

	PROCESO PARA EL DESARROLLO DE LAS ENSEÑANZAS DE LA FACULTAD DE CIENCIAS DE LA UEx		 FACULTAD DE CIENCIAS <small>[UEx]</small>
	Curso académico: 2024-25	Código: P/CL009_FC_D002	

PLAN DOCENTE DE LA ASIGNATURA

Identificación y características de la asignatura			
Código	501731	Créditos ECTS	6
Denominación (español)	Modelos Lineales		
Denominación (inglés)	Linear Models		
Titulaciones	Grado en Estadística		
Centro	Facultad de Ciencias		
Semestre	6	Carácter	Obligatoria
Módulo	Formación Obligatoria		
Materia	Estadística		
Profesor/es			
Nombre	Despacho	Correo-e	Página web
Manuel Mota Medina	B36	mota@unex.es	Campus Virtual
Área de conocimiento	Estadística e Investigación Operativa		
Departamento	Matemáticas		
Profesor coordinador (si hay más de uno)			
Competencias			
CB1: Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.			
CB2: Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.			
CB3: Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.			
CB4: Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.			
CB5: Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.			
CG1: Desarrollar las capacidades de análisis, abstracción, intuición, organización y síntesis, así como el razonamiento lógico, riguroso y crítico.			
CG2: Capacitar al alumno para utilizar los conocimientos teóricos y prácticos adquiridos en la			

	PROCESO PARA EL DESARROLLO DE LAS ENSEÑANZAS DE LA FACULTAD DE CIENCIAS DE LA UEx		
	Curso académico: 2024-25	Código: P/CL009_FC_D002	

definición y planteamiento de problemas, así como en la búsqueda de sus soluciones tanto en contextos académicos como profesionales.
CG3: Preparar al alumno para el trabajo en equipos multidisciplinares, capacitándolo para entender los razonamientos de especialistas de otros campos y comunicar sus propios razonamientos y conclusiones.
CG4: Promover la curiosidad y el interés por los métodos y técnicas que estudia la Estadística y la Investigación Operativa, animándolo a mantenerlos y transmitirlos una vez finalizados sus estudios.
CG5: Mostrar la importancia, necesidad y utilidad de la metodología estadística en otras ciencias (ciencias experimentales, ciencias de la salud, ciencias sociales y humanas, etc.)
CG6: Dotar al alumno de los conocimientos necesarios para que pueda continuar estudios posteriores en otras disciplinas tanto científicas como tecnológicas.
CT4: Prepararse para el aprendizaje autónomo de nuevos conocimientos, métodos y técnicas; y para emprender estudios posteriores con un alto grado de autonomía.
CT5: Dominar las Tecnologías de la Información y las Comunicaciones mediante el uso de aplicaciones informáticas de análisis estadístico, cálculo numérico y simbólico, visualización gráfica, tratamiento de datos, optimización, y el desarrollo de programas que resuelvan problemas estadísticos utilizando para cada caso el entorno computacional adecuado.
CE3: Estudiar y resolver problemas en situaciones de incertidumbre, sabiendo construir y validar modelos probabilísticos para la descripción de tales situaciones.
CE6: Realizar estudios comparativos entre poblaciones y detectar posibles relaciones entre variables.
CE7: Aplicar correctamente la metodología estadística en análisis de datos e interpretar en sus justos términos los resultados obtenidos.
CE8: Identificar y analizar estadísticamente la información relevante contenida en problemas reales, así como aplicar técnicas estadísticas específicas para su resolución.
CE12: Diseñar, programar e implementar software estadístico y de gestión de bases de datos.
Contenidos
Breve descripción del contenido
Distribuciones de probabilidad de interés en modelos lineales. Estimación y contraste de hipótesis en modelos lineales. Regresión múltiple. Análisis de la covarianza. Diseño de experimentos. Introducción a los modelos lineales generalizados.
Temario de la asignatura
Denominación del tema 1: Distribuciones de probabilidad y formas cuadráticas. Contenidos del tema 1: 1.1 Introducción. 1.2 Resultados algebraicos. 1.3 Distribuciones de probabilidad asociadas al modelo lineal. 1.4 Formas cuadráticas. Descripción de las actividades prácticas del tema 1: Ninguna
Denominación del tema 2: Modelo Lineal de Rango Completo. Contenidos del tema 2: 2.1 Modelo Lineal: definiciones y ejemplos. 2.2 Estimación puntual en el Modelo Lineal Básico. 2.3 Estimación en el Modelo Lineal Normal. 2.4 Contraste de hipótesis en el Modelo Lineal Normal. 2.5 Regresión Lineal, Regresión Polinómica y Análisis de la Covarianza. Descripción de las actividades prácticas del tema 2: Ajuste de modelos de Regresión Lineal,

	PROCESO PARA EL DESARROLLO DE LAS ENSEÑANZAS DE LA FACULTAD DE CIENCIAS DE LA UEx		 FACULTAD DE CIENCIAS <small>[UEx]</small>
	Curso académico: 2024-25	Código: P/CL009_FC_D002	

Regresión Polinómica y Análisis de la Covarianza haciendo uso de software estadístico.
Métodos de selección de variables en modelos de regresión.

Denominación del tema 3: **Modelo Lineal de Rango No Completo.**

Contenidos del tema 3: 3.1 Introducción. 3.2 Estimación en el Modelos Lineal de rango no completo. 3.3 Contraste de hipótesis en el Modelo Lineal de rango no completo.

Descripción de las actividades prácticas del tema 3: Ninguna

Denominación del tema 4: **Modelos de Diseño de Experimentos.**

Contenidos del tema 4: 4.1 Introducción al Diseño de Experimentos. 4.2 Experimentos con un factor. Efectos fijos. 4.3 Experimentos con dos factores. Efectos fijos. 4.4 Modelos de efectos aleatorios y mixtos.

Descripción de las actividades prácticas del tema 4: Ajuste de modelos de Diseño de Experimentos con uno o dos factores haciendo uso de software estadístico.

Denominación del tema 5: **Modelos Lineales Generalizados.**

Contenidos del tema 5: 5.1 Introducción a los Modelos Lineales Generalizados. 5.2 Estimación en Modelos Lineales Generalizados. 5.3 Contraste de hipótesis en Modelos Lineales Generalizados. 5.4 Variables dicotómicas y regresión logística.

Descripción de las actividades prácticas del tema 5: Ajuste de distintos tipos de Modelos Lineales Generalizados haciendo uso de software estadístico.

Actividades formativas

Horas de trabajo del alumno por tema		Horas	Horas actividades prácticas				Horas actividad de seguimiento	Horas. No presencial
Tema	Total		GG	CH	L	O		
1	17	6			0			11
2	45	14			8			23
3	13	6			0			7
4	30	10			4			16
5	16	6			2			8
Evaluación	29	3			1			25
TOTAL	150	45			15			90

GG: Grupo Grande (100 estudiantes).

CH: prácticas clínicas hospitalarias (7 estudiantes)

L: prácticas laboratorio o campo (15 estudiantes)

O: prácticas sala ordenador o laboratorio de idiomas (20 estudiantes)

S: clases problemas o seminarios o casos prácticos (40 estudiantes).

TP: Tutorías Programadas (seguimiento docente, tipo tutorías ECTS).

EP: Estudio personal, trabajos individuales o en grupo, y lectura de bibliografía.

	PROCESO PARA EL DESARROLLO DE LAS ENSEÑANZAS DE LA FACULTAD DE CIENCIAS DE LA UEx		
	Curso académico: 2024-25	Código: P/CL009_FC_D002	

Metodologías docentes

1. Explicación y discusión de los contenidos.
2. Resolución, análisis y discusión de problemas. Realización, exposición y defensa de trabajos/proyectos.
3. Actividades experimentales como prácticas en laboratorios, aulas de informática y trabajos de campo.
4. Actividades de seguimiento individual o por grupos del aprendizaje.
5. Trabajo autónomo del estudiante.

Resultados de aprendizaje

Al completar la materia ESTADÍSTICA, el estudiante:

- Conoce y comprende los principales conceptos de la inferencia estadística básica: estimador, intervalo de confianza, contrastes de hipótesis unilaterales y bilaterales y p-valor.
- Es capaz de resolver problemas de inferencia estadística (estimación puntual y por intervalos de confianza y contrastes de hipótesis) para la media y la varianza de una población normal, para una proporción, para la comparación de las medias de dos poblaciones normales, para la comparación de dos proporciones y en el modelo lineal normal.
- Es capaz de plantear de manera clara el modelo estadístico a considerar para la resolución de un problema de relación entre variables o de un problema de comparación entre grupos.
- Sabe plantear el modelo estadístico a considerar para resolver un problema de regresión o análisis de la varianza multivariante y es capaz de construir estimadores y contrastes de hipótesis adecuados para dichos modelos.
- Puede, tras la aplicación de las distintas metodologías estudiadas, ser capaz de extraer las conclusiones estadísticas más relevantes y de redactarlas de manera que resulten comprensible en el ámbito científico.
- Sabe distinguir entre inferencia paramétrica e inferencia no paramétrica.
- Conoce y sabe aplicar distinto software estadístico para las metodologías estadísticas estudiadas.

Sistemas de evaluación

Criterios de evaluación:

- Demostrar la adquisición y comprensión de los principales conceptos teóricos de la asignatura.
- Aplicar de manera eficiente los conocimientos teóricos en la resolución de ejercicios y/o problemas.
- Aplicar de manera eficiente los conocimientos teóricos en la modelización de problemas prácticos reales.
- Participar activamente en la resolución de problemas (teórico-prácticos) en la clase.
- Realizar, exponer y defender con suficiencia los trabajos de propuestos.

	PROCESO PARA EL DESARROLLO DE LAS ENSEÑANZAS DE LA FACULTAD DE CIENCIAS DE LA UEx		 FACULTAD DE CIENCIAS <small>[UEx]</small>
	Curso académico: 2024-25	Código: P/CL009_FC_D002	

Instrumentos para la evaluación:

Tanto en la convocatoria ordinaria como en la extraordinaria, el estudiante podrá elegir, en las condiciones que establezca la normativa de evaluación vigente, entre el sistema de evaluación continua o el sistema de evaluación con una única prueba final de carácter global. De no realizar esta elección, se entenderá que opta por la evaluación continua.

Para aquellos **estudiantes que opten por el sistema de evaluación continua**, la evaluación se realizará mediante:

- a. Un examen que constará de una parte teórica y una parte práctica que deberá resolverse haciendo uso de software estadístico. Cada parte se calificará con una nota entre 0 y 10. En la nota global del examen se ponderará con un 60% la teoría y con un 40% la parte práctica.
- b. La entrega de un trabajo con la resolución de una serie de problemas teóricos y de otro trabajo con un análisis de datos mediante los modelos vistos en teoría y haciendo uso de software estadístico. Cada uno de estos trabajos se calificará entre 0 y 10. La nota de este apartado será la media de las calificaciones de los distintos trabajos. Estas actividades tendrán carácter no recuperable.

Para superar la asignatura la puntuación de la parte de teoría a que hace referencia el apartado a. no podrá ser inferior a 4. Una vez cumplido este requisito, la calificación final de la asignatura se obtendrá multiplicando por 0.8 la nota resultante del apartado a. y por 0.2 la nota resultante del apartado b.

Para aquellos **estudiantes que opten por una prueba final de carácter global**, la evaluación se realizará mediante el examen que se indica en el apartado a. anterior. Para superar la asignatura la puntuación de la parte de teoría a que hace referencia el apartado a. no podrá ser inferior a 4.

Bibliografía (básica y complementaria)

- Dobson, A. (1990). "An introduction to Generalized Linear Models". Chapman-Hall.
- Faraway, J.J. (2005). "Linear Model with R". Chapman-Hall.
- Graybill, F.A. (1961). "An Introduction to Linear Statistical Models. Vol. I". McGraw-Hill.
- Graybill, F.A. (2000). "Theory and Applications of the Linear Model". Duxbury Classic.
- Montanero, J. (2008). "Modelos Lineales". Manuales Uex ON-LINE 56.
- Montgomery, D.C. (2004). "Design and Analysis of Experiments. 6th Edition". Wiley.
- Peña, D. (1987). "Estadística: Modelo y Métodos. Vol. II". Alianza Universidad Textos.
- Peña, D. (2002). "Regresión y Diseño de Experimentos". Alianza Universidad Textos.
- Wood, S.N. (2006). "Generalized Additive Models. An Introduction with R". Chapman-Hall.
- Yáñez, I. y Martín, M. (1991) "Diseño de Experimentos y Teoría de Muestras". UNED.
- Página web del programa R: www.r-project.org

	PROCESO PARA EL DESARROLLO DE LAS ENSEÑANZAS DE LA FACULTAD DE CIENCIAS DE LA UEx		 FACULTAD DE CIENCIAS [UEx]
	Curso académico: 2024-25	Código: P/CL009_FC_D002	

Otros recursos y materiales docentes complementarios

Tema 1: Distribuciones y formas cuadráticas

1.1 Introducción

En este tema se proporcionan una serie de herramientas algebraicas y probabilísticas que se manejarán a lo largo de toda la asignatura. Más concretamente se estudiarán resultados sobre matrices y sobre distribuciones de probabilidad, particularmente la normal multivariante, y las distribuciones asociadas a formas cuadráticas.

1.2 Resultados algebraicos

Consideraremos vectores y matrices con coeficientes en \mathbb{R} . Los denotaremos con letra negrita. Los vectores serán considerados como matrices columna, es decir de orden $n \times 1$. De este modo, si $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Denotaremos como \mathbf{I}_n a la matriz identidad de orden n y como $\mathbf{0}$ a la matriz o vector cuyos elementos son todos iguales a 0 (no se especificará en la notación la dimensión de dicha matriz o vector).

Una matriz cuadrada de orden n , \mathbf{A} , se dice *ortogonal* si su traspuesta y su inversa coinciden, es decir, si $\mathbf{A}\mathbf{A}^t = \mathbf{A}^t\mathbf{A} = \mathbf{I}_n$.

Traza

La *traza* de una matriz cuadrada de orden n , $\mathbf{A} = (a_{ij})$ es la suma de los elementos de su diagonal principal, es decir, $\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}$. A continuación enumeramos algunas propiedades de la traza:

- a) $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$, $\text{tr}(k\mathbf{A}) = k \text{tr}(\mathbf{A})$ siendo k escalar.
- b) $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^t)$.
- c) $\text{tr}(\mathbf{I}_n) = n$.
- d) $\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{A})$.
- e) $\text{tr}(\mathbf{A}\mathbf{B}\mathbf{C}) = \text{tr}(\mathbf{C}\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{C}\mathbf{A})$.
- f) si \mathbf{P} invertible $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{P}^{-1}\mathbf{A}\mathbf{P})$.

Rango

El *rango* de una matriz \mathbf{A} es el máximo número de filas o columnas linealmente independientes. Lo denotaremos por $r(\mathbf{A})$. A continuación enumeramos algunas propiedades del rango:

- Si \mathbf{A}, \mathbf{C} invertibles, entonces $r(\mathbf{AB}) = r(\mathbf{BC}) = r(\mathbf{B})$.
- $r(\mathbf{A}^t\mathbf{A}) = r(\mathbf{AA}^t) = r(\mathbf{A}) = r(\mathbf{A}^t)$.
- Una matriz cuadrada de orden n , \mathbf{A} , es invertible si y sólo si $r(\mathbf{A}) = n$.

Autovalores

Si \mathbf{A} es una matriz cuadrada de orden n con n autovalores (no necesariamente distintos).

- El determinante de \mathbf{A} es igual al producto de sus autovalores.
- La traza de \mathbf{A} es igual a la suma de sus autovalores.
- El rango de \mathbf{A} es igual al número de sus autovalores distintos de 0.

Matrices idempotentes

Una matriz cuadrada \mathbf{P} se dice *idempotente* si $\mathbf{P}^2 = \mathbf{P}$. Algunas propiedades de las matrices idempotentes:

- Si \mathbf{P} es idempotente de orden n entonces $\mathbf{I}_n - \mathbf{P}$ también es idempotente.
- Si \mathbf{P} es una matriz idempotente entonces sus únicos posibles autovalores, tanto reales como complejos, son 0 ó 1.
- Si \mathbf{P} es una matriz simétrica de orden n entonces \mathbf{P} es idempotente de rango r si y sólo si el autovalor 1 de \mathbf{P} tiene multiplicidad r y el autovalor 0 tiene multiplicidad $n - r$.
- Si \mathbf{P} simétrica idempotente entonces $r(\mathbf{P}) = \text{tr}(\mathbf{P})$.

Formas cuadráticas

Si \mathbf{A} es una matriz cuadrada simétrica de orden n , $\mathbf{x} = (x_1, \dots, x_n)^t \in \mathbb{R}^n$, se denomina *forma cuadrática asociada a \mathbf{A}* a la aplicación $Q_{\mathbf{A}} : \mathbb{R}^n \rightarrow \mathbb{R}$ definida por:

$$Q_{\mathbf{A}}(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x} = \sum_{i,j=1}^n a_{ij} x_i x_j = \sum_{i=1}^n a_{ii} x_i^2 + 2 \sum_{i<j} a_{ij} x_i x_j.$$

Cada forma cuadrática $Q_{\mathbf{A}}$ queda definida unívocamente por la matriz simétrica \mathbf{A} .

Matrices y formas cuadráticas definidas y semidefinidas positivas

- Diremos $\mathbf{x}^t \mathbf{A} \mathbf{x}$ es *semidefinida positiva* si $\mathbf{x}^t \mathbf{A} \mathbf{x} \geq 0$ para todo $\mathbf{x} \in \mathbb{R}^n$.
- Diremos que una forma cuadrática $\mathbf{x}^t \mathbf{A} \mathbf{x}$ es *definida positiva* si

$$\mathbf{x}^t \mathbf{A} \mathbf{x} \geq 0 \quad \text{para todo } \mathbf{x} \in \mathbb{R}^n \quad ; \quad \mathbf{x}^t \mathbf{A} \mathbf{x} = 0 \iff \mathbf{x} = \mathbf{0}$$

La matriz simétrica \mathbf{A} es (semi)definida positiva si su forma cuadrática asociada lo es.

Propiedades

- a) Si \mathbf{P} es una matriz no singular y \mathbf{A} simétrica, entonces \mathbf{A} es definida (resp. semidefinida) positiva si y sólo si $\mathbf{P}^t \mathbf{A} \mathbf{P}$ es definida (resp. semidefinida) positiva.
- b) Si \mathbf{A} es definida positiva de orden n entonces existe una matriz no singular \mathbf{P} tal que $\mathbf{P}^t \mathbf{A} \mathbf{P} = \mathbf{I}_n$.
- c) Una matriz es definida (resp. semidefinida) positiva si y sólo si sus autovalores son mayores (resp. mayores o iguales) que 0. En particular, las matrices idempotentes son semidefinidas positivas.

Descomposición de matrices definidas positivas

- Si \mathbf{A} es una matriz simétrica de orden n y rango r ($r \leq n$) con coeficientes reales entonces existe una matriz \mathbf{L} de orden $n \times r$ (con coeficientes complejos) tal que $\mathbf{A} = \mathbf{L} \mathbf{L}^t$. Esta matriz \mathbf{L} no es única.
- Si además \mathbf{A} es semidefinida positiva entonces \mathbf{L} tiene coeficientes reales.
- Si \mathbf{A} es definida positiva, entonces \mathbf{L} es cuadrada y de rango n (invertible). Además podemos tomar \mathbf{L} simétrica. A una matriz \mathbf{L} cuadrada e invertible verificando $\mathbf{A} = \mathbf{L} \mathbf{L}$ la denominaremos $\mathbf{L} = \mathbf{A}^{1/2}$ y a su inversa $\mathbf{A}^{-1/2}$.

Podemos decir que la matriz $\mathbf{A}^{1/2}$ es una *raíz cuadrada* de \mathbf{A} .

Matrices $\mathbf{A} \mathbf{A}^t$ y $\mathbf{A}^t \mathbf{A}$

Sea \mathbf{A} una matriz de orden $m \times n$ entonces se verifica que tanto $\mathbf{A}^t \mathbf{A}$ como $\mathbf{A} \mathbf{A}^t$ son matrices semidefinidas positivas. Además,

- Si $r(\mathbf{A}) = m$ entonces $\mathbf{A} \mathbf{A}^t$ es definida positiva.
- Si $r(\mathbf{A}) = n$ entonces $\mathbf{A}^t \mathbf{A}$ es definida positiva.

1.3 Distribuciones de probabilidad asociadas al modelo lineal

Vectores aleatorios

Si (Ω, \mathcal{A}, P) es un espacio de probabilidad, denominaremos *vector aleatorio* a toda función medible $\mathbf{Y} : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}^n$. El vector aleatorio \mathbf{Y} se escribiría $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^t$ y sus componentes son variables aleatorias.

