisión x_{ij} debe ser no negativa. La solución de este problema dará el plan óptimo de distribución (transporte) para la empresa.

En resumen, la formulación completa del modelo de PL para el problema de transporte es

Minimizar:

$$C = 240x_{11} + 300x_{12} + 160x_{13} + 500x_{14} + 360x_{15} + 420x_{21} + 440x_{22} + 300x_{23} + 200x_{24} + 220x_{25} + 300x_{31} + 340x_{32} + 300x_{33} + 480x_{34} + 400x_{35}$$

Sujeta a:

$$\begin{aligned} \text{Restricciones} \quad & \sum_{i=1}^3 x_{i1} = 50 \\ & \sum_{i=1}^3 x_{i2} = 10 \\ \text{de} \quad & \sum_{i=1}^3 x_{i3} = 60 \\ \text{demanda} \quad & \sum_{i=1}^3 x_{i4} = 30 \\ & \sum_{i=1}^3 x_{i5} = 20 \\ \\ \text{Restricciones} \quad & \sum_{j=1}^5 x_{1j} \leq 100 \\ \text{de} \quad & \sum_{j=1}^5 x_{2j} \leq 60 \\ \text{capacidad} \quad & \sum_{j=1}^5 x_{3j} \leq 50 \end{aligned}$$

$$\begin{aligned} \text{Restricciones de} \\ \text{no-negatividad} \quad & x_{ij} \geq 0 \end{aligned}$$

Example 101 *Un problema de mezclas. Varios grados de gasolina se obtienen de combinar ciertas mezclas de gasolinas que salen directamente de las operaciones de refinación. En una operación de refinación, hay muchas mezclas de gasolinas, muchas gasolinas como producto final (por ejemplo, varios grados de gasolina para avión y motores) y muchas características que se consideran importantes en la composición química de varios grados de gasolina (incluyendo, por ejemplo grado de octanaje, presión del vapor, contenido de azufre y contenido de goma). En este ejemplo simplificado, se asume que una refinería tiene disponible solo dos tipos de mezcla de gasolinas, cuyas características se muestran en la tabla 3.*

Tabla 3 Características de las mezclas de gasolinas

<i>Combinaciones disponibles</i>	<i>Grado de octanaje</i>	<i>Presión del vapor</i>	<i>Disponibilidad</i>
<i>Mezcla tipo 1</i>	<i>104</i>	<i>5</i>	<i>30,000</i>
<i>Mezcla tipo 2</i>	<i>94</i>	<i>9</i>	<i>70,000</i>

Estas mezclas de gasolinas pueden combinarse para producir dos productos finales, las gasolinas para aviación y para motores. Las características requeridas para estos productos finales se muestran en la tabla 4.

Tabla 4 Características de los productos finales de gasolinas

	<i>Grado de octanaje</i>	<i>Presión del vapor</i>	<i>Niveles de venta</i>	<i>Precio de venta</i>
<i>Productos finales</i>	<i>mínimo</i>	<i>máxima</i>	<i>máximos</i>	<i>(U.M. por barril)</i>
<i>Gasolina para avión</i>	<i>102</i>	<i>6</i>	<i>20,000</i>	<i>45.10</i>
<i>Gasolina para motor</i>	<i>96</i>	<i>8</i>	<i>Cualquier cantidad</i>	<i>32.40</i>

Cuando las gasolinas se combinan, la mezcla resultante tiene un octanaje y una presión de vapor en proporción al volumen de cada gasolina. Por ejemplo, si 1,000 barriles de mezcla de gasolina 1 se combinan con 1,000 barriles de mezcla de gasolina 2, la gasolina resultante tendrá un grado de

octanaje de 99 porque:

$$\frac{1,000 \cdot 104 + 1,000 \cdot 94}{2,000} = 99$$

y una presión de vapor de 7 porque:

$$\frac{1,000 \cdot 5 + 1,000 \cdot 9}{2,000} = 7$$

La empresa desea maximizar los ingresos por la venta de las gasolinas finales.

Formulación: Denotamos a las variables de decisión como

x_1 = cantidad de barriles de la mezcla 1 usada en la gasolina para aviación.

x_2 = cantidad de barriles de la mezcla 2 usada en la gasolina para aviación.

x_3 = cantidad de barriles de la mezcla 1 usada en la gasolina para motores.

x_4 = cantidad de barriles de la mezcla 2 usada en la gasolina para motores.

Entonces, la función objetivo es R = ingresos totales y el problema trata de maximizar a R , donde

$$\begin{aligned} R &= 45.10(x_1 + x_2) + 32.40(x_3 + x_4) \\ &= 45.10x_1 + 45.10x_2 + 32.40x_3 + 32.40x_4 \end{aligned}$$

Una de las restricciones que afectan cómo la refinería combinará las gasolinas, trata sobre las ventas o la demanda, el hecho de que no más de 20,000 barriles de gasolina para aviación pueden venderse (ver tabla 4). Lo anterior, se representa mediante la expresión

$$x_1 + x_2 \leq 20,000$$

Un segundo conjunto de restricciones se relaciona con la disponibilidad de las mezclas de gasolinas (ver tabla 3), lo cual se expresa con

$$x_1 + x_3 \leq 30,000$$

$$x_2 + x_4 \leq 70,000$$

Un tercer conjunto de restricciones se refiere a los grados de octanaje del producto final. Recordemos que la cantidad total de gasolina para aviación es

$$x_1 + x_2$$

y su octanaje se determina con la fórmula

$$\text{Grado de octanaje gasolina avión} = \frac{104x_1 + 94x_2}{x_1 + x_2}$$

Este grado debe ser cuando menos 102, así que la restricción se expresa como

$$\frac{104x_1 + 94x_2}{x_1 + x_2} \geq 102$$

Reordenamos la expresión como una restricción lineal como sigue:

$$\begin{aligned} 104x_1 + 94x_2 &\geq 102(x_1 + x_2) \\ 2x_1 - 8x_2 &\geq 0 \end{aligned}$$

Similarmente, el grado de octanaje para el motor de gasolina produce la restricción:

$$\begin{aligned} 104x_3 + 94x_4 &\geq 96(x_3 + x_4) \\ 8x_3 - 2x_4 &\geq 0 \end{aligned}$$

Un cuarto conjunto de restricciones se refiere a los requerimientos de presión de vapor del producto final. Para la gasolina de aviación, se tiene

$$\begin{aligned} 5x_1 + 9x_2 &\leq 6(x_1 + x_2) \\ -x_1 + 3x_2 &\leq 0 \end{aligned}$$

Mientras que los requerimientos de presión de vapor para la gasolina de motor es

$$\begin{aligned} 5x_3 + 9x_4 &\leq 8(x_3 + x_4) \\ -3x_3 + x_4 &\leq 0 \end{aligned}$$

En resumen, la formulación completa del modelo de PL para el problema de mezclas es

Maximizar:	$R = 45.10x_1 + 45.10x_2 + 32.40x_3 + 32.40x_4$
Sujeta a:	
Restricciones de demanda	$x_1 + x_2 \leq 20,000$
Restricciones de disponibilidad	$x_1 + x_3 \leq 30,000$ $x_2 + x_4 \leq 70,000$
Restricciones de grado de octanaje	$2x_1 - 8x_2 \geq 0$ $8x_3 - 2x_4 \geq 0$
Restricciones de presión de vapor	$-x_1 + 3x_2 \leq 0$ $-3x_3 + x_4 \leq 0$
Restricciones de no-negatividad	$x_1, x_2, x_3, x_4 \geq 0$

Mezclas de gasolinas fue una de las aplicaciones iniciales de la programación lineal a los problemas de negocios. Este ejemplo es una presentación del problema real muy simplificada, pero muestra los elementos esenciales. Este tipo de problema surge en muchos contextos como la preparación de alimentos para animales, en los que se requiere satisfacer requerimientos mínimos de nutrientes.

Los tres problemas anteriores se plantean en un periodo de tiempo fijo y por ello se les llama problemas estáticos. La programación lineal también se ha aplicado a problemas dinámicos, aquellos que se extienden sobre varios periodos de tiempo como el siguiente ejemplo.

Example 102 *Un problema de programación de actividades. Una empresa firma un contrato de suministro de un producto durante los siguientes seis meses. Los costos de producción varían mensualmente debido a cambios anticipados en los costos de los materiales. La capacidad de producción de la empresa es de 100 unidades mensuales en tiempo normal y hasta 15 unidades adicionales mensuales con tiempo extra.*

La siguiente tabla contiene los costos mensuales de los requerimientos de suministro y producción. El costo de mantener almacenada una unidad no vendida es de 2 UM por mes. El problema de la empresa es determinar el número de unidades a producir mensualmente en tiempo normal y en tiempo extra, para cumplir con los requerimientos al mínimo costo. La empresa no tiene unidades disponibles al inicio del primer mes y no desea tener unidades sobrantes al final del sexto mes.

Formulación: Las variables de decisión e intermedias son las siguientes:

X_i = número de unidades producidas en tiempo normal en el i -ésimo mes

Y_i = número de unidades producidas en tiempo extra en el i -ésimo mes

I_i = número de unidades en existencia (no vendidas), al final del i -ésimo mes

con $i = 1, 2, 3, 4, 5, 6$

El objetivo es minimizar el costo total por producir en tiempo normal, en tiempo extra y por almacenar.

Minimizar:

$$C = 30(X_1 + X_2) + 32(X_3 + X_4) + 31X_5 + 32X_6 + 35(Y_1 + Y_2) + 37(Y_3 + Y_4) + 36Y_5 + 37Y_6 + 2(I_1 + I_2 + I_3 + I_4 + I_5 + I_6)$$

Sujeto a las restricciones de producción en tiempo normal, en tiempo extra, de balance o enlace para asegurar que se cumplan los requerimientos de suministro, de inventario final nulo y de no-negatividad.

Las restricciones de producción en tiempo normal se expresan como

$$X_i \leq 100$$

donde $i = 1, 2, 3, 4, 5, 6$.

Las restricciones de producción en tiempo extra como

$$Y_i \leq 15$$

con $i = 1, 2, 3, 4, 5, 6$.

Las restricciones de balance o enlace se necesitan para enlazar los periodos de tiempo y asegurar que los compromisos se cumplan. Estas restricciones tienen la forma

$$\text{unidades en existencia} = \text{gasto de unidades}$$

ó

$$\left(\begin{array}{c} \text{inventario} \\ \text{inicial} \end{array} \right) + \left(\begin{array}{c} \text{producción} \\ \text{tiempo normal} \end{array} \right) + \left(\begin{array}{c} \text{producción} \\ \text{tiempo extra} \end{array} \right) = \left(\begin{array}{c} \text{compromiso} \\ \text{suministro} \end{array} \right) + \left(\begin{array}{c} \text{inventario} \\ \text{final} \end{array} \right)$$

Para el mes 1, la relación que se obtiene es

$$0 + X_1 + Y_1 = 95 + I_1$$

puesto que no hay inventario inicial. Reordenando, se obtiene la ecuación

$$0 + X_1 + Y_1 - I_1 = 95$$

Similarmente, para el mes 2

$$I_1 + X_2 + Y_2 - I_2 = 85$$

El mes 3

$$I_2 + X_3 + Y_3 - I_3 = 110$$

El mes 4

$$I_3 + X_4 + Y_4 - I_4 = 115$$

El mes 5

$$I_4 + X_5 + Y_5 - I_5 = 90$$

El mes 6

$$I_5 + X_6 + Y_6 - I_6 = 105$$

Finalmente, como el inventario final debe ser cero, la última restricción es

$$I_6 = 0$$

En resumen, la formulación completa del modelo de PL para el problema de programación de actividades es

Minimizar:

$$C = 30(X_1 + X_2) + 32(X_3 + X_4) + 31X_5 + 32X_6 + 35(Y_1 + Y_2) + 37(Y_3 + Y_4) + 36Y_5 + 37Y_6 + 2(I_1 + I_2 + I_3 + I_4 + I_5 + I_6)$$

Sujeta a:

	$0 + X_1 + Y_1 - I_1 = 95$
	$I_1 + X_1 + Y_1 - I_2 = 85$
restricciones de	$I_2 + X_1 + Y_1 - I_3 = 110$
balance de inventario	$I_3 + X_1 + Y_1 - I_4 = 115$
	$I_4 + X_1 + Y_1 - I_5 = 90$
	$I_5 + X_1 + Y_1 - I_6 = 105$
restricciones de producción en tiempo normal	$X_i \leq 100$, para toda $i = 1, 2, 3, 4, 5, 6$.
restricciones de producción en tiempo extra	$Y_i \leq 15$, para toda $i = 1, 2, 3, 4, 5, 6$.
restricciones sobre el inventario final	$I_6 = 0$
restricciones de no-negatividad	$X_i, Y_i, I_i \geq 0$, para toda $i = 1, 2, 3, 4, 5, 6$.

4.15 Forma estándar de un problema de programación lineal

	Minimizar:	$c_1x_1 + \dots + c_nx_n$
	Sujeta a:	$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1$
		$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2$
		\vdots
		$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m$
	y	$x_1, x_2, \dots, x_n \geq 0$
Forma matricial		
	Minimizar:	$c^T x$
	Sujeta a:	$Ax = b$
	y	$x \geq 0$

Los problemas de programación lineal son susceptibles de escribirse en la forma estándar mediante el uso de variables de holgura, de excedencia y variables libres, como se muestra en la tabla siguiente:

Problema PL original	Variables	Problema PL estándar
Maximizar: $c^T x$ Sujeta a: $Ax \leq b$ y $x \geq 0$	De holgura $y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$	Maximizar: $c^T x$ Sujeta a: $[A e_j y_j] x = b$ $x \geq 0, y \geq 0$ y con e_j vector unitario cuya j-ésima componente es 1, y las restantes, cero
Minimizar: $c^T x$ Sujeta a: $Ax \geq b$ y $x \geq 0$	De excedencia $y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$	Minimizar: $c^T x$ Sujeta a: $[A -e_j y_j] x = b$ $x \geq 0, y \geq 0$ y Minimizar: $c^T x$ Sujeta a: $Ax \leq b$ $x \geq 0$ y con $A_{(m-1) \times (n-1)}$
Minimizar: $c^T x$ Sujeta a: $Ax \geq b$ y para $j = 2, \dots, n$	Libres x_1 Supongamos que $a_{i1} \neq 0$. Expresamos a x_1 en función de las restantes variables usando la i-ésima ecuación. Sustituimos en las restantes ecuaciones la expresión obtenida para x_1 .	y con $x = \begin{bmatrix} x_2 \\ \vdots \\ x_n \end{bmatrix}$

5 Teorema fundamental de la Programación Lineal

Dado un problema de programación lineal en la forma estándar

$$\begin{array}{ll} \text{Minimizar:} & c^T x \\ \text{Sujeta a:} & Ax = b \\ \text{y} & x \geq 0 \end{array}$$

Donde A es una matriz de dimensión $m \times n$ y rango m .

1. Si hay una solución factible, hay una solución factible básica.
2. Si hay una solución factible óptima, hay una solución factible básica óptima.

Proof. De la propiedad 1. Sean

$$A^1, A^2, \dots, A^n$$

las n columnas de A .

Supóngase que

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

es una solución factible, entonces esa solución satisface

$$A^1 x_1 + A^2 x_2 + \dots + A^n x_n = b$$

Supóngase que exactamente p de las variables x_j son positivas, con $0 \leq p \leq n$. Por conveniencia, supongamos que son las primeras p variables. Entonces, se cumple

$$\begin{array}{l} A^1 x_1 + A^2 x_2 + \dots + A^p x_p = b \\ \text{y} \\ x_{p+1} = x_{p+2} = \dots = x_n = 0 \end{array}$$

Respecto a la independencia lineal de las p columnas hay dos posibilidades, son linealmente independientes o linealmente dependientes.

Caso 1. Supóngase que A^1, A^2, \dots, A^n son linealmente independientes. Entonces, es claro que $p \leq n$, dado que m es el rango de A . Si $p = m$, la solución factible es básica y se completa la demostración. Si $p < m$, se pueden hallar $m - p$ columnas de las restantes $n - p$, de modo que el conjunto resultante con m columnas sea linealmente independiente, puesto que el rango de A es m . Por tanto, asignando el valor de cero a esas $m - p$ variables se obtiene una solución factible básica (degenerada).

Caso 2. Supóngase que A^1, A^2, \dots, A^n , son linealmente dependientes. Entonces existe una combinación lineal no trivial de esas columnas que es cero. Sean y_1, y_2, \dots, y_p , constantes no todas nulas, y al menos una positiva, tales que

$$A^1 y_1 + A^2 y_2 + \dots + A^p y_p = 0$$

De esta igualdad y del hecho de que x es solución factible, se obtiene la ecuación

$$(x_1 - \varepsilon y_1) A^1 + (x_2 - \varepsilon y_2) A^2 + \dots + (x_p - \varepsilon y_p) A^p = b$$

con ε un número real. Dependiendo del valor de ε , se puede violar la desigualdad

$$x_j - \varepsilon y_j \geq 0$$

Además, haciendo

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_p \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

El n -vector $x - \varepsilon y$ es solución del sistema $Ax = b$.

Si $\varepsilon = 0$, entonces se obtiene la solución factible original. Si ε aumenta desde cero, esto es toma valores positivos cada vez mayores, las componentes de $x - \varepsilon y$ aumentarán, disminuirán o permanecerán constantes en la medida en que la y_j correspondiente sea negativa, positiva o cero, respectivamente.

Como al menos una componente y_j es positiva, por lo menos una componente de $x - \varepsilon y$ decrecerá conforme ε aumente su valor. Si se incrementa el valor de ε hasta el punto donde una o más componentes de $x - \varepsilon y$ sean cero, definimos

$$\varepsilon_0 = \min \left\{ \frac{x_j}{y_j} \mid y_j > 0 \right\}$$

Para este valor, $x - \varepsilon_0 y$ es una solución factible y tiene a lo más $p - 1$ variables positivas.

Al repetir este proceso, en caso necesario, se podrían eliminar las variables positivas hasta tener una solución factible con las columnas correspondientes que son linealmente independientes y en este punto aplicar el caso 1. ■

Proof. De la propiedad 2. Sea

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

una solución factible óptima.

Como en la demostración anterior, supóngase que hay exactamente p variables positivas. De nuevo se tienen dos casos respecto a la independencia lineal, como antes.

En el caso 2, la parte de que existe una solución factible básica es igual. Con respecto a que sea óptima, para cada valor de ε se concluye al observar que el valor de la función objetivo en esa solución es

$$c^T x - \varepsilon c^T y$$

Para una ε de magnitud suficientemente pequeña

$$x - \varepsilon y$$

es solución factible para valores tanto positivos como negativos, por lo que

$$c^T y = 0$$

ya que, en caso contrario, se violaría el hecho de que x es solución óptima, escogiendo los signos apropiados para ε

Una vez establecido que la nueva solución factible, con menos componentes positivos, también es óptima, el resto de la demostración se establece como en la parte 1. ■

6 ALGORITMO SIMPLEX

El enfoque general del método simplex es obtener una sucesión de soluciones factibles básicas que mejoran el valor de la función objetivo hasta que alcanza una solución óptima.