Al vector $\boldsymbol{\mu} = E[\mathbf{Y}] = (E[Y_1], E[Y_2], \dots, E[Y_n])^t$ se le denomina *vector de medias* del vector aleatorio \mathbf{Y} . Algunas propiedades del vector de medias:

- Si $\mathbf{b} \in \mathbb{R}^n$, entonces $E[\mathbf{Y} + \mathbf{b}] = E[\mathbf{Y}] + \mathbf{b}$. En particular, $E[\mathbf{Y} - \boldsymbol{\mu}] = \mathbf{0}$.
- Si \mathbf{A} es una matriz $k \times n$, entonces $E[\mathbf{A}\mathbf{Y}] = \mathbf{A}E[\mathbf{Y}]$.

Si $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^t$, entonces se define la *covarianza de las variables* Y_i e Y_j ,

$$\sigma_{ij} = \text{Cov}[Y_i, Y_j] = E[(Y_i - \mu_i)(Y_j - \mu_j)].$$

Obviamente, si $i = j$

$$\sigma_{ii} = \text{Var}[Y_i]$$

A la matriz

$$\mathbf{V} = \text{Cov}[\mathbf{Y}] = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{pmatrix}$$

se le denomina *matriz de covarianzas* de \mathbf{Y} . Algunas propiedades de la matriz de covarianzas:

- $\text{Cov}[\mathbf{Y}] = E[(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})^t]$.
- La matriz $\text{Cov}[\mathbf{Y}]$ es simétrica y semidefinida positiva.
- Si $\mathbf{b} \in \mathbb{R}^n$, entonces $\text{Cov}[\mathbf{Y} + \mathbf{b}] = \text{Cov}[\mathbf{Y}]$.
- Si \mathbf{A} es una matriz $k \times n$, entonces $\text{Cov}[\mathbf{A}\mathbf{Y}] = \mathbf{A} \text{Cov}[\mathbf{Y}] \mathbf{A}^t$. En particular, si denotamos $\mathbf{V} = \text{Cov}[\mathbf{Y}]$ y \mathbf{V} es definida positiva, $\text{Cov}[\mathbf{V}^{-1/2}\mathbf{Y}] = \mathbf{I}_n$.

La *función de distribución* de un vector aleatorio se define como

$$F_{\mathbf{Y}}(\mathbf{y}) = F_{\mathbf{Y}}(y_1, \dots, y_n) = P(Y_1 \leq y_1, \dots, Y_n \leq y_n) \quad , \quad \mathbf{y} = (y_1, \dots, y_n)^t \in \mathbb{R}^n$$

Diremos que el vector aleatorio \mathbf{Y} es continuo si existe una función $f_{\mathbf{Y}} : \mathbb{R}^n \rightarrow [0, \infty)$ tal que

$$F_{\mathbf{Y}}(\mathbf{y}) = \int_{-\infty}^{y_1} \dots \int_{-\infty}^{y_n} f_{\mathbf{Y}}(x_1, \dots, x_n) dx_1 \dots dx_n \quad , \quad \mathbf{y} = (y_1, \dots, y_n)^t \in \mathbb{R}^n$$

A la función $f_{\mathbf{Y}}(\mathbf{y})$ se le denomina *función de densidad del vector aleatorio* \mathbf{Y} .

Se define la *función característica del vector aleatorio* \mathbf{Y} (al que supondremos continuo) como

$$\varphi_{\mathbf{Y}}(\mathbf{s}) = E[e^{i\mathbf{s}^t\mathbf{Y}}] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{i \sum_{k=1}^n s_k y_k\right\} f_{\mathbf{Y}}(y_1, \dots, y_n) dy_1 \cdots dy_n \quad , \quad \mathbf{s} = (s_1, \dots, s_n)^t \in \mathbb{R}^n.$$

Análogamente, se define la *función generatriz de momentos del vector aleatorio* \mathbf{Y} (caso de existir)

$$M_{\mathbf{Y}}(\mathbf{s}) = E[e^{s^t\mathbf{Y}}] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{\sum_{k=1}^n s_k y_k\right\} f_{\mathbf{Y}}(y_1, \dots, y_n) dy_1 \cdots dy_n \quad , \quad \mathbf{s} = (s_1, \dots, s_n)^t \in \mathbb{R}^n.$$

Distribuciones marginales

Supongamos que $\mathbf{Y}^t = (\mathbf{Y}_1^t, \mathbf{Y}_2^t)$ con $\mathbf{Y}_1 = (Y_1, \dots, Y_k)^t$, $k < n$, e $\mathbf{Y}_2 = (Y_{k+1}, \dots, Y_n)^t$.

La función de distribución marginal de \mathbf{Y}_1 se puede calcular,

$$F_{\mathbf{Y}_1}(\mathbf{y}_1) = \lim_{\mathbf{y}_2 \rightarrow \infty} F_{\mathbf{Y}}(\mathbf{y}_1, \mathbf{y}_2) \quad , \quad \mathbf{y}_1 \in \mathbb{R}^k,$$

entendiendo que $\mathbf{y}_2 \in \mathbb{R}^{n-k}$ tiende a ∞ si todas sus coordenadas tienden a infinito.

Análogamente, la función de densidad de \mathbf{Y}_1 se puede calcular,

$$f_{\mathbf{Y}_1}(\mathbf{y}_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{Y}}(y_1, \dots, y_k, y_{k+1}, \dots, y_n) dy_{k+1} \cdots dy_n \quad , \quad \mathbf{y}_1 = (y_1, \dots, y_k)^t \in \mathbb{R}^k.$$

En cuanto a la función característica o generatriz de momentos de \mathbf{Y}_1 , se pueden calcular

$$\varphi_{\mathbf{Y}_1}(\mathbf{s}_1) = \varphi_{\mathbf{Y}}\left(\begin{matrix} \mathbf{s}_1 \\ \mathbf{0} \end{matrix}\right) \quad , \quad M_{\mathbf{Y}_1}(\mathbf{s}_1) = M_{\mathbf{Y}}\left(\begin{matrix} \mathbf{s}_1 \\ \mathbf{0} \end{matrix}\right) \quad , \quad \mathbf{s}_1 \in \mathbb{R}^k.$$

Algunas propiedades de las distribuciones marginales:

- Los vectores de medias de \mathbf{Y}_1 e \mathbf{Y}_2 son, respectivamente,

$$\boldsymbol{\mu}_1 = (\mu_1, \dots, \mu_k)^t \quad , \quad \boldsymbol{\mu}_2 = (\mu_{k+1}, \dots, \mu_n)^t$$

y sus matrices de covarianzas

$$\mathbf{V}_{11} = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1k} \\ \vdots & \ddots & \vdots \\ \sigma_{k1} & \cdots & \sigma_{kk} \end{pmatrix} \quad , \quad \mathbf{V}_{22} = \begin{pmatrix} \sigma_{k+1k+1} & \cdots & \sigma_{k+1n} \\ \vdots & \ddots & \vdots \\ \sigma_{nk+1} & \cdots & \sigma_{nn} \end{pmatrix}$$

- La matriz de covarianzas de \mathbf{Y}_1 e \mathbf{Y}_2 es

$$\mathbf{V}_{12} = \text{Cov}[\mathbf{Y}_1, \mathbf{Y}_2] = \begin{pmatrix} \sigma_{1k+1} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{kk+1} & \cdots & \sigma_{kn} \end{pmatrix} = E[(\mathbf{Y}_1 - \boldsymbol{\mu}_1)(\mathbf{Y}_2 - \boldsymbol{\mu}_2)^t].$$

- Si \mathbf{A} y \mathbf{B} matrices $k \times n$ y $(n - k) \times n$, respectivamente, entonces

$$\mathbf{V}_{12} = \text{Cov}[\mathbf{A}\mathbf{Y}_1, \mathbf{B}\mathbf{Y}_2] = \mathbf{A} \text{Cov}[\mathbf{Y}_1, \mathbf{Y}_2] \mathbf{B}^t.$$

Podemos notar que el vector de medias $\boldsymbol{\mu}$ y la matriz de covarianzas \mathbf{V} de un vector aleatorio quedan particionados de la forma:

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad , \quad \mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{12}^t & \mathbf{V}_{22} \end{pmatrix}.$$

Independencia

Supongamos que $\mathbf{Y}^t = (\mathbf{Y}_1^t, \mathbf{Y}_2^t)$ con $\mathbf{Y}_1 = (Y_1, \dots, Y_k)^t$ e $\mathbf{Y}_2 = (Y_{k+1}, \dots, Y_n)^t$. Los vectores aleatorios \mathbf{Y}_1 e \mathbf{Y}_2 son *independientes* si y sólo si

$$F_{\mathbf{Y}}(\mathbf{y}) = F_{\mathbf{Y}_1}(\mathbf{y}_1) \cdot F_{\mathbf{Y}_2}(\mathbf{y}_2) \quad , \quad \mathbf{y}_1 \in \mathbb{R}^k \quad , \quad \mathbf{y}_2 \in \mathbb{R}^{n-k} \quad , \quad \mathbf{y} = (\mathbf{y}_1^t, \mathbf{y}_2^t)^t$$

o equivalentemente

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{Y}_1}(\mathbf{y}_1) \cdot f_{\mathbf{Y}_2}(\mathbf{y}_2) \quad , \quad \mathbf{y}_1 \in \mathbb{R}^k \quad , \quad \mathbf{y}_2 \in \mathbb{R}^{n-k} \quad , \quad \mathbf{y} = (\mathbf{y}_1^t, \mathbf{y}_2^t)^t$$

o equivalentemente

$$\varphi_{\mathbf{Y}}(\mathbf{s}) = \varphi_{\mathbf{Y}_1}(\mathbf{s}_1) \cdot \varphi_{\mathbf{Y}_2}(\mathbf{s}_2) \quad , \quad \mathbf{s}_1 \in \mathbb{R}^k \quad , \quad \mathbf{s}_2 \in \mathbb{R}^{n-k} \quad , \quad \mathbf{s} = (\mathbf{s}_1^t, \mathbf{s}_2^t)^t$$

Se verifica que si \mathbf{Y}_1 e \mathbf{Y}_2 son vectores aleatorios independientes entonces $\text{Cov}[\mathbf{Y}_1, \mathbf{Y}_2] = \mathbf{0}$. El recíproco en general no es cierto.

Distribución normal unidimensional

Sea $\mu \in \mathbb{R}$ y $\sigma^2 > 0$. Diremos que la variable aleatoria Y tiene *distribución normal de parámetros μ y σ^2* si su función de densidad es

$$f(y) = \frac{1}{\sqrt{2\pi} \sigma} \exp\left\{-\frac{(y - \mu)^2}{2\sigma^2}\right\} \quad , \quad y \in \mathbb{R}$$

En tal caso denotaremos $Y \sim N(\mu, \sigma^2)$.

Propiedades

- Las funciones característica y generatriz de momentos de la distribución $N(\mu, \sigma^2)$ son:

$$\varphi_Y(s) = \exp\{i\mu s - \frac{1}{2}\sigma^2 s^2\} \quad , \quad M_Y(s) = \exp\{\mu s + \frac{1}{2}\sigma^2 s^2\} \quad , \quad s \in \mathbb{R}$$

- La media de Y es μ y su varianza σ^2 .
- Si $Z = \frac{Y - \mu}{\sigma}$, entonces $Z \sim N(0, 1)$.
- Si $Y \sim N(\mu, \sigma^2)$ y $a, b \in \mathbb{R}$, entonces $aY + b \sim N(a\mu + b, a^2\sigma^2)$.

Distribución normal multivariante

Sea $\mathbf{V} = (\sigma_{ij})$ una matriz cuadrada de orden n , con coeficientes reales, simétrica y definida positiva; y sea $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^t \in \mathbb{R}^n$

DEFINICIÓN. Diremos que el vector aleatorio n -dimensional $\mathbf{Y} = (Y_1, \dots, Y_n)^t$ sigue *distribución Normal n -dimensional de parámetros $\boldsymbol{\mu}$ y \mathbf{V}* ($\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \mathbf{V})$) si tiene la función de densidad conjunta:

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{n/2} |\mathbf{V}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^t \mathbf{V}^{-1} (\mathbf{y} - \boldsymbol{\mu})\right\} \quad , \quad \mathbf{y} = (y_1, \dots, y_n)^t \in \mathbb{R}^n.$$

PROPOSICIÓN. La función característica y la función generatriz de momentos de \mathbf{Y} son, respectivamente:

$$\varphi_{\mathbf{Y}}(\mathbf{s}) = \exp\{i\mathbf{s}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{s}'\mathbf{V}\mathbf{s}\} \quad , \quad M_{\mathbf{Y}}(\mathbf{s}) = \exp\{\mathbf{s}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{s}'\mathbf{V}\mathbf{s}\} \quad , \quad \mathbf{s} \in \mathbb{R}^n$$

Propiedades

1. El vector de medias de \mathbf{Y} es $\boldsymbol{\mu}$ y su matriz de covarianzas es \mathbf{V} .
2. Si $\mathbf{Z} = \mathbf{A}\mathbf{Y} + \mathbf{b}$, con \mathbf{A} matriz $k \times n$ de rango k ($k \leq n$) y $\mathbf{b} \in \mathbb{R}^k$, entonces $\mathbf{Z} \sim N_k(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\mathbf{V}\mathbf{A}'$).
3. Las distribuciones marginales de una distribución Normal Multivariante son Normales. Concretamente, si $\mathbf{Y}_1 = (Y_1, \dots, Y_k)'$, $k < n$, entonces $\mathbf{Y}_1 \sim N_k(\boldsymbol{\mu}_1, \mathbf{V}_{11})$.
4. La distribución de $\mathbf{Y}_1 = (Y_1, \dots, Y_k)'$ condicionada a $Y_{k+1} = y_{k+1}, \dots, Y_n = y_n$ es $N_k(\boldsymbol{\mu}_1 + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}(\mathbf{y}_2 - \boldsymbol{\mu}_2), \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{12}')$, siendo $\mathbf{y}_2 = (y_{k+1}, \dots, y_n)'$.
5. Si $\mathbf{Y}' = (\mathbf{Y}'_1, \mathbf{Y}'_2)$ con $\mathbf{Y}_1 = (Y_1, \dots, Y_k)'$ e $\mathbf{Y}_2 = (Y_{k+1}, \dots, Y_n)'$ entonces \mathbf{Y}_1 y \mathbf{Y}_2 son independientes si y sólo si $\text{Cov}[\mathbf{Y}_1, \mathbf{Y}_2] = \mathbf{0}$.
6. $\mathbf{Y} = (Y_1, \dots, Y_n)' \sim N_n(\boldsymbol{\mu}, \mathbf{V})$ si y sólo si toda combinación lineal de Y_1, \dots, Y_n sigue distribución Normal (i.e. para todo $\boldsymbol{\lambda} \in \mathbb{R}^n$, $\boldsymbol{\lambda}'\mathbf{Y} \sim N(\boldsymbol{\lambda}'\boldsymbol{\mu}, \boldsymbol{\lambda}'\mathbf{V}\boldsymbol{\lambda})$).
7. Las variables aleatorias Y_1, \dots, Y_n son independientes e idénticamente distribuidas con distribución $N(0, \sigma^2)$ si y sólo si el vector aleatorio $\mathbf{Y} = (Y_1, \dots, Y_n)' \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$.

Distribuciones de probabilidad asociadas a la Normal

Distribución Chi-cuadrado no central

DEFINICIÓN. Si $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \mathbf{I}_n)$, llamaremos **Chi-cuadrado no central** con n grados de libertad y parámetro de descentralización $\mu^* = \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\mu}$ a la distribución de la variable aleatoria $Q = \mathbf{Y}'\mathbf{Y}$ y escribiremos $Q \sim \chi^2(n, \mu^*)$.

Propiedades

1. Q tiene función de densidad

$$f(x) = \sum_{k=0}^{\infty} \exp\{-\mu^*\} \frac{\mu^{*k} x^{\frac{n+2k}{2}-1} \exp\{-\frac{x}{2}\}}{k! \Gamma(\frac{n+2k}{2}) 2^{\frac{n+2k}{2}}} \quad \text{si } x > 0; \quad 0 \quad \text{si } x \leq 0$$

2. La función generatriz de momentos de Q es

$$M_Q(s) = (1 - 2s)^{-n/2} \exp\{-\mu^*(1 - \frac{1}{1-2s})\} \quad , \quad s \text{ en un entorno de } 0$$

3. $\chi^2(n, 0) \equiv \chi^2(n)$.

4. Si $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$, entonces $\mathbf{Y}^t \mathbf{Y} / \sigma^2 \sim \chi^2(n, \boldsymbol{\mu}^*)$, siendo $\boldsymbol{\mu}^* = \frac{1}{2\sigma^2} \boldsymbol{\mu}^t \boldsymbol{\mu}$.
5. Sean $Q_i \sim \chi^2(n_i, \mu_i)$, $i = 1, \dots, k$, variables aleatorias independientes. Entonces

$$\sum_{i=1}^k Q_i \sim \chi^2\left(\sum_{i=1}^k n_i, \sum_{i=1}^k \mu_i\right)$$

Distribución F-Snedecor no central

DEFINICIÓN. Si $Q_1 \sim \chi^2(n_1, \mu)$ y $Q_2 \sim \chi^2(n_2, 0)$ son variables aleatorias independientes, llamaremos **F-Snedecor no central** con n_1, n_2 grados de libertad y parámetro de descentralización μ a la distribución de la variable aleatoria

$$Z = \frac{Q_1/n_1}{Q_2/n_2}$$

y escribiremos $Z \sim F(n_1, n_2, \mu)$.

Z tiene función de densidad

$$f(x) = \sum_{k=0}^{\infty} e^{-\mu} \frac{\mu^k}{k!} \frac{\binom{n_1}{n_2}^{\frac{n_1+2k}{2}} \Gamma\left(\frac{n_1+n_2+2k}{2}\right)}{\Gamma\left(\frac{n_1+2k}{2}\right) \Gamma\left(\frac{n_2}{2}\right)} x^{\frac{n_1+2k}{2}-1} \left(1 + \frac{n_1}{n_2} x\right)^{-\frac{n_1+n_2+2k}{2}} \quad \text{si } x > 0; \quad 0 \quad \text{si } x \leq 0$$

OBSERVACIÓN:

- $F(n_1, n_2, 0) \equiv F(n_1, n_2)$.
- Si $T \sim t(n)$ entonces $T^2 \sim F(1, n)$.

1.4 Formas cuadráticas.

Sea $\mathbf{A} = (a_{ij})$ una matriz cuadrada, simétrica de orden n y sea $\mathbf{Y} = (Y_1, \dots, Y_n)^t$ un vector aleatorio. Llamaremos *forma cuadrática* asociada al vector aleatorio \mathbf{Y} a la variable aleatoria:

$$\mathbf{Y}^t \mathbf{A} \mathbf{Y} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} Y_i Y_j = \sum_{i=1}^n a_{ii} Y_i^2 + \sum_{i < j} 2a_{ij} Y_i Y_j$$

Los siguientes resultados nos relacionan la distribución y los momentos de $\mathbf{Y}^t \mathbf{A} \mathbf{Y}$ con los de \mathbf{Y} . Supongamos que \mathbf{Y} tiene vector de medias $\boldsymbol{\mu}$ y matriz de covarianzas \mathbf{V} . Entonces se verifica:

- a) $E[\mathbf{Y}^t \mathbf{A} \mathbf{Y}] = \text{tr}(\mathbf{A} \mathbf{V}) + \boldsymbol{\mu}^t \mathbf{A} \boldsymbol{\mu}$
Si además $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \mathbf{V})$, entonces
- b) $\text{Var}[\mathbf{Y}^t \mathbf{A} \mathbf{Y}] = 2 \text{tr}((\mathbf{A} \mathbf{V})^2) + 4 \boldsymbol{\mu}^t \mathbf{A} \mathbf{V} \mathbf{A} \boldsymbol{\mu}$.
- c) $\text{Cov}[\mathbf{Y}, \mathbf{Y}^t \mathbf{A} \mathbf{Y}] = 2 \mathbf{V} \mathbf{A} \boldsymbol{\mu}$.

Distribuciones de las formas cuadráticas.

Podemos establecer el siguiente resultado:

TEOREMA 1. Sea $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \mathbf{V})$ y sea \mathbf{A} una matriz simétrica de orden n . Entonces se verifica:

$$\mathbf{Y}^t \mathbf{A} \mathbf{Y} \sim \chi^2(r(\mathbf{A}), \frac{1}{2} \boldsymbol{\mu}^t \mathbf{A} \boldsymbol{\mu}) \text{ cualquiera que sea } \boldsymbol{\mu} \iff \mathbf{A} \mathbf{V} \text{ es idempotente.}$$

Independencia de las formas cuadráticas.