6.1 Resolviendo el problema para una solución factible básica

Consideremos el sistema de ecuaciones $Ax = b$. Supongamos que x es un vector con n componentes, b un vector con m componentes y A una matriz de dimensión $m \times n$. Por facilidad supondremos que $m < n$. De las columnas de A se seleccionan un conjunto de m columnas linealmente independientes con las que se forma una matriz B . Estas existen si el rango de A es m . Entonces, la matriz B es no singular y la ecuación $Bx = b$ tiene solución única, digamos x_B , multiplicando por la izquierda se tiene

$$\begin{aligned} B^{-1}Bx_B &= B^{-1}b \\ x_B &= B^{-1}b \end{aligned}$$

Construimos ahora una solución a partir de la solución anterior, agregando $n - m$ componentes a la solución anterior, todas ellas iguales a cero

$$x = \begin{bmatrix} x_B \\ 0 \end{bmatrix}$$

Este vector, claramente satisface $Ax = b$. Además, sea c_B el vector cuyas componentes son los coeficientes de la función objetivo (incluyendo ceros para las variables de holgura), de los elementos correspondientes de x_B . El valor de la función objetivo para esta solución básica es

$$c_B x_B = c_B B^{-1}b$$

Definition 103 Para el sistema $Ax = b$ con m ecuaciones lineales simultáneas en n indeterminadas, sea B cualquier submatriz de A , de dimensión $m \times m$, no singular. Entonces, si todas las $n - m$ componentes de x no asociadas a las columnas de B se igualan a cero, la solución del conjunto resultante de ecuaciones se llama solución básica de sistema, respecto a la base B . Las componentes de x asociadas a las columnas de B se denominan variables básicas.

Remark 104 B es una base del espacio E^m , porque es una matriz formada con m columnas linealmente independientes. De ahí que, el vector $b \in E^m$, se expresa como una combinación lineal de las columnas de B .

Remark 105 Para evitar redundancia o restricciones contradictorias, se asume la hipótesis de rango completo que dice que las m columnas de A , matriz de dimensión $m \times n$, son linealmente independientes.

Conforme a la hipótesis de rango completo, el sistema $Ax = b$ tiene siempre, al menos, una solución básica. Si alguna o algunas variables básicas de la solución básica son nulas, la solución se denomina solución básica degenerada.

En el caso de una solución básica no degenerada, tanto las variables básicas como la base B , se identifican por medio de sus componentes positivas. En cambio, en el caso de una solución básica degenerada, hay ambigüedad, porque se pueden intercambiar las variables básicas y no básicas de valor cero.

Consideremos el problema de programación lineal siguiente:

$$\begin{aligned} \text{Maximizar: } & z = c_1x_1 + c_2x_2 + \dots + c_nx_n \\ \text{Sujeta a:} & \\ & a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \leq b_1 \\ & a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \leq b_2 \\ & \vdots \\ & a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \leq b_m \end{aligned}$$

donde todos los lados derechos de las restricciones de desigualdad son positivos.

Introducimos variables de holgura no negativas para obtener el problema con restricciones de igualdad.

El problema consiste en

$$\begin{aligned} \text{Maximizar: } & z - c_1x_1 - c_2x_2 - \dots - c_nx_n = 0 \\ \text{Sujeta a:} & \\ & a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + x_{n+1} = b_1 \\ & a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n + 0 + x_{n+2} = b_2 \\ & \vdots \\ & a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n + 0 + \dots + 0 + x_{n+m} = b_m \\ \text{con} & x_{n+1}, \dots, x_{n+m} \geq 0 \end{aligned}$$

Construimos una tabla estándar (standard simplex tableau) con los coeficientes de las variables no-básicas x_1, \dots, x_n y las básicas x_{n+1}, \dots, x_{n+m} .

	x_1	x_2	\dots	x_n	x_{n+1}	x_{n+2}	\dots	x_{n+m}	LD
z	$-c_1$	$-c_2$	\dots	$-c_n$	0	0	\dots	0	0
x_{n+1}	a_{11}	a_{12}	\dots	a_{1n}	1	0	\dots	0	b_1
x_{n+2}	a_{21}	a_{22}	\dots	a_{2n}	0	1	\dots	0	b_2
\vdots	\vdots	\vdots		\vdots	\vdots	\vdots		\vdots	\vdots
x_{n+m}	a_{m1}	a_{m2}	\dots		0	0	\dots	1	b_m

Cualquier modificación que se haga durante el algoritmo, dejará una matriz de dimensión $m \times n$.

El procedimiento consiste en hacer operaciones elementales con las filas, siguiendo los siguientes cinco pasos básicos.

1. Escoger el coeficiente de mayor valor absoluto, en el renglón de z , supongamos que es c_k . Si todas las entradas de este renglón son positivas, entonces el máximo se ha obtenido. En caso contrario, identificamos la columna k .

2. Evaluar los cocientes

$$\frac{b_i}{a_{ik}}$$

para todo coeficiente positivo a_{ik} con $i = 1, \dots, m$. Seleccionamos el menor de ellos, supongamos que es

$$\frac{b_l}{a_{lk}}$$

por lo que identificamos el renglón l .

3. Reemplazamos la variable básica x_{n+l} por la no-básica x_k en la columna izquierda, de las variables básicas, con lo que se hace un cambio de base.
4. En el renglón l dividimos por a_{lk} . Esto quiere decir que el pivote es el coeficiente que se encuentra en la intersección del renglón l y la columna k , identificados en los pasos 1 y 2.
5. Usar la eliminación gaussiana, restando a cada fila el múltiplo de la fila pivote que anula los coeficientes en la columna k , excepto al elemento lk , el cual es uno.

Los cinco pasos se repiten hasta que se logra el máximo, esto es, cuando no hay coeficientes negativos en el renglón de z .

Example 106 Una empresa produce dos tipos de tableros para circuitos $R1$ y $R2$, contruidos como se indica:

$R1$ contiene 3 resistencias, un condensador, 2 transistores y 2 inductancias.

$R2$ contiene 4 resistencias, 2 condensadores y 3 transistores.

Los recursos disponibles para la producción diaria son de 2400 resistencias, 900 condensadores, 1600 transistores y 1200 inductancias.

La empresa requiere saber cuántos tableros del tipo $R1$ y cuántos del tipo $R2$ debe producir para maximizar la utilidad total, si la utilidad por un tablero del tipo $R1$ es 5 UM y por un tablero $R2$, es de 9 UM.

Solution 107 Representamos mediante x_1 la cantidad desconocida de tableros del tipo $R1$ a producir, con x_2 la cantidad desconocida de tableros del tipo $R2$ a producir. El modelo de programación lineal que resuelve el problema de la empresa es

Maximizar: $z = 5x_1 + 9x_2$

Sujeta a:

$$3x_1 + 4x_2 \leq 2400$$

$$x_1 + 2x_2 \leq 900$$

$$2x_1 + 3x_2 \leq 1600$$

$$2x_1 \leq 1200$$

$$x_1, x_2 \geq 0$$

Introducimos las variables de holgura al construir la tabla inicial

	x_1	x_2	x_3	x_4	x_5	x_6	LD
z	-5	-9	0	0	0	0	0
x_3	3	4	1	0	0	0	2400
x_4	1	2	0	1	0	0	900
x_5	2	3	0	0	1	0	1600
x_6	2	0	0	0	0	1	1200

Paso 1. Seleccionamos la columna de x_2 , porque contiene el coeficiente de mayor valor absoluto, de los términos de la función objetivo z .

	x_1	x_2	x_3	x_4	x_5	x_6	LD
z	-5	-9	0	0	0	0	0
x_3	3	4	1	0	0	0	2400
x_4	1	2	0	1	0	0	900
x_5	2	3	0	0	1	0	1600
x_6	2	0	0	0	0	1	1200

Paso 2. Evaluamos los cocientes, dividiendo los lados derechos entre cada coeficiente no nulo. Seleccionamos el renglón que contenga al menor cociente. El coeficiente 2, es el pivote.

	x_1	x_2	x_3	x_4	x_5	x_6	LD	
z	-5	-9	0	0	0	0	0	
x_3	3	4	1	0	0	0	2400	$\frac{2400}{4} = 600$
x_4	1	2	0	1	0	0	900	$\frac{900}{2} = 450$
x_5	2	3	0	0	1	0	1600	$\frac{1600}{3} = 533.33$
x_6	2	0	0	0	0	1	1200	

Paso 3. Intercambiamos la variable no-básica x_2 , por la básica x_4 .

	x_1	x_2	x_3	x_4	x_5	x_6	LD
z	-5	-9	0	0	0	0	0
x_3	3	4	1	0	0	0	2400
x_2	1	2	0	1	0	0	900
x_5	2	3	0	0	1	0	1600
x_6	2	0	0	0	0	1	1200

Paso 4. Dividimos por 2 al renglón seleccionado.

	x_1	x_2	x_3	x_4	x_5	x_6	LD
z	-5	-9	0	0	0	0	0
x_3	3	4	1	0	0	0	2400
x_2	$\frac{1}{2}$	1	0	$\frac{1}{2}$	0	0	450
x_5	2	3	0	0	1	0	1600
x_6	2	0	0	0	0	1	1200

Paso 5. Mediante operaciones elementales, se anulan los coeficientes de la columna de x_2 , es decir, sumando o restando el múltiplo conveniente de la fila seleccionada a las restantes filas.

	x_1	x_2	x_3	x_4	x_5	x_6	LD
z	$-\frac{1}{2}$	0	0	$\frac{9}{2}$	0	0	4050
x_3	1	0	1	-2	0	0	600
x_2	$\frac{1}{2}$	1	0	$\frac{1}{2}$	0	0	450
x_5	$\frac{1}{2}$	0	0	$-\frac{3}{2}$	1	0	250
x_6	2	0	0	0	0	1	1200

Repetimos los cinco pasos de nuevo porque aún se tiene un coeficiente negativo en el renglón de z .

Pasos 1 y 2

	x_1	x_2	x_3	x_4	x_5	x_6	LD	
z	$-\frac{1}{2}$	0	0	$\frac{9}{2}$	0	0	4050	
x_3	1	0	1	-2	0	0	600	$\frac{600}{1} = 600$
x_2	$\frac{1}{2}$	1	0	$\frac{1}{2}$	0	0	450	$\frac{450}{1/2} = 900$
x_1	$\frac{1}{2}$	0	0	$-\frac{3}{2}$	1	0	250	$\frac{250}{1/2} = 500$
x_6	2	0	0	0	0	1	1200	$\frac{1200}{2} = 600$

Pasos 3, 4 y 5

	x_1	x_2	x_3	x_4	x_5	x_6	LD
z	0	0	0	3	1	0	4300
x_3	0	0	1	1	-2	0	100
x_2	0	1	0	2	-1	0	200
x_1	1	0	0	-3	2	0	500
x_6	0	0	0	6	-4	1	200

Por lo tanto, la solución óptima es

$$\begin{aligned}
 x_1 &= 500 \\
 x_2 &= 200 \\
 x_3 &= 100 \\
 x_4 &= 0 \\
 x_5 &= 0 \\
 x_6 &= 200
 \end{aligned}$$

y el valor óptimo es

$$z = 4300$$

Las variables de holgura, variables básicas que quedan en la columna de la izquierda, toman los valores 500, 200, 100 y 200. Mientras que las restantes variables de holgura, x_4 y x_5 , se anulan, para satisfacer la segunda y tercera restricciones. Finalmente, concluimos que el plan de producción que maximiza la utilidad total, a 4300 UM, es producir 500 tableros del tipo R1 y 200 tableros del tipo R2.

7 PROBLEMA DUAL Y ANÁLISIS DE SENSIBILIDAD

7.1 Origen del problema dual

La teoría de dualidad se basa principalmente en la comprensión de que en cada iteración del método simplex, los correspondientes coeficientes de las variables de holgura, revelan las operaciones elementales realizadas (a la matriz identidad), es decir, expresan la forma cómo se obtuvo cada ecuación.

La tabla de parámetros de ambos problemas, en la forma generalizada simétrica, es

		Primal							
		x_1	x_2	x_3	\dots	x_n		LD	
Dual	λ_1	a_{11}	a_{12}	a_{13}	\dots	a_{1n}	\leq	b_1	Coef de w a minimizar
	λ_2	a_{21}	a_{22}	a_{23}	\dots	a_{2n}	\leq	b_2	
	\vdots	\vdots	\vdots	\vdots	\dots	\vdots	\vdots	\vdots	
	λ_m	a_{m1}	a_{m2}	a_{m3}	\dots	a_{mn}	\leq	b_m	
		\geq	\geq	\geq	\dots	\geq			
	LD	c_1	c_2	c_3	\dots	c_n			Coef de z a maximizar

En particular, los problemas primal y dual, presentados en forma matricial, para el ejemplo de la empresa Wyndor Glass, son

		x_1	x_2			LD
λ_1	1	0	\leq	4		
λ_2	0	2	\leq	12		
λ_3	3	2	\leq	18		
		\geq	\geq			
	3	5				

7.1.1 Relaciones generales entre los problemas primal y dual

1. Los parámetros de una restricción funcional en uno de los problemas son los coeficientes de la variable en el otro problema.
2. Los coeficientes en la FO de un problema son los lados derechos de las restricciones funcionales para el otro problema.

La correspondencia directa entre las entidades de los dos problemas se expresa como sigue

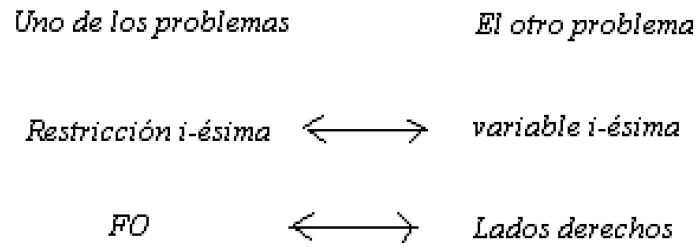


Figure 13

Estas relaciones son la clave para las aplicaciones de la teoría de dualidad, inclusive para el análisis de sensibilidad.

7.2 Formulación de problemas duales

Sean x y λ soluciones factibles de los respectivos problemas primal y dual.

Forma simétrica

Primal	Dual
Minimizar: $c^T x$	Maximizar: $\lambda^T b$
Sujeto a:	Sujeto a:
$Ax \geq b$	$\lambda^T A \leq c$
$x \geq 0$	$\lambda \geq 0$

Forma asimétrica

Primal	Dual
Minimizar: $c^T x$	Maximizar: $\lambda^T b$
Sujeto a:	Sujeto a:
$Ax = b$	$\lambda^T A \leq c$
$x \geq 0$	λ no restringido

7.2.1 Construcción del problema dual

Forma asimétrica Dado el problema primal:

$$\begin{aligned}
 &\text{Minimizar: } c^T x \\
 &\text{Sujeto a:} \\
 &\quad Ax = b \\
 &\quad x \geq 0
 \end{aligned}$$

Paso 1. Sustituir la igualdad $Ax = b$, por desigualdades

$$\begin{aligned} Ax &\leq b \\ -Ax &\leq -b \end{aligned}$$

Nuevo problema primal:

$$\begin{aligned} &\text{Minimizar: } c^T x \\ &\text{Sujeto a:} \\ &\quad Ax \leq b \\ &\quad -Ax \leq -b \\ &\quad x \geq 0 \end{aligned}$$

La matriz de coeficientes del nuevo sistema de restricciones es

$$\begin{bmatrix} A \\ -A \end{bmatrix}$$

El correspondiente vector dual es

$$\begin{bmatrix} u \\ -v \end{bmatrix}$$

con las restricciones de no negatividad

$$u \geq 0, v \geq 0$$

Paso 2. Formular el problema dual correspondiente

$$\begin{aligned} &\text{Maximizar: } (u^T - v^T) b = u^T b - v^T b \\ &\text{Sujeto a:} \\ &\quad (u^T - v^T) A \leq c \\ &\quad u \geq 0 \\ &\quad v \geq 0 \end{aligned}$$

Paso 3. Substituir

$$\lambda = u - v$$

en el problema dual anterior, para obtener finalmente

$$\begin{aligned} &\text{Maximizar: } \lambda^T b \\ &\text{Sujeto a:} \\ &\quad \lambda^T A \leq c \\ &\quad \lambda \text{ no restringido} \end{aligned}$$

En resumen, si algunas de las desigualdades lineales del problema primal (1) se cambian por igualdades, entonces las componentes correspondientes de λ , en el problema dual se convierten en variables libres (no restringidas). Si algunas de las componentes de x en el problema primal son libres, entonces las desigualdades correspondientes se cambian por igualdades en el problema dual.

Problema dual de la dieta Problema primal. Seleccionar una combinación de alimentos (la dieta) que cumpla ciertos requerimientos nutricionales al mínimo costo.

Problema dual. Determinar los precios unitarios de los componentes nutricionales de una píldora (sintéticos) que maximicen la utilidad, mientras se mantengan competitivos con los alimentos reales, es decir, no ser mayores que los precios de los alimentos en el mercado.

Forma simétrica

Primal	Dual
Minimizar: $c^T x$	Maximizar: $\lambda^T b$
Sujeto a:	Sujeto a:
$Ax \geq b$	$\lambda^T A \leq c$
$x \geq 0$	$\lambda \geq 0$
	λ_i precio unitario del i -ésimo nutriente
	c_j costo del alimento j -ésimo en el mercado
	a_{ij} es el i -ésimo nutriente, en el j -ésimo alimento
	b_j unidades del j -ésimo nutriente serán compradas.

Example 108 Problema dual del transporte

Problema primal: Al que se enfrenta un productor es seleccionar el modelo de envío de productos entre varios orígenes y destinos fijos, de modo que se minimice el costo de transportación mientras se satisfaga la demanda y bajo el supuesto que el total de productos enviados es igual al total de productos recibidos.

Problema dual: Para interpretar el problema dual en el contexto anterior, suponga que un empresario considera que puede hacer los envíos más eficientemente, así que le propone al productor comprar su producto en la planta (origen) y vendérsela a los almacenistas (destinos). El precio de cada producto, en estas transacciones, variará de un punto a otro, y será determinado por el empresario por anticipado. Así que, el empresario deberá escoger los precios, de modo que su oferta sea atractiva para el productor.

Es decir, el empresario seleccionará entonces los precios para los m orígenes y los n destinos. Para ser competitivo, con las formas usuales de transportación, sus precios deberán satisfacer que la cantidad neta que el productor deberá pagar por cada unidad de producto vendida en un origen i y comprada en un destino j , deberá ser menor o igual que el costo de transportación correspondiente.

Problema Primal (forma estándar)

Minimizar: $\sum_{i,j} c_{ij}x_{ij}$

Sujeto a:

$$\sum_{j=i}^n x_{ij} = a_i \text{ para todo } i = 1, \dots, m$$

$$\sum_{i=1}^m x_{ij} = b_j \text{ para todo } j = 1, \dots, n$$

$$x_{ij} \geq 0 \text{ para todo } i = 1, \dots, m; j = 1, \dots, n$$

Se asume que la cantidad total enviada es igual a la cantidad total recibida

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$$

Condición de optimalidad

Sean $\lambda_1, \lambda_2, \dots, \lambda_m$ las variables de decisión del problema dual y

$$z_j = [\lambda_1 \dots \lambda_m] \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix} = \sum_{i=1}^m a_{ij} \lambda_i$$

el j -ésimo lado izquierdo de la restricción funcional del problema dual, donde

$$z_j \geq c_j$$

Entonces la *condición de optimalidad* queda formulada como

$$z_j - c_j \geq 0$$

para toda $j = 1, 2, \dots, n$

$$\lambda_i \geq 0$$

para toda $i = 1, 2, \dots, m$

De modo que el *método simplex* es la búsqueda de los valores λ_i tales que

$$W = \sum_{i=1}^m b_i \lambda_i$$

sujeta a

$$\sum_{i=1}^m a_{ij} \lambda_i \geq c_j$$

para toda $j = 1, 2, \dots, n$

$$\lambda_i \geq 0$$

para toda $i = 1, 2, \dots, m$

Donde W es el valor de z en la iteración correspondiente.