Se verifican los siguientes resultados:

TEOREMA 2. Sea $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \mathbf{V})$. Sea \mathbf{A} una matriz simétrica y semidefinida positiva de orden n y \mathbf{B} una matriz $s \times n$ de rango s . Entonces se verifica:

$$\mathbf{Y}^t \mathbf{A} \mathbf{Y} \text{ y } \mathbf{B} \mathbf{Y} \text{ son independientes cualquiera que sea } \boldsymbol{\mu} \iff \mathbf{B} \mathbf{V} \mathbf{A} = \mathbf{0}.$$

TEOREMA 3. Sea $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \mathbf{V})$. Sean \mathbf{A} y \mathbf{B} dos matrices simétricas y semidefinidas positivas de orden n . Entonces se verifica:

$$\mathbf{Y}^t \mathbf{A} \mathbf{Y} \text{ e } \mathbf{Y}^t \mathbf{B} \mathbf{Y} \text{ son independientes cualquiera que sea } \boldsymbol{\mu} \iff \mathbf{A} \mathbf{V} \mathbf{B} = \mathbf{0} (\iff \mathbf{B} \mathbf{V} \mathbf{A} = \mathbf{0}).$$

Tema 2: Modelo Lineal de Rango Completo

En este tema, partimos de un vector aleatorio $\mathbf{Y} = (Y_1, \dots, Y_n)^t$ que representa un conjunto de datos tomados de cierta variable bajo estudio. Si asumiéramos que sus coordenadas, Y_1, \dots, Y_n , son independientes e idénticamente distribuidas, constituirían una muestra aleatoria simple que es la base de la teoría básica en inferencia estadística.

Pero la inferencia estadística no se desarrolla siempre bajo este supuesto. Podemos asumir, por ejemplo, que la media μ_i de cada Y_i depende de observaciones de otros parámetros con un significado práctico en el estudio que se está llevando a cabo. Si esta dependencia es lineal, nos encontraremos ante un Modelo Lineal. El estudio de estos modelos es el objeto de esta asignatura.

2.1 Modelos lineales: definición y ejemplos

En este apartado vamos a proporcionar la definición de Modelo Lineal, junto con una doble clasificación de los Modelos Lineales que marca la estructura de los contenidos de esta asignatura. Por último, presentaremos algunos ejemplos de modelos clásicos para el análisis de datos que pueden ser enmarcados dentro de la teoría de los Modelos Lineales.

Definición 1 Sea $\mathbf{Y} = (Y_1, \dots, Y_n)^t$ un vector aleatorio n -dimensional y \mathbf{X} una matriz de orden $n \times p$ ($p < n$). Diremos que \mathbf{Y} se ajusta a un Modelo Lineal si existe $\boldsymbol{\beta} \in \mathbb{R}^p$ tal que

$$E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta},$$

Si el vector aleatorio \mathbf{Y} se ajusta a un Modelo Lineal,

$$\mathbf{Y} = E[\mathbf{Y}] + \mathbf{Y} - E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta} + (\mathbf{Y} - E[\mathbf{Y}]).$$

Por tanto, podemos también decir que \mathbf{Y} se ajusta a un Modelo Lineal si

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mathcal{E}}, \quad (1)$$

siendo $\boldsymbol{\mathcal{E}} = \mathbf{Y} - E[\mathbf{Y}]$ un vector aleatorio n dimensional con $E[\boldsymbol{\mathcal{E}}] = \mathbf{0}$.

En la práctica, las coordenadas de \mathbf{Y} representan las observaciones de la variable (aleatoria) de interés en el estudio estadístico que se lleva a cabo. En este caso, no estamos suponiendo que Y_1, \dots, Y_n sean independientes e idénticamente distribuidas como se asumía en una muestra aleatoria simple. La ecuación (1) establece que este vector aleatorio se puede descomponer como suma de los siguientes términos:

- a) Una combinación lineal de las columnas de la matriz \mathbf{X} determinada por el vector de parámetros $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^t \in \mathbb{R}^p$. Los coeficientes de la matriz \mathbf{X} , denominada **matriz del modelo**, son valores conocidos, bien porque provengan de ciertas variables presentes en el estudio estadístico, bien porque definen ciertas hipótesis que se asumen sobre la variable de interés. El vector de parámetros $\boldsymbol{\beta}$ se supone desconocido. Sobre dicho vector se centrará principalmente el estudio inferencial que desarrollaremos en este tema.
- b) Un vector aleatorio $\boldsymbol{\mathcal{E}}$ cuyas coordenadas representan los “errores” (aleatorios) que se cometen al representar los datos incluidos en \mathbf{Y} como combinación lineal de las columnas de \mathbf{X} . Dichos errores se asumen con media cero, si bien son necesarias hipótesis adicionales para poder desarrollar la teoría inferencial. En la práctica estos errores no son observables, por lo que es difícil contrastar las hipótesis teóricas que se asuman sobre ellos.

Podemos establecer dos clasificaciones de los modelos lineales:

- Atendiendo al **rango de la matriz X**:
 - Diremos que el modelo es de rango completo si $r(\mathbf{X}) = p$.
 - Diremos que el modelo es de rango no completo: $r(\mathbf{X}) < p$.
- Según las hipótesis que se asuman sobre la **distribución del vector de errores \mathcal{E}** , o equivalentemente sobre la distribución del vector \mathbf{Y} .
 - Diremos que \mathbf{Y} se ajusta a un **Modelo Lineal Básico**, si asumimos que existe $\sigma^2 > 0$ tal que $\text{Cov}[\mathbf{Y}] = \text{Cov}[\mathcal{E}] = \sigma^2 \mathbf{I}_n$.
 - Diremos que \mathbf{Y} se ajusta a un **Modelo Lineal Normal** si $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$ para algún $\sigma^2 > 0$ (equivalentemente $\mathcal{E} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$).

En este tema estudiaremos los Modelos Lineales de rango completo, centrando dicho estudio en la inferencia estadística sobre los parámetros de interés en el modelo: σ^2 , $\boldsymbol{\beta}$ y las combinaciones lineales de las coordenadas de $\boldsymbol{\beta}$. Comenzaremos dicho estudio abordando el Modelo Lineal Básico, donde sólo es factible la estimación de los parámetros. Esta teoría de estimación se completará en el Modelo Lineal Normal, obteniendo propiedades relativas a la distribución de los estimadores, que posteriormente utilizaremos para desarrollar el estudio de la estimación por intervalos y regiones de confianza y la teoría de contraste de hipótesis.

Previamente veamos unos ejemplos de estudios estadísticos sencillos que se enmarcan en el contexto de los Modelos Lineales.

Ejemplo 1

Consideremos la relación entre la altura h y el peso w de los habitantes de una población. Sabemos que dicha relación puede existir pero no es determinística. Si asumimos que la relación entre ambas variables es lineal, podemos escribir la ecuación que la define del siguiente modo:

$$w = \beta_0 + \beta_1 h + \mathcal{E}$$

Se trata de un modelo de Regresión Lineal Simple. Para obtener información sobre los parámetros β_0 y β_1 , se procederá midiendo las alturas $\mathbf{H} = (H_1, \dots, H_n)^t$ de n individuos y sus pesos respectivos $\mathbf{W} = (W_1, \dots, W_n)^t$. Asumiendo que el modelo es correcto, estos datos deberían ajustarse al mismo, es decir,

$$W_i = \beta_0 + \beta_1 H_i + \mathcal{E}_i, \quad i = 1, \dots, n$$

Escribiendo estas igualdades matricialmente, obtenemos

$$\mathbf{W} = \mathbf{1}_n \beta_0 + \mathbf{H} \beta_1 + \mathcal{E} = (\mathbf{1}_n | \mathbf{H}) \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \mathcal{E}$$

siendo $\mathbf{1}_n = (1, \dots, 1)^t$ y $\mathcal{E} = (\mathcal{E}_1, \dots, \mathcal{E}_n)^t$. Se trata de un modelo lineal con matriz $\mathbf{X} = (\mathbf{1}_n | \mathbf{H})$. Si no todas las alturas son iguales, $r(\mathbf{X}) = 2$, por lo tanto el Modelo de Regresión Lineal Simple puede ser considerado un caso particular de Modelo Lineal de rango completo.

Ejemplo 2

Consideremos la posible influencia que sobre la producción de trigo (Y) tienen 3 fertilizantes F_1 , F_2 y F_3 . Para ello se tratan n_i parcelas con el fertilizante F_i , obteniéndose unas producciones de Y_{i1}, \dots, Y_{in_i} , $i = 1, 2, 3$.

Si denotamos como μ a la producción media de trigo si ninguna causa de variabilidad estuviera presente, τ_i al efecto que sobre la producción media tiene el fertilizante F_i , $i = 1, 2, 3$ entonces podemos suponer que

$$E[Y_{ij}] = \mu + \tau_i \quad i = 1, 2, 3 \quad j = 1, 2, \dots, n_i$$

Este tipo de modelos se denominan Modelos de Diseño de Experimentos. Si denotamos como

$$\mathbf{Y} = (Y_{11}, \dots, Y_{1n_1}, Y_{21}, \dots, Y_{2n_2}, Y_{31}, \dots, Y_{3n_3})^t$$

al vector con todos los datos de producción, podemos escribir matricialmente las ecuaciones anteriores del siguiente modo:

$$E[\mathbf{Y}] = \begin{pmatrix} 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix} = \mathbf{X}\boldsymbol{\beta}$$

Por lo tanto, el vector \mathbf{Y} se ajusta a un Modelo Lineal, siendo el vector de parámetros $\boldsymbol{\beta} = (\mu, \tau_1, \tau_2, \tau_3)^t$. En este caso $r(\mathbf{X}) = 3$, menor que el número de columnas, por lo que el Modelo de Diseño de Experimentos es un Modelo Lineal de rango no completo.

Ejercicio propuesto:

Si en el Ejemplo 2 anterior suponemos que se toman también los datos de humedad de las parcelas $\mathbf{H} = (H_{11}, \dots, H_{1n_1}, H_{21}, \dots, H_{2n_2}, H_{31}, \dots, H_{3n_3})^t$ y asumimos que la relación entre humedad y producción es lineal, define un modelo lineal que describa adecuadamente a los datos.

2.2 Modelo Lineal Básico: estimación puntual

Consideraremos un vector aleatorio \mathbf{Y} que se ajusta al Modelo Lineal de la Definición 1. Asumiremos que dicho modelo es de rango completo, es decir, $r(\mathbf{X}) = p$ y que se dan las hipótesis del Modelo Lineal Básico, es decir, $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$, $E[\boldsymbol{\mathcal{E}}] = \mathbf{0}$, $\text{Cov}[\mathbf{Y}] = \text{Cov}[\boldsymbol{\mathcal{E}}] = \sigma^2 \mathbf{I}_n$ para algún $\sigma^2 > 0$.

Nuestro objetivo en este apartado es desarrollar la teoría de estimación puntual para los principales parámetros del modelo: la varianza de observaciones y errores, σ^2 , el vector de parámetros, $\boldsymbol{\beta}$, y los funcionales lineales $\psi = \boldsymbol{\lambda}'\boldsymbol{\beta}$ con $\boldsymbol{\lambda} \in \mathbb{R}^p$. Comenzaremos nuestro estudio con la estimación puntual de $\boldsymbol{\beta}$, que es el principal parámetro de este modelo.

Estimación de $\boldsymbol{\beta}$

Para obtener un estimador de $\boldsymbol{\beta}$ utilizaremos el método de mínimos cuadrados, es decir, proporcionaremos como estimador de $\boldsymbol{\beta}$ el valor que minimice la suma de las distancias al cuadrado entre las observaciones y sus medias. Se trata por tanto de minimizar la siguiente función de $\boldsymbol{\beta}$:

$$f(\boldsymbol{\beta}) = \sum_{i=1}^n (Y_i - E[Y_i])^2 = (\mathbf{Y} - E[\mathbf{Y}])'(\mathbf{Y} - E[\mathbf{Y}]) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}). \quad (2)$$

Proposición 1 *El estimador de mínimos cuadrados de $\boldsymbol{\beta}$ es:*

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

y verifica las siguientes propiedades:

- $E[\widehat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$, es decir, $\widehat{\boldsymbol{\beta}}$ es un estimador insesgado para $\boldsymbol{\beta}$.
- $\text{Cov}[\widehat{\boldsymbol{\beta}}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$
- El error cuadrático medio del estimador $\widehat{\boldsymbol{\beta}}$ viene dado por $ECM(\widehat{\boldsymbol{\beta}}) = \sigma^2 \text{tr}((\mathbf{X}'\mathbf{X})^{-1})$.

Demostración

Si denotamos $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$, $\mathbf{X}'\mathbf{Y} = (b_1, \dots, b_p)'$ y $\mathbf{X}'\mathbf{X} = (a_{ij})_{1 \leq i, j \leq p}$, obtenemos que la expresión a minimizar, (2), puede escribirse del siguiente modo:

$$f(\boldsymbol{\beta}) = \mathbf{Y}'\mathbf{Y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{Y}'\mathbf{Y} - 2 \sum_{i=1}^p b_i \beta_i + \sum_{i=1}^p a_{ii} \beta_i^2 + 2 \sum_{i < j} a_{ij} \beta_i \beta_j$$

que es un polinomio de grado 2 en β_1, \dots, β_p y por tanto infinitamente diferenciable. El mínimo absoluto de esta función, de existir, debe ser un extremo relativo. Calculamos las derivadas parciales de f :

$$\frac{\partial f}{\partial \beta_k} = -2b_k + 2a_{kk}\beta_k + 2 \sum_{i \neq k} a_{ki}\beta_i = -2b_k + 2 \sum_{i=1}^p a_{ki}\beta_i, \quad k = 1, \dots, p.$$

Denotando $\nabla f = \left(\frac{\partial f}{\partial \beta_1}, \dots, \frac{\partial f}{\partial \beta_p} \right)'$ al vector de las derivadas parciales de f , obtenemos

$$\nabla f(\boldsymbol{\beta}) = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

Si $\widehat{\beta}$ es un extremo relativo de dicha función entonces

$$\nabla f(\widehat{\beta}) = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\widehat{\beta} = \mathbf{0}$$

o equivalentemente $\widehat{\beta}$ debe ser solución del sistema de ecuaciones lineales:

$$\mathbf{X}'\mathbf{X}\widehat{\beta} = \mathbf{X}'\mathbf{Y} \quad (3)$$

La matriz de los coeficientes $\mathbf{X}'\mathbf{X}$ es cuadrada de orden p y $r(\mathbf{X}'\mathbf{X}) = r(\mathbf{X}) = p$, luego es invertible y el sistema (3) es compatible determinado, siendo su única solución:

$$\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

Para comprobar que es mínimo relativo calculamos las derivadas parciales de orden 2:

$$\frac{\partial^2 f}{\partial \beta_k \partial \beta_l} = 2a_{kl}, \quad k, l = 1, \dots, p.$$

de donde se obtiene que la matriz Hessiana de f es

$$\mathbf{H}_f(\beta) = 2\mathbf{X}'\mathbf{X}.$$

Como $\mathbf{X}'\mathbf{X}$ es semidefinida positiva e invertible, es definida positiva, luego \mathbf{H}_f es definida positiva en cualquier punto, en particular $\mathbf{H}_f(\widehat{\beta})$ es definida positiva y $\widehat{\beta}$ es un mínimo relativo.

Además, como el gradiente de f no se anula en ningún otro punto, $\widehat{\beta}$ es un mínimo absoluto de f y por tanto es el estimador de mínimos cuadrados de β .

La demostración de las propiedades a) y b) del estimador $\widehat{\beta}$ queda como ejercicio. En cuanto a la propiedad c), basta tener en cuenta que, puesto que $\widehat{\beta} = (\widehat{\beta}_1, \dots, \widehat{\beta}_p)'$ es insesgado,

$$ECM(\widehat{\beta}) = \sum_{i=1}^n \text{Var}[\widehat{\beta}_i] = \text{tr}(\text{Cov}[\widehat{\beta}]) = \sigma^2 \text{tr}((\mathbf{X}'\mathbf{X})^{-1}).$$

Ejercicio propuesto:

Demostrar las propiedades a) y b) de la Proposición 1.

Una vez obtenido el estimador $\widehat{\boldsymbol{\beta}}$ y sus propiedades, vamos a compararlo con otros estimadores de $\boldsymbol{\beta}$, concretamente con los estimadores lineales insesgados.

Definición 2 Se dice que \mathbf{T} es un estimador lineal de $\boldsymbol{\beta}$ si $\mathbf{T} = \mathbf{A}\mathbf{Y}$ para alguna matriz \mathbf{A} de orden $p \times n$.

El estimador $\widehat{\boldsymbol{\beta}} = ((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y}$ es lineal insesgado para $\boldsymbol{\beta}$. En el siguiente resultado, probamos que $\widehat{\boldsymbol{\beta}}$ minimiza el error cuadrático medio dentro de la clase de estimadores lineales insesgados de $\boldsymbol{\beta}$.

Teorema 2 (Gauss-Markov) $\widehat{\boldsymbol{\beta}}$ es el estimador lineal insesgado de mínima varianza de $\boldsymbol{\beta}$, es decir, si \mathbf{T} es otro estimador lineal insesgado de $\boldsymbol{\beta}$, entonces $ECM(\widehat{\boldsymbol{\beta}}) \leq ECM(\mathbf{T})$ y se da igualdad si y sólo si $\mathbf{T} = \widehat{\boldsymbol{\beta}}$.

Demostración

Supongamos que $\mathbf{T} = \mathbf{A}\mathbf{Y}$ es otro estimador lineal insesgado de $\boldsymbol{\beta}$. Si descomponemos $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}$, entonces

$$\boldsymbol{\beta} = E[\mathbf{T}] = E[\widehat{\boldsymbol{\beta}} + \mathbf{B}\mathbf{Y}] = \boldsymbol{\beta} + \mathbf{B}E[\mathbf{Y}] = \boldsymbol{\beta} + \mathbf{B}\mathbf{X}\boldsymbol{\beta}$$

Luego $\mathbf{B}\mathbf{X} = \mathbf{0}$. Teniendo esto en cuenta, podemos calcular la matriz de covarianzas de \mathbf{T} del siguiente modo:

$$\begin{aligned} \text{Cov}[\mathbf{T}] &= \mathbf{A} \text{Cov}[\mathbf{Y}]\mathbf{A}' = \sigma^2 \mathbf{A}\mathbf{A}' = \sigma^2((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}' \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{B}\mathbf{X})' + \sigma^2\mathbf{B}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2\mathbf{B}\mathbf{B}' = \sigma^2((\mathbf{X}'\mathbf{X})^{-1} + \mathbf{B}\mathbf{B}') \end{aligned}$$

De este modo, si $\mathbf{B} = (b_{ij})_{1 \leq i \leq p, 1 \leq j \leq n}$, obtenemos que

$$ECM(\mathbf{T}) = \text{tr}(\text{Cov}[\mathbf{T}]) = \sigma^2 \text{tr}((\mathbf{X}'\mathbf{X})^{-1}) + \sigma^2 \text{tr}(\mathbf{B}\mathbf{B}') = ECM(\widehat{\boldsymbol{\beta}}) + \sigma^2 \sum_{i=1}^p \sum_{j=1}^n b_{ij}^2$$

Por tanto, $ECM(\mathbf{T}) \geq ECM(\widehat{\boldsymbol{\beta}})$ y se dará la igualdad si y sólo si $b_{ij} = 0$ para todo $i = 1, \dots, p, j = 1, \dots, n$, es decir, si y sólo si $\mathbf{B} = \mathbf{0}$ o equivalentemente, si y sólo si $\mathbf{T} = \widehat{\boldsymbol{\beta}}$.

Valores ajustados y residuos. Estimación de σ^2

Una vez determinado el estimador $\widehat{\boldsymbol{\beta}}$, podemos calcular los valores ajustados por el modelo lineal y también los residuos del modelo, que pueden interpretarse como estimaciones de los errores cometidos en dicho ajuste.

Definición 3 (valores ajustados y residuos) Llamaremos valores ajustados por el modelo al vector aleatorio $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$. El vector $\mathbf{e} = \mathbf{Y} - \widehat{\mathbf{Y}}$ será denominado vector de residuos.

Los valores ajustados y los residuos jugarán un papel fundamental en el análisis de un modelo lineal. Intuitivamente, los valores ajustados pueden ser vistos como estimadores de $\boldsymbol{\mu} = E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$, es decir, del valor medio del vector de observaciones. Constituyen por tanto las observaciones que cabría esperar si el modelo lineal fuera adecuado. Los residuos son por tanto la diferencia entre lo observado y lo “predicho” por el modelo y por ello nos permiten conocer hasta qué punto el modelo nos proporciona un buen ajuste.