Entonces, las soluciones factibles del problema dual son las que satisfacen la *condición de optimalidad* del problema primal. De ahí que, la solución óptima del problema primal corresponde a una solución factible del problema dual.

En consecuencia, el valor óptimo de z en el problema primal es el valor mínimo factible de W en el dual.

Por lo que, el problema dual es una reformulación de la meta del método simplex, esto es, obtener una solución del problema primal que satisfaga la prueba de optimalidad.

Antes de que se obtenga el valor óptimo, los correspondientes valores de λ_i en el renglón de la FO (los coeficientes de las variables de holgura de la tabla de simplex), son **no factibles** para el problema dual.

Cuando se logra el objetivo, los λ_i proveen una solución óptima del problema dual (el vector solución se denota por λ^*), porque es una solución factible que se logra en el valor mínimo factible de W .

Además, λ^* proporciona los precios sombra para el problema primal y

$$W^* = z^*$$

Esta última igualdad, implica que

$$cx \leq \lambda b$$

para cualquier par de soluciones factibles x y λ , de sus respectivos problemas primal y dual.

La discusión anterior se ilustra con el ejemplo de la empresa Wyndor Glass, para el cual, las modificaciones al renglón de la función objetivo, en las iteraciones del simplex, se presentan en la siguiente tabla.

Iteración	Problema primal						Problema dual					
	x_1	x_2	x_3	x_4	x_5	Z	λ_1	λ_2	λ_3	$z_1 - c_1$	$z_2 - c_2$	W
0	-3	-5	0	0	0	0	0	0	0	-3	-5	0
1	-3	0	0	5/2	0	30	0	5/2	0	-3	0	30
2	0	0	0	3/2	1	36	0	3/2	1	0	0	36

Figure 14

- La solución inicial no es factible porque las variables de excedencia son negativas.
- La primera iteración elimina una de ellas, pero no la otra.
- La segunda iteración elimina a ambas.
- La condición de optimalidad se satisface para el problema primal porque todas las variables duales y de excedencia son no negativas.
- La solución $[0, \frac{3}{2}, 1]$ es óptima para el problema dual y el valor óptimo es $Z^* = W^* = 36$.

En resumen, las relaciones entre los problemas primal y dual se caracterizan mediante las siguientes propiedades:

Proposition 109 *Dualidad Débil.* Si x es una solución factible del problema primal y λ es una solución factible del problema dual, entonces

$$cx \leq \lambda b$$

Proposition 110 *Dualidad fuerte.* Si x^* y λ^* son soluciones óptimas para sus respectivos problemas, primal y dual, entonces

$$cx^* = \lambda^* b$$

Proposition 111 Si x_0 y λ_0 son soluciones factibles de sus respectivos problemas, primal y dual, y si

$$c^T x_0 = \lambda_0 b$$

entonces x_0 y λ_0 son soluciones óptimas, de los respectivos problemas primal y dual.

Theorem 112 *De dualidad de la Programación Lineal.*

(i) Si alguno de los problemas primal o dual tiene una solución óptima finita, también el otro la tiene y los valores de la función objetivo correspondientes son iguales.

(ii) Si alguno de los problemas tiene función objetivo no acotada, entonces el otro no tiene solución factible.

Theorem 113 Si el problema de PL estándar tiene una solución óptima básica para la base B , entonces el vector λ que satisface

$$\lambda^T = c_B^T B^{-1}$$

es una solución óptima del problema dual correspondiente. Los valores óptimos de ambos problemas son iguales.

Propiedad de simetría Para cualquier problema primal y su problema dual, todas las relaciones entre ellos deben ser simétricas porque el dual del dual es el problema primal.

Teorema de dualidad Las siguientes son las únicas relaciones posibles entre los problemas primal y dual.

1. Si un problema tiene soluciones factibles y una función objetivo acotada (con lo cual, tiene una solución óptima), entonces lo mismo sucede con el otro problema. Por lo que, las propiedades de dualidad débil y fuerte son aplicables.
2. Si uno de los problemas tiene soluciones factibles y la función objetivo no es acotada (es decir, el PL no tiene soluciones óptimas), entonces el otro problema no tiene soluciones factibles.
3. Si un problema no tiene soluciones factibles, entonces el otro problema o no tiene soluciones factibles o tiene una función objetivo no acotada.

Relación con el método simplex En cada iteración, el método simplex provee un par de soluciones de ambos problemas donde la solución al primal es factible pero la solución dual no es factible (con excepción de la última iteración).

Propiedad de las soluciones complementarias En cada iteración, el método simplex identifica simultáneamente una solución extrema factible del problema primal (CPF) y una solución complementaria del problema dual (situada en el renglón de la FO, en las columnas asignadas a las variables de holgura), donde

$$cx = \lambda b$$

Si x no es óptima para el problema primal, entonces λ no es factible para el problema dual.

Propiedad de las soluciones complementarias óptimas En la iteración final, el método simplex identifica simultáneamente una solución óptima x^* del problema primal y una solución óptima complementaria λ^* del problema dual (situada en el renglón de la FO, en las columnas de asignadas a las variables de holgura), donde

$$cx^* = \lambda^* b$$

Las componentes del vector λ^* , proveen los valores de los precios sombra del problema primal.

En resumen, si la matriz A del sistema de restricciones contiene a la matriz identidad I de dimensión $m \times m$, correspondiente a las variables de holgura (excedencia), la inversa de la base B , esto es B^{-1} , aparecerá en la última matriz del método simplex, en la posición de la matriz identidad I .

Las componentes del último renglón bajo B^{-1}

$$c_1^T - c_B^T B^{-1}$$

donde c_1 es el m -vector de coeficientes de la función objetivo, correspondientes a las variables de holgura. Restando c_1 de los elementos de la última fila, se obtiene el inverso aditivo de la solución al problema dual, esto es

$$\lambda^T = c_B^T B^{-1}$$

Método simplex (primal) revisado Considere el problema

$$\begin{aligned} \text{Maximizar: } z &= c^T x \\ \text{Sujeto a: } & Ax \leq b \\ & x \geq 0 \end{aligned}$$

Donde las dimensiones de la matriz A y los vectores x , c y b , son $m \times n$, $n \times 1$, $n \times 1$ y $m \times 1$, respectivamente.

Las restricciones funcionales del problema de PL, en forma estándar y aumentada, debida a la introducción de variables de holgura

$$x_s = \begin{bmatrix} x_{n+1} \\ \vdots \\ x_{n+m} \end{bmatrix}$$

son

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} & 1 & 0 & \cdots & 0 \\ a_{21} & \cdots & a_{2n} & 0 & 1 & 0 & \vdots \\ \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ x_{n+1} \\ \vdots \\ x_{n+m} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

Donde, las dimensiones de la matriz de coeficientes y las variables de decisión, son $m \times (n + m)$ y $n + m$, respectivamente.

El enfoque general del método simplex es obtener una sucesión de soluciones básicas factibles que mejoren el valor de la función objetivo, hasta que se obtiene una solución óptima. Una de las características clave del método simplex revisado incluye la forma en la cual resuelve para cada nueva solución básica, después de identificar las variables básicas y no-básicas. Dadas esas variables, la solución básica resultante es la solución de las m ecuaciones

$$[A \quad I] \begin{bmatrix} x \\ x_s \end{bmatrix} = b$$

donde, las n -variables no básicas de las $n + m$ componentes de

$$\begin{bmatrix} x \\ x_s \end{bmatrix}$$

se hacen cero, con lo cual se eliminan esas n -variables y se obtiene un sistema de m ecuaciones en m desconocidas (las variables básicas), denotado por

$$Bx_B = b$$

donde el vector de variables básicas

$$x_B = \begin{bmatrix} x_{B1} \\ \vdots \\ x_{Bm} \end{bmatrix}$$

se obtiene al eliminar las variables no-básicas en

$$\begin{bmatrix} x \\ x_s \end{bmatrix}$$

La matriz básica

$$B = \begin{bmatrix} B_{11} & \cdots & B_{1n} \\ \vdots & \ddots & \vdots \\ B_{m1} & \cdots & B_{mm} \end{bmatrix}$$

se obtiene de eliminar las columnas correspondientes a los coeficientes de las variables no-básicas, en la matriz

$$[A \quad I]$$

Además, los elementos de x_B y las columnas de B , pueden resultar en orden diferente cuando se ha aplicado el método simplex. El método simplex sólo introduce variables básicas de modo que B es no singular, es decir, existe B^{-1} . De ahí que, se resuelve

$$Bx_B = b$$

multiplicando ambos lados de la igualdad por B^{-1} .

De donde,

$$\begin{aligned} B^{-1}Bx_B &= B^{-1}b \\ x_B &= B^{-1}b \end{aligned}$$

Es la solución para las variables básicas.

Sea c_B el vector cuyos elementos son los coeficientes de la función objetivo (incluyendo ceros para las variables de holgura), correspondientes a los elementos de x_B . el valor de la función objetivo para esta solución básica es

$$z = c_B x_B = c_B B^{-1} b$$

Por lo que $c_B B^{-1}$, proporciona la variación de z respecto a la variación de b .

Es decir,

$$\Delta z = (c_B B^{-1})_i$$

es el precio sombra, asociado al lado derecho de la i -ésima restricción.

Forma matricial del conjunto de restricciones Se calcula la inversa de una nueva base, a partir de otra base, siempre que las dos bases difieran sólo en un vector columna. El procedimiento se ajusta muy bien para los cálculos del método simplex, ya que las bases sucesivas difieren exactamente en una columna (vector), como resultado del intercambio de los vectores entrante y saliente.

Dada una base B , la siguiente base de la iteración próxima diferirá de B sólo en una columna. El procedimiento de forma producto calcula entonces la inversa de la siguiente base, premultiplicando la actual por una matriz E construida especialmente como sigue. Para las matrices identidad I , B y su inversa B^{-1} , el vector P_r (saliente) de B se reemplaza por un nuevo vector P_j (entrante).

Lo anterior lo representamos así

$$\alpha^j = B^{-1}P_j$$

La nueva inversa es

$$B_{sig}^{-1} = EB^{-1}$$

donde

$$E = [e_1 \quad \cdots \quad e_{r-1} \quad \xi \quad e_{r+1} \quad \cdots \quad e_m]$$

y el j -ésimo vector columna de la matriz identidad I es

$$e_j = [0, \dots, 0, 1, 0, \dots, 0]$$

cuya j -ésima componente es 1.

Ahora se define el vector

$$\xi = \left[-\frac{\alpha_1^j}{\alpha_r^j}, \dots, \frac{1}{\alpha_r^j}, \dots, \frac{\alpha_m^j}{\alpha_r^j} \right]$$

cuyo r -ésimo componentes es

$$\frac{1}{\alpha_r^j}$$

siempre y cuando

$$\alpha_r^j \neq 0$$

En caso contrario, B^{-1} no existe.

En resumen, para el problema original dado en forma matricial

$$\begin{bmatrix} 1 & -c & \mathbf{0} \\ \mathbf{0} & A & I \end{bmatrix} \begin{bmatrix} z \\ x \\ x_s \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix}$$

en cualquier iteración se tiene

$$\begin{aligned} \begin{bmatrix} z \\ x_B \end{bmatrix} &= \begin{bmatrix} 1 & c_B B^{-1} \\ \mathbf{0} & B^{-1} \end{bmatrix} \begin{bmatrix} 0 \\ b \end{bmatrix} = \begin{bmatrix} c_B B^{-1} b \\ B^{-1} b \end{bmatrix} \\ \begin{bmatrix} 1 & c_B B^{-1} \\ \mathbf{0} & B^{-1} \end{bmatrix} \begin{bmatrix} 1 & -c & \mathbf{0} \\ \mathbf{0} & A & I \end{bmatrix} &= \begin{bmatrix} 1 & c_B B^{-1} A - c & c_B B^{-1} \\ 0 & B^{-1} A & B^{-1} \end{bmatrix} \\ \begin{bmatrix} 1 & c_B B^{-1} A - c & c_B B^{-1} \\ 0 & B^{-1} A & B^{-1} \end{bmatrix} \begin{bmatrix} z \\ x \\ x_s \end{bmatrix} &= \begin{bmatrix} c_B B^{-1} \\ B^{-1} \end{bmatrix} \end{aligned}$$

Example 114 *Considérese siguiente matriz y su inversa.*

$$B = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 4 & 0 & 1 \end{bmatrix}$$

$$B^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 \\ -2 & 1 & 1 \end{bmatrix}$$

El vector columna

$$P_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

se cambia a

$$Q_3 = \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}$$

Por lo que, la matriz B siguiente se calcula como sigue

$$B^{-1}Q_3 = \left[\begin{array}{ccc} \frac{3}{4} & \frac{1}{2} & 2 \end{array} \right] = \left[\begin{array}{ccc} \alpha_1^3 & \alpha_2^3 & \alpha_3^3 \end{array} \right]$$

De donde

$$\xi = \left[\begin{array}{ccc} -\frac{3}{8} & -\frac{1}{4} & \frac{1}{2} \end{array} \right]$$

puesto que $r = 3$.

Finalmente, como la matriz E es el resultado de remplazar la tercera columna de la matriz identidad por ξ , la matriz inversa de siguiente matriz B es

$$B_{sig}^{-1} = EB^{-1} = \begin{bmatrix} 1 & 0 & -\frac{3}{8} \\ 0 & 1 & -\frac{1}{4} \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 \\ -2 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{5}{4} & -\frac{5}{8} & -\frac{3}{8} \\ \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ -1 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

7.2.2 Teoremas de holgura complementaria

Forma asimétrica

Theorem 115 Sean x y λ soluciones factibles de los respectivos problemas primal y dual. Una condición necesaria y suficiente para que x y λ sean soluciones óptimas es que para toda j se cumplan las siguientes implicaciones.

Primal	Implicaciones	Dual
$Min \ c^T x$ Sujeto a $Ax = b$ $x \geq 0$	$x_j > 0 \Rightarrow \lambda^T A^j = c_j$ $x_j = 0 \Leftarrow \lambda^T A^j < c_j$	$Max \ \lambda^T b$ Sujeto a $\lambda^T A \leq c^T$

Donde A^j es la j -ésima columna de A .

Forma Simétrica

Theorem 116 Sean x y λ soluciones factibles de los respectivos problemas primal y dual. Una condición necesaria y suficiente para que x y λ sean soluciones óptimas es que para toda i y toda j , se cumplan las siguientes implicaciones.

Primal	Implicaciones	Dual
$Min \ c^T x$ Sujeto a $Ax \geq b$ $x \geq 0$	$x_j > 0 \Rightarrow \lambda^T A^j = c_j$ $x_j = 0 \Leftarrow \lambda^T A^j < c_j$ $A_i x = b_i \Leftarrow \lambda_i > 0$ $A_i x > b_i \Rightarrow \lambda_i = 0$	$Max \ \lambda^T b$ Sujeto a $\lambda^T A \leq c^T$ $\lambda \geq 0$

Donde A_i es el i -ésimo renglón y A^j es la j -ésima columna, de A .

Example 117 En la tabla siguiente se anotan los pasos consecutivos para hacer las inferencias y resolver ambos problemas.

<i>Primal</i>	<i>Solución</i>	<i>Implicaciones</i>	<i>Dual</i>
$Min w = 2x_1 - x_2$ <i>Sujeto a</i> $2x_1 - x_2 - x_3 \geq 3$ $x_1 - x_2 + x_3 \geq 2$ $x_1 \geq 0, x_2 \geq 0$ $x_3 \geq 0$	<i>factible</i> $x_1 = \frac{5}{3} > 0$ $x_2 = 0$ $x_3 = \frac{1}{3} > 0$ <i>1</i>	\Rightarrow \Leftarrow \Rightarrow	$2\lambda_1 + \lambda_2 = 2$ $-\lambda_1 - \lambda_2 \leq -1$ $-\lambda_1 + \lambda_2 = 0$ <i>2</i>
			\Downarrow $2\lambda_1 + \lambda_2 = 2$ $-\lambda_1 + \lambda_2 = 0$ <i>3</i>
	$x_2 = 0$ $2x_1 - x_3 = 3$ $x_1 + x_3 = 2$ <i>5</i>	\Leftarrow \Leftarrow	$3\lambda_1 = 2$ $\lambda_1 = \frac{2}{3} > 0$ $\lambda_2 = \frac{2}{3} > 0$ $u = \frac{10}{3}$ <i>4</i>
	$3x_1 = 5$ $x_1 = \frac{5}{3}$ $x_3 = \frac{1}{3}$ $w = \frac{10}{3}$ <i>6</i>		$w = u$ <i>la solución es óptima</i> <i>7</i>
			<i>factible</i>

7.3 Tópicos especiales de programación matemática⁸

La categoría de programación matemática incluye una amplia variedad de técnicas utilizadas para resolver problemas de optimización. El tomador de decisiones está tratando de maximizar o minimizar algún objetivo, sujeto a un conjunto de restricciones. La técnica más ampliamente utilizada en esta categoría es por su puesto la programación lineal.

7.3.1 Programación entera

Muchos de los problemas de negocios pueden formularse y resolverse apropiadamente mediante la programación a excepción de que se requieren soluciones enteras. Por ejemplo, una variable x en un modelo de programación lineal se puede referir a si o no construir una nueva planta. Si $x = 1$, la planta se construye, $x = 0$ significa que no se construye la planta. Cualquier valor entre 0 y 1 no tiene sentido en este problema. El proceso de solución estándar no garantiza que se encontrará una solución entera óptima (con excepción de los problemas de programación lineal formulados para las redes, los problemas de transporte y ruta crítica son ejemplos de tales problemas).

⁸Bonini, C.P., Asuman, W. H., y Bierman, H. Jr. QUANTITATIVE ANALYSIS FOR MANAGEMENT Chicago: McGraw Hill. IRWIN Series, 1997(9), capítulo 4.

Para manejar este tipo de problemas se ha desarrollado un conjunto de técnicas llamado programación entera. Algunos problemas pueden requerir que todas las variables en el problema sean enteros – solución entera. Un caso especial de solución entera ocurre cuando todas las variables pueden tomar solo los valores 0 y 1, denominado problema de programación binaria.

$$\begin{aligned}
 & \text{Max :} \\
 & P = 4x_1 + 5x_2 \\
 & \text{Sujeta a :} \\
 & \frac{2}{3}x_1 + x_2 \leq 1 \\
 & x_1 \leq 1 \\
 & x_1, x_2 \geq 0 \\
 & x_1, x_2 \text{ enteros}
 \end{aligned}$$

Las soluciones enteras factibles están representadas en la siguiente gráfica por los pequeños puntos en las intersecciones de la cuadrícula correspondiente a los valores enteros de las variables.

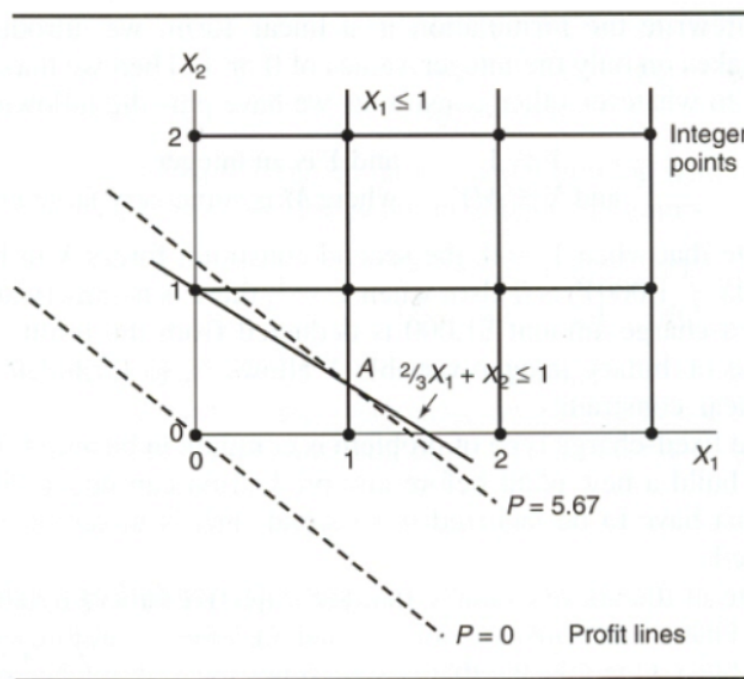


Figure 15

Formulación de problemas de programación entera En general, un problema de programación lineal entera se formula de la misma forma que los problemas de programación lineal estándares con la adición de que una o más de las variables se asumen con valores enteros. Esto es,

se establece una función objetivo lineal que se busca maximizar o minimizar sujeta a restricciones lineales. Sin embargo, se usan variables binarias o cero/uno que son únicas en la programación entera. El valor de 1 puede usarse para establecer la presencia de algo (una fábrica, por ejemplo), mientras que el 0 indica su ausencia.

El problema del cargo fijo Este ejemplo ilustra el uso de las variables binarias. Supongamos que un producto dado, en caso de que se decida producir, contribuye al beneficio con 50 UM por unidad. Pero, se incurriría en gastos por la correspondiente corrida del equipo de producción en una sola ocasión por la cantidad de 1,000 UM. Si no se produce, estos gastos serían 0.

Sea x el número de unidades producidas. Entonces el problema puede formularse como sigue

$$\text{Contribución} = \begin{cases} 50x - 1000 & \text{si } x > 0 \\ 0 & \text{si } x = 0 \end{cases}$$

La función *Contribución* no es una función lineal de x en esta formulación. Para reescribir la formulación en una forma lineal, introducimos una nueva variable y , binaria. Entonces el problema consiste en maximizar $50xy - 1000y$ sujeta a las restricciones que se hagan más otras dos adicionales:

$$\begin{aligned} y &\leq 1 \\ x &< My \end{aligned}$$

donde y es una variable entera y M es un número muy grande.

Cuando $y = 0$, la segunda restricción hace que x sea igual a 0 y la *Contribución* también sería 0. Cuando $y = 1$, prácticamente, no hay un límite para el valor de x ; el cargo fijo de 1,000 UM se descuenta del beneficio. De ahí que, el uso de una variable entera binaria y permite formular este tipo de problema con restricciones lineales.

Otras aplicaciones del tipo cargo fijo son comunes en los negocios. Hay costos de inversión para construir una nueva planta antes de que ocurra la producción. Usualmente, existen costos fijos en los que se incurren cuando se planea un segundo turno o se abre un nuevo almacén.

Nótese que en el ejemplo expuesto, la función objetivo y las restricciones son lineales, esto es, solo participan términos múltiples constantes de las variables, no así otras potencias o productos entre variables. Un error frecuente en la formulación de problemas como el anterior sucede cuando se expresa a la función objetivo como

$$50xy - 1000y$$

que también hace una contribución de

$$50x - 1000$$

cuando $y = 1$ y 0 cuando $y = 0$. Pero el término xy no es lineal porque es el producto de las variables x y y . Por lo tanto, esta formulación no sería un problema de programación lineal entera válida.

Otro enfoque para formular el costo fijo en una hoja de cálculo usa la función condicional "SI". Supóngase que el valor de la variable x se introducirá en la celda $D2$, entonces la función contribución podría escribirse como

$$SI(D2 > 0, 50 * D2 - 1,000, 0)$$

Esta formulación es lógicamente válida, pero no es una función lineal y no sería apropiada para un modelo lineal.

El problema del tamaño del lote Un problema similar es uno en el cual algún nivel mínimo se requiere antes de que se realice una actividad. Por ejemplo, una empresa tiene que comprar una cantidad mínima de por lo menos 50 unidades de un cierto producto. Sea x el número de unidades de ese producto. Entonces $x = 0$ ó $x \geq 50$. Esta situación se puede formular con una variable binaria y mediante las restricciones:

$$\begin{aligned}x &\leq My \\x &\geq 50y\end{aligned}$$

donde M es un número muy grande. Nótese que si $y = 0$, entonces, de la primera restricción x debe ser 0. Si $y = 1$, entonces la segunda restricción hace que x sea al menos 50.

Restricciones del tipo uno u otra, pero no ambas Algunas veces, en situaciones de decisión, una u otra restricción debe cumplirse, pero no ambas. Por ejemplo:

$$5x_1 + 2x_2 \leq 10$$

ó

$$3x_1 - 4x_2 \leq 24$$

pero no ambas.

Se puede manejar las alternativas anteriores con la programación entera usando la variable binaria y y modificando ambas restricciones como sigue

$$\begin{aligned}5x_1 + 2x_2 &\leq 10 + My \\3x_1 - 4x_2 &\leq 24 + M(1 - y)\end{aligned}$$

Cuando $y = 0$, la primera restricción es la obligada, pero el lado derecho de la segunda es un valor muy grande, de ahí que no sea restrictiva. Y recíprocamente, cuando $y = 1$, la segunda restricción obliga mientras que la primera no, puesto que My es un valor muy grande.

8 ANÁLISIS DE SENSIBILIDAD

El propósito del estudio de la programación lineal es ayudar a guiar la decisión final del gerente al proporcionarle entendimiento sobre las consecuencias probables de seguir varias opciones administrativas bajo una diversidad de suposiciones sobre las condiciones futuras.

Este análisis usualmente conocido como análisis de sensibilidad (o qué pasaría si...); porque hace la consideración de qué le sucedería a la solución óptima si diferentes suposiciones se asumieran sobre las condiciones futuras.

Los ejemplos y problemas de los problemas de programación lineal considerados hasta ahora han proporcionado los datos necesarios para determinar precisamente todas las cantidades requeridas en la tabla de parámetros. En cambio, las aplicaciones reales van más allá de esa situación, es decir, el estudio se realiza bajo la condición de disponer sólo de burdas estimaciones de los parámetros del modelo.

Por ejemplo, en el estudio de caso de la empresa Wyndor⁹, dos de los parámetros clave del modelo son los coeficientes de la función objetivo que representan las utilidades unitarias de los

⁹Hillier, F., Hillier, M. and Lieberman, G. (2000) Introduction to Introduction to Operations Research. Boston: McGraw-Hill. Higher Education.

dos productos (puertas y ventanas). Estos parámetros fueron estimados en 300 y 500 unidades monetarias. Sin embargo, lo que estas unidades podrían representar depende de muchos factores – los costos de las materias primas, la producción, los envíos, la publicidad, y así sucesivamente.

Beneficios al realizar el análisis de sensibilidad. Típicamente, muchos de los parámetros de un modelo de programación lineal son simplemente estimaciones que no pueden determinarse precisamente en un momento dado. El análisis de sensibilidad muestra cuán precisas deben ser las estimaciones para evitar una solución óptima errónea, y de ahí identificar precisamente los parámetros sensibles donde una valoración cuidadosa sería necesaria para refinar las estimaciones.

Si las condiciones cambian una vez que el estudio se ha completado (lo que es muy frecuente), el análisis de sensibilidad deja indicadores que muestran (sin resolver de nuevo el modelo) si un cambio en un parámetro del modelo cambia la solución óptima.

La discusión se centra en el estudio de cómo los cambios en los parámetros de un modelo de PL afectan a la solución óptima. El análisis de que tal si... se le conoce como análisis de sensibilidad porque involucra una revisión de qué tan sensible es la solución óptima a los cambios en cada uno de los parámetros.

Continuando con el caso Wyndor-Glass, sobre la discusión entre los gerentes para obtener la combinación de producción de puertas y ventanas, óptima, se plantean las siguientes cuestiones:

”¿... Se tiene una manera de comprobar qué tan lejos se pueden hacer las estimaciones de los parámetros sin cambiar la combinación óptima de producción?”

Si, ... podemos encontrar rápidamente lo que denominamos rango de optimalidad de cada utilidad unitaria.

¿Y si queremos saber qué sucede si ambas estimaciones están fuera de lugar?

También, podemos proporcionar una manera de comprobar si la combinación óptima podría cambiar para cualquier nueva combinación de las utilidades unitarias que pensaras que fuera cierta.”

Resumen de las preguntas del gerente:

1. ¿Qué pasaría si la estimación de la utilidad unitaria de uno de los productos de Wyndor fuera imprecisa?
2. ¿Qué sucedería si las estimaciones de las utilidades unitarias de ambos productos de Wyndor fueran imprecisas?
3. ¿Qué sucedería si se hiciera un cambio en el número de horas del tiempo de producción (los miembros derechos de las restricciones funcionales o nivel disponible del recurso) por semana disponible para los nuevos productos de Wyndor en una de las plantas?
4. ¿Qué sucedería si se hicieran cambios simultáneos en el número de horas de tiempo de producción por semana para los nuevos productos de Wyndor en todas las plantas?

9 Rango de optimalidad del análisis de sensibilidad

Pregunta 1. - ¿Qué pasaría si la estimación de la utilidad unitaria de uno de los productos de Wyndor fuera imprecisa?

Sea P_D = la utilidad unitaria para la nueva clase de puerta. Debido a que el valor original de $P_D = 300$ puede cambiar considerablemente en ambas direcciones sin cambiar la solución óptima, se dice que P_D no es un parámetro sensible. Así que no es necesario identificar con precisión su estimación para tener confianza de que el modelo proporciona una solución óptima correcta.

El rango de valores de P_D sobre los cuales la solución factible $(D, W) = (2, 6)$ permanece como solución óptima se denomina rango permisible de P_D . De acuerdo con el reporte de sensibilidad de la página 34, el rango de optimalidad para P_D es $[300 - 300, 300 + 450] = [0, 750]$.

Cuando un pequeño cambio en el valor de ciertos coeficientes de la función objetivo produce cambios en la solución óptima, dicho coeficiente se dice que es un parámetro sensible.

Los indicadores (extremos del rango de optimalidad) proveen el segundo beneficio del análisis de sensibilidad, es decir, indicadores de cuando un cambio en un parámetro produce cambios en la solución óptima del modelo.

9.1 Cambios simultáneos en los coeficientes de la función objetivo

Los datos que proporciona el análisis de sensibilidad sobre los niveles permitidos de incrementos o reducciones de los coeficientes de la función objetivo pueden utilizarse para analizar el efecto de hacer cambios simultáneos en esos coeficientes.

9.1.1 Regla del 100%

La regla del 100% para cambios simultáneos en los coeficientes de la función objetivo dice que si se hacen cambios simultáneos sobre dichos coeficientes, calcúlese el porcentaje del cambio permisible (incremento o reducción) para los coeficientes que se mantengan en su rango de optimalidad. Si la suma de los porcentajes no excede al 100%, la solución óptima original definitivamente continuará siéndolo. En caso contrario no se puede asegurar que siga siendo óptima.

Example 118 Considerando de nuevo el problema de la empresa Wyndor, el siguiente reporte de sensibilidad permitirá saber si los cambios a las utilidades unitarias de 300 a 500 para las puertas y de 500 a 400 para las ventanas no modifican la solución óptima.

Celdas variables		Valor	Gradiente	Coeficiente	Aumento	Reducción
Celda	Nombre	igual	reducido	objetivo	permisible	permisible
\$B\$10	Solución puertas	2	0	300	450	300
\$C\$10	Solución ventanas	6	0	500	1E+30	300

Restricciones		Valor	Precio	Restriccion	Aumento	Reducción
Celda	Nombre	igual	sombra	lado derecho	permisible	permisible
\$D\$6	Planta 1 Totales	2	0	4	1E+30	2
\$D\$7	Planta 2 Totales	12	150	12	6	6

De acuerdo con el reporte, el aumento de la utilidad unitaria de 300 a 500, esto es, de 200, está dentro del aumento permisible. También, la reducción de 500 a 400, de 100, es permisible. Entonces, los porcentajes de aumento y reducción correspondientes se calculan como sigue.

$$P_D : 300 \rightarrow 500$$

Porcentaje de incremento permisible

$$100 \left(\frac{500 - 300}{450} \right) = 44.4\bar{4}$$

$$P_W : 500 \rightarrow 400$$

Porcentaje de reducción permisible

$$100 \left(\frac{500 - 400}{300} \right) = 33.3\bar{3}$$

El porcentaje total de ambos cambios es de $77.7\bar{7}$. Por ser un porcentaje menor al 100%, se asegura que la solución óptima original sigue siendo óptima.

En resumen, la regla del 100% se usa para determinar qué tanto se pueden cambiar los coeficientes de la función objetivo antes de que la solución óptima cambie.

- Cuando se tiene gran cantidad de variables de decisión, es poco práctico analizar todas las combinaciones posibles de cambios, por lo que la regla del 100% simplifica el análisis, pues solo se requiere dividir cada cambio en cada coeficiente por su correspondiente aumento o reducción permisible y determinar si no se excede al 100%.
- Aún cuando se den cambios posteriores al estudio, la regla del 100% indica rápidamente si la solución óptima original lo sigue siendo.

9.2 Análisis del precio sombra para los lados derechos de las restricciones

Ahora nos centraremos en el estudio de los efectos de los cambios en los lados derechos de las restricciones funcionales de los problemas de programación lineal. Estos parámetros representan en general algunas decisiones sobre políticas administrativas de los gerentes, mas qué cantidades que están fuera de control del gerente.

El análisis del precio sombra, que es parte del análisis de sensibilidad, proporciona una valiosa guía al gerente acerca de qué efecto tendría alterar las decisiones sobre esas políticas administrativas. Este análisis se hace para cada restricción funcional, una a la vez para determinar qué efecto tendrá en el valor óptimo de la función objetivo, si se hicieran cambios pequeños en el lado derecho de una de las restricciones.

Dada una solución óptima y el valor correspondiente de la función objetivo de un problema de programación lineal, el precio sombra asociado a una restricción funcional es la razón a la cual el valor de la función objetivo puede incrementarse o reducirse cuando se incrementa o reduce el lado derecho de una restricción por una pequeña cantidad (en general en una unidad).

Un precio sombra de una restricción nos permite revisar inmediatamente cuál sería el efecto de alterar una decisión de política administrativa que cambia el lado derecho de una restricción. Siempre que el cambio no sea muy grande, el cambio resultante en el valor óptimo de la función objetivo será justamente tantas veces como se cambie el lado derecho, esto es el cambio (positivo o negativo) multiplicado por el precio sombra.

De ahí que, el precio sombra para una restricción funcional informa al gerente qué tanto el beneficio total de producir dos productos como en el caso Wyndor, podría incrementarse por cada hora adicional disponible del tiempo de producción en cada planta.

9.2.1 Rango de factibilidad del análisis de sensibilidad

En el ejemplo de la empresa Wyndor, el rango de factibilidad de la primera restricción es $[4 - 2, 4 + \infty) = [2, \infty)$.

9.2.2 Cambios simultáneos en los lados derechos de las restricciones

El rango de factibilidad para el lado derecho de una restricción funcional identifica el rango de valores sobre el cual el precio sombra para esta restricción puede aplicarse para evaluar los efectos de los cambios en este lado derecho. Esto es muy útil cuando el gerente quiere saber cual es el

impacto de un cambio en una política, esto es, un cambio en el lado derecho de una restricción. Pero, en general las decisiones administrativas consideran cambios simultáneos en los lados derechos de las restricciones. Analicemos el siguiente escenario:

$RHS_2 : 12 \rightarrow 13$ implica que el cambio en la utilidad total = precio sombra = 150 UM

$RHS_3 : 18 \rightarrow 17$ implica que el cambio en la utilidad total = - precio sombra = -100 UM

Por lo tanto, el incremento neto en la utilidad total es de 50 UM.

Pero, no sabemos si ambos precios sombra siguen siendo válidos, cuando ambos lados derechos (RHS) se cambian como se hizo. Por lo que es necesario valerse de una regla del 100% similar a la anterior.

Regla del 100% para los lados derechos de las restricciones Con cambios simultáneos en los lados derechos, debemos enfocarnos en el porcentaje de incremento o reducción permisibles, como se ilustra en el siguiente ejemplo.

Si $RHS_2 : 12 \rightarrow 13$ entonces el porcentaje de incremento permisible es

$$100 \left(\frac{13 - 12}{6} \right) = 16.6\bar{6}\%$$

Si $RHS_3 : 18 \rightarrow 17$ entonces el porcentaje de reducción permisible es

$$100 \left(\frac{18 - 17}{6} \right) = 16.6\bar{6}\%$$

Por lo tanto, el porcentaje total de 33.33% no excede al 100%, así que los precios sombra son definitivamente válidos para predecir los efectos de los cambios como se usaron, para dar el incremento neto en la utilidad total.

10 MÉTODO DE MÍNIMOS CUADRADOS (Gauss):

Problem 119 *Ajustar una recta a través de puntos conocidos, de modo que la suma de los cuadrados de las distancias de esos puntos a la recta sea mínima (las distancias se miden en la dirección vertical para hacer la comparación entre el valor ajustado y el valor observado).*

Suposición general: Todos los valores de x_i en la muestra

$$\{(x_1, y_1), \dots, (x_n, y_n)\}$$

son diferentes.

Para la muestra de tamaño n , se calculan las distancias verticales de cada punto de la muestra a la recta de ajuste

$$y = ax + b$$

esto es

$$|y - a - bx|$$

La suma de los cuadrados de las n distancias

$$d = \sum_{i=1}^n [y_i - a - bx_i]^2$$

que se desea minimizar es una función de los parámetros a y b . Como se sabe la condición necesaria para que d sea mínima se formula con las ecuaciones

$$\begin{aligned}\frac{\partial d}{\partial a} &= 0 \\ \frac{\partial d}{\partial b} &= 0\end{aligned}$$

Al resolver el sistema de ecuaciones parciales se llega a la formulación de la recta de regresión

$$y - \bar{y} = b(x - \bar{x})$$

de los valores y_i , de la muestra, sobre los valores x_i de la misma, donde

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i \\ \bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i\end{aligned}$$

La pendiente b es el coeficiente de regresión de y sobre x .

Procedimiento de solución del sistema:

$$\begin{aligned}\frac{\partial d}{\partial a} &= -2 \sum_{i=1}^n [y_i - a - bx_i] = 0 \Rightarrow na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ \frac{\partial d}{\partial b} &= -2 \sum_{i=1}^n x_i [y_i - a - bx_i] = 0 \Rightarrow a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i\end{aligned}$$

Resolviendo el sistema de ecuaciones para a y b , por la regla de Cramer

$$\begin{aligned}a &= \bar{y} - b\bar{x} \\ b &= \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n(n-1) s_x^2}\end{aligned}$$

Donde, el determinante del sistema es

$$n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 = n(n-1) s_x^2$$

debido a que la varianza de la muestra es

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right]$$

De acuerdo a la suposición general el determinante no es nulo y por tanto el sistema tiene solución única.

El coeficiente de regresión se formula de manera simple mediante la expresión

$$b = \frac{s_{xy}}{s_x^2}$$

el numerador es la covarianza de la muestra

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \left[\sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right]$$

Problem 120 Utilice el método de mínimos cuadrados para encontrar los coeficientes de la recta de regresión de y (dureza Brinell) sobre x (deformación) de cierto tipo de acero para los datos siguientes:

x (mm)	6	9	11	13	22	26	28	33	35
y (kg/mm ²)	68	67	65	53	44	40	37	34	32

Part IV

MUESTREO E INFERENCIA ESTADÍSTICA

10.1 ¿Qué es la estadística?