Es sencillo comprobar que ambos vectores son transformaciones lineales de \mathbf{Y} :

$$\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{H}\mathbf{Y}, \quad \mathbf{e} = \mathbf{Y} - \widehat{\mathbf{Y}} = (\mathbf{I}_n - \mathbf{H})\mathbf{Y} \quad \text{siendo } \mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad (4)$$

La matriz \mathbf{H} está asociada a la proyección ortogonal sobre el subespacio de \mathbb{R}^n generado por las columnas de \mathbf{X} e $\mathbf{I}_n - \mathbf{H}$ está asociada a la proyección ortogonal sobre el subespacio ortogonal de dicho subespacio. Más concretamente, podemos establecer el siguiente resultado:

Lema 3 Se verifican las siguientes propiedades:

a) Las matrices \mathbf{H} e $\mathbf{I}_n - \mathbf{H}$ son simétricas e idempotentes.

b) $(\mathbf{I}_n - \mathbf{H})\mathbf{X} = \mathbf{0}$

Demostración

Se deja como ejercicio.

Teniendo en cuenta que los residuos nos proporcionan una medida del error cometido por el modelo, la suma de cuadrados de los residuos (o del error) puede sernos útil para calibrar la bondad de ajuste del modelo. Se define del siguiente modo:

$$SCE = \mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})$$

SCE es una forma cuadrática asociada a \mathbf{Y} . En efecto, aplicando el Lema 3, $\mathbf{I}_n - \mathbf{H}$ es simétrica e idempotente, y por tanto,

$$SCE = (\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \mathbf{Y}'(\mathbf{I}_n - \mathbf{H})^2\mathbf{Y} = \mathbf{Y}'(\mathbf{I}_n - \mathbf{H})\mathbf{Y}. \quad (5)$$

En el siguiente resultado se calcula el valor esperado de SCE y se obtiene un estimador insesgado para σ^2 :

Proposición 4 Se verifica que

$$E[SCE] = (n - p)\sigma^2,$$

y en consecuencia, un estimador insesgado de σ^2 es

$$\widehat{\sigma}^2 = \frac{1}{n - p}(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}).$$

Demostración

Puesto que SCE es una forma cuadrática asociada a \mathbf{Y} , utilizando la fórmula para el cálculo de la esperanza de una forma cuadrática (véase propiedad a) de las formas cuadráticas en la última sección del Tema 1) con $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, $\mathbf{V} = \sigma^2\mathbf{I}_n$ y $\mathbf{A} = \mathbf{I}_n - \mathbf{H}$, se obtiene que

$$E[SCE] = E[\mathbf{Y}'(\mathbf{I}_n - \mathbf{H})\mathbf{Y}] = \sigma^2 \text{tr}(\mathbf{I}_n - \mathbf{H}) + \boldsymbol{\beta}'\mathbf{X}'(\mathbf{I}_n - \mathbf{H})\mathbf{X}\boldsymbol{\beta}.$$

Aplicando el Lema 3, $(\mathbf{I}_n - \mathbf{H})\mathbf{X} = \mathbf{0}$, y utilizando ahora las propiedades de la traza,

$$E[SCE] = \sigma^2(\text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')) = \sigma^2(n - \text{tr}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X})) = \sigma^2(n - \text{tr}(\mathbf{I}_p)) = \sigma^2(n - p).$$

Estimación de funcionales lineales de $\boldsymbol{\beta}$

En este apartado, se estudia la estimación de funcionales lineales $\psi = \boldsymbol{\lambda}'\boldsymbol{\beta}$, $\boldsymbol{\lambda} \in \mathbb{R}^p$. Puesto que ya sabemos que $\widehat{\boldsymbol{\beta}}$ es un estimador de $\boldsymbol{\beta}$ con buenas propiedades, un estimador apropiado para ψ será:

$$\widehat{\psi} = \boldsymbol{\lambda}'\widehat{\boldsymbol{\beta}}.$$

En el siguiente resultado, proporcionamos algunas propiedades del estimador $\widehat{\psi}$.

Proposición 5 El estimador $\widehat{\psi}$ es insesgado para ψ y su error cuadrático medio (varianza) es $ECM(\widehat{\psi}) = \text{Var}(\widehat{\psi}) = \sigma^2\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\lambda}$. Además, se verifica que $\widehat{\psi}$ es el estimador lineal insesgado de mínima varianza de ψ , en el sentido de que si T es otro estimador lineal insesgado de ψ , entonces $ECM(\widehat{\psi}) \leq ECM(T)$ y se da igualdad si y sólo si $T = \widehat{\psi}$.

Demostración

Queda como ejercicio propuesto.

Observación: No es posible calcular exactamente el error cuadrático medio de los estimadores $\widehat{\beta}$ y $\widehat{\psi}$ ya que desconocemos el valor del parámetro σ^2 en las expresiones de $\text{Cov}(\widehat{\beta})$ y $\text{Var}(\widehat{\psi})$. No obstante, podemos estimar dicho error sustituyendo σ^2 por $\widehat{\sigma}^2$ en dichas expresiones.

Ejercicio propuesto: Demostrar el Lema 3 y la Proposición 5.

2.3. Modelo Lineal Normal: estimación puntual

En este apartado, asumiremos que el vector aleatorio \mathbf{Y} se ajusta a un Modelo Lineal Normal, es decir $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$, siendo \mathbf{X} una matriz $n \times p$ ($p < n$) y $r(\mathbf{X}) = p$, $\boldsymbol{\beta} \in \mathbb{R}^p$ y $\sigma^2 > 0$.

En estas condiciones, la función de densidad de \mathbf{Y} , la de la normal multivariante que vimos en el tema 1, es

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \quad \mathbf{y} \in \mathbb{R}^n$$

Si vemos esta expresión como una función de los parámetros $\boldsymbol{\beta}$ y σ^2 , obtenemos la función de verosimilitud del Modelo Lineal Normal:

$$L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y}) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right\}.$$

En este contexto, calcularemos los estimadores de $\boldsymbol{\beta}$ y σ^2 por el método de máxima verosimilitud, para lo cual hemos de maximizar la función $L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y})$ en el conjunto $\mathbb{R}^p \times \mathbb{R}_+$.

Proposición 6 *Los estimadores de máxima verosimilitud de $\boldsymbol{\beta}$ y σ^2 son:*

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad \widehat{\sigma}^2 = \frac{1}{n}(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}).$$

Además, la verosimilitud máxima del modelo es

$$L(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2; \mathbf{Y}) = \left(\frac{e^{-1}}{2\pi \widehat{\sigma}^2} \right)^{n/2}.$$

Demostración

En primer lugar, notemos que, puesto que el logaritmo es una función creciente, los valores de $\boldsymbol{\beta}$ y σ^2 que maximizan L son los mismos que los que maximizan la log-verosimilitud $\ell(\boldsymbol{\beta}, \sigma^2; \mathbf{Y})$, definida del siguiente modo:

$$\ell(\boldsymbol{\beta}, \sigma^2; \mathbf{Y}) = \log L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \quad (6)$$

De la expresión (6) se observa que, fijado el valor de $\sigma^2 > 0$, la función de $\boldsymbol{\beta}$, $\ell(\boldsymbol{\beta}, \sigma^2; \mathbf{Y})$ alcanzará su máximo cuando la función $f(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ sea mínima. En la Proposición 1, probamos que dicho mínimo se alcanza en el estimador de mínimos cuadrados de $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

Sólo resta obtener el máximo de la función $\ell(\widehat{\boldsymbol{\beta}}, \sigma^2; \mathbf{Y})$ en \mathbb{R}_+ para obtener el estimador de máxima verosimilitud de σ^2 . Puesto que es una función infinitamente derivable en \mathbb{R}_+ el máximo, caso de existir, será un máximo relativo. Calculemos la derivada de $\ell(\widehat{\boldsymbol{\beta}}, \sigma^2; \mathbf{Y})$ respecto a σ^2 :

$$\frac{d \ell(\widehat{\boldsymbol{\beta}}, \sigma^2; \mathbf{Y})}{d\sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})$$

Si denotamos por $\widehat{\sigma}^2$ al valor donde se alcanza el máximo, por ser extremo relativo dicho valor debe verificar

$$-\frac{n}{2} \frac{1}{\widehat{\sigma}^2} + \frac{1}{2(\widehat{\sigma}^2)^2}(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = 0$$

o equivalentemente

$$n\widehat{\sigma}^2 - (\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = 0$$

de donde se deduce, que el único candidato a máximo relativo es

$$\tilde{\sigma}^2 = \frac{1}{n}(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})$$

Para concluir la demostración de la primera parte del enunciado, basta probar que la derivada segunda de $\ell(\widehat{\boldsymbol{\beta}}, \sigma^2; \mathbf{Y})$ respecto a σ^2 en el punto $\tilde{\sigma}^2$ es menor que 0, lo cual se deja como ejercicio.

En cuanto a la verosimilitud máxima, su valor es

$$L(\widehat{\boldsymbol{\beta}}, \tilde{\sigma}^2; \mathbf{Y}) = \frac{1}{(2\pi)^{n/2}(\tilde{\sigma}^2)^{n/2}} \exp\left\{-\frac{1}{2\tilde{\sigma}^2}(n\tilde{\sigma}^2)\right\} = \frac{1}{(2\pi)^{n/2}(\tilde{\sigma}^2)^{n/2}} e^{-n/2} = \left(\frac{e^{-1}}{2\pi\tilde{\sigma}^2}\right)^{n/2}$$

lo cual concluye la demostración.

Observaciones:

- El estimador de $\boldsymbol{\beta}$ por máxima verosimilitud, $\widehat{\boldsymbol{\beta}}$, coincide con el de mínimos cuadrados, y por tanto conserva las propiedades de éste último.
- El estimador de σ^2 por máxima verosimilitud, $\tilde{\sigma}^2$, no coincide con el estimador $\widehat{\sigma}^2$ obtenido en el Modelo Lineal Básico. La relación entre ambos viene dada por:

$$\widehat{\sigma}^2 = \frac{n-p}{n} \tilde{\sigma}^2$$

Por tanto, el estimador $\tilde{\sigma}^2$ no es insesgado para σ^2 .

Proposición 7 (Distribución de los estimadores)

- $\widehat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$.
- $\frac{(n-p)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p)$.
- $\widehat{\boldsymbol{\beta}}$ y $\widehat{\sigma}^2$ son independientes.

Demostración

a) El estimador $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ tiene distribución normal p dimensional porque es una transformación lineal de \mathbf{Y} , que tiene distribución normal n dimensional. El vector de medias y la matriz de covarianzas fueron calculados en la Proposición 1.

b) Haciendo uso de la ecuación (5), podemos escribir $(n-p)\widehat{\sigma}^2/\sigma^2$ como una forma cuadrática del siguiente modo:

$$\frac{(n-p)\widehat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2}(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \frac{1}{\sigma^2}\mathbf{Y}'(\mathbf{I}_n - \mathbf{H})\mathbf{Y} \quad (7)$$

donde, recordemos, $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Para poder aplicar el Teorema 1 sobre formas cuadráticas, visto en el Tema 1, y deducir que la distribución de (7) es $\chi^2(r(\mathbf{A}), \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}/2)$, bastará probar que $\mathbf{A}\mathbf{V}$ es idempotente, siendo

$$\mathbf{A} = \frac{1}{\sigma^2}(\mathbf{I}_n - \mathbf{H}), \quad \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}, \quad \mathbf{V} = \sigma^2\mathbf{I}_n$$

Como $\mathbf{AV} = \mathbf{I}_n - \mathbf{H}$ y dicha matriz es idempotente (Lema 3), se satisface la hipótesis de Teorema 1 del Tema 1 para formas cuadráticas, luego (7) tiene distribución chi cuadrado y además

$$r(\mathbf{A}) = r(\mathbf{I}_n - \mathbf{H}) = \text{tr}(\mathbf{I}_n - \mathbf{H}) = n - p$$

y puesto que, aplicando de nuevo el Lema 3, $(\mathbf{I}_n - \mathbf{H})\mathbf{X} = \mathbf{0}$,

$$\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} = \boldsymbol{\beta}'\mathbf{X}'(\mathbf{I}_n - \mathbf{H})\mathbf{X}\boldsymbol{\beta} = 0$$

lo cual concluye la demostración de este apartado.

c) Según hemos visto en los apartados anteriores, $\widehat{\boldsymbol{\beta}}$ es una transformación lineal de \mathbf{Y} y $(n-p)\widehat{\sigma}^2$ es una forma cuadrática asociada a \mathbf{Y} . Aplicando el Teorema 2 sobre formas cuadráticas, visto en el Tema 1, para ver que son independientes bastará probar que $\mathbf{BVA} = \mathbf{0}$, siendo en este caso

$$\mathbf{A} = \mathbf{I}_n - \mathbf{H}, \quad \mathbf{V} = \sigma^2\mathbf{I}_n, \quad \mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

En efecto, utilizando de nuevo que $(\mathbf{I}_n - \mathbf{H})\mathbf{X} = \mathbf{0}$ (Lema 3), se obtiene que

$$\mathbf{BVA} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}_n - \mathbf{H}) = \mathbf{0}$$

lo cual concluye la demostración.

Proposición 8 (Suficiencia y completitud. Estimador insesgado de mínima varianza)

a) El estadístico $\mathbf{T} = \begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\sigma}^2 \end{pmatrix}$ es suficiente y completo para el parámetro $\begin{pmatrix} \boldsymbol{\beta} \\ \sigma^2 \end{pmatrix}$.

b) $\widehat{\boldsymbol{\beta}}$ y $\widehat{\sigma}^2$ son los estimadores insesgados de mínima varianza de $\boldsymbol{\beta}$ y σ^2 respectivamente.

Demostración

a) Probaremos en primer lugar que el estadístico $\mathbf{S} = \begin{pmatrix} \mathbf{X}'\mathbf{Y} \\ \mathbf{Y}'\mathbf{Y} \end{pmatrix}$ es suficiente y completo para el parámetro $p+1$ dimensional $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\beta} \\ \sigma^2 \end{pmatrix}$.

Para ello, tengamos en cuenta que la función de densidad de \mathbf{Y} se puede escribir del siguiente modo:

$$f_{\boldsymbol{\theta}}(\mathbf{Y}) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right\} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{Y}'\mathbf{Y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y})\right\} = g_{\boldsymbol{\theta}}(\mathbf{S})$$

Aplicando el Teorema 17 de factorización de Neyman-Halmos-Savage, el estadístico \mathbf{S} es suficiente. Además, si escribimos $\mathbf{X}'\mathbf{Y} = (S_1, \dots, S_p)'$ e $\mathbf{Y}'\mathbf{Y} = S_{p+1}$, podemos observar que

$$g_{\boldsymbol{\theta}}(\mathbf{S}) = C(\boldsymbol{\theta}) \exp\left\{-\frac{1}{2\sigma^2}S_{p+1} + \sum_{i=1}^p \frac{\beta_i}{\sigma^2}S_i\right\} = C(\boldsymbol{\theta}) \exp\left\{Q_{p+1}(\boldsymbol{\theta})\mathbf{Y}'\mathbf{Y} + \sum_{i=1}^p Q_i(\boldsymbol{\theta})S_i\right\}$$

siendo

$$C(\boldsymbol{\theta}) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right\}, \quad Q_{p+1}(\boldsymbol{\theta}) = -\frac{1}{2\sigma^2}, \quad Q_i(\boldsymbol{\theta}) = \frac{\beta_i}{\sigma^2} \quad i = 1, \dots, p.$$

Puesto que la distribución Normal multidimensional pertenece a la familia exponencial, el Teorema 18 de completitud en esta familia y obtenemos que el estadístico \mathbf{S} también es completo.

Para demostrar que \mathbf{T} es también suficiente y completo, bastará probar que es una transformación biyectiva de \mathbf{S} . En efecto,

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ \widehat{\sigma}^2 &= \frac{1}{n-p}(\mathbf{Y}'\mathbf{Y} - \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}) = \frac{1}{n-p}(\mathbf{Y}'\mathbf{Y} - (\mathbf{X}'\mathbf{Y})'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y})\end{aligned}$$

y por tanto, $\widehat{\boldsymbol{\beta}}$ y $\widehat{\sigma}^2$ se pueden obtener como una transformación de \mathbf{S} . Además dicha transformación es biyectiva, puesto que

$$\begin{aligned}\mathbf{X}'\mathbf{Y} &= \mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} \\ \mathbf{Y}'\mathbf{Y} &= (n-p)\widehat{\sigma}^2 + \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}}\end{aligned}$$

b) Puesto que $\widehat{\boldsymbol{\beta}}$ y $\widehat{\sigma}^2$ son estimadores insesgados de $\boldsymbol{\beta}$ y σ^2 , respectivamente, el Teorema 19 de Lehmann-Scheffé garantiza que

$$\begin{aligned}E[\widehat{\boldsymbol{\beta}}|\mathbf{T}] &= \widehat{\boldsymbol{\beta}} \\ E[\widehat{\sigma}^2|\mathbf{T}] &= \widehat{\sigma}^2\end{aligned}$$

son los estimadores insesgados de mínima varianza de $\boldsymbol{\beta}$ y σ^2 , respectivamente, lo cual concluye la demostración.

Corolario 9 Sea $\psi = \boldsymbol{\lambda}'\boldsymbol{\beta}$, $\boldsymbol{\lambda} \in \mathbb{R}^p$ un funcional lineal de $\boldsymbol{\beta}$ y sea $\widehat{\psi} = \boldsymbol{\lambda}'\widehat{\boldsymbol{\beta}}$ su estimador insesgado. Se verifica que:

- $\widehat{\psi} \sim N(\psi, \sigma^2 \boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\lambda})$.
- $\widehat{\psi}$ y $\widehat{\sigma}^2$ son variables aleatorias independientes.
- $\widehat{\psi}$ es el estimador insesgado de mínima varianza de ψ .

Demostración

Queda como ejercicio propuesto.

Ejercicio propuesto:

Demostrar el Corolario 8.

2.4. Modelo Lineal Normal: intervalos y regiones de confianza

Intervalo de confianza al nivel $1 - \alpha$ para σ^2

Proposición 10 Sea $0 < \alpha < 1$ y sean $\chi_{n-p, \alpha/2}^2$ y $\chi_{n-p, 1-\alpha/2}^2$ los cuantiles de orden $1 - \alpha/2$ y $\alpha/2$, respectivamente, de una distribución $\chi^2(n-p)$. Entonces

$$\left[\frac{(n-p)\widehat{\sigma}^2}{\chi_{n-p, \alpha/2}^2}, \frac{(n-p)\widehat{\sigma}^2}{\chi_{n-p, 1-\alpha/2}^2} \right]$$

es un intervalo de confianza al nivel $1 - \alpha$ para σ^2 .

Demostración

Utilizando el apartado b) de la Proposición 7, obtenemos que

$$P\left(\chi_{n-p, 1-\alpha/2}^2 \leq \frac{(n-p)\widehat{\sigma}^2}{\sigma^2} \leq \chi_{n-p, \alpha/2}^2\right) = 1 - \alpha$$

o equivalentemente

$$P\left(\frac{(n-p)\widehat{\sigma}^2}{\chi_{n-p, \alpha/2}^2} \leq \sigma^2 \leq \frac{(n-p)\widehat{\sigma}^2}{\chi_{n-p, 1-\alpha/2}^2}\right) = 1 - \alpha,$$

lo cual concluye la demostración.

Intervalos de confianza al nivel $1 - \alpha$ para funcionales lineales de β .

En primer lugar, proporcionaremos los intervalos de confianza para los funcionales de la forma $\psi = \beta_i$, $i = 1, \dots, p$, es decir, para las coordenadas de β . Posteriormente, dejaremos como ejercicio la deducción de los intervalos de confianza para el caso general de funcionales lineales de la forma $\psi = \lambda^t \beta$ con $\lambda \in \mathbb{R}^p$.

Proposición 11 Sea $0 < \alpha < 1$ y sea $t_{n-p, \alpha/2}$ el cuantil de orden $1 - \alpha/2$ de una distribución $t(n-p)$. Entonces

$$\left[\widehat{\beta}_i - \widehat{\sigma} \sqrt{c_{ii}} t_{n-p, \alpha/2}, \widehat{\beta}_i + \widehat{\sigma} \sqrt{c_{ii}} t_{n-p, \alpha/2} \right]$$

con c_{ii} , $i = 1, \dots, p$ el i -ésimo coeficiente de la diagonal de $(\mathbf{X}^t \mathbf{X})^{-1}$, es un intervalo de confianza al nivel $1 - \alpha$ para β_i , $i = 1, \dots, p$.

Demostración

Por el apartado a) de la Proposición 7, sabemos que $\widehat{\beta} \sim N_p(\beta, \sigma^2(\mathbf{X}^t \mathbf{X})^{-1})$ y por tanto, su marginal i -ésima también es normal, concretamente $\widehat{\beta}_i \sim N(\beta_i, \sigma^2 c_{ii})$. Tipificando obtenemos

$$\frac{\widehat{\beta}_i - \beta_i}{\sigma \sqrt{c_{ii}}} \sim N(0, 1)$$

Por otro lado, utilizando los apartados b) y c) de la Proposición 7, obtenemos que $(n-p)\widehat{\sigma}^2/\sigma^2 \sim \chi^2(n-p)$ y que dicha variable es independiente de $\widehat{\beta}$ y por tanto, también de $\widehat{\beta}_i$. De todo ello se deduce que

$$\frac{(\widehat{\beta}_i - \beta_i)/(\sigma \sqrt{c_{ii}})}{\sqrt{(n-p)\widehat{\sigma}^2/((n-p)\sigma^2)}} = \frac{\widehat{\beta}_i - \beta_i}{\widehat{\sigma} \sqrt{c_{ii}}} \sim t(n-p)$$

En consecuencia, si $t_{n-p,\alpha/2}$ es el cuantil de orden $1 - \alpha/2$ de la distribución $t(n - p)$, se verifica

$$P\left(-t_{n-p,\alpha/2} \leq \frac{\widehat{\beta}_i - \beta_i}{\widehat{\sigma} \sqrt{c_{ii}}} \leq t_{n-p,\alpha/2}\right) = 1 - \alpha.$$

Despejando β_i obtenemos que

$$P\left(\widehat{\beta}_i - \widehat{\sigma} \sqrt{c_{ii}} t_{n-p,\alpha/2} \leq \beta_i \leq \widehat{\beta}_i + \widehat{\sigma} \sqrt{c_{ii}} t_{n-p,\alpha/2}\right) = 1 - \alpha,$$

y la demostración está completa.

Proposición 12 Sea $\psi = \lambda' \beta$, $\lambda \in \mathbb{R}^p$ un funcional lineal de β . Si $0 < \alpha < 1$ y $t_{n-p,\alpha/2}$ es el cuantil de orden $1 - \alpha/2$ de una distribución $t(n - p)$, entonces

$$I = \left[\lambda' \widehat{\beta} - \widehat{\sigma} \sqrt{\lambda' (\mathbf{X}' \mathbf{X})^{-1} \lambda} t_{n-p,\alpha/2}, \lambda' \widehat{\beta} + \widehat{\sigma} \sqrt{\lambda' (\mathbf{X}' \mathbf{X})^{-1} \lambda} t_{n-p,\alpha/2} \right]$$

es un intervalo de confianza al nivel $1 - \alpha$ para ψ .

Demostración

Queda propuesta como ejercicio.

Región de confianza al nivel $1 - \alpha$ para β

Proposición 13 Sean $0 < \alpha < 1$ y $F_{p,n-p,\alpha}$ el cuantil de orden $1 - \alpha$ de una distribución $F(p, n - p)$. Definimos el siguiente subconjunto de \mathbb{R}^p :

$$\mathcal{R} = \{z \in \mathbb{R}^p : (z - \widehat{\beta})' \mathbf{X}' \mathbf{X} (z - \widehat{\beta}) \leq p \widehat{\sigma}^2 F_{p,n-p,\alpha}\}.$$

Se verifica que $P(\beta \in \mathcal{R}) = 1 - \alpha$ y en consecuencia, \mathcal{R} es una región de confianza al nivel $1 - \alpha$ para β .

Demostración

Por el apartado a) de la Proposición 7, sabemos que $\widehat{\beta} \sim N_p(\beta, \sigma^2 (\mathbf{X}' \mathbf{X})^{-1})$. Como $\mathbf{X}' \mathbf{X}$ es definida positiva, existe una matriz simétrica e invertible $(\mathbf{X}' \mathbf{X})^{1/2}$ tal que $\mathbf{X}' \mathbf{X} = (\mathbf{X}' \mathbf{X})^{1/2} (\mathbf{X}' \mathbf{X})^{1/2}$. Se verifica además que

$$\mathbf{Z} = \frac{1}{\sigma} (\mathbf{X}' \mathbf{X})^{1/2} (\widehat{\beta} - \beta) \sim N_p(\mathbf{0}, \mathbf{I}_n)$$

y por tanto

$$\mathbf{Z}' \mathbf{Z} = \frac{1}{\sigma^2} (\widehat{\beta} - \beta)' (\mathbf{X}' \mathbf{X}) (\widehat{\beta} - \beta) \sim \chi^2(p)$$

Por otro lado, utilizando los apartados b) y c) de la Proposición 7, obtenemos que $(n - p) \widehat{\sigma}^2 / \sigma^2 \sim \chi^2(n - p)$ y que dicha variable es independiente de $\widehat{\beta}$ y por tanto, también de $\mathbf{Z}' \mathbf{Z}$.

De todo ello se deduce que

$$\frac{(\widehat{\beta} - \beta)' (\mathbf{X}' \mathbf{X}) (\widehat{\beta} - \beta) / (p \sigma^2)}{(n - p) \widehat{\sigma}^2 / ((n - p) \sigma^2)} = \frac{1}{p \widehat{\sigma}^2} (\widehat{\beta} - \beta)' (\mathbf{X}' \mathbf{X}) (\widehat{\beta} - \beta) \sim F(p, n - p)$$

En consecuencia, si $F_{p,n-p,\alpha}$ es el cuantil de orden $1 - \alpha$ de la distribución $F(p, n - p)$, se verifica

$$P\left(\frac{1}{p \widehat{\sigma}^2} (\widehat{\beta} - \beta)' (\mathbf{X}' \mathbf{X}) (\widehat{\beta} - \beta) \leq F_{p,n-p,\alpha}\right) = P\left((\widehat{\beta} - \beta)' (\mathbf{X}' \mathbf{X}) (\widehat{\beta} - \beta) \leq p \widehat{\sigma}^2 F_{p,n-p,\alpha}\right) = 1 - \alpha$$

y de este modo concluye la demostración.

Observación: La región \mathcal{R} es un elipsoide centrado en $\widehat{\boldsymbol{\beta}}$. Si $\lambda_1 \geq \dots \geq \lambda_p > 0$ son los valores propios de $(\mathbf{X}'\mathbf{X})$ y $\mathbf{e}_1, \dots, \mathbf{e}_p$, los correspondientes vectores propios que forman una base ortonormal, entonces los ejes de \mathcal{R} están en las direcciones de $\mathbf{e}_1, \dots, \mathbf{e}_p$. Además la longitud del eje i -ésimo viene dada por

$$\sqrt{\frac{p \widehat{\sigma}^2 F_{p, n-p, \alpha}}{\lambda_i}} \quad i = 1, \dots, p.$$

Ejercicio propuesto:

Demostrar la Proposición 12.

2.5. Modelo Lineal Normal: Contraste de Hipótesis

En este apartado, bajo las hipótesis del Modelo Lineal Normal, presentaremos los contrastes de hipótesis más habituales en Modelos Lineales de Rango completo y proporcionaremos test de hipótesis que nos permitan decidir en dichos contrastes.

2.5.1. Contraste $H_0 : \beta = \beta^*$ (β^* vector de constantes conocidas).

En la hipótesis nula de este contraste se asume que el vector de parámetros β toma un valor conocido de \mathbb{R}^p . El siguiente resultado nos proporciona un test de hipótesis no aleatorio basado en la región de confianza calculada en la Proposición 13:

Proposición 14 Dado α , $0 < \alpha < 1$, un test de hipótesis de extensión α para el contraste $H_0 : \beta = \beta^*$ es

$$\Phi(\mathbf{Y}) = \begin{cases} 1 & \text{si } F > F_{p,n-p,\alpha} \\ 0 & \text{si } F \leq F_{p,n-p,\alpha} \end{cases}$$

siendo

$$F = \frac{(\widehat{\beta} - \beta^*)^t (\mathbf{X}'\mathbf{X})(\widehat{\beta} - \beta^*)}{p\widehat{\sigma}^2}$$

Es decir, el test Φ propone rechazar H_0 al nivel de significación α si $F > F_{p,n-p,\alpha}$.

Demostración

Consideramos la región de confianza \mathcal{R} calculada en la Proposición 13 y proponemos el siguiente test no aleatorio:

$$\Phi(\mathbf{Y}) = \begin{cases} 1 & \text{si } \beta^* \notin \mathcal{R} \\ 0 & \text{si } \beta^* \in \mathcal{R} \end{cases}$$

Es inmediato comprobar que $\beta^* \in \mathcal{R}$ si y sólo si $F \leq F_{p,n-p,\alpha}$. En cuanto a la extensión de Φ , puesto que es Φ no aleatorio y \mathcal{R} es una región de confianza al nivel $1 - \alpha$, tenemos que:

$$Ext \Phi = P_{\beta=\beta^*}(\Phi = 1) = P_{\beta=\beta^*}(\beta^* \notin \mathcal{R}) = \alpha$$

lo cual concluye la demostración.

2.5.2. Contraste $H_0 : \Psi = \Psi_0$, siendo $\Psi = \lambda'\beta$ con $\lambda \in \mathbb{R}^p$ conocido y $\Psi_0 \in \mathbb{R}$ valor conocido.

En este caso, la hipótesis nula es que una combinación lineal de las coordenadas de β toma un valor conocido. El siguiente resultado nos proporciona un test de hipótesis no aleatorio basado en el intervalo de confianza calculado en la Proposición 12:

Proposición 15 Dado α , $0 < \alpha < 1$, un test de hipótesis de extensión α para el contraste $H_0 : \Psi = \Psi_0$ es

$$\Phi(\mathbf{Y}) = \begin{cases} 1 & \text{si } |T| > t_{n-p,\alpha/2} \\ 0 & \text{si } |T| \leq t_{n-p,\alpha/2} \end{cases}$$

siendo

$$T = \frac{\widehat{\Psi} - \Psi_0}{\widehat{\sigma} \sqrt{\lambda'(\mathbf{X}'\mathbf{X})^{-1}\lambda}}$$

Es decir, el test Φ propone rechazar H_0 al nivel de significación α si

$$\frac{|\widehat{\Psi} - \Psi_0|}{\widehat{\sigma} \sqrt{\lambda'(\mathbf{X}'\mathbf{X})^{-1}\lambda}} \geq t_{n-p, \alpha/2}.$$

Demostración

Es análoga a la de la Proposición 14, considerando en este caso el intervalo de confianza I calculado en la Proposición 12.

Un caso particular de interés es el contraste $H_0 : \beta_i = 0$, siendo i un índice fijo. En este caso, el test calculado en la Proposición 15 propone rechazar H_0 al nivel de significación α si

$$\frac{|\widehat{\beta}_i|}{\widehat{\sigma} \sqrt{c_{ii}}} \geq t_{n-p, \alpha/2},$$

siendo c_{ii} el i -ésimo coeficiente en la diagonal de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$.

2.5.3 Contraste $H_0 : \beta_1 = \dots = \beta_k = 0, k < p$.

Proporcionaremos un test de hipótesis, por el procedimiento de la razón de verosimilitudes.

Estadístico de la razón de verosimilitudes: Este estadístico, al que denotaremos Λ , se define del siguiente modo:

$$\Lambda = \frac{\sup_{\beta_1 = \dots = \beta_k = 0} L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y})}{\sup_{\beta \in \mathbb{R}^p, \sigma^2 > 0} L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y})} \quad (8)$$

Según vimos en la Proposición 6, el supremo del numerador se calcula como sigue:

$$\sup_{\beta \in \mathbb{R}^p, \sigma^2 > 0} L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y}) = L(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2; \mathbf{Y}) = \left(\frac{e^{-1}}{2\pi \widehat{\sigma}^2} \right)^{n/2}.$$

Para calcular el supremo del denominador, dividamos la matriz $\mathbf{X} = (\mathbf{X}_1 | \mathbf{X}_2)$, donde \mathbf{X}_1 y \mathbf{X}_2 contienen, respectivamente, a las k primeras y $p-k$ últimas columnas de \mathbf{X} y denotemos $\boldsymbol{\gamma}_1 = (\beta_1, \dots, \beta_k)'$ y $\boldsymbol{\gamma}_2 = (\beta_{k+1}, \dots, \beta_p)'$. El modelo lineal bajo $H_0 : \beta_1 = \dots = \beta_k = 0$ se expresa del siguiente modo:

$$E[\mathbf{Y}] = \mathbf{X}_2 \boldsymbol{\gamma}_2 \quad (9)$$

Nos referiremos al modelo (9) como Modelo Reducido por H_0 . No es difícil comprobar que $r(\mathbf{X}_2) = p-k$ y por tanto el Modelo Reducido (9) es un Modelo Lineal de Rango Completo. Aplicando de nuevo la Proposición 6 obtenemos que los estimadores de máxima verosimilitud de $\boldsymbol{\gamma}_2$ y σ^2 en este modelo son, respectivamente,

$$\widehat{\boldsymbol{\gamma}}_2 = (\mathbf{X}_2' \mathbf{X}_2)^{-1} \mathbf{X}_2' \mathbf{Y}, \quad \widehat{\sigma}^2 = \frac{1}{n} (\mathbf{Y} - \mathbf{X}_2 \widehat{\boldsymbol{\gamma}}_2)' (\mathbf{Y} - \mathbf{X}_2 \widehat{\boldsymbol{\gamma}}_2)$$

y la verosimilitud máxima en el modelo reducido es

$$\left(\frac{e^{-1}}{2\pi \widehat{\sigma}^2} \right)^{n/2}$$

El estadístico de la razón de verosimilitudes Λ definido en (8) queda como sigue:

$$\Lambda = \frac{\left(\frac{e^{-1}}{2\pi \tilde{\sigma}^2}\right)^{n/2}}{\left(\frac{e^{-1}}{2\pi \tilde{\sigma}^2}\right)^{n/2}} = \left(\frac{\tilde{\sigma}^2}{\tilde{\sigma}^2}\right)^{n/2} = \left(\frac{SCE}{SCE_R}\right)^{n/2}$$

siendo SCE la suma de cuadrados de los residuos en el modelo original y SCE_R la suma de cuadrados de los residuos en el modelo reducido por H_0 . Utilizando la matriz \mathbf{H} definida en (4) y que también se puede calcular para el modelo reducido, tenemos que

$$\Lambda^{2/n} = \frac{\mathbf{Y}^t(\mathbf{I}_n - \mathbf{H})\mathbf{Y}}{\mathbf{Y}^t(\mathbf{I}_n - \mathbf{H}_2)\mathbf{Y}}$$

siendo $\mathbf{H} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$ y $\mathbf{H}_2 = \mathbf{X}_2(\mathbf{X}_2^t\mathbf{X}_2)^{-1}\mathbf{X}_2^t$. El test de razón de verosimilitudes se expresa como sigue:

$$\Phi(\mathbf{Y}) = \begin{cases} 1 & \text{si } \Lambda^{2/n} < C_\alpha \\ 0 & \text{si } \Lambda^{2/n} \geq C_\alpha \end{cases} \quad (10)$$

siendo C_α una constante que hemos de determinar para que el test tenga extensión α ($0 < \alpha < 1$)

Distribución del estadístico: Para determinar completamente el test (10), nos resta calcular la constante C_α de modo que dicho test tenga extensión α . Para ello necesitamos conocer la distribución de probabilidad de $\Lambda^{2/n}$.

Aunque las formas cuadráticas del numerador y denominador tienen asociadas distribuciones chi cuadrado, no son variables aleatorias independientes, y por tanto no es posible obtener la distribución asociada a dicho estadístico. Para poder determinar C_α , hemos de expresar el test (10) en función de otro estadístico cuya distribución sea conocida. Para ello reescribimos el estadístico de la razón de verosimilitudes del siguiente modo:

$$\Lambda^{2/n} = \frac{\mathbf{Y}^t(\mathbf{I}_n - \mathbf{H})\mathbf{Y}}{\mathbf{Y}^t(\mathbf{I}_n - \mathbf{H}_2)\mathbf{Y}} = \frac{\mathbf{Y}^t(\mathbf{I}_n - \mathbf{H})\mathbf{Y}}{\mathbf{Y}^t(\mathbf{I}_n - \mathbf{H})\mathbf{Y} + \mathbf{Y}^t(\mathbf{H} - \mathbf{H}_2)\mathbf{Y}} = \frac{Q_0}{Q_0 + Q_1}, \quad (11)$$

donde

$$Q_0 = \mathbf{Y}^t(\mathbf{I}_n - \mathbf{H})\mathbf{Y} = SCE \quad \text{y} \quad Q_1 = \mathbf{Y}^t(\mathbf{H} - \mathbf{H}_2)\mathbf{Y}.$$

A partir de Q_0 y de Q_1 , el siguiente resultado nos proporciona un estadístico cuya distribución es conocida:

Proposición 16 *Con las notaciones anteriores, se verifica que:*

a) Se verifica que $Q_1/\sigma^2 \sim \chi^2(k, \frac{1}{2\sigma^2}\boldsymbol{\gamma}_1^t\mathbf{X}_1^t(\mathbf{I}_n - \mathbf{H}_2)\mathbf{X}_1\boldsymbol{\gamma}_1)$.

b) Q_0 y Q_1 son independientes.

c)

$$F = \frac{n-p}{k} \frac{Q_1}{Q_0} \sim F(k, n-p, \frac{1}{2\sigma^2}\boldsymbol{\gamma}_1^t\mathbf{X}_1^t(\mathbf{I}_n - \mathbf{H}_2)\mathbf{X}_1\boldsymbol{\gamma}_1).$$

Demostración

a) Aplicamos de nuevo el Teorema 1 sobre formas cuadráticas, visto en el Tema 1. Para deducir que la distribución de Q_1/σ^2 es $\chi^2(r(\mathbf{A}), \boldsymbol{\mu}^t\mathbf{A}\boldsymbol{\mu}/2)$, bastará probar que $\mathbf{A}\mathbf{V}$ es idempotente, siendo

$$\mathbf{A} = \frac{1}{\sigma^2}(\mathbf{H} - \mathbf{H}_2), \quad \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}, \quad \mathbf{V} = \sigma^2\mathbf{I}_n$$

Como $\mathbf{AV} = \mathbf{H} - \mathbf{H}_2$, bastará probar que dicha matriz es idempotente o, equivalentemente, que $\mathbf{I}_n - \mathbf{H} + \mathbf{H}_2$ es idempotente. Ahora bien, teniendo en cuenta el Lema 3, $\mathbf{I}_n - \mathbf{H}$ y \mathbf{H}_2 son idempotentes y $(\mathbf{I}_n - \mathbf{H})\mathbf{H}_2 = (\mathbf{I}_n - \mathbf{H})\mathbf{X}_2(\mathbf{X}_2^t\mathbf{X}_2)^{-1}\mathbf{X}_2^t = \mathbf{0}$, ya que \mathbf{X}_2 son las últimas $p - k$ columnas de \mathbf{X} . Por tanto, se obtiene que

$$(\mathbf{I}_n - \mathbf{H} + \mathbf{H}_2)^2 = (\mathbf{I}_n - \mathbf{H})^2 + \mathbf{H}_2^2 + (\mathbf{I}_n - \mathbf{H})\mathbf{H}_2 + \mathbf{H}_2(\mathbf{I}_n - \mathbf{H}) = \mathbf{I}_n - \mathbf{H} + \mathbf{H}_2$$

es decir, $\mathbf{I}_n - \mathbf{H} + \mathbf{H}_2$ y $\mathbf{H} - \mathbf{H}_2$ son idempotentes y en consecuencia Q_1 tiene distribución chi cuadrado no central.