Para Tanur, et.al (1989) es un campo de estudio, un grupo de datos que resume el estado de la economía nacional, el desempeño de una inversión, las condiciones particulares de un negocio, una guía a lo desconocido. La estadística es la ciencia de los datos, la cual involucra diversas acciones como recolectar, clasificar, resumir, organizar, analizar e interpretar información numérica, tomar decisiones...

Example 121 *El ingeniero muestrea las características de calidad de un producto, digamos fusibles, junto con otras variables controlables del proceso de producción, para facilitar la identificación de las variables que están relacionadas con dicha calidad. Toma muestras del producto, recién fabricado, antes de su envío, para decidir si se entregan o se retienen ciertos lotes de ese producto... (p.1)*

10.1.1 Definiciones de la estadística

En la literatura se encuentran diversas definiciones de la estadística, algunas de ellas dicen que la estadística

1. Es una rama de las matemáticas que trata de la recopilación, el análisis, la interpretación y la presentación de una gran cantidad de datos numéricos (The New Collegiate Dictionary Webster).
2. Trata con métodos para obtener conclusiones a partir de los resultados de los experimentos o procesos (Fraser, 1958).
3. Es la tecnología del método científico que trata del diseño de experimentos e investigaciones y la inferencia estadística (Mood, Graybill, y Boes, 1974).
4. Es una rama del método científico que trata de los datos reunidos al contar o medir las propiedades de alguna población (Kendal y Stuart, 1977).
5. Abarca el conocimiento relacionado con la toma de decisiones en situaciones de incertidumbre (Freund y Walpole, 1987).

En las definiciones anteriores se identifican algunos elementos en común. Cada definición implica una recopilación de datos teniendo como objetivo la inferencia. Cada una requiere la selección de un subconjunto de una gran colección de datos existente o conceptual, con el propósito de formular inferencias con respecto a las características del conjunto completo. Por lo que, la Estadística es una teoría de la información que tiene como objetivo efectuar inferencias sobre la población a partir de un muestreo.

Definition 122 *El gran conjunto de datos que es el centro de nuestro interés se denomina población y el subconjunto de ahí seleccionado es una muestra.*

A continuación, se listan otros elementos que ayudan a conceptualizar el quehacer estadístico.
... la Estadística trata del diseño de experimentos o encuestas mediante muestras para obtener una cantidad determinada de información a un costo mínimo y del uso óptimo de esta información para hacer inferencias con respecto a una población.

...El objetivo de la Estadística es deducir inferencias con respecto a una población a partir de la información contenida en una muestra y proporcionar una medida correspondiente para la exactitud de la inferencia.

... Un experimento estadístico involucra la observación de una muestra seleccionada de un conjunto mayor de datos, real o teórico, denominada población. Las mediciones en la muestra, consideradas como las observaciones de una o más variables aleatorias, se utilizan después para hacer una inferencia con respecto a las características de la población en estudio.

... la probabilidad de una muestra observada desempeña un papel muy importante al hacer una inferencia y en la evaluación de su confiabilidad.

... es claro que la manera de obtener la muestra afectará la probabilidad de un resultado particular de una muestra.

... la manera de hacer el muestreo, conocida como diseño de un experimento, afecta tanto la cantidad de información en una muestra como la probabilidad de observar un resultado específico de la muestra. De ahí que se deba describir con detalle cada procedimiento de muestreo si se desea hacer inferencias válidas para la población a partir de la muestra.

10.1.2 Uso de software y recolección de datos

En el mercado existen variados programas estadísticos, algunos tienen disponibilidad de recursos en la Web, como los siguientes:

- Excel, <http://office.microsoft.com/es-es/excel/HP100908423082.aspx?pid=CH100648513082>
- Fathom, http://www.keypress.com/x4138.xml?version=Fathom_2.1_Windows
- Maple12, <http://www.maplesoft.com/Products/Maple/>
- Minitab, <http://www.minitab.com/products/minitab/14/>
- SPSS, <http://www.spss.com/>
- Statgraphics, <http://www.statgraphics.com/>
- TI-89 Titanium, <http://education.ti.com/educationportal/sites/US/productDetail/usti89ti.html>

También son de interés los sitios que proporcionan bases de datos:

- Data and Story Library, <http://lib.stat.cmu.edu/DASL/>
- StatLib, <http://lib.stat.cmu.edu/>

10.1.3 Distribuciones de probabilidad

Definition 123 Una distribución de probabilidad o masa de probabilidad P definida para una variable aleatoria Y discreta, que asume los valores $y_i \in S$, con $i = 1, \dots, n$, tiene las siguientes propiedades:

1. La probabilidad puntual p es no negativa y no mayor a 1, es decir

$$0 \leq p(y_i) \leq 1$$

para cada $y_i \in S$.

2. Para cada valor $y \notin S$, se asume

$$p(y) = 0$$

3. La probabilidad de que la variable aleatoria tome un valor particular se denota por

$$P(Y = y_i) = p(y_i)$$

con $y_i \in S$.

- 4.

$$P(Y) = \sum_{i=1}^n p(y_i)$$

5. La suma de las probabilidades, de todos los valores puntuales posibles del espacio muestral S , es uno.

$$P(S) = \sum_{y \in S} p(y) = 1$$

Example 124 A continuación se ilustra la distribución de probabilidad de una variable discreta.

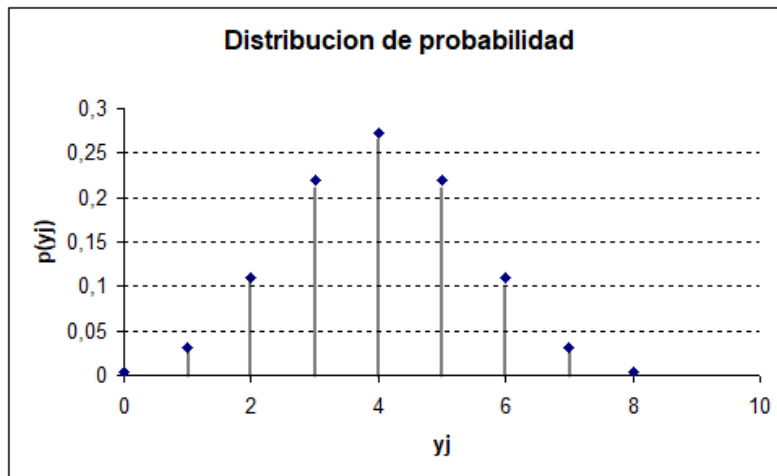


Figure 16

Los valores que asume la variable discreta Y son

$$\{0, 1, 2, 3, 4, 5, 6, 7, 8\}$$

Sus correspondientes probabilidades puntuales son

y_j	$p(y_j)$
0	0.0039
1	0.0313
2	0.1094
3	0.2188
4	0.2734
5	0.2188
6	0.1094
7	0.0313
8	0.0039

Exercise 125 Represente gráficamente la distribución de probabilidad binomial dada en la tabla.

x_i	$p(x_i)$
0	0.0625
1	0.25
2	0.375
3	0.25
4	0.0625

Definition 126 Una variable aleatoria X y su distribución correspondiente son del tipo continuo o simplemente continuas, si la función de distribución de X ,

$$F(x) = P(X \leq x)$$

puede representarse por la integral

$$F(x) = \int_{-\infty}^x f(u) du$$

donde f es una función no negativa, continua, excepto en posiblemente un número finito de valores de u , a la que se denomina densidad de probabilidad o simplemente densidad de la distribución.

Remark 127 La función de densidad es cero para los valores no asumidos por la variable aleatoria X .

Remark 128 La probabilidad puntual es

$$P(X = x) = f(x)$$

Remark 129 En la definición se requiere la existencia de la integral y la convergencia de la serie correspondiente.

Remark 130 Por el teorema fundamental del cálculo, se sabe que

$$F'(x) = f(x)$$

para toda x donde la función de densidad $f(x)$ es continua, de ahí que la función de densidad es la derivada de la función de distribución.

Proposition 131 De las propiedades de la función de densidad y de la integral, se cumple que

$$\int_{-\infty}^{\infty} f(u) du = 1$$

y

$$P(a \leq x \leq b) = F(b) - F(a) = \int_a^b f(u) du$$

Esta última expresión tiene una interpretación geométrica. Si $[a, b]$ es un intervalo real acotado, con $a \neq b$, la probabilidad de que la variable aleatoria X tome valores en ese intervalo se evalúa mediante el área de la región limitada por la curva de la densidad $f(x)$, y las rectas verticales $x = a$ y $x = b$.

$$P(a \leq x \leq b)$$

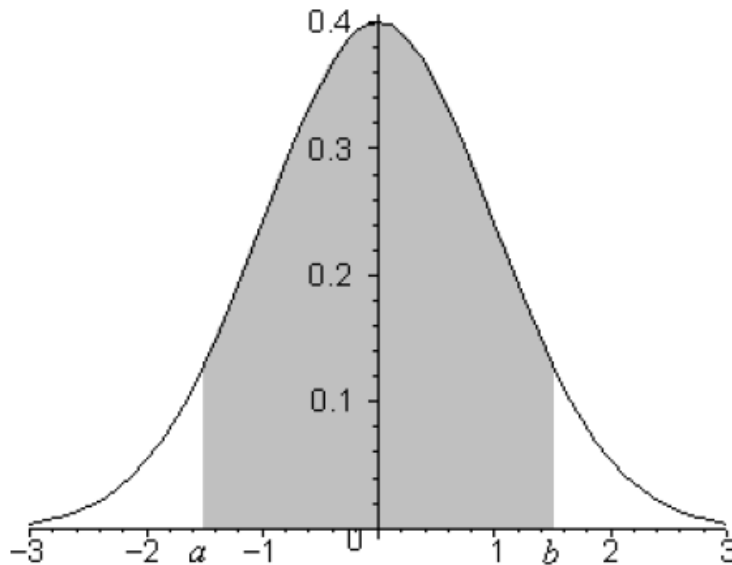


Figure 17

Remark 132 Cuando la variable X es continua, se cumplen las siguientes igualdades

$$P(a \leq x \leq b) = P(a < x < b) = P(a < x \leq b) = P(a \leq x < b)$$

Pero, cuando la variable aleatoria X es discreta, las igualdades no se cumplen, es decir, la inclusión o no de los datos en los extremos de los intervalos, afecta el cálculo de las probabilidades correspondientes.

Definition 133 El valor medio o media de una distribución con función de densidad f se denota por μ y se define como

$$\mu = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

La media se conoce también como la esperanza matemática de la variable X , lo cual se denota por $E(X)$, siempre y cuando la integral existe y la serie converge. En caso contrario, se dice que la distribución no tiene media.

Definition 134 Una distribución f es simétrica con respecto a c si

$$f(c+x) = f(c-x)$$

Theorem 135 Media de una distribución simétrica. Si una distribución es simétrica con respecto a c , y tiene media μ , entonces

$$\mu = c$$

Definition 136 La varianza de una distribución σ^2 mide la dispersión de los valores que puede tomar la variable aleatoria correspondiente, su valor se obtiene de:

$$\sigma^2 = \sum_i f(y_i) (y_i - \mu)^2$$

en el caso de una distribución discreta, o

$$\sigma^2 = \int_{-\infty}^{\infty} f(x) (x - \mu)^2 dx$$

cuando la distribución es continua, la serie converge absolutamente y la integral existe.

Remark 137 Observemos que, si la función de densidad se define como

$$f(x) = \begin{cases} 1 & \text{si } x \in S \\ 0 & \text{si } x \notin S \end{cases}$$

para un espacio muestral S , se tiene

$$\sigma^2 = 0$$

lo cual no tiene interés práctico y por tanto siempre se considera que la varianza es positiva.

$$\sigma^2 > 0$$

Definition 138 La raíz cuadrada positiva σ , de la varianza, se denomina desviación estándar.

Example 139 *Considérese la distribución uniforme con función de densidad*

$$f(x) = \frac{1}{b-a}$$

para

$$a < x < b$$

En este caso la media, la varianza y la desviación estándar son:

$$\begin{aligned}\mu &= \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \cdot \frac{x^2}{2} \Big|_a^b = \frac{a+b}{2} \\ \sigma^2 &= \int_a^b \frac{(x-\mu)^2}{b-a} dx = \frac{1}{b-a} \cdot \frac{(x-\mu)^3}{3} \Big|_a^b = \frac{1}{12} (b-a)^2\end{aligned}$$

Theorem 140 *Transformación lineal. Si una variable X tiene media μ y varianza σ^2 , entonces la variable aleatoria ,*

$$X^* = c_1 x + c_2$$

con $c_1 \neq 0$, tiene media

$$\overline{X^*} = c_1 \mu + c_2$$

y varianza

$$\sigma^{*2} = c_1^2 \sigma^2$$

Theorem 141 *Variable estandarizada. Si una variable aleatoria X tiene media μ y varianza σ^2 , entonces la variable*

$$z = \frac{X - \mu}{\sigma}$$

tiene media 0 y varianza 1. Esta variable se denomina variable estandarizada correspondiente a X .

Definition 142 *Si X es cualquier variable aleatoria y $g(X)$ es cualquier función continua definida para toda X real, entonces la esperanza matemática de $g(X)$ es*

$$E(g(X)) = \sum_i g(x_i) f(x_i)$$

cuando la variable X es discreta y f es su función de masa de probabilidad.

Definition 143 *La esperanza matemática de $g(X)$ es*

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

para X es continua y f , su función de densidad.

Definition 144 *Para una variable aleatoria X (discreta o continua) y*

$$g(X) = X^k$$

las expresiones

$$E(X^k) = \sum_i x_i^k f(x_i)$$

y

$$E(X^k) = \int_{-\infty}^{\infty} x^k f(x) dx$$

se denominan k -ésimo momento de X , respectivamente.

Definition 145 Si

$$g(X) = (X - \mu)^k$$

entonces

$$E\left([X - \mu]^k\right) = \sum_i (x_i - \mu)^k f(x_i)$$

y

$$E\left([X - \mu]^k\right) = \int_{-\infty}^{\infty} (x - \mu)^k f(x) dx$$

se denominan k -ésimo momento central de X (en el caso discreto y en el caso continuo, respectivamente).

Remark 146 Nótese que

$$\begin{aligned} E(X^0) &= 1 \\ E(X) &= \mu \\ E\left([X - \mu]^2\right) &= \sigma^2 \end{aligned}$$

10.1.4 Distribución de probabilidad para una variable aleatoria normal¹⁰

Definition 147 La función de densidad de una variable aleatoria normal x , con media μ y varianza σ^2 , es

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-0.5\left[\frac{x-\mu}{\sigma}\right]^2}$$

y su función de distribución es

$$N(\mu, \sigma)(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-0.5\left[\frac{u-\mu}{\sigma}\right]^2} du$$

Definition 148 La distribución estándar normal es una distribución normal con $\mu = 0$ y $\sigma = 1$, cuya variable se denota con la letra z y se denomina, variable aleatoria normal estándar.

Proposition 149 Si x es una variable aleatoria normal con media μ y desviación estándar σ , entonces la variable aleatoria z , estandarización de x o z -score, definida por la fórmula

$$z = \frac{x - \mu}{\sigma}$$

¹⁰McClave, James and Benson, P. George. 2001. Statistics for Business and Economics. New York: Macmillan Publishing Company, capítulo 5.

tiene una distribución normal estándar.

El valor de z describe el número de desviaciones estándares entre x y μ , es decir, representa la distancia entre una medición dada x y la media μ , es por esa razón una medida de la posición relativa de una medición.

Pasos para encontrar una probabilidad correspondiente a una variable aleatoria normal.

1. Dibujar la distribución normal e indicar la media de la variable aleatoria x . Iluminar después el área correspondiente a la probabilidad que se quiere encontrar.
2. Convertir las fronteras del área iluminada de los valores de x a los valores de la variable aleatoria normal estándar z usando la fórmula

$$z = \frac{x - \mu}{\sigma}$$

3. Mostrar los valores de z bajo los correspondientes valores de x en el dibujo.
4. Usar la tabla de áreas bajo la curva normal para encontrar las áreas correspondientes a los valores de z . Si es necesario, usar la simetría de la distribución normal para encontrar las áreas correspondientes a los valores negativos de z y el hecho de que el área total, a cada lado de la media es igual a 0.5, para convertir las áreas de la tabla a las probabilidades del evento que se ha marcado.

Example 150 Si z es la variable aleatoria normal estándar, ¿qué probabilidad tiene de que su valor esté entre -1.33 y 1.33 ?

Solution 151 Usando una tabla convencional de áreas bajo la curva normal estándar y usando la simetría de la curva, se tiene que

$$P(-1.33 \leq z \leq 1.33) = 2P(0 \leq z \leq 1.33) = 2(0.4082) = 0.8164$$

Example 152 Encontrar la probabilidad de que una variable aleatoria normal estándar exceda el valor de 1.64 ; esto es, encontrar

$$P(z > 1.64)$$

Solution 153 La probabilidad solicitada se calcula como sigue

$$P(z > 1.64) = 0.5 - P(z \leq 1.64) = 0.5 - 0.4495 = 0.0505$$

Example 154 Determinar la probabilidad de que una variable aleatoria normal estándar caiga a la izquierda de 0.67 , es decir, que $z < 0.67$.

Solution 155 La probabilidad que se solicita es

$$P(-\infty < z < 0.67) = P(-\infty < z < 0) + P(0 \leq z < 0.67) = 0.5 + 0.2486 = 0.7486$$

Example 156 Encontrar la probabilidad de que el valor absoluto de una variable aleatoria normal estándar exceda a 1.96 .

Solution 157 Si

$$|z| > 1.96$$

entonces, de la definición de valor absoluto, se satisfacen las siguientes desigualdades

$$\begin{aligned} z &> 1.96 \\ z &< -1.96 \end{aligned}$$

Así que la cuestión es determinar

$$P(|z| > 1.96) = P(z < -1.96) + P(z > 1.96)$$

O bien,

$$\begin{aligned} P(|z| > 1.96) &= P(z < -1.96) + [0.5 - P(0 \leq z < 1.96)] \\ &= 2[0.5 - P(0 \leq z < 1.96)] \end{aligned}$$

usando las propiedades de la función de densidad normal estándar. Finalmente, se obtiene

$$P(|z| > 1.96) = 2(0.025) = 0.05$$

10.1.5 Probabilidad condicional

Definition 158 Considérese dos sucesos o eventos, E_1 y E_2 . La probabilidad de que ocurra E_2 dado que ha ocurrido E_1 , se denota por

$$P(E_2|E_1)$$

Esta probabilidad se denomina condicional de que ocurra E_2 cuando E_1 se ha presentado.

Definition 159 Si la ocurrencia de E_1 no afecta la probabilidad de ocurrencia de E_2 , entonces

$$P(E_2|E_1) = P(E_2)$$

y se dice que los eventos E_1 y E_2 son sucesos independientes. En caso contrario, los eventos son dependientes. Esto es, cuando

$$P(E_2|E_1) \neq P(E_2)$$

Definition 160 El evento compuesto E_1 y E_2 , denotado por E_1E_2 , tiene probabilidad

$$P(E_1E_2) = P(E_1)P(E_2|E_1)$$

En particular, si los eventos son independientes

$$P(E_1E_2) = P(E_1)P(E_2)$$

Proposition 161 La probabilidad condicional de que ocurra E_2 dado que E_1 ha ocurrido es

$$P(E_2|E_1) = \frac{P(E_1E_2)}{P(E_1)}$$

10.1.6 Teorema de Bayes

Definition 162 Si un espacio muestral S es dividido en los n eventos siguientes E_1, \dots, E_n , los cuales son mutuamente excluyentes y colectivamente exhaustivos, el conjunto $\{E_1, \dots, E_n\}$, se denomina una partición del espacio S .