Vamos determinar los parámetros de esta distribución:

$$r(\mathbf{A}) = r(\mathbf{H} - \mathbf{H}_2) = \text{tr}(\mathbf{H} - \mathbf{H}_2) = \text{tr}(\mathbf{H}) - \text{tr}(\mathbf{H}_2) = p - (p - k) = k$$

Por otro lado, utilizando de nuevo que $(\mathbf{I} - \mathbf{H})\mathbf{X} = \mathbf{0}$, se obtiene que

$$\frac{1}{2}\boldsymbol{\mu}^t\mathbf{A}\boldsymbol{\mu} = \frac{1}{2\sigma^2}\boldsymbol{\beta}^t\mathbf{X}^t(\mathbf{H} - \mathbf{H}_2)\mathbf{X}\boldsymbol{\beta} = -\frac{1}{2\sigma^2}\boldsymbol{\beta}^t\mathbf{X}^t(\mathbf{I}_n - \mathbf{H})\mathbf{X}\boldsymbol{\beta} + \frac{1}{2\sigma^2}\boldsymbol{\beta}^t\mathbf{X}^t(\mathbf{I}_n - \mathbf{H}_2)\mathbf{X}\boldsymbol{\beta} = \frac{1}{2\sigma^2}\boldsymbol{\beta}^t\mathbf{X}^t(\mathbf{I}_n - \mathbf{H}_2)\mathbf{X}\boldsymbol{\beta}$$

y teniendo en cuenta que $\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_1\boldsymbol{\gamma}_1 + \mathbf{X}_2\boldsymbol{\gamma}_2$ y, por el Lema 3, $(\mathbf{I}_n - \mathbf{H}_2)\mathbf{X}_2 = \mathbf{0}$, se obtiene que el parámetro de no centralidad es

$$\frac{1}{2}\boldsymbol{\mu}^t\mathbf{A}\boldsymbol{\mu} = \frac{1}{2\sigma^2}\boldsymbol{\gamma}_1^t\mathbf{X}_1^t(\mathbf{I}_n - \mathbf{H}_2)\mathbf{X}_1\boldsymbol{\gamma}_1 \quad (12)$$

b) Aplicamos el Teorema 3 sobre independencia de formas cuadráticas, visto en el Tema 1. Para deducir la independencia de Q_0 y Q_1 , bastará probar que $\mathbf{BVA} = \mathbf{0}$, siendo

$$\mathbf{A} = \mathbf{H} - \mathbf{H}_2, \quad \mathbf{V} = \sigma^2\mathbf{I}_n, \quad \mathbf{B} = \mathbf{I}_n - \mathbf{H}$$

En efecto, teniendo en cuenta que $(\mathbf{I}_n - \mathbf{H})\mathbf{H} = \mathbf{0}$ y $(\mathbf{I}_n - \mathbf{H})\mathbf{H}_2 = \mathbf{0}$ (Lema 3), se obtiene que:

$$\mathbf{BVA} = \sigma^2(\mathbf{I}_n - \mathbf{H})(\mathbf{H} - \mathbf{H}_2) = \sigma^2(\mathbf{I}_n - \mathbf{H})\mathbf{H} - \sigma^2(\mathbf{I}_n - \mathbf{H})\mathbf{H}_2 = \mathbf{0}$$

como queríamos probar.

c) Es consecuencia inmediata de a) y b) y de la definición de la distribución F de Snedecor no central, que aparece en el Tema 1.

Extensión del test: Utilizando la Proposición 16 podemos calcular el valor crítico del test Φ definido en (10). En efecto, partiendo de (11) podemos escribir:

$$\Lambda^{2/n} = \frac{Q_0}{Q_0 + Q_1} = \frac{1}{1 + Q_1/Q_0} = \frac{1}{1 + \frac{k}{n-p}F}$$

y teniendo en cuenta que la función $f(x) = 1/(1 + \frac{k}{n-p}x)$ es estrictamente decreciente en \mathbb{R}^+ , tenemos que existe una constante C'_α tal que

$$\Lambda < C_\alpha \quad \text{si y sólo si} \quad F > C'_\alpha$$

Por tanto, el test de razón de verosimilitudes (10) se puede reescribir en términos del estadístico F del siguiente modo:

$$\Phi(\mathbf{Y}) = \begin{cases} 1 & \text{si } F > C'_\alpha \\ 0 & \text{si } F \leq C'_\alpha \end{cases}$$

Resta calcular el valor crítico C'_α para que el test Φ tenga extensión α . Para ello, asumiendo H_0 cierta, se verifica que $\gamma_1 = (\beta_1, \dots, \beta_k)' = \mathbf{0}$ y teniendo en cuenta (12), el parámetro de no centralidad de la distribución de F también es igual a 0. En consecuencia,

$$\alpha = \text{Ext}\Phi = \sup_{\gamma_1=\mathbf{0}} P(F > C'_\alpha) = P(F(k, n-p) > C'_\alpha)$$

Por tanto, para que el test Φ tenga extensión α , $C'_\alpha = F_{k, n-p, \alpha}$, siendo $F_{k, n-p, \alpha}$ el cuantil de orden $1 - \alpha$ de la distribución $F(k, n-p)$.

En resumen, el test de razón de verosimilitudes propone rechazar H_0 al nivel de significación α si $F \geq F_{k, n-p, \alpha}$.

2.5.4 Contraste $H_0 : \psi_1 = \dots = \psi_k = 0, k < p$, y ψ_i funcional lineal de $\beta, i = 1, \dots, k$.

Este es el contraste más general que abordaremos en esta sección. Sin pérdida de generalidad, podemos suponer que ψ_1, \dots, ψ_k son linealmente independientes. Un caso particular de interés es $H_0 : \beta_1 = \dots = \beta_k$, es decir, que k coordenadas de β son iguales, lo que es equivalente a $H_0 : \beta_1 - \beta_k = \dots = \beta_{k-1} - \beta_k = 0$, que tiene la forma indicada con $\psi_i = \beta_i - \beta_k, i = 1, \dots, k-1$.

Si consideramos $\psi_i = \lambda_i' \beta$, con $\lambda_i \in \mathbb{R}^p, i = 1, \dots, k$, linealmente independientes, podemos reformular el contraste $H_0 : \psi_1 = \dots = \psi_k = 0$ como $H_0 : \mathbf{U}_1 \beta = \mathbf{0}$, donde

$$\mathbf{U}_1 = \begin{pmatrix} \lambda_1' \\ \vdots \\ \lambda_k' \end{pmatrix}$$

es una matriz $k \times p$ y $r(\mathbf{U}_1) = k$.

El test que obtendremos, es una generalización del test de hipótesis calculado en el apartado 2.5.3 Para ello, reformularemos el modelo en términos de nuevos parámetros. Sean $\lambda_{k+1}, \dots, \lambda_p \in \mathbb{R}^p$ linealmente independientes de $\lambda_1, \dots, \lambda_k \in \mathbb{R}^p$. Denotaremos

$$\mathbf{U}_2 = \begin{pmatrix} \lambda_{k+1}' \\ \vdots \\ \lambda_p' \end{pmatrix} \quad \text{y} \quad \mathbf{U} = \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{pmatrix}$$

La matriz \mathbf{U} es cuadrada de orden p y $r(\mathbf{U}) = p$ por lo que es invertible. Además, denotando también, para $i = k+1, \dots, p, \psi_i = \lambda_i \beta$ y denotando $\Psi = (\psi_1, \dots, \psi_k, \psi_{k+1}, \dots, \psi_p)'$, se verifica que $\Psi = \mathbf{U}\beta$. Si $\mathbf{V} = \mathbf{U}^{-1}$, el modelo lineal se puede reescribir como

$$E[\mathbf{Y}] = \mathbf{X}\beta = \mathbf{X}\mathbf{V}\mathbf{U}\beta = \mathbf{Z}\Psi$$

siendo $\mathbf{Z} = \mathbf{X}\mathbf{V}$ una matriz $n \times p$ y $r(\mathbf{Z}) = r(\mathbf{X}) = p$. Se trata de un modelo lineal normal de rango completo. El contraste de $H_0 : \psi_1 = \dots = \psi_k = 0$ en este modelo es del mismo tipo que el contraste estudiado en el apartado 2.5.3, por lo tanto, el test que desarrollamos en este apartado se puede utilizar aquí.

Más concretamente, si tomamos \mathbf{Z}_2 , la matriz formada por las últimas $p-k$ columnas de \mathbf{Z} , entonces $\mathbf{Z}_2 = \mathbf{X}\mathbf{V}_2$ es la matriz que agrupa a las $p-k$ últimas columnas de \mathbf{Z} y, de acuerdo a lo visto en el apartado 2.5.3, podemos definir el estadístico

$$F = \frac{n-p}{k} \frac{\mathbf{Y}'(\mathbf{H} - \mathbf{H}_2)\mathbf{Y}}{\mathbf{Y}'(\mathbf{I}_n - \mathbf{H})\mathbf{Y}}$$

siendo $\mathbf{H} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ y $\mathbf{H}_2 = \mathbf{Z}_2(\mathbf{Z}_2'\mathbf{Z}_2)^{-1}\mathbf{Z}_2'$. Se rechaza H_0 al nivel de significación α si $F \geq F_{k, n-p, \alpha}$.

Queda por probar que el estadístico F no depende de los vectores $\lambda_{k+1}, \dots, \lambda_p \in \mathbb{R}^p$ que hemos elegido para completar la base.

Respecto al denominador, es sencillo verificar (queda propuesto como ejercicio) que

$$\mathbf{H} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad (13)$$

Por tanto, la forma cuadrática del denominador, es la suma de cuadrados de los residuos, la misma que que apareció en el apartado 2.5.3 y que fue denotada por Q_0 .

Respecto al numerador, si tomamos \mathbf{V}_2 , la matriz formada por las últimas $p - k$ columnas de \mathbf{V} , entonces $\mathbf{Z}_2 = \mathbf{X}\mathbf{V}_2$. Si suponemos que hemos elegido otro conjunto de vectores distintos $\delta_{k+1}, \dots, \delta_p \in \mathbb{R}^p$ y denotamos

$$\mathbf{U}'_2 = \begin{pmatrix} \delta_{k+1}' \\ \vdots \\ \delta_p' \end{pmatrix} \quad y \quad \mathbf{U}' = \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{U}'_2 \end{pmatrix}$$

Puesto que $\lambda_1, \dots, \lambda_p$ forman una base de \mathbb{R}^p , se verifica que

$$\delta_i = a_{i1}\lambda_1 + \dots + a_{ik}\lambda_k + a_{i,k+1}\lambda_{k+1} + \dots + a_{ip}\lambda_p, \quad i = k+1, \dots, p$$

o equivalentemente

$$\mathbf{U}' = \left(\begin{array}{c|c} \mathbf{I}_k & \mathbf{0} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{array} \right) \mathbf{U}$$

siendo

$$\mathbf{A}_{21} = \begin{pmatrix} a_{k+11} & \dots & a_{k+1k} \\ \vdots & \ddots & \vdots \\ a_{p1} & \dots & a_{pk} \end{pmatrix}, \quad \mathbf{A}_{22} = \begin{pmatrix} a_{k+1,k+1} & \dots & a_{k+1p} \\ \vdots & \ddots & \vdots \\ a_{p,k+1} & \dots & a_{pp} \end{pmatrix}$$

Por lo tanto, si $\mathbf{W} = (\mathbf{U}')^{-1}$, la matriz del modelo con esta otra transformación sería $\mathbf{Z}' = \mathbf{X}\mathbf{W}$. La relación entre \mathbf{V} y \mathbf{W} viene dada por

$$\mathbf{W} = \mathbf{V} \left(\begin{array}{c|c} \mathbf{I}_k & \mathbf{0} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{22}^{-1} \end{array} \right)$$

Si denotamos por \mathbf{W}_2 a la matriz formada por las $p - k$ últimas columnas de \mathbf{W} , se obtiene que $\mathbf{W}_2 = \mathbf{V}_2\mathbf{A}_{22}^{-1}$. De este modo, si \mathbf{Z}'_2 es la matriz que agrupa las $p - k$ últimas columnas de \mathbf{Z}' , se verifica que

$$\mathbf{Z}'_2 = \mathbf{X}\mathbf{W}_2 = \mathbf{X}\mathbf{V}_2\mathbf{A}_{22}^{-1} = \mathbf{Z}_2\mathbf{A}_{22}^{-1}.$$

Entonces, procediendo igual que en (13) se verifica que

$$\mathbf{Z}'_2((\mathbf{Z}'_2)'_2\mathbf{Z}'_2)^{-1}(\mathbf{Z}'_2)'_2 = \mathbf{Z}_2(\mathbf{Z}_2'\mathbf{Z}_2)^{-1}\mathbf{Z}_2' = \mathbf{H}_2,$$

lo cual demuestra la invarianza de \mathbf{H}_2 ante “cambios de base”.

Apéndice: Resultados de Inferencia Estadística

Teorema 17 (de factorización de Neyman-Halmos-Savage) Sea $\mathbf{Y} = (Y_1, \dots, Y_n)^t$ una muestra aleatoria con función de densidad conjunta $f_{\boldsymbol{\theta}}(\mathbf{y})$, $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^s$ y $\mathbf{T} : \mathbb{R}^n \rightarrow \Theta$ un estadístico.

Entonces \mathbf{T} es suficiente para el parámetro $\boldsymbol{\theta} \in \Theta$ si y sólo si existen $g_{\boldsymbol{\theta}} : \mathbb{R}^s \rightarrow \mathbb{R}_+$ y $h : \mathbb{R}^n \rightarrow \mathbb{R}_+$ tales que

$$f_{\boldsymbol{\theta}}(\mathbf{Y}) = g_{\boldsymbol{\theta}}(\mathbf{T}(\mathbf{Y}))h(\mathbf{Y}).$$

Teorema 18 (de completitud en familias exponenciales) Sea $\mathbf{Y} = (Y_1, \dots, Y_n)^t$ una muestra aleatoria con función de densidad conjunta $f_{\boldsymbol{\theta}}(\mathbf{y})$, $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^s$ y $\mathbf{T} : \mathbb{R}^n \rightarrow \Theta$ un estadístico. Supongamos que $f_{\boldsymbol{\theta}}(\mathbf{y})$ pertenece a la familia exponencial de distribuciones y que Θ tiene interior no vacío en \mathbb{R}^s .

Entonces $\mathbf{T} = (T_1, \dots, T_s)^t$ es completo si y sólo si existen funciones $C : \Theta \rightarrow \mathbb{R}_+$, $Q_1, \dots, Q_s : \Theta \rightarrow \mathbb{R}$ y $h : \mathbb{R}^n \rightarrow \mathbb{R}_+$ tales que

$$f_{\boldsymbol{\theta}}(\mathbf{Y}) = C(\boldsymbol{\theta}) \exp \left\{ \sum_{j=1}^s Q_j(\boldsymbol{\theta}) T_j(\mathbf{Y}) \right\} h(\mathbf{Y}).$$

Teorema 19 (de Lehmann-Scheffé) Sea \mathbf{T} un estadístico suficiente y completo y sea $\widehat{\boldsymbol{\theta}}$ un estimador insesgado del parámetro $\boldsymbol{\theta}$. Entonces $E[\widehat{\boldsymbol{\theta}} | \mathbf{T}]$ es el estimador insesgado de mínima varianza de $\boldsymbol{\theta}$.

Tema 3: Modelo Lineal de Rango No Completo

3.1 Introducción

En este tema consideraremos un vector aleatorio n dimensional $\mathbf{Y} = (Y_1, \dots, Y_n)^t$ que se ajusta un modelo lineal, es decir, que verifica

$$E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$$

siendo \mathbf{X} una matriz de orden $n \times p$ ($p < n$) de constantes conocidas y $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^t$ es un vector de parámetros desconocidos. Equivalentemente, \mathbf{Y} se puede escribir

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mathcal{E}},$$

siendo $\boldsymbol{\mathcal{E}}$ un vector aleatorio n -dimensional con $E[\boldsymbol{\mathcal{E}}] = \mathbf{0}$.

La novedad que presenta este modelo respecto al estudiado en el Tema 2 es que supondremos que se trata de un modelo de rango no completo, es decir, $r(\mathbf{X}) = k < p$.

En función de las hipótesis que asumamos sobre la distribución de \mathbf{Y} (o equivalentemente sobre la distribución del vector de errores $\boldsymbol{\mathcal{E}}$), podemos distinguir:

- Modelo Lineal Básico: asumiremos $\text{Cov}[\mathbf{Y}] = \text{Cov}[\boldsymbol{\mathcal{E}}] = \sigma^2 \mathbf{I}_n$.
- Modelo Lineal Normal: asumiremos $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$ (equivalentemente $\boldsymbol{\mathcal{E}} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$).

Nuestro objetivo en ambas situaciones será realizar inferencias sobre el vector de parámetros, $\boldsymbol{\beta}$, y sobre combinaciones lineales de dichos parámetros, $\boldsymbol{\psi} = \boldsymbol{\lambda}'\boldsymbol{\beta}$, con $\boldsymbol{\lambda} \in \mathbb{R}^p$. También estamos interesados en la varianza de las observaciones y de los errores, σ^2 . Sin embargo, veremos que la estimación del parámetro $\boldsymbol{\beta}$ no va a ser factible utilizando procedimientos similares a los del Tema 2.

Por ello, la forma de proceder que seguiremos será la de realizar un cambio de parámetros en el modelo (**reparametrización**) de modo que obtengamos un Modelo Lineal de Rango Completo al que sí podamos aplicar la teoría estudiada en el Tema 2. Esto lo haremos en el punto 3.2, y más concretamente en el apartado 3.2.1 que es el más importante de este tema.

3.2. Estimación en el Modelo Lineal de Rango No Completo.

3.2.1 Estimación puntual en el Modelo Lineal Básico.

En primer lugar, al igual que en el Tema 2, intentaremos obtener estimadores de $\boldsymbol{\beta}$ (ya veremos que no obtendremos buenos resultados). Buscamos el valor de $\boldsymbol{\beta}$ que minimice la expresión:

$$f(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).$$

Aplicando el método de mínimos cuadrados al igual que hicimos en el Modelo Lineal de Rango Completo (derivando parcialmente respecto a β_1, \dots, β_p e igualando a 0 dichas derivadas parciales), obtenemos que los estimadores $\widehat{\boldsymbol{\beta}}$ de mínimos cuadrados para $\boldsymbol{\beta}$ en el Modelo Lineal de Rango No Completo deberían ser las soluciones del sistema de ecuaciones normales

$$\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}$$

Ahora bien

$$k = r(\mathbf{X}) = r(\mathbf{X}'\mathbf{X}) \leq r(\mathbf{X}'\mathbf{X}|\mathbf{X}'\mathbf{Y}) = r(\mathbf{X}'(\mathbf{X}|\mathbf{Y})) \leq r(\mathbf{X}') = k$$

Por tanto, el rango de la matriz de los coeficientes, $\mathbf{X}'\mathbf{X}$, es igual al rango de la matriz ampliada $(\mathbf{X}'\mathbf{X}|\mathbf{X}'\mathbf{Y})$ por lo que, aplicando el Teorema de Rouché-Fröbenius, el sistema es compatible. Pero dicho rango es igual a $k < p$ que es el número de incógnitas, por lo que dicho sistema es compatible indeterminado y por tanto tiene infinitas soluciones.

Esto no significa que β tenga infinitos estimadores. De hecho β no tiene estimadores lineales insesgados. En efecto, si \mathbf{AY} , con \mathbf{A} matriz $p \times n$, fuera un estimador lineal insesgado de β tendríamos que

$$\beta = E[\mathbf{AY}] = \mathbf{AE}[\mathbf{Y}] = \mathbf{AX}\beta, \quad \beta \in \mathbb{R}^p$$

Por tanto, como la igualdad es cierta para todo $\beta \in \mathbb{R}^p$, debe ocurrir que

$$\mathbf{I}_p = \mathbf{AX}$$

lo cual es imposible porque

$$r(\mathbf{I}_p) = p \quad \text{y} \quad r(\mathbf{AX}) \leq r(\mathbf{X}) = k < p.$$

Por tanto en un Modelo Lineal de Rango no Completo, no todas las transformaciones lineales de β tienen estimadores lineales insesgados. El propio parámetro β no los tiene. Nuestro interés se centrará en funciones lineales de β para las que existan estimadores lineales insesgados, a las que denominaremos “funciones lineales estimables”, que definiremos a continuación.

Definición 1 (funciones lineales estimables). Diremos que $\psi = \lambda'\beta$, es una función lineal estimable (f.l.e.) si tiene algún estimador lineal insesgado, es decir, si existe un vector $\mathbf{c} \in \mathbb{R}^n$, tal que $T = \mathbf{c}'\mathbf{Y}$ es insesgado:

$$E[T] = E[\mathbf{c}'\mathbf{Y}] = \lambda'\beta \quad \text{para todo } \beta \in \mathbb{R}^p.$$

De la definición anterior se deduce que $\psi = \lambda'\beta$ f.l.e. si y solo si

$$\psi = \lambda'\beta = \mathbf{c}'E[\mathbf{Y}] = \mathbf{c}'\mathbf{X}\beta \quad \text{para todo } \beta \in \mathbb{R}^p$$

por tanto, si $\mathbf{c} = (c_1, \dots, c_n)'$ y $\mathbf{X}_1, \dots, \mathbf{X}_n$ son las filas de \mathbf{X} lo anterior es equivalente a que

$$\lambda = \mathbf{X}'\mathbf{c} = c_1\mathbf{X}_1 + \dots + c_n\mathbf{X}_n$$

Hemos probado que

Proposición 1. $\psi = \lambda'\beta$ es f.l.e. si y sólo si es λ es combinación lineal de las filas de \mathbf{X} . En particular, $\mathbf{X}_1'\beta, \dots, \mathbf{X}_n'\beta$ son f.l.e. Como además, $r(\mathbf{X}) = k$, el número máximo de f.l.e. linealmente independientes es k .