Theorem 163 De Bayes. Sea A un evento en S y $\{E_1, \dots, E_n\}$ una partición del espacio S , entonces

$$P(E_i|A) = \frac{P(E_i) P(A|E_i)}{P(E_1) P(A|E_1) + \dots + P(E_n) P(A|E_n)}$$

Example 164 Una empresa cuenta con cuatro áreas de producción. En la primera trabaja el 40% de los empleados especializados, en la segunda área se desempeña el 28%, en la tercera 20% y en la cuarta el 12% restante. Cada área de producción cuenta con tres módulos en los que labora el número de personas que se indica en la siguiente tabla.

Área	E_i	Personas en módulos de producción			Total de personas por área	
		Personal especializado	Módulo 1	Módulo 2		Módulo 3
1	E_1	40%	3	5	2	10
2	E_2	28%	2	3	2	7
3	E_3	20%	1	2	2	5
4	E_4	12%	1	1	1	3

25

Si se elige aleatoriamente un área de producción y un empleado especializado y ese empleado labora en el módulo 3, ¿cuál es la probabilidad de que ese empleado pertenezca al área 1?

Solution 165 El evento A , en este caso, es que el empleado labore en el módulo 3. Luego, la probabilidad de que ese empleado pertenezca al área 1, dado que labora en el módulo 3, se expresa y calcula como sigue.

$$\begin{aligned} P(E_1|\text{Modulo}_3) &= \frac{P(E_1) P(\text{Modulo}_3|E_1)}{P(E_1) P(\text{Modulo}_3|E_1) + \dots + P(E_4) P(\text{Modulo}_3|E_4)} \\ &= \frac{(0.40)(0.2)}{(0.40)(0.2) + (0.28)(0.2857142857) + (0.2)(0.4) + (0.12)(0.3)} \\ &= 0.2857142857 \end{aligned}$$

10.2 Técnicas de muestreo

Definition 166 Sean N y n el número de elementos en la población y en la muestra, respectivamente. Si se hace el muestreo de tal manera que cada una de las muestras tiene la misma probabilidad de ser escogida, el muestreo se denomina aleatorio y el resultado es una muestra aleatoria.

Para una población con N elementos y una muestra de tamaño n , el muestreo aleatorio es aleatorio si cualquiera de las

$$\frac{N!}{n!(N-n)!}$$

posibles muestras, tiene igual oportunidad de ser seleccionada.

10.2.1 Definiciones básicas¹¹

1. Un elemento es un objeto en el cual se toman mediciones
2. Una población es una colección de elementos acerca de los cuales deseamos hacer alguna inferencia.
3. Las unidades de muestreo son colecciones no traslapadas de elementos de la población que cubren la población completa.
4. Un marco es una lista de unidades de muestreo.
5. Una muestra es una colección de unidades seleccionadas de un marco o de varios marcos.

Definition 167 *Si un tamaño de muestra n es seleccionado de una población de tamaño N de tal manera que cada muestra posible de tamaño n tiene la misma probabilidad de ser seleccionada, el procedimiento de muestreo se denomina muestreo irrestricto aleatorio. A la muestra así obtenida se le llama muestra irrestricta aleatoria.*

Definition 168 *Una muestra aleatoria estratificada es la que se obtiene mediante la separación de los elementos de la población en grupos que no presenten traslapes, llamados estratos, y la selección posterior de una muestra irrestricta aleatoria simple de cada estrato.*

Los motivos principales para utilizar una muestreo aleatorio estratificado en lugar del muestreo irrestricto aleatorio son los siguientes:

1. La estratificación puede producir un límite más pequeño para el error de estimación que el que se generaría por una muestra irrestricta aleatoria del mismo tamaño. Este resultado es particularmente cierto si las mediciones dentro de los estratos son homogéneas.
2. El costo por observación en la encuesta puede reducirse mediante la estratificación de los elementos de la población en grupos convenientes.
3. Se pueden obtener estimaciones de parámetros poblacionales para subgrupos de la población. Los subgrupos deben de ser entonces estratos identificables.

Definition 169 *Una muestra obtenida al seleccionar aleatoriamente un elemento de los primeros k elementos en el marco y después cada k -ésimo elemento se denomina muestra sistemática de 1 en k .*

El muestreo sistemático proporciona una opción útil para el muestreo irrestricto aleatorio por las siguientes razones:

1. El muestreo sistemático es más fácil de llevar a cabo en el campo, y por lo tanto, a diferencia de las muestras irrestrictas aleatorias y las muestras aleatorias estratificadas, está menos expuesto a los errores de selección que cometen los investigadores de campo.
2. El muestreo sistemático puede proporcionar mayor información que la que puede proporcionar el muestreo irrestricto aleatorio por unidad de costo.

¹¹Sheaffer, R., Mendenhall, W. Y Ott, L. Elementos de muestreo. Trads.Gilberto Rendón S. y José Roberto Gómez A. México: Grupo Editorial Iberoamérica, 1986, pp. 321

Definition 170 Una muestra por conglomerados es una muestra aleatoria en la cual cada unidad es una colección, o conglomerado, de elementos.

El muestreo por conglomerados es menos costoso que el muestreo aleatorio estratificado o ir-restricto, si el costo por obtener un marco que liste todos los elementos poblacionales es muy alto o si el costo por obtener observaciones se incrementa con la distancia que separa los elementos.

Es un diseño efectivo para obtener una cantidad específica de información al costo mínimo bajo las siguientes condiciones:

1. No se encuentra disponible o es muy costoso obtener un buen marco que liste los elementos de la población, mientras que se puede lograr fácilmente un marco que liste los conglomerados.
2. El costo por obtener observaciones se incrementa con la distancia que separa los elementos.

10.2.2 Muestreo con y sin reemplazo¹²

Example 171 *Distribución binomial.* De una caja que contiene N objetos, M son defectuosos, si se extrae un tornillo al azar, la probabilidad de obtener un defectuoso es

$$p = \frac{M}{N}$$

Al extraer una muestra de n tornillos con reemplazo, la probabilidad de que precisamente x de los n tornillos sean defectuosos es

$$f(x) = \binom{n}{x} \left(\frac{M}{N}\right)^x \left(1 - \frac{M}{N}\right)^{n-x}$$

con

$$x = 0, 1, \dots, n$$

Esta distribución binomial tiene media

$$\mu = n \frac{M}{N}$$

y varianza

$$\sigma^2 = n \frac{M}{N} \left(1 - \frac{M}{N}\right)$$

Example 172 *Distribución hipergeométrica.* En el muestreo sin reemplazo, la probabilidad de obtener un defectuoso es

$$f(x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$

para

$$x = 0, 1, \dots, n$$

¹²Kreyszig, Erwin (1990) Matemáticas avanzadas para Ingeniería. Vol. 2. México: LIMUSA, 551.

Esta distribución hipergeométrica tiene media

$$\mu = n \frac{M}{N}$$

y varianza

$$\sigma^2 = \frac{nM(N-M)(N-n)}{N^2(N-1)}$$

respectivamente.

Donde

$$\binom{N}{n}$$

es el número de maneras distintas de tomar n cosas de un total de N .

$$\binom{M}{x}$$

el número de formas distintas de tomar x defectuosas de un total de M defectuosas.

Y ,

$$\binom{N-M}{n-x}$$

es la cantidad de maneras distintas de tomar $n-x$ no defectuosas de un total de $N-M$, no defectuosas.

10.3 Distribuciones de muestreo

El conocimiento de las variables aleatorias y sus distribuciones de probabilidad nos permite construir modelos teóricos de poblaciones. Debido a que las mediciones en la muestra son valores observados de variables aleatorias, el valor de un estadístico de la muestra variará de una manera aleatoria de una muestra a otra. En otras palabras, puesto que los estadísticos de la muestra son variables aleatorias, poseen distribuciones de probabilidad que son discretas o continuas. En esta sección se explorarán las distribuciones de probabilidad de esos estadísticos, denominados distribuciones de muestreo. Se aprenderá el porqué, muchas distribuciones de muestreo tienden a ser aproximadamente normales, y veremos cómo las distribuciones de muestreo pueden usarse para evaluar la confiabilidad de las inferencias usando estadísticos. A continuación recordemos algunas definiciones básicas.

Definition 173 *Un parámetro es una medida descriptiva de una población, su valor es casi siempre desconocido, por lo que suele estimarse con base en las observaciones realizadas en la población, .*

Definition 174 *Un estadístico de una muestra es una medida numérica descriptiva de la muestra, que se calcula a partir de las observaciones realizadas en la muestra.*

Definition 175 *La distribución de muestreo de un estadístico de una muestra, calculado con base en una muestra con n mediciones, es la distribución de probabilidad del estadístico.*

10.3.1 Propiedades de la distribución de muestreo de \bar{x}

La media de la distribución de muestreo de \bar{x} es igual a la media de la población muestreada. Esto es,

$$\mu_{\bar{x}} = E(\bar{x}) = \mu$$

La desviación estándar de una distribución de muestreo de \bar{x} o error estándar de la media, es igual a

$$\frac{\text{Desviación estándar de la población muestreada}}{\text{Raíz cuadrada del tamaño de la muestra}}$$

Esto es,

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Si el tamaño de la muestra n es relativamente grande comparada con el tamaño N de la población, por ejemplo un 5% o más, entonces el cociente debe multiplicarse por un factor de corrección

$$\sqrt{\frac{N-n}{N-1}}$$

Para la mayoría de las situaciones de muestreo, este factor de corrección será cercano a 1 y puede ignorarse.

Theorem 176 *Si una muestra aleatoria de n observaciones se selecciona de una población con una distribución normal, la distribución de muestreo de \bar{x} , será una distribución normal.*

Theorem 177 *Del Límite Central. Considere una muestra aleatoria de n observaciones seleccionada de una población (cualquier población), con media μ y desviación estándar σ . Entonces, cuando n es suficientemente grande, la distribución de muestreo de \bar{x} , será aproximadamente una distribución normal con media*

$$\mu_{\bar{x}} = \mu$$

y desviación estándar

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Entre más grande sea el tamaño de la muestra, mejor será la aproximación de la distribución de muestreo de \bar{x} .

10.4 Estimación

Es frecuente que la única información que se puede obtener de una población es a partir de un estudio de muestras tomadas al azar de la población. La inferencia estadística, o teoría de la estimación estadística es el procedimiento mediante el cual los parámetros de la población se estiman a partir del estudio de muestras tomadas de esa población. Una estimación es un valor particular del estimador asociado a una muestra.

Definition 178 *Un estimador por punto de un parámetro poblacional es una regla o fórmula que nos dice como usar los datos de la muestra para calcular un solo número que puede usarse como un estimado del parámetro poblacional.*

Definition 179 *Un estadístico es una función de los elementos de una muestra que no contiene parámetros desconocidos.*

Example 180 *La media y la varianza muestrales, \bar{x} y s^2 , son ejemplos de estimadores puntuales de los respectivos parámetros μ y σ^2 .*

Examinando la distribución de muestreo podemos determinar que tan grande podría ser la diferencia entre un estimado y el valor verdadero del parámetro (llamado error de estimación). También podemos decir si un estimador es más probable que sobrestime o subestime al parámetro.

Definition 181 *Si la distribución de muestreo de un estadístico de una muestra tiene una media igual al parámetro de la población del estadístico que se pretende estimar, el estadístico se dice que es una estimación insesgada del parámetro. Si la media de la distribución de muestreo no es igual al parámetro, se dice que el estadístico es un estimador sesgado del parámetro.*

Los estimadores deben de cumplir las siguientes propiedades, entre otras, para considerárseles buenos estimadores puntuales.

1. La propiedad de insesgamiento. Por ejemplo, el promedio a largo plazo $E(\bar{x}) = \mu$.
2. Varianza mínima. El estimador puntual de mínima varianza posee la varianza más pequeña que cualquier otro estimador.

Proposition 182 \bar{x} y s^2 son estimadores insesgados de μ y σ^2 , respectivamente.

Proof. En el caso de la media, se tiene

$$\begin{aligned} E(\bar{x}) &= E\left(\frac{\sum x_i}{n}\right) \\ &= \frac{1}{n} \sum E(y_i) \\ &= \frac{1}{n} \sum \mu \\ &= \mu \end{aligned}$$

Para la varianza,

$$\begin{aligned}
 E(s^2) &= E\left(\frac{\sum (x_i - \bar{x})^2}{n-1}\right) \\
 &= \frac{1}{n-1} E\left(\sum (x_i - \bar{x})^2\right) \\
 &= \frac{1}{n-1} E(SS) \\
 &= \sigma^2
 \end{aligned}$$

Donde, la suma corregida SS , cumple con lo siguiente

$$\begin{aligned}
 SS &= \sum (x_i - \bar{x})^2 \\
 E(SS) &= E\left(\sum (x_i - \bar{x})^2\right) \\
 &= E\left(\sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2)\right) \\
 &= E\left(\sum x_i^2 - 2\bar{x} \sum x_i + \sum \bar{x}^2\right) \\
 &= E\left(\sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right) \\
 &= E\left(\sum x_i^2 - n\bar{x}^2\right) \\
 &= \sum (\mu^2 + \sigma^2) - n\left(\mu^2 + \frac{\sigma^2}{n}\right) \\
 &= (n-1)\sigma^2
 \end{aligned}$$

donde $n-1$, son los grados de libertad de la suma de cuadrados. ■

De manera general decimos que si X es una variable aleatoria con varianza σ^2 y la suma corregida

$$SS = \sum (x_i - \bar{x})^2$$

tiene ν grados de libertad, entonces:

$$E\left(\frac{SS}{\nu}\right) = \sigma^2$$

Es decir, los grados de libertad de una suma de cuadrados es igual al número de términos independientes en dicha suma. En el caso de

$$E(SS) = E\left(\sum (x_i - \bar{x})^2\right)$$

los términos $x_i - \bar{x}$, no todos son independientes porque

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

, sólo $n - 1$ son independientes.

La media teórica de una distribución de muestreo de varianzas es igual a

$$\frac{n-1}{n}\sigma^2$$

donde σ^2 es la varianza de la población, n el tamaño de la muestra.

La varianza de una muestra es un estimador sesgado de la varianza de la población, pero es un estimador insesgado de

$$\frac{n-1}{n}\sigma^2$$

De otro modo,

$$\frac{n}{n-1}\sigma_s^2$$

donde σ_s^2 es la varianza de una muestra, o la media de una cantidad de varianzas muestrales, es un estimador insesgado de la varianza de la población σ^2 .

10.4.1 Momentos con respecto a la media

Consideremos un conjunto discreto de n datos agrupados en k clases, cuyos representantes de clase son x_i , y sus respectivas frecuencias f_i , la media se define como

$$\bar{x} = \frac{\sum_{i=1}^k f_i x_i}{n}$$

El primer momento respecto a la media es

$$m_1 = \frac{\sum f_i (x_i - \bar{x})^1}{n} = 0$$

El segundo momento es igual a la varianza

$$m_2 = \frac{\sum f_i (x_i - \bar{x})^2}{n} = \sigma^2$$

En general, el r -ésimo momento respecto a la media es

$$m_r = \frac{\sum f (x - \bar{x})^r}{n}$$

Los primeros cuatro momentos respecto al origen son m'_1, m'_2, m'_3 y m'_4 .

10.4.2 k-estadísticos

Los momentos respecto a la media, calculados con una muestra individual, son aproximaciones de los momentos de la población. Si la muestra no es grande, las funciones denominadas k-estadísticos son mejores aproximaciones.

k-estadístico de la muestra	Estimación del estadístico de la población
$k_1 = \bar{x} = m_1'$	μ_1'
$k_2 = \frac{N}{N-1}\sigma^2 = \frac{N}{N-1}m_2$	μ_2
$k_3 = \frac{N^2}{(N-1)(N-2)}m_3$	μ_3
$k_4 = \frac{N^2}{(N-1)(N-2)(N-3)} [(N+1)m_4 - 3(N-1)m_2^2]$	$\mu_4 - 3\mu_2^2$

La media y la mediana son ejemplos de estimadores insesgados de la media de la población, por el hecho de que la distribución de muestreo de las medias es una curva normal.

La varianza de la muestra es un estimador sesgado de la varianza de la población.

10.4.3 Consistencia

Los estimadores consistentes de un parámetro son estimaciones que se hacen más precisos conforme se aumenta el tamaño de la muestra. Es decir, una estimación es consistente si la probabilidad de que difiera del verdadero valor del parámetro en una cantidad pequeña dada, tiende a uno conforme el tamaño de la muestra es tan grande como se quiera.

10.4.4 Estimaciones puntuales y por intervalos

Las estimaciones de los parámetros dadas por un valor se denominan puntuales, mientras que las estimaciones por intervalos proveen un rango de variación del parámetro, como estimación. Esta última estimación es más satisfactoria porque no solo indica un valor probable sino que proporciona la confiabilidad de la estimación.

10.4.5 Intervalo de confianza para la media de la población, con muestras grandes

Para muestras grandes, $n \geq 30$, la distribución de muestreo de la media es aproximadamente normal, por lo que es posible encontrar un valor de la media dentro del rango

$$[\mu - \sigma, \mu + \sigma]$$

en 68% de las muestras.

Para cualquier otro porcentaje y con ayuda de una tabla de áreas bajo la curva normal, se determina el rango en el que se espera encontrar la media de una muestra. El porcentaje seleccionado se denomina nivel de confianza. Por ejemplo, con un nivel de confianza del 80% el intervalo de confianza es

$$[\mu - 1.281551566\sigma, \mu + 1.281551566\sigma]$$

¹³, cuyos límites, inferior y superior, son los valores

$$\mu - 1.281551566\sigma$$

¹³El valor 1.281551566, se obtuvo con Maple6, usando la secuencia de comandos

`with(stats) : statevalf[icdf, normald[0, 1]](.9);`

y

$$\mu + 1.281551566\sigma$$

respectivamente.

El factor 1.281551566, se denomina coeficiente de confianza z_c .

Actividad. Encuentra los niveles de confianza o los coeficientes de confianza que faltan en la siguiente tabla.

Nivel de confianza	Coficiente de confianza z_c
99.73%	3.00
99%	2.58
98%	
	2.17
96%	
95.45%	
	1.96
	1.64
80%	
75%	
	1.00
50%	

¿Qué estimación se puede hacer de la media de una población y que nivel de confianza se le asigna? La media de una distribución de medias muestrales es un estimador no sesgado de la población, en el caso de una población infinita la media de la población se estima por

$$\bar{x} \pm z_c \frac{\sigma}{\sqrt{n}}$$

donde \bar{x} es una media muestral, n su tamaño y σ es la desviación estándar de la población. Si se desconoce el valor de σ , se usa la estimación $\frac{s}{\sqrt{n}}$.

Cuando la población es finita, de tamaño N , la formula se ajusta como sigue

$$\bar{x} \pm z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

donde N , es el tamaño de la población.

Example 183 *Determine que estimaciones son más confiables*

- a) 3.25 ± 0.13 ó 7.23 ± 0.13
- b) 3.25 ± 0.03 ó 3.25 ± 0.13
- c) Compare la respuesta al punto b, con la situación 1 ± 1 ó 100 ± 1 .