Ejercicio propuesto. Probar que la combinación lineal de f.l.e. es una f.l.e.

Caracterización de las f.l.e. y cálculo de estimadores

Veamos otra forma de comprobar que $\psi = \lambda^t \beta$ es f.l.e.

Proposición 2. $\psi = \lambda^t \beta$ es f.l.e. si y sólo el sistema de ecuaciones $\mathbf{X}^t \mathbf{X} \mathbf{r} = \lambda$ es compatible (el vector $\mathbf{r} \in \mathbb{R}^p$ contiene a las incógnitas de dicho sistema).

Demostración. Si $\psi = \lambda^t \beta$ es f.l.e., hemos visto anteriormente que $\lambda = \mathbf{X}^t \mathbf{c}$ para algún $\mathbf{c} \in \mathbb{R}^n$. Apliquemos de nuevo el Teorema de Rouché-Fröbenius para ver que el sistema $\mathbf{X}^t \mathbf{X} \mathbf{r} = \lambda$ es compatible indeterminado:

$$k = r(\mathbf{X}^t \mathbf{X}) \leq r(\mathbf{X}^t \mathbf{X} | \lambda) = r(\mathbf{X}^t \mathbf{X} | \mathbf{X}^t \mathbf{c}) = r(\mathbf{X}^t (\mathbf{X} | \mathbf{c})) \leq r(\mathbf{X}^t) = k$$

Es decir, el rango de la matriz de los coeficientes es igual al rango de la matriz ampliada, k , que es menor que el número de incógnitas, p , y por tanto el sistema $\mathbf{X}^t \mathbf{X} \mathbf{r} = \lambda$ es compatible indeterminado.

Recíprocamente, si suponemos que dicho sistema es compatible, podemos tomar una solución particular \mathbf{r}_0 que verifica (por ser solución) que $\mathbf{X}^t \mathbf{X} \mathbf{r}_0 = \lambda$, o trasponiendo, $\lambda^t = \mathbf{r}_0^t \mathbf{X}^t \mathbf{X}$.

Si tomamos $\mathbf{c} = \mathbf{X} \mathbf{r}_0$, demostraremos a continuación que $\mathbf{c}^t \mathbf{Y}$ es un estimador insesgado de ψ y, aplicando la Definición 1, obtendremos que ψ es una f.l.e.

En efecto, como $\mathbf{c}^t = \mathbf{r}_0^t \mathbf{X}^t$,

$$E[\mathbf{c}^t \mathbf{Y}] = \mathbf{r}_0^t \mathbf{X}^t E[\mathbf{Y}] = \mathbf{r}_0^t \mathbf{X}^t \mathbf{X} \beta = \lambda^t \beta = \psi$$

□

Observación: En el resultado anterior, se prueba además que si \mathbf{r}_0 es una solución de $\mathbf{X}^t \mathbf{X} \mathbf{r} = \lambda$, entonces $\widehat{\psi} = \mathbf{r}_0^t \mathbf{X}^t \mathbf{Y}$ es un estimador lineal insesgado de ψ . El problema que plantea dicho estimador es que, en principio, depende de la solución \mathbf{r}_0 del sistema $\mathbf{X}^t \mathbf{X} \mathbf{r} = \lambda$ que tomemos. El siguiente resultado da solución a dicho problema.

Proposición 3. Si $\widehat{\beta}_0$ es una solución de las ecuaciones normales, $\mathbf{X}^t \mathbf{X} \widehat{\beta}_0 = \mathbf{X}^t \mathbf{Y}$, se verifica que $\widehat{\psi} = \lambda^t \widehat{\beta}_0$. Por tanto el valor de $\widehat{\psi}$ no depende de la solución \mathbf{r}_0 elegida para el sistema $\mathbf{X}^t \mathbf{X} \mathbf{r} = \lambda$.

Demostración. Supongamos que β_0 es solución de las ecuaciones normales, es decir, verifica que

$$\mathbf{X}^t \mathbf{X} \widehat{\beta}_0 = \mathbf{X}^t \mathbf{Y}$$

Entonces, como $\mathbf{X}^t \mathbf{X} \mathbf{r}_0 = \lambda$ (o trasponiendo $\mathbf{r}_0^t \mathbf{X}^t \mathbf{X} = \lambda^t$) se tiene que

$$\widehat{\psi} = \mathbf{r}_0^t \mathbf{X}^t \mathbf{Y} = \mathbf{r}_0^t \mathbf{X}^t \mathbf{X} \widehat{\beta}_0 = \lambda^t \widehat{\beta}_0$$

De este modo, si \mathbf{r}_1 es otra solución del sistema $\mathbf{X}^t \mathbf{X} \mathbf{r} = \lambda$, realizando los mismo cálculos que en la ecuación anterior se obtendría que

$$\mathbf{r}_1^t \mathbf{X}^t \mathbf{Y} = \lambda^t \widehat{\beta}_0 = \mathbf{r}_1^t \mathbf{X}^t \mathbf{Y}$$

Por tanto la definición de $\widehat{\psi}$ no depende de la solución del sistema $\mathbf{X}^t \mathbf{X} \mathbf{r} = \lambda$ que elijamos para calcular dicho estimador. □

Corolario 4. Si $\mathbf{U}_1, \dots, \mathbf{U}_p$ son las filas (ó columnas) de $\mathbf{X}^t \mathbf{X}$ entonces $\psi_1 = \mathbf{U}_1^t \beta, \dots, \psi_p = \mathbf{U}_p^t \beta$ son f.l.e.

Demostración. Se deja como ejercicio. □

Proposición 5 (Propiedades de los estimadores).

a) Sea $\psi = \lambda' \beta$ y sea $\widehat{\psi}$ su estimador lineal insesgado. Entonces $\text{Var}[\widehat{\psi}] = \sigma^2 \mathbf{r}_0' \mathbf{X}' \mathbf{X} \mathbf{r}_0 = \sigma^2 \mathbf{r}_0' \lambda$, siendo \mathbf{r}_0 cualquier solución del sistema $\mathbf{X}' \mathbf{X} \mathbf{r} = \lambda$.

b) Si $\psi_1 = \lambda_1' \beta$ y $\psi_2 = \lambda_2' \beta$ son f.l.e. y $\widehat{\psi}_1$ y $\widehat{\psi}_2$ son sus respectivos estimadores lineales insesgados, entonces

$$\text{Cov}[\widehat{\psi}_1, \widehat{\psi}_2] = \sigma^2 \mathbf{r}_1' \mathbf{X}' \mathbf{X} \mathbf{r}_2 = \sigma^2 \mathbf{r}_1' \lambda_2 = \sigma^2 \mathbf{r}_2' \lambda_1,$$

siendo \mathbf{r}_1 y \mathbf{r}_2 soluciones de los sistemas $\mathbf{X}' \mathbf{X} \mathbf{r} = \lambda_1$ y $\mathbf{X}' \mathbf{X} \mathbf{r} = \lambda_2$, respectivamente.

Demostración.

Apartado a)

Puesto que $\widehat{\psi} = \mathbf{r}_0' \mathbf{X}' \mathbf{Y}$ y $\text{Cov}[\mathbf{Y}] = \sigma^2 \mathbf{I}_n$ (modelo lineal básico), se obtiene que

$$\text{Var}[\widehat{\psi}] = \text{Cov}[\widehat{\psi}] = \text{Cov}[\mathbf{r}_0' \mathbf{X}' \mathbf{Y}] = \mathbf{r}_0' \mathbf{X}' \text{Cov}[\mathbf{Y}] \mathbf{X} \mathbf{r}_0 = \sigma^2 \mathbf{r}_0' \mathbf{X}' \mathbf{X} \mathbf{r}_0$$

Apartado b)

Queda propuesto como ejercicio. Se realiza de forma análoga a la demostración del Apartado a). □

La Proposición 2 nos proporcionó una forma de calcular el estimador insesgado de una f.l.e. $\psi = \lambda' \beta$. Basta encontrar una solución particular del sistema $\mathbf{X}' \mathbf{X} \mathbf{r} = \lambda$ y a partir de ella se obtiene el estimador y su varianza. La matriz $\mathbf{X}' \mathbf{X}$ y el vector $\mathbf{X}' \mathbf{Y}$ pueden obtenerse de las ecuaciones normales del modelo. La Proposición 3, nos indica que el estimador de ψ también puede obtenerse calculando una solución particular $\widehat{\beta}$ de las ecuaciones normales ($\widehat{\psi} = \lambda' \widehat{\beta}$). Sin embargo este procedimiento no permite el cálculo de la varianza de $\widehat{\psi}$.

Ejercicio propuesto. Probar el Corolario 4 y el apartado c) de la Proposición 5.

Teorema de Gauss-Markov y ejemplo

Proposición 6 (Teorema de Gauss-Markov). *Se verifica que $\widehat{\psi}$ es el estimador lineal insesgado de mínima varianza de ψ , es decir, si $T = \mathbf{c}'\mathbf{Y}$ es otro estimador lineal insesgado de ψ , entonces $\text{Var}[T] \geq \text{Var}[\widehat{\psi}]$ y se da la igualdad si y solo si $T = \widehat{\psi}$.*

Demostración. Tomemos $\mathbf{a} \in \mathbb{R}^n$ tal que $T = \widehat{\psi} + \mathbf{a}'\mathbf{Y}$ (en realidad $\mathbf{a} = \mathbf{c} - \mathbf{X}\mathbf{r}_0$). Por ser T insesgado para ψ se obtiene que

$$\psi = E[T] = E[\widehat{\psi}] + \mathbf{a}'E[\mathbf{Y}] = \psi + \mathbf{a}'\mathbf{X}\boldsymbol{\beta} \quad \boldsymbol{\beta} \in \mathbb{R}^p$$

Por tanto,

$$\mathbf{a}'\mathbf{X}\boldsymbol{\beta} = 0 \quad \text{para todo } \boldsymbol{\beta} \in \mathbb{R}^p$$

y en consecuencia, $\mathbf{a}'\mathbf{X} = \mathbf{0}$ (equivalentemente $\mathbf{X}'\mathbf{a} = \mathbf{0}$). Utilizando esto último y que $\widehat{\psi} = \mathbf{r}_0'\mathbf{X}'\mathbf{Y}$ y $\text{Cov}[\mathbf{Y}] = \sigma^2\mathbf{I}_n$ (modelo lineal básico)

$$\begin{aligned} \text{Var}[T] &= \text{Var}[\widehat{\psi}] + \text{Var}[\mathbf{a}'\mathbf{Y}] + \text{Cov}[\widehat{\psi}, \mathbf{a}'\mathbf{Y}] = \text{Var}[\widehat{\psi}] + \text{Cov}[\mathbf{a}'\mathbf{Y}] + \text{Cov}[\mathbf{r}_0'\mathbf{X}'\mathbf{Y}, \mathbf{a}'\mathbf{Y}] \\ &= \text{Var}[\widehat{\psi}] + \sigma^2\mathbf{a}'\mathbf{a} + \sigma^2\mathbf{r}_0'\mathbf{X}'\mathbf{a} = \text{Var}[\widehat{\psi}] + \sigma^2\mathbf{a}'\mathbf{a} \end{aligned}$$

Si $\mathbf{a} = (a_1, \dots, a_n)'$, entonces $\mathbf{a}'\mathbf{a} = \sum_{i=1}^n a_i^2 \geq 0$ luego

$$\text{Var}[T] = \text{Var}[\widehat{\psi}] + \sigma^2 \sum_{i=1}^n a_i^2 \geq \text{Var}[\widehat{\psi}]$$

Además se da la igualdad si y solo si $\sum_{i=1}^n a_i^2 = 0$, o equivalentemente si $\mathbf{a} = \mathbf{0}$. Como $T = \widehat{\psi} + \mathbf{a}'\mathbf{Y}$, se dará la igualdad si y sólo si $T = \widehat{\psi}$. □

Ejemplo 1. Consideremos el modelo lineal básico definido por las ecuaciones

$$\begin{aligned} E[Y_1] &= \beta_1 + \beta_2 + \beta_3 \\ E[Y_2] &= \beta_1 + \beta_3 \\ E[Y_3] &= \beta_2 \\ E[Y_4] &= 2\beta_1 - 3\beta_2 + 2\beta_3 \end{aligned}$$

- Hallar las ecuaciones normales del modelo.
- Probar que $\psi_1 = \beta_1 - 2\beta_2 + \beta_3$ y $\psi_2 = \beta_1 + \beta_3$ son funciones lineales estimables y hallar sus respectivos estimadores lineales insesgados de mínima varianza, $\widehat{\psi}_1$ y $\widehat{\psi}_2$.
- Calcular $\text{Var}[\widehat{\psi}_1]$, $\text{Var}[\widehat{\psi}_2]$ y $\text{Cov}[\widehat{\psi}_1, \widehat{\psi}_2]$.

Apartado a)

En primer lugar es sencillo comprobar que

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 2 & -3 & 2 \end{pmatrix} \quad \mathbf{X}'\mathbf{X} = \begin{pmatrix} 6 & -5 & 6 \\ -5 & 11 & -5 \\ 6 & -5 & 6 \end{pmatrix} \quad \mathbf{X}'\mathbf{Y} = \begin{pmatrix} Y_1 - Y_2 + 2Y_4 \\ Y_1 + Y_3 - 3Y_4 \\ Y_1 - Y_2 + 2Y_4 \end{pmatrix}$$

Las ecuaciones normales del modelo son por tanto:

$$\begin{aligned} 6\widehat{\beta}_1 - 5\widehat{\beta}_2 + 6\widehat{\beta}_3 &= Y_1 - Y_2 + 2Y_4 \\ -5\widehat{\beta}_1 + 11\widehat{\beta}_2 - 5\widehat{\beta}_3 &= Y_1 + Y_3 - 3Y_4 \\ 6\widehat{\beta}_1 - 5\widehat{\beta}_2 + 6\widehat{\beta}_3 &= Y_1 - Y_2 + 2Y_4 \end{aligned}$$

Suprimiendo la tercera ecuación, que es redundante y tomando, por ejemplo, $\widehat{\beta}_3 = 0$, obtenemos una solución particular de este sistema:

$$\widehat{\beta}_1 = \frac{1}{41}(16Y_1 - 11Y_2 + 5Y_3 + 7Y_4) \quad \widehat{\beta}_2 = \frac{1}{41}(11Y_1 - 5Y_2 + 6Y_3 - 8Y_4) \quad \widehat{\beta}_3 = 0$$

Apartado b)

Si $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$, entonces $\psi_1 = \lambda_1' \boldsymbol{\beta}$ y $\psi_2 = \lambda_2' \boldsymbol{\beta}$ con

$$\lambda_1 = (1, -2, 1)' \quad \text{y} \quad \lambda_2 = (1, 0, 1)'$$

Para probar que ψ_1 y ψ_2 son f.l.e. basta aplicar la Proposición 1 comprobando que λ_1 y λ_2 son combinación lineal de las filas de \mathbf{X} , lo cual queda como ejercicio.

Para calcular $\widehat{\psi}_1$ y $\widehat{\psi}_2$ tenemos dos opciones (el resultado es el mismo):

- Calcular $\widehat{\psi}_1 = \lambda_1' \widehat{\boldsymbol{\beta}}$ y $\widehat{\psi}_2 = \lambda_2' \widehat{\boldsymbol{\beta}}$ con $\widehat{\boldsymbol{\beta}}$ solución particular de las ecuaciones normales.
- Obtener soluciones particulares, \mathbf{r}_1 y \mathbf{r}_2 , de los sistemas de ecuaciones $\mathbf{X}'\mathbf{X}\mathbf{r} = \lambda_1$ y $\mathbf{X}'\mathbf{X}\mathbf{r} = \lambda_2$ y calcular $\widehat{\psi}_1 = \mathbf{r}_1' \mathbf{X}'\mathbf{Y}$, $\widehat{\psi}_2 = \mathbf{r}_2' \mathbf{X}'\mathbf{Y}$.

Por cualquiera de estos procedimientos se obtiene que

$$\widehat{\psi}_1 = \frac{1}{41}(-6Y_1 - Y_2 - 7Y_3 + 23Y_4) \quad \widehat{\psi}_2 = \frac{1}{41}(16Y_1 - 11Y_2 + 5Y_3 + 7Y_4)$$

Queda como ejercicio realizar los cálculos por los procedimientos indicados.

Apartado c)

Este apartado queda como ejercicio. Como indicación, se puede utilizar la Proposición 5 de propiedades de los estimadores o, más directamente, si denotamos $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)'$, como $\text{Cov}[\mathbf{Y}] = \sigma^2 \mathbf{I}_4$, entonces $\text{Var}[\mathbf{c}'\mathbf{Y}] = \sigma^2 \mathbf{c}'\mathbf{c}$ y $\text{Cov}[\mathbf{c}_1'\mathbf{Y}, \mathbf{c}_2'\mathbf{Y}] = \sigma^2 \mathbf{c}_1'\mathbf{c}_2$.

Ejercicios propuestos.

1. En el apartado b) del ejemplo, probar que λ_1 y λ_2 son combinación lineal de las filas de \mathbf{X} .
2. Realizar los cálculos necesarios para obtener las expresiones de $\widehat{\psi}_1$ y $\widehat{\psi}_2$ que se indican en el apartado b) del ejemplo.
3. Calcular las varianzas y covarianza propuestas en el apartado c) del ejemplo.

Reparametrización

El concepto de f.l.e. puede ser utilizado para realizar un cambio en los parámetros del modelo que lo transforme en un modelo de rango completo. Surge así el concepto de reparametrización que pasamos a definir.

Definición 2 (Reparametrización). Llamaremos **reparametrización del modelo lineal de rango k** a una colección $\gamma_1 = \lambda_1^t \beta, \dots, \gamma_k = \lambda_k^t \beta$ de funciones lineales estimables linealmente independientes.

Si denotamos

$$\boldsymbol{\gamma} = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_k \end{pmatrix}, \quad \mathbf{L} = \begin{pmatrix} \lambda_1^t \\ \vdots \\ \lambda_k^t \end{pmatrix}$$

se verifica que \mathbf{L} es una matriz de orden $k \times p$ y $r(\mathbf{L}) = k$ y que $\boldsymbol{\gamma} = \mathbf{L}\boldsymbol{\beta}$. También llamaremos reparametrización al vector $\boldsymbol{\gamma}$.

Si $\boldsymbol{\gamma} = \mathbf{L}\boldsymbol{\beta}$ es una reparametrización del modelo y $\lambda_1, \dots, \lambda_k$ son las filas de \mathbf{L} , entonces $\gamma_1 = \lambda_1^t \beta, \dots, \gamma_k = \lambda_k^t \beta$ son f.l.e. linealmente independientes. Como el número máximo de f.l.e. linealmente independientes es k , cualquier otra f.l.e. se puede escribir como combinación lineal de $\gamma_1, \dots, \gamma_k$. En particular, si denotamos $\mathbf{X}_1, \dots, \mathbf{X}_n$ a las filas de \mathbf{X} se verifica que

$$\begin{aligned} \mathbf{X}_1^t \boldsymbol{\beta} &= z_{11}\gamma_1 + \dots + z_{1k}\gamma_k \\ \mathbf{X}_2^t \boldsymbol{\beta} &= z_{21}\gamma_1 + \dots + z_{2k}\gamma_k \\ &\vdots \\ \mathbf{X}_n^t \boldsymbol{\beta} &= z_{n1}\gamma_1 + \dots + z_{nk}\gamma_k \end{aligned}$$

o escrito matricialmente

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{Z}\boldsymbol{\gamma}, \quad \text{siendo } \mathbf{Z} = \begin{pmatrix} z_{11} & \dots & z_{1k} \\ \vdots & \ddots & \vdots \\ z_{n1} & \dots & z_{nk} \end{pmatrix}$$

Teniendo en cuenta la igualdad anterior, \mathbf{Y} se ajusta a un modelo lineal que depende de los nuevos parámetros:

$$E[\mathbf{Y}] = \mathbf{Z}\boldsymbol{\gamma}$$

Nos referiremos a este modelo como **modelo reparametrizado**.

La matriz de este modelo, \mathbf{Z} , es de orden $n \times k$. Como $\mathbf{X} = \mathbf{Z}\mathbf{L}$, se verifica que $r(\mathbf{Z}) = k$. En efecto, si $r(\mathbf{Z}) < k$, entonces

$$r(\mathbf{X}) = r(\mathbf{Z}\mathbf{L}) \leq r(\mathbf{Z}) < k$$

lo cual contradice la hipótesis de que $r(\mathbf{X}) = k$.

Por tanto, el modelo reparametrizado es un modelo lineal de rango completo. Aplicando la teoría estudiada en el tema anterior para el Modelo Lineal Básico, el estimador de mínimos cuadrados de $\boldsymbol{\gamma}$ y el estimador de σ^2 tienen las siguientes expresiones:

$$\widehat{\boldsymbol{\gamma}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} \quad \text{y} \quad \widehat{\sigma}^2 = \frac{1}{n-k}(\mathbf{Y} - \mathbf{Z}\widehat{\boldsymbol{\gamma}})'(\mathbf{Y} - \mathbf{Z}\widehat{\boldsymbol{\gamma}})$$

Ambos estimadores son insesgados. Además se verifica el siguiente resultado:

Proposición 7. Si $\widehat{\beta}_0$ es una solución particular de las ecuaciones normales $\mathbf{X}'\mathbf{X}\widehat{\beta} = \mathbf{X}'\mathbf{Y}$, entonces $\widehat{\gamma} = \mathbf{L}\widehat{\beta}_0$ y en consecuencia

$$\widehat{\sigma}^2 = \frac{1}{n-k}(\mathbf{Y} - \mathbf{X}\widehat{\beta}_0)'(\mathbf{Y} - \mathbf{X}\widehat{\beta}_0)$$

Demostración. Puesto que las ecuaciones normales del modelo reparametrizado, $\mathbf{Z}'\mathbf{Z}\widehat{\gamma} = \mathbf{Z}'\mathbf{Y}$, son un sistema compatible determinado y por tanto tienen solución única, para establecer la primera igualdad bastará probar que $\mathbf{L}\widehat{\beta}$ es solución de dicho sistema.

Sabemos que $\widehat{\beta}_0$ verifica $\mathbf{X}'\mathbf{X}\widehat{\beta}_0 = \mathbf{X}'\mathbf{Y}$ y como $\mathbf{X} = \mathbf{ZL}$ se obtiene que

$$\mathbf{L}'\mathbf{Z}'\mathbf{ZL}\widehat{\beta}_0 = \mathbf{L}'\mathbf{Z}'\mathbf{Y}$$

o equivalentemente

$$\mathbf{L}'(\mathbf{Z}'\mathbf{ZL}\widehat{\beta}_0 - \mathbf{Z}'\mathbf{Y}) = \mathbf{0}$$

Si consideramos $\mathbf{L}'\mathbf{x} = \mathbf{0}$ como un sistema homogéneo de n ecuaciones con k incógnitas, puesto que $r(\mathbf{L}') = k$ es compatible determinado y por tanto su única solución es $\mathbf{x} = \mathbf{0}$. En consecuencia, se deduce que

$$\mathbf{Z}'\mathbf{ZL}\widehat{\beta}_0 - \mathbf{Z}'\mathbf{Y} = \mathbf{0}$$

o equivalentemente

$$\mathbf{Z}'\mathbf{Z}(\mathbf{L}\widehat{\beta}_0) = \mathbf{Z}'\mathbf{Y}$$

es decir, $\mathbf{L}\widehat{\beta}_0$ es solución de las ecuaciones normales del modelo reparametrizado.

Además, $\mathbf{X}\widehat{\beta}_0 = \mathbf{ZL}\widehat{\beta}_0 = \mathbf{Z}\widehat{\gamma}$ y por tanto

$$\widehat{\sigma}^2 = \frac{1}{n-k}(\mathbf{Y} - \mathbf{Z}\widehat{\gamma})'(\mathbf{Y} - \mathbf{Z}\widehat{\gamma}) = \frac{1}{n-k}(\mathbf{Y} - \mathbf{X}\widehat{\beta}_0)'(\mathbf{Y} - \mathbf{X}\widehat{\beta}_0)$$

□

En cuanto a la estimación de f.l.e., también puede realizarse en el modelo reparametrizado. En efecto, si $\psi = \lambda'\beta$ es una f.l.e., tiene que ser combinación lineal de $\gamma_1, \dots, \gamma_k$, es decir, $\psi = \rho'\gamma$ para cierto $\rho \in \mathbb{R}^k$. Además se verifica que

$$\lambda'\beta = \psi = \rho'\gamma = \rho'\mathbf{L}\beta, \quad \beta \in \mathbb{R}^p$$

y por tanto $\lambda' = \rho'\mathbf{L}$ (equivalentemente $\lambda = \mathbf{L}'\rho$).

De este modo, si $\widehat{\beta}_0$ es una solución de las ecuaciones normales $\mathbf{X}'\mathbf{X}\widehat{\beta}_0 = \mathbf{X}'\mathbf{Y}$, teniendo en cuenta la Proposición 7, el estimador de mínimos cuadrados de ψ se calcula del siguiente modo:

$$\widehat{\psi} = \lambda'\widehat{\beta}_0 = \rho'\mathbf{L}\widehat{\beta}_0 = \rho'\widehat{\gamma}$$

y por tanto coincide con el estimador lineal insesgado de mínima varianza en el modelo reparametrizado.

Ejercicio propuesto. Indicar una reparametrización del modelo lineal del Ejemplo 1.

3.2.2. Estimación puntual en el Modelo Lineal Normal.

Sea \mathbf{Y} vector aleatorio n dimensional que se ajusta a un Modelo Lineal Normal de rango no completo, es decir, $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$ con \mathbf{X} matriz $n \times p$ y $r(\mathbf{X}) = k < p$ y $\boldsymbol{\beta} \in \mathbb{R}^p$ (equivalentemente, $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mathcal{E}}$ con $\boldsymbol{\mathcal{E}} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$).

Sea $\boldsymbol{\gamma} = \mathbf{L}\boldsymbol{\beta}$ una reparametrización e $\mathbf{Y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\mathcal{E}}$ el modelo lineal de rango completo reparametrizado. Sean $\widehat{\boldsymbol{\gamma}}$ y $\widehat{\sigma}^2$ los estimadores insesgados de $\boldsymbol{\gamma}$ y σ^2 , respectivamente. Puesto que estamos en un Modelo Lineal de Rango Completo, podemos aplicar las Proposiciones 1 y 2 del Tema 2 (apartado 2.2.2) para deducir las siguientes propiedades:

- $\widehat{\boldsymbol{\gamma}} \sim N_k(\boldsymbol{\gamma}, \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1})$.
- $\frac{(n-k)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2(n-k)$.
- $\widehat{\boldsymbol{\gamma}}$ y $\widehat{\sigma}^2$ son independientes.
- El estadístico $T = \begin{pmatrix} \widehat{\boldsymbol{\gamma}} \\ \widehat{\sigma}^2 \end{pmatrix}$ es suficiente y completo.

A partir de estas propiedades, podemos deducir las correspondientes de las funciones lineales estimables. Sea $\psi = \boldsymbol{\lambda}'\boldsymbol{\beta}$ una función lineal estimable y sea $\boldsymbol{\rho} \in \mathbb{R}^k$ tal que $\psi = \boldsymbol{\rho}'\boldsymbol{\gamma}$ y sea $\widehat{\psi}$ el estimador insesgado de ψ calculado en la sección anterior.

Proposición 8. *Se verifican los siguientes enunciados:*

- $\widehat{\psi} \sim N(\psi, \sigma^2\mathbf{r}_0'(\mathbf{X}'\mathbf{X})\mathbf{r}_0)$, siendo \mathbf{r}_0 cualquier solución del sistema $\mathbf{X}'\mathbf{X}\mathbf{r} = \boldsymbol{\lambda}$.
- $\widehat{\psi}$ y $\widehat{\sigma}^2$ son independientes.
- $\widehat{\psi}$ es el estimador insesgado de mínima varianza de ψ .

Demostración.

Apartado a)

Sabemos que $\widehat{\psi} = \mathbf{r}_0'\mathbf{X}'\mathbf{Y}$ e \mathbf{Y} tiene distribución Normal. Luego $\widehat{\psi}$ es la transformación lineal de un vector Normal y por tanto Normal.

Apartado b)

Sabemos que $\widehat{\psi} = \boldsymbol{\rho}'\widehat{\boldsymbol{\gamma}}$ y, por las propiedades anteriores, $\widehat{\boldsymbol{\gamma}}$ es independiente de $\widehat{\sigma}^2$. Luego $\widehat{\psi}$ es función de $\widehat{\boldsymbol{\gamma}}$ y por tanto también independiente de $\widehat{\sigma}^2$.

Apartado c)

Como $\widehat{\psi}$ es insesgado para ψ y $T = (\widehat{\boldsymbol{\gamma}}', \widehat{\sigma}^2)'$ suficiente y completo, aplicando el Teorema de Lehmann-Scheffé obtenemos que

$$E[\widehat{\psi}|T] = E[\boldsymbol{\rho}'\widehat{\boldsymbol{\gamma}}|T] = \boldsymbol{\rho}'\widehat{\boldsymbol{\gamma}} = \widehat{\psi}$$

es el estimador insesgado de mínima varianza de ψ . □

3.2.3 Intervalos de confianza en el modelo lineal Normal.

Utilizando las propiedades del apartado anterior podemos deducir intervalos de confianza para σ^2 y para una f.l.e. ψ . Queda como ejercicio deducir este último.

- a) Intervalo de confianza para σ^2 al nivel de confianza $1 - \alpha$:

$$\left[\frac{(n-k)\widehat{\sigma}^2}{\chi_{n-k,\alpha/2}^2}, \frac{(n-k)\widehat{\sigma}^2}{\chi_{n-k,1-\alpha/2}^2} \right]$$

siendo $\chi_{n-k,\alpha/2}^2$ y $\chi_{n-k,1-\alpha/2}^2$ cuantiles de orden $1 - \alpha/2$ y $\alpha/2$, respectivamente, de la distribución $\chi^2(n-k)$.

- b) Intervalo de confianza para una f.l.e. $\psi = \lambda' \boldsymbol{\beta}$ al nivel de confianza $1 - \alpha$

$$\left[\widehat{\psi} - t_{n-k,\alpha/2} \sqrt{\widehat{\sigma}^2 \mathbf{r}_0' (\mathbf{X}' \mathbf{X}) \mathbf{r}_0}, \widehat{\psi} + t_{n-k,\alpha/2} \sqrt{\widehat{\sigma}^2 \mathbf{r}_0' (\mathbf{X}' \mathbf{X}) \mathbf{r}_0} \right]$$

siendo \mathbf{r}_0 solución del sistema $\mathbf{X}' \mathbf{X} \mathbf{r} = \boldsymbol{\lambda}$ y $t_{n-k,\alpha/2}$ cuantil de orden $1 - \alpha/2$ de una distribución $t(n-k)$.

Ejercicio propuesto: Deducir el intervalo de confianza al nivel $1 - \alpha$ para ψ utilizando las propiedades estudiadas en el apartado 3.2.2.

3.3. Contraste de hipótesis en el Modelo Lineal de Rango No Completo.

Nos planteamos el contraste de la hipótesis $H_0 : \psi_1 = \dots = \psi_s = 0$, con $s < k$ y $\psi_i = \lambda_i^t \beta$, $i = 1, \dots, s$, funciones lineales estimables linealmente independientes.

Si consideramos una colección de f.l.e. linealmente independientes, $\psi_{s+1} = \lambda_{s+1}^t \beta, \dots, \psi_k = \lambda_k^t \beta$ que, a su vez, linealmente independientes de ψ_1, \dots, ψ_s , entonces

$$\Psi = \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_s \\ \psi_{s+1} \\ \vdots \\ \psi_k \end{pmatrix} = \mathbf{L}\beta, \quad \text{con } \mathbf{L} = \begin{pmatrix} \lambda_1^t \\ \vdots \\ \lambda_s^t \\ \lambda_{s+1}^t \\ \vdots \\ \lambda_k^t \end{pmatrix}$$

es una reparametrización del modelo, es decir, $\mathbf{Y} \sim N_n(\mathbf{Z}\Psi, \sigma^2\mathbf{I}_n)$ para una matriz \mathbf{Z} de orden $n \times k$ y $r(\mathbf{Z}) = k$, que es un modelo lineal de rango completo.

El contraste de $H_0 : \psi_1 = \dots = \psi_s = 0$, planteado en dicho modelo, es el mismo que ya resolvimos en el apartado 2.4.3. del Tema 2. Por tanto podemos utilizar el test de hipótesis calculado en dicho apartado.

El estadístico de contraste se calcula del siguiente modo:

$$F = \frac{n-k}{s} \frac{Q_1}{Q_0}$$

siendo $Q_0 = (\mathbf{Y} - \mathbf{X}\widehat{\beta}_0)^t(\mathbf{Y} - \mathbf{X}\widehat{\beta}_0) = \mathbf{Y}^t(\mathbf{I}_n - \mathbf{H})\mathbf{Y} = SCE$, es decir, la suma de cuadrados de los residuos del modelo original, con $\widehat{\beta}_0$ solución de las ecuaciones normales $\mathbf{X}^t\mathbf{X}\widehat{\beta} = \mathbf{X}^t\mathbf{Y}$ y $\mathbf{H} = \mathbf{Z}(\mathbf{Z}^t\mathbf{Z})^{-1}\mathbf{Z}^t$.

Para calcular Q_1 , tengamos en cuenta que el modelo reducido por H_0 es $\mathbf{Y} \sim N_n(\mathbf{Z}_2\Psi_2, \sigma^2\mathbf{I}_n)$ con $\Psi_2 = (\psi_{s+1}, \dots, \psi_k)^t$ y \mathbf{Z}_2 la matriz formada por las últimas $k - s$ columnas de \mathbf{Z} . De este modo,

$$Q_1 = \mathbf{Y}^t(\mathbf{H} - \mathbf{H}_2)\mathbf{Y} = \mathbf{Y}^t(\mathbf{I}_n - \mathbf{H}_2)\mathbf{Y} - \mathbf{Y}^t(\mathbf{I}_n - \mathbf{H})\mathbf{Y} = SCE_R - SCE$$

siendo SCE_R la suma de cuadrados de los residuos en el modelo reducido por H_0 y $\mathbf{H}_2 = \mathbf{Z}_2(\mathbf{Z}_2^t\mathbf{Z}_2)^{-1}\mathbf{Z}_2^t$.

Con estas notaciones, se rechaza H_0 al nivel de significación α si $F \geq F_{s, n-k, \alpha}$.

Ejercicio propuesto: Sea ψ función lineal estimable. Deducir, a partir del intervalo de confianza para ψ , un test de hipótesis de extensión α para contrastar la hipótesis $H_0 : \psi = \psi_0$ con ψ_0 valor conocido.

Tema 4: Modelos de Diseño de Experimentos

4.1 Introducción al Diseño de Experimentos

El objetivo de un experimento es estudiar el efecto que sobre una variable de interés, que llamaremos **variable respuesta**, tienen otras variables que llamaremos variables experimentales o **factores**. Supondremos que la variable respuesta es continua y que los factores se fijan durante el experimento a ciertos **niveles** determinados (se les suele considerar como variables cualitativas, aunque en ocasiones vengan expresados en escala numérica). El experimento consiste en seleccionar ciertas unidades experimentales, fijar los niveles de los factores y observar el valor de la variable respuesta en cada unidad experimental. El número total de datos es el tamaño del experimento.

El **diseño estadístico de experimentos** es el proceso de planificación de un experimento para obtener datos apropiados, que pueden ser analizados mediante métodos estadísticos para extraer conclusiones válidas y objetivas.

La metodología estadística es el único enfoque objetivo para analizar un problema que involucre datos sujetos a errores experimentales. Por tanto, el diseño del experimento y el análisis estadístico de los datos están estrechamente relacionados, ya que el método de análisis depende directamente del diseño empleado.

En este tema veremos una serie de modelos que se adaptan a unos diseños experimentales bastante comunes. Como hemos dicho antes, en todos estos modelos aparecerán dos elementos comunes:

- La **variable respuesta** (Y): la característica que puede ser observada, medida y analizada. Supondremos que es cuantitativa continua y con distribución normal.
- El **factor o factores**: un factor es una variable cuyo posible efecto sobre la respuesta se quiere estudiar; los niveles de un factor son los valores que puede tomar el factor. El factor será tratado como una variable cualitativa.

La forma de diseñar un experimento es distribuir a las unidades experimentales entre los distintos niveles del factor para posteriormente medir el valor que la variable respuesta toma en dichas unidades experimentales.

- Si el experimentador determina los niveles del factor que serán considerados se dice que el modelo es de **efectos fijos**
- Si el experimentador toma una serie de niveles al azar entre todos los niveles del factor se dice que el modelo es de **efectos aleatorios**
- Un modelo con más de un factor donde los niveles de algunos factores hayan sido fijados y los de otros factores se hayan seleccionado aleatoriamente se denomina **modelo mixto**

4.2 Experimentos con un factor. Efectos fijos

4.2.1 Diseño completamente aleatorizado

Llamaremos *diseño completamente aleatorizado* a aquel en el que la asignación de los niveles del factor a cada una de las unidades experimentales consideradas se hace totalmente al azar.

Ejemplo. Supongamos que un agricultor quiere comparar el rendimiento de 5 fertilizantes en la producción de trigo. Un procedimiento posible sería tomar 20 parcelas de tierra y sembrar en ellas el trigo, escogiendo el fertilizante A para cuatro de ellas, el B para otras cuatro, y así sucesivamente. La asignación de los fertilizantes a las parcelas se haría mediante un procedimiento aleatorio y de cada una de ellas se obtendría el dato de la producción obtenida en dicha parcela.

El Modelo Lineal que se corresponde con esta definición es:

$$Y_{ij} = \mu + \tau_i + \mathcal{E}_{ij}, \quad i = 1, \dots, k; \quad j = 1, \dots, n_i \quad (N = \sum_{i=1}^k n_i)$$

En este modelo, el vector de observaciones es $\mathbf{Y} = (Y_{11}, \dots, Y_{kn_k})^t$, el vector de parámetros es $\boldsymbol{\beta} = (\mu, \tau_1, \dots, \tau_k)^t$, por tanto $p = k + 1$, y la matriz del modelo es:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{pmatrix} \quad \mathbf{X}'\mathbf{Y} = \begin{pmatrix} Y_{..} \\ Y_{1.} \\ \vdots \\ Y_{k.} \end{pmatrix}$$

siendo $Y_{i.} = \sum_{j=1}^{n_i} Y_{ij}$, $i = 1, \dots, k$ e $Y_{..} = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}$. La matriz \mathbf{X} tiene N filas. En la columna 2 hay n_1 unos, en la columna 3 hay n_2 unos y así sucesivamente hasta llegar a la última columna donde hay n_k unos. Se verifica que $r(\mathbf{X}) = k < p = k + 1$, por tanto es un **modelo de rango no completo**.

Observaciones.

- En el modelo de *efectos fijos*, τ_i , $i = 1, \dots, k$, son parámetros desconocidos.
- Si $n_1 = \dots = n_k$, el modelo se dice *balanceado*. En caso contrario se dice que es *no balanceado*.
- Asumiremos, en general, que $\mathcal{E} \sim N_N(\mathbf{0}, \sigma^2 \mathbf{I}_N)$, siendo $\mathcal{E} = (\mathcal{E}_{11}, \dots, \mathcal{E}_{kn_k})^t$, aunque para algún resultado la normalidad no sea necesaria.

Ecuaciones Normales del Modelo. Para obtener las ecuaciones normales de este modelo ($\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}$), podemos utilizar que $E[\mathbf{X}'\mathbf{Y}] = \mathbf{X}'\mathbf{X}\boldsymbol{\beta}$. De este modo, teniendo en cuenta que

$$E[Y_{..}] = N\mu + n_1\tau_1 + \cdots + n_k\tau_k \quad E[Y_{i.}] = n_i\mu + n_i\tau_i, \quad i = 1, \dots, k$$

obtenemos las ecuaciones normales del modelo:

$$\begin{aligned} N\widehat{\mu} + n_1\widehat{\tau}_1 + \cdots + n_k\widehat{\tau}_k &= Y_{..} \\ n_i\widehat{\mu} + n_i\widehat{\tau}_i &= Y_{i.}, \quad i = 1, \dots, k \end{aligned}$$

Una solución particular de estas ecuaciones, que se obtiene imponiendo la condición $\sum_{i=1}^k \tau_i = 0$, es

$$\widehat{\boldsymbol{\beta}} = (\widehat{\mu}, \widehat{\tau}_1, \dots, \widehat{\tau}_k)' = (\bar{Y}_{..}, \bar{Y}_{1.} - \bar{Y}_{..}, \dots, \bar{Y}_{k.} - \bar{Y}_{..})'$$

donde $\bar{Y}_{i.} = Y_{i.}/n_i$, $i = 1, \dots, k$, e $\bar{Y}_{..} = Y_{..}/N$.

Estimación puntual.

En primer lugar, vamos a considerar el conjunto de k funciones lineales estimables (f.l.e.) linealmente independientes que nos proporcionan las filas de \mathbf{X} , $\{\mu + \tau_i, i