Example 184 *Si la estimación de la media de una población, basada en una muestra con 100 elementos es*

$$6.36 \pm 1.5$$

¿cuál sería la estimación, del mismo parámetro de la misma población, basada en una muestra con 200 elementos?

Solution 185 El intervalo de confianza, con el mismo nivel de confianza, basado en una muestra con 100 elementos, para la media se expresa como

$$\bar{x} \pm z_c \frac{\sigma}{\sqrt{N}}$$

de ahí que, conservando para la misma media muestral,

$$\begin{aligned} z_c \frac{\sigma}{\sqrt{100}} &= 2.5 \\ z_c \sigma &= 2.5\sqrt{100} \end{aligned}$$

De donde, la estimación por intervalos de la media de la misma población, basada en una muestra de tamaño 200 es

$$\begin{aligned} 6.36 \pm z_c \frac{\sigma}{\sqrt{N}} \\ 6.36 \pm 2.5 \frac{\sqrt{100}}{\sqrt{200}} \\ 6.36 \pm 1.767766952 \end{aligned}$$

Example 186 Si el error probable de un cierto estadístico que se distribuye normalmente es 2.0, ¿cuál es el intervalo del 90% de confianza?

10.4.6 Intervalo de confianza para la media de la población, con muestras pequeñas

Con muestras pequeñas nos enfrentamos a los dos siguientes problema para los que además se da una solución.

Problem 187 La forma de la distribución de muestreo para la media de muestreo \bar{x} (y el estadístico z) ahora depende de la forma de la población de la cual se muestrea. Ya no podemos suponer que la distribución de muestreo de \bar{x} es aproximadamente normal, debido a que, el Teorema del Límite Central asegura la normalidad solo para las muestras que son suficientemente grandes.

Solution 188 De acuerdo al teorema que dice: si una muestra aleatoria de n observaciones es seleccionada de una población con una distribución normal, la distribución de muestreo de \bar{x} será una distribución normal, la distribución de muestreo de (\bar{x} y de z) es exactamente normal aún para muestras relativamente pequeñas si la población de la que es muestreada es normal. Y, es aproximadamente normal si la población de la que es muestreada es aproximadamente normal.

Problem 189 La desviación estándar de la población σ es casi siempre desconocida. Aunque todavía es verdadero que

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

la desviación estándar de la muestra s puede proveer un aproximación pobre de σ , cuando el tamaño de la muestra es pequeño.

Solution 190 *En lugar de usar el estadístico normal estándar*

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma} \sqrt{n}$$

que requiere del conocimiento de una buena aproximación de σ , definimos y usamos el estadístico

$$t = \frac{\bar{x} - \mu}{s} \sqrt{n}$$

donde, la desviación estándar de la muestra s , reemplaza a la desviación estándar de la población σ .

Para una muestra pequeña de menos de 30 elementos, los intervalos de confianza, construidos en la sección anterior, sobreestiman la confiabilidad de la estimación, por lo que se les reemplaza como sigue.

Definition 191 *Intervalo de confianza del $100(1 - \alpha)\%$ para la media con una muestra pequeña*

$$\bar{x} \pm t_{\alpha/2} \frac{\sigma}{\sqrt{n-1}}$$

donde $t_{\alpha/2}$ está basado en $(n - 1)$ grados de libertad y se obtiene de la distribución $t - Student$.

Suposiciones: Una muestra aleatoria es seleccionada de una población con una distribución de frecuencias relativas que es aproximadamente normal.

Problem 192 *¿Qué se hace cuando la distribución de frecuencias relativas de la población se aleja enormemente de la normalidad?*

Solution 193 *Se usan los métodos estadísticos no paramétricos.*

Muchas técnicas estadísticas suponen que la variable aleatoria está normalmente distribuida, lo cual está justificado por el teorema del límite central que establece

Theorem 194 *del Límite Central. Si x_1, x_2, \dots, x_n es una sucesión de n variables aleatorias independientes con*

$$\begin{aligned} E(x) &= \mu \\ V(x_i) &= \sigma^2 \end{aligned}$$

finitas, y

$$x = x_1 + x_2 + \dots + x_n$$

entonces

$$z_n = \frac{x - n\mu}{\sqrt{n\sigma^2}}$$

tiene aproximadamente una distribución normal, en el sentido de que si

$$F_n(z)$$

es la función de distribución de z_n y si $\phi(z)$ es la función de distribución de la variable aleatoria normal estándar, $N(0, 1)$, entonces

$$\lim_{n \rightarrow \infty} \frac{F_n(z)}{\phi(z)} = 1$$

Remark 195 Este teorema dice, de manera breve, que la suma de n variables aleatorias independientes tiene, aproximadamente, una distribución normal.

En muchos casos la aproximación es buena si $n < N$, mientras que en otros se requiere que $n > 100$.

Frecuentemente, se cree que el error en un experimento es consecuencia de varias fuentes independientes que se combinan en forma aditiva, de ahí que la distribución normal sea un modelo convincente, en el caso del error experimental combinado.

10.4.7 Distribuciones muestrales que se definen en términos de variables aleatorias normales

Sean z_1, z_2, \dots, z_n , n -variables aleatorias normales, con media 0 y varianza 1. Supongamos que ν de esas variables son independientes. La variable

$$\chi^2 = z_1^2 + z_2^2 + \dots + z_\nu^2$$

tiene la distribución Ji o Chi, cuadrada con ν grados de libertad.

La definición de Chi cuadrada es equivalente a

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

dado que

$$z = \frac{x - \mu}{\sigma}$$

y $\nu = n - 1$.

Definition 196 Proposition 197 Definition 198 La función de densidad Chi cuadrada es

$$f(\chi^2) = \frac{1}{s^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)} (\chi^2)^{\left(\frac{\nu}{2}-1\right)} e^{-\frac{\chi^2}{s^2}}$$

donde $\chi^2 > 0$.

Proposition 199 Esta distribución es asimétrica o sesgada con media ν y varianza 2ν .

Example 200 Considérese la muestra aleatoria x_1, x_2, \dots, x_n , con distribución $N(\mu, \sigma^2)$, entonces la variable aleatoria

$$\frac{SS}{\sigma^2} = \frac{\sum (x_i - \bar{x})^2}{\sigma^2} \sim \chi_{n-1}^2$$

es decir, tiene una distribución Ji cuadrada, con $n - 1$ grados de libertad.

Remark 201 Nótese que si las observaciones en la muestra se distribuyen normalmente y son independientes, lo que denotamos con $NID(\mu, \sigma^2)$, entonces la distribución de s^2 es

$$\frac{\sigma^2}{n-1} \chi_{n-1}^2$$

porque

$$s^2 = \frac{SS}{n-1}$$

y

$$SS \sim \sigma^2 \chi_{n-1}^2$$

Lo que significa que si la población se distribuye normalmente, la distribución muestral de la varianza de las muestras independientes es una constante multiplicada por una distribución Ji cuadrada.

Definition 202 Si z y χ_{n-1}^2 son las variables aleatorias independientes normal estándar y Ji cuadrada, respectivamente, entonces la variable aleatoria

$$t_k = \frac{z}{\sqrt{\frac{\chi_k^2}{k}}}$$

tiene una distribución t -student con k grados de libertad. La función de densidad correspondiente es

$$f(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\Gamma\left(\frac{k}{2}\right)} \frac{1}{\left[\frac{t^2}{k} + 1\right]^{\frac{k+1}{2}}}$$

con

$$-\infty < t < \infty$$

Esta función tiene media 0 y varianza $\frac{k}{k-2}$ para $k > 2$.¹⁴

Exercise 203 Para la muestra aleatoria x_1, x_2, \dots, x_n , de tarifas salariales dada en la siguiente tabla, con distribución $N(\mu, \sigma^2)$, la variable

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

tiene una distribución t -student con $n - 1$ grados de libertad. Complete la tabla, para los valores grados de libertad, señalados.

¹⁴La función gamma evaluada en $\frac{k}{2}$ esta dada por la ecuación

$$\Gamma\left(\frac{k}{2}\right) = \left(\frac{k}{2} - 1\right)!$$

Flores, R. y Lozano, H. (1998). Estadística aplicada para administración. Mexico: Grupo Editorial Iberoamerica, p. 360.

Tarifas salariales

X	χ_k^2	z	$t(3)$	$t(5)$	$t(15)$
43.97	$3.10424E - 05$	1.512319214			
41.47	$7.99181E - 05$	1.245044055			
37.35	0.000365061	0.661315108			
36.01	0.000591229	0.661315108			
35.19	0.000791598	0.573648856			
34.87	0.000886483	0.539437635			
31	0.003371619	0.125695689			
28.92	0.006719108	-0.096677243			
27.97	0.009137417	-0.198241804			
27.21	0.011642429	-0.279493452			
22.17	0.052783079	-0.818320173			
22.06	0.054439064	-0.83008028			
20.25	0.08914949	-1.023587495			
9.1	0.765349729	-2.215634704			

10.5 Inferencias basadas en una muestra

10.5.1 Prueba de hipótesis

La información de una muestra se usa para probar qué valor puede tomar un parámetro de una población, como la media o la proporción y medir la confiabilidad de la inferencia. Por ejemplo, cuando se quiere determinar si el tiempo de espera promedio en una fila es menor de un cierto valor, digamos cinco minutos. Otro ejemplo, si se quiere determinar la opinión promedio sobre un tópico. En ambos casos se está interesado en hacer una inferencia acerca del valor de un parámetro respecto de uno específico. Si es menor, igual o mayor que ese valor específico. Este tipo de inferencia se denomina prueba de hipótesis (contraste de hipótesis).

Los elementos de una prueba de hipótesis son los siguientes:

1. Hipótesis nula (H_o): Es una afirmación o teoría acerca de los valores de uno o más parámetros. La teoría en general representa el status quo, la cual se adopta hasta que se prueba su falsedad.
2. Hipótesis alternativa o de investigación (H_a): Es una afirmación o teoría que contradice la hipótesis nula. Teoría que se adopta solo si existe evidencia suficiente para establecer su veracidad.
3. Nivel de significancia de la prueba α . Usualmente el valor de α se elige pequeño (0.01, 0.05 o 0.10).
4. Estadístico de prueba: Un estadístico muestral que se usa para decidir cuando rechazar la hipótesis nula.
5. Región de rechazo: Conjunto de valores numéricos del estadístico de prueba, para los cuales la hipótesis nula se rechaza. La región de rechazo se determina por el valor crítico o teórico del estadístico de prueba, tal que, la probabilidad de que contenga al estadístico de prueba, cuando la hipótesis nula es verdadera, sea α .
Es decir, que la probabilidad de cometer un error del Tipo I, rechazar la hipótesis nula, siendo ésta verdadera, sea igual a α .
6. Suposiciones: Establecer claramente, las suposiciones que se hagan sobre la población que se muestrea.
7. Experimento y cálculo del valor empírico del estadístico de prueba. Realizar un experimento de muestreo y determinar el valor numérico del estadístico de prueba con los datos provenientes del muestreo.
8. Conclusión: Si el valor empírico del estadístico de prueba cae en la región de rechazo, rechazar la hipótesis nula y concluir que la hipótesis alternativa se acepta como verdadera.

10.5.2 Regiones de rechazo de la hipótesis nula

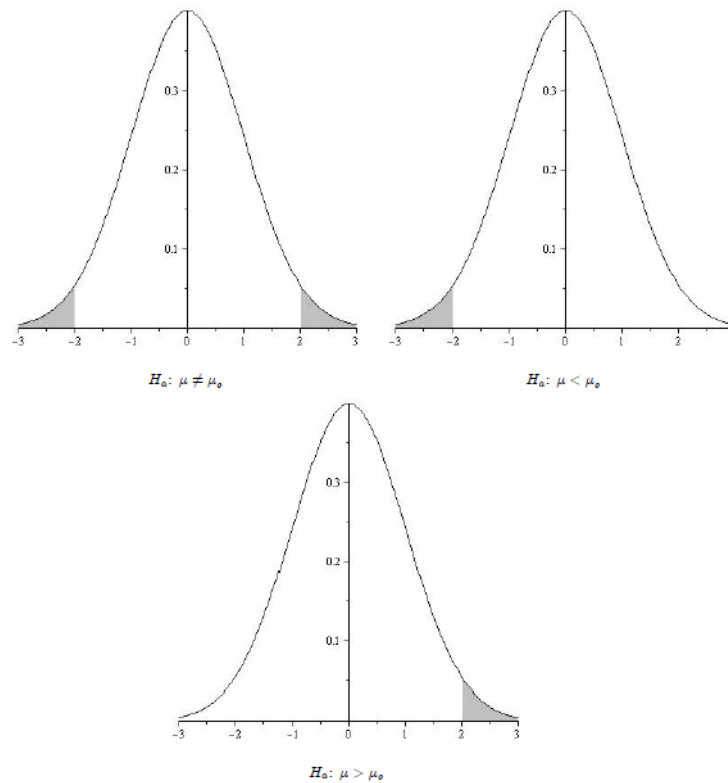


Figure 18

10.5.3 Tipos de errores. Conclusiones y consecuencias para una prueba de hipótesis

Se sabe que el proceso de prueba de hipótesis puede conducir a una conclusión incorrecta, cometer un error del Tipo I, en $100\alpha\%$ de las veces, cuando se rechaza H_0 , siendo que es verdadera. Si el estadístico de prueba no cae en la región de rechazo, es decir, cae en la región de aceptación, entonces no se rechaza la hipótesis nula. Se reserva el juicio sobre qué hipótesis es verdadera. No se concluye que la hipótesis nula es verdadera, porque, en general, no conocemos la probabilidad β , de que el procedimiento de prueba lleva a una aceptación incorrecta de la hipótesis nula, lo que se denomina cometer un error del Tipo II.

En muchas aplicaciones prácticas de la prueba de hipótesis, en los negocios, el no rechazo origina que los administradores se conduzcan como si la hipótesis nula hubiera sido aceptada. La distinción entre aceptación y no rechazo, frecuentemente se hace confusa en la práctica.

Los cuatro resultados posibles de una prueba de hipótesis se resumen en la tabla siguiente. El estado verdadero de la naturaleza se refiere al hecho de que alguna de las hipótesis nula o alternativa, es verdadera. El verdadero estado de la naturaleza es desconocido para el investigador que conduce

la prueba. Los renglones que contienen decisiones, se refieren a la acción del investigador, de asumir como verdadera a H_o ó H_a , basada en los resultados del experimento de muestreo.

El criterio es tomar una decisión sólo cuando se conoce la probabilidad de cometer el error que corresponde a la decisión.

Puesto que el nivel de significancia de la prueba, α , lo especifica usualmente el investigador, por lo general se tendrá la posibilidad de rechazar H_o y de aceptar H_a , cuando la evidencia empírica, de la muestra, apoye tal decisión.

Sin embargo, debido a que β no se especifica, se evitará la decisión de aceptar H_o , prefiriendo en su lugar, decir que la evidencia de la muestra no es suficiente para rechazar a H_o , cuando el estadístico de prueba no cae en la región de rechazo.

	Verdadero estado	de la naturaleza
Conclusión	Ho verdadera	Ha verdadera
Aceptar Ho (asumiendo Ho verdadera)	Decisión correcta	Error del tipo II (probabilidad β)
Rechazar Ho (asumiendo Ha verdadera)	Error del tipo I (probabilidad α)	Decisión correcta

Definition 204 *La probabilidad de cometer un error del Tipo I es*

$$P(\text{Error del Tipo I}) = P(\text{rechazar } H_o | H_o \text{ es verdadera}) = \alpha$$

Definition 205 *La probabilidad de cometer un error del Tipo II es*

$$P(\text{Error del Tipo II}) = P(\text{no rechazar } H_o | H_o \text{ es falsa}) = \beta$$

Summary 206 *A continuación se hace un resumen de las fórmulas para estimación y prueba de hipótesis.*

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$H_0: (\mu_1 - \mu_2) = D_0$$

$$H_a: (\mu_1 - \mu_2) < D_0$$

$$H_a: (\mu_1 - \mu_2) > D_0$$

$$H_a: (\mu_1 - \mu_2) \neq D_0$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$z < -z_{\alpha} \quad , \quad t < -t_{\alpha}$$

$$|z| > z_{\alpha/2} \quad , \quad |t| > t_{\alpha/2}$$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

$$\approx (\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

$$\sigma_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

$$H_0: (p_1 - p_2) = D_0$$

$$H_a: (p_1 - p_2) < D_0$$

$$H_a: (p_1 - p_2) > D_0$$

$$H_a: (p_1 - p_2) \neq D_0$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sigma_{(\hat{p}_1 - \hat{p}_2)}}$$

$$\sigma_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \approx \sqrt{\frac{\hat{p} \hat{q}}{n_1} + \frac{\hat{p} \hat{q}}{n_2}} = \sqrt{\hat{p} \hat{q} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$\text{donde } \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Figure 19

10.6 Regresión y correlación

El siguiente diagrama, resume los puntos esenciales de un análisis de regresión lineal simple y de correlación.

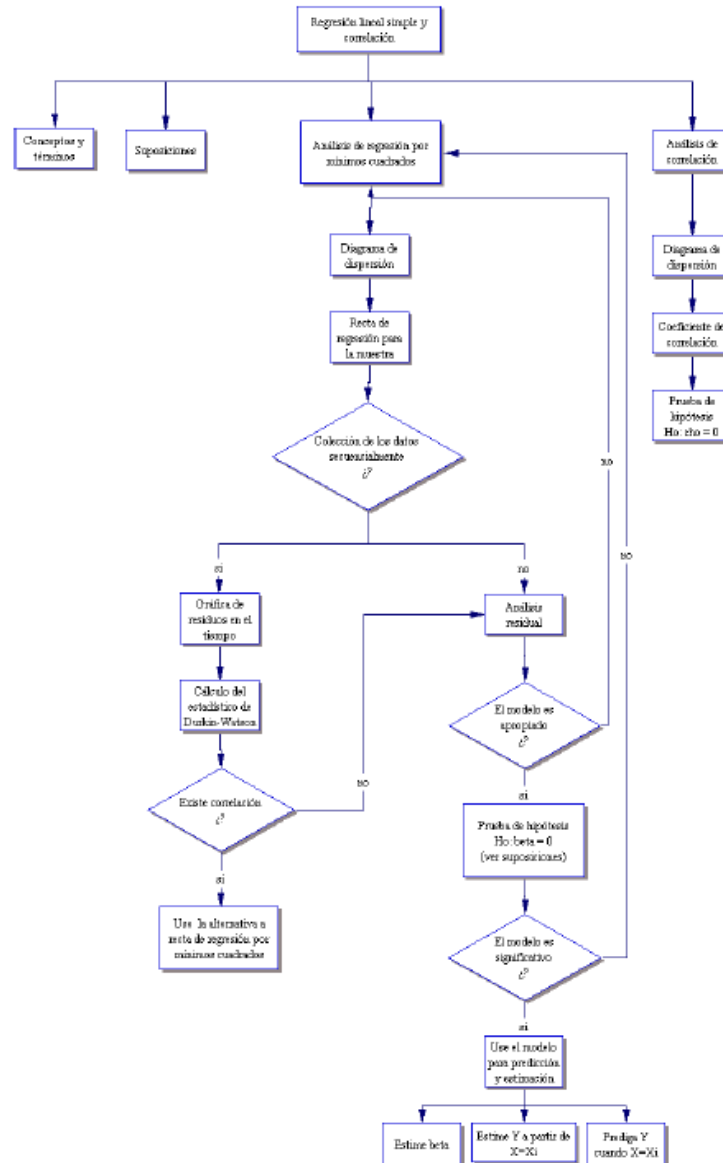


Figure 20

Un análisis completo de la construcción y análisis de un modelo de regresión lineal simple, se discute en el siguiente ejemplo.

Example 207 La siguiente tabla¹⁵ resume las ventas (y) en millones de dólares y el número (x) en miles de empleados por compañía para una muestra aleatoria de 20 compañías clasificadas en la categoría 500 (con una fortuna denominada 500). Para ese conjunto de datos se tiene $SS_{yy} = 1,756,002,651.75$; $SS_{xx} = 28,385.22$; $SS_{xy} = 3,881,560.4$; $\bar{y} = 7,740.75$ y $\bar{x} = 39$.

Compañías	Ventas y (millones de dólares)	Empleados x (miles de empleados)
Procter & Gamble	27,406	94.0
H. B. Fuller	855	5.6
Jonson & Johnson	12,447	82.7
Mattel	1,650	12.5
Intel	4,779	25.1
New York Times	1,703	10.1
GAF	926	4.1
Coltec Industries	1,373	11.4
General Mills	7,153	108.1
ITT Rayonier	979	3.0
Texaco	37,551	40.2
Westinghouse Electric	12,794	113.7
Tribune	2,035	12.9
Allied-Signal	11,882	98.3
Polaroid	2,096	12.0
Legget & Platt	1,082	10.4
Gillette	4,706	31.2
Georgia-Pacific	11,524	57.0
Compaq Computer	3,271	10.0
Merck	8,603	37.7

Figure 21

1. Realice un análisis de regresión que siga los cinco pasos presentados en el capítulo 10 de McClave, et.al. (2001), y señale claramente, las suposiciones que se hagan.
2. De un intervalo de predicción para una compañía de la misma clase con 50,000 empleados.

Solution 208 Los cinco pasos se exponen a continuación.

Paso 1. Asumimos un modelo probabilístico lineal para relacionar las ventas con el número de empleados.

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

Paso 2. Determinamos los coeficientes de la recta de regresión por mínimos cuadrados:

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = 136.7458276 ; \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 2407.662724$$

¹⁵McClave, 1999, 7^a. ed.

Por lo tanto, la ecuación de regresión de y sobre x , por el método de mínimos cuadrados es

$$\hat{y} = 2407.663 + 136.746x$$

A continuación, se representa gráficamente la recta de regresión por mínimos cuadrados y los datos.

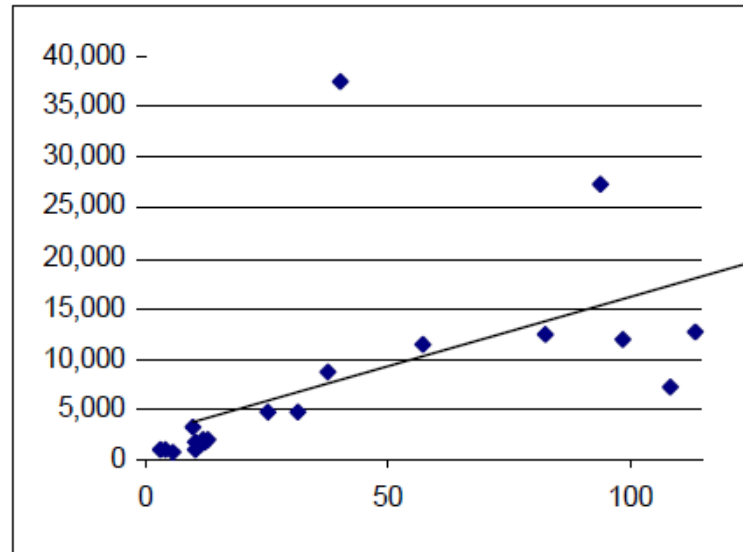


Figure 22

Estimamos $\sigma_{\hat{\beta}}^2$ por medio de $s_{\hat{\beta}}^2$.

$$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy} = 1,756,002,652 - 136.7458276 (3,881,560.4) = 1,225,215,463$$

$$s^2 = \frac{SSE}{n-2} = \frac{1,225,215,463}{20-2} = 68,067,525.72$$

$$s = \sqrt{s^2} = 8,250.3046$$

Paso 3. Enunciamos los siguientes supuestos.

1. La media de la distribución de probabilidad de ε es cero.
2. La varianza de la distribución de probabilidad de ε es constante para todos los valores de x .
3. La distribución de probabilidad de ε es normal.
4. Los errores asociados a cualquier par de observaciones diferentes son independientes.

Paso 4. Para determinar si un modelo de línea recta proporciona información útil sobre la relación entre el número de empleados y ventas, probamos:

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

La región de rechazo requiere $\frac{\alpha}{2} = 0.025$ en cada cola de la distribución t con $20 - 2 = 18$ grados de libertad. De la tabla correspondiente se obtiene $t_{0.025} = 2.101$. Por tanto la región de rechazo es $|t| > 2.101$.

El estadístico de prueba es

$$t = \frac{\hat{\beta}_1 - 0}{s_{\hat{\beta}_1}} = \frac{136.7458276}{\frac{8,250.3046}{\sqrt{28,385.22}}} = 2.79$$

El valor observado del estadístico de prueba 2.79 cae en la región de rechazo, de ahí que se rechaza la hipótesis nula. Es decir, existe suficiente evidencia para indicar que el modelo de la recta de mínimos cuadrados proporciona información útil sobre la relación entre el número de empleados y las ventas con $\alpha = .05$.

El p-valor para cada coeficiente β_i , que la mayoría del software estadístico reporta, es uno de dos direcciones, de ahí que se ajuste cuando se conduce una prueba de hipótesis de una sola dirección, como se indica enseguida.

$$\begin{aligned} \text{Cuando } H_a : \beta_1 > 0 \quad p\text{-valor} &= \begin{cases} \frac{p}{2} & \text{si } t > 0 \\ 1 - \frac{p}{2} & \text{si } t < 0 \end{cases} \\ \text{Cuando } H_a : \beta_1 < 0 \quad p\text{-valor} &= \begin{cases} \frac{p}{2} & \text{si } t < 0 \\ 1 - \frac{p}{2} & \text{si } t > 0 \end{cases} \end{aligned}$$

donde p es el p-valor generado con el software y t es el estadístico de prueba.

Para obtener información adicional sobre la relación entre x y y , construimos el intervalo del 95% de confianza, para β_1 . Cuando $\alpha = .05$, $\frac{\alpha}{2} = 0.025$ y $t_{0.025,18} = 2.101$. Así que el intervalo de confianza es $[33.862, 239.630]$ porque

$$\hat{\beta}_1 \pm t_{0.025,18} s_{\hat{\beta}_1} = \hat{\beta}_1 \pm t_{0.025,18} \frac{s}{\sqrt{SS_{xx}}} = 136.7458276 \pm 2.101 \left(\frac{8,250.3046}{\sqrt{28,385.22}} \right) = 136.7458276 \pm 102.884$$

Se tiene un 95% de confianza, de que el intervalo de $[33.862, 239.630]$ incluye al parámetro β_1 .

Una vez obtenida la recta de regresión de y sobre x , $\hat{y} = 2407.663 + 136.746x$, la cual expresa la relación bivariada entre las variables x y y , es necesario obtener una medida descriptiva del grado de la dependencia lineal de y sobre x , una de las más utilizadas es el coeficiente de correlación de Pearson, el cual es un coeficiente de correlación de producto-momento, calculado como sigue.

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{3,881,560.4}{\sqrt{28,385.22(1,756,002,652)}} = 0.5498$$

El resultado indica que la correlación entre el número de empleados y las ventas es positiva, pero no es fuerte debido a que no toma un valor cercano a 1.

Otra medida de la utilidad del modelo de regresión, es el coeficiente de determinación r^2 , que mide la contribución de la variable independiente x en la predicción de y . Si x contribuye escasamente o no informa en la predicción de y , las sumas de los cuadrados de las desviaciones de la media o del valor estimado,

$$SS_{yy} = \sum (y_i - \bar{y})^2 \quad SSE = \sum (y_i - \hat{y}_i)^2$$

son casi iguales. Por el contrario, si x aporta información en la predicción de y , el SSE será menor que SS_{yy} . De hecho, si todos los puntos caen en la recta de mínimos cuadrados, $SSE = 0$. Entonces

la reducción en el valor de SSE será atribuible a x , lo cual se expresa como una proporción de SS_{yy}

$$r^2 = \frac{SS_{yy} - SSE}{SS_{yy}}$$

Nótese además, que SS_{yy} es la "variación total" de las observaciones en torno a su media y que SSE es la restante "variabilidad de la muestra sin explicar" que es atribuible a la relación lineal con x .

$$r^2 = \frac{\text{variabilidad de la muestra explicada}}{\text{variabilidad total de la muestra}}$$

Es decir, el coeficiente de determinación informa de la proporción total de la variabilidad total de la muestra explicada por la relación lineal entre las variables. En la regresión lineal simple, dicha proporción es igual al cuadrado del coeficiente de correlación lineal simple. En el ejemplo que nos ocupa, se tiene

$$r^2 = 0.5498^2 = 0.3023$$

nos informa que el 30.23% de la variabilidad en las ventas se explica por la relación lineal entre las ventas y el número de empleados.

Paso 5. Encontrar un intervalo de predicción del 95% para las ventas, cuando la compañía tiene 50 mil empleados. Ese intervalo tiene la forma

$$\hat{y} \pm t_{\alpha/2, \nu} \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}}$$

donde

$$\hat{y} = 2407.663 + 136.746(50) = 9,244.963$$

Se hace la sustitución de los datos y se tiene que el intervalo del 95% confianza, del valor de \hat{y} .

$$9,244.963 \pm 2.101 \sqrt{1 + \frac{1}{20} + \frac{(50 - 39)^2}{28,385.22}}$$

Al realizar las operaciones, se tiene que, el límite inferior del intervalo es 9242.805750 y el límite superior, 9247.120250.

Remark 209 *El problema de mínimos cuadrados para encontrar la ecuación de una recta que se ajuste óptimamente, minimizando la suma de los cuadrados de las desviaciones, se formula como sigue*

10.6.1 Problema de mínimos cuadrados

Este problema consiste en encontrar la solución mediante mínimos cuadrados, a un sistema lineal sobredeterminado $Ax = b$, donde los vectores x y b representan datos (provenientes de observaciones), contenidos en un subespacio. El vector b contiene desviaciones (errores) de valores promedio.

De manera más precisa, consideremos a un vector $b \in \mathbb{R}^m$ y un subespacio $S \subset \mathbb{R}^m$, el problema de mínimos cuadrados trata de hallar, un vector $p \in S$ tal que p sea el más cercano a b . Es posible

encontrar a p , si el espacio se describe mediante una base. El caso más sencillo, se tiene cuando la base es ortogonal, donde la solución geométrica es

$$p = \text{proy}_S b$$

Para el sistema $Ax = b$, donde A tiene dimensión $m \times n$, $x \in \mathbb{R}^n$ y $b \in \mathbb{R}^m$, se tiene o no solución según b esté o no en el espacio columna de A .

En el segundo caso, el sistema es inconsistente. Para un sistema inconsistente, se busca minimizar el error promedio, generado con las m ecuaciones.

10.6.2 Problema de mínimos cuadrados con una variable

Cuando $n = 1$, se trata de una sola indeterminada, definimos el error promedio como

$$\|ax - b\|^2 = \sum_{i=1}^m (a_i x - b_i)^2$$

en el sistema

$$Ax = b$$

Nótese que

$$E^2 = (ax - b)^T (ax - b) = a^T a x^2 - 2a^T b x + b^T b$$

que toma su valor mínimo cuando

$$\frac{dE^2}{dx} = 2a^T a x - 2a^T b = 0$$

con $a \neq 0$.

La solución

$$x = \frac{a^T b}{a^T a}$$

por mínimos cuadrados, del sistema $Ax = b$, es idéntica a la proyección

$$p = xa$$

que es el punto en la recta, en cuestión, que pasa por a más cercano a b .

Observemos que,

$$a^T (b - xa) = a^T b - \frac{a^T b}{a^T a} a^T a = 0$$

La recta que une b con p , es perpendicular a la recta que contiene a a .

Example 210 Encontrar la mejor solución por mínimos cuadrados al sistema

$$\begin{aligned} 3x &= 10 \\ 4x &= 5 \end{aligned}$$

Example 211 El peso observado de un paciente durante cuatro ocasiones ha sido

$$b_1 = 75, b_2 = 76.5, b_3 = 75 \text{ y } b_4 = 75.5$$

¿Cuál es el mejor valor, en el sentido de mínimos cuadrados, asociado al peso de ese paciente?

10.6.3 Problema de mínimos cuadrados con varias variables

Como se dijo antes, el problema se resuelve proyectando el vector b sobre un subespacio definido por la matriz $A_{m \times n}$. En el sistema $Ax = b$, sea $m > n$, entonces puede ocurrir que el sistema sea inconsistente, es decir, que el vector b no sea combinación lineal de la columnas de A . Se busca \bar{x} que minimice el error

$$E = \|Ax - b\|$$

por mínimos cuadrados.

Geoméricamente, p es la proyección de b sobre el espacio columna de A .

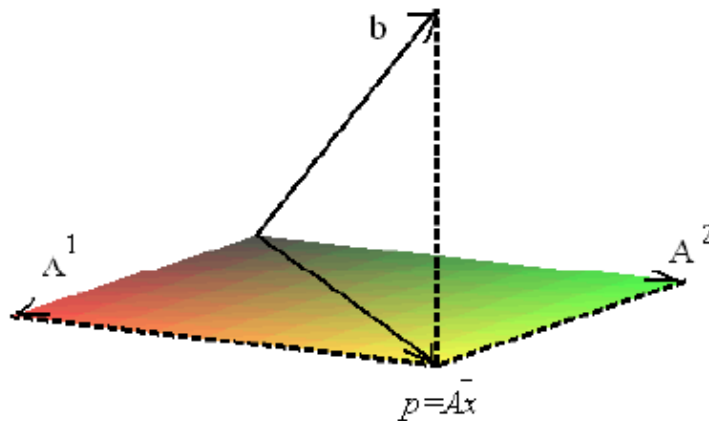


Figure 23

El vector error $Ax - b$ debe ser ortogonal a ese espacio. Es decir,

$$(Ay)^T (A\bar{x} - b) = 0$$

para todo vector $y \in E^n$.

O bien, $y^T [A^T A\bar{x} - A^T b] = 0$, para todo vector $y \in E^n$, lo que implica que

$$A^T A\bar{x} - A^T b = 0$$

De donde, la solución del sistema inconsistente, satisface

$$A^T A\bar{x} = A^T b$$

Si las columnas de A son linealmente independientes, entonces AA^T es una matriz simétrica invertible y la solución única por mínimos cuadrados es

$$\bar{x} = (A^T A)^{-1} A^T b$$

y, la proyección es

$$p = A\bar{x} = A (A^T A)^{-1} A^T b$$

La matriz

$$B = (A^T A)^{-1} A^T$$

es una inversa izquierda de la matriz A , porque se cumple

$$BA = (A^T A)^{-1} A^T A = I$$

Example 212 Sean

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 0 & 0 \end{bmatrix} \quad y \quad b = \begin{bmatrix} 4 \\ 3 \\ 9 \end{bmatrix}$$

Nótese que espacio columna de A , tiene dimensión dos. Calculamos la solución por mínimos cuadrados del sistema

$$Ax = b$$

como sigue

$$\begin{aligned} A^T A &= \begin{bmatrix} 1 & 1 & 0 \\ 2 & 5 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 7 \\ 7 & 29 \end{bmatrix} \\ (A^T A)^{-1} &= \frac{1}{9} \begin{bmatrix} 29 & -7 \\ -7 & 2 \end{bmatrix} \\ (A^T A)^{-1} A^T &= \frac{1}{9} \begin{bmatrix} 29 & -7 \\ -7 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 2 & 5 & 0 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} 15 & -6 & 0 \\ -3 & 3 & 0 \end{bmatrix} \\ \bar{x} &= (A^T A)^{-1} A^T b = \frac{1}{9} \begin{bmatrix} 15 & -6 & 0 \\ -3 & 3 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \\ 9 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} 42 \\ -3 \end{bmatrix} \end{aligned}$$

Por tanto, la mejor solución del sistema

$$\begin{aligned} x + 2y &= 4 \\ x + 5y &= 3 \\ 0x + 0y &= 9 \end{aligned}$$

se obtiene ignorando la tercera ecuación, con un error en esa ecuación de 9.

La proyección es

$$p = A\bar{x} = \frac{1}{9} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 42 \\ -3 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} 36 \\ 27 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 0 \end{bmatrix}$$

En resumen, la proyección p , el vector b y el espacio columna de A se mostró en la figura 23.

Cuando b pertenece al espacio columna de A entonces del hecho de que

$$b = Ax$$

se concluye que

$$p = A (A^T A)^{-1} A^T Ax = Ax = b$$

Es decir, el punto más cercano a b es el mismo b .

Si b es ortogonal a todas las columnas de A , b no pertenece al subespacio y es perpendicular al subespacio,

$$A^T b = 0$$

de donde

$$\bar{x} = (A^T A)^{-1} A^T b = 0$$

y

$$p = A\bar{x} = 0$$

Exercise 213 *Encontrar la mejor solución por mínimos cuadrados, de cada sistema y la proyección sobre el espacio columna, correspondiente.*

$$1. \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

$$2. \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} [x] = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}$$

$$3. \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}$$

11 Bibliografía básica

References

- [1] Barbolla, R. Cerdá, E. y Sanz, P. (2001) Optimización. Cuestiones, ejercicios y aplicaciones a la economía. Madrid: Prentice Hall.
- [2] Bertsekas, D., et.al. (2003) Convex Analysis and Optimization, Athena Scientific. Thomson.
- [3] Bertsekas, D. And Tsiriklis, J. (1997) Introduction to linear optimization. Athena Scientific.
- [4] Luenberger, D.G. (1973) Introduction to linear and nonlinear Programming; Addison Wesley.
- [5] McClave, James T., Benson, P. George and Sincich, Terry (2001) Statistics for Business and Economics. New Jersey: Prentice Hall, pp. 1067.
- [6] Murthy, D., Page, N. & Rodin, E. (1990) Mathematical Modelling. A tool for Problem Solving Engineering, Physical, Biological and Social Sciences. N.Y.: Pergamon Press.
- [7] Strang. G. (2006) Linear algebra and its applications. Belmont, California: Thomson-Brooks/Cole.

12 Bibliografía complementaria

References

- [1] Feller, William (2007) Introducción a la Teoría de Probabilidades y sus Aplicaciones. Volumen I. México: LIMUSA.
- [2] Hillier, F., Hillier, M. and Lieberman, G. (2000) Introduction to Introduction to Operations Research. Boston: McGraw-Hill. Higher Education.
- [3] Lang, S. (1993) Introduction to Linear Algebra”. 2^a edición. Springer-Verlag.
- [4] Kenett, Ron S. and Zacks, Shelemyahu (2000) Estadística Industrial Moderna. Diseño y control de la calidad y la confiabilidad. México: Thomson, pp. 821.
- [5] Levine, M. David, Dephan, David, Krehbiel, Timothy C., and Berenson, Mark L. (2002) Statistics for Managers using Microsoft Excel®. N.J.: Prentice Hall, pp. 885.
- [6] Sheaffer, R., Mendenhall, W. Y Ott, L. (1986) Elementos de muestreo. Trads. Gilberto Rendón S. y José Roberto Gómez A. México: Grupo Editorial Iberoamérica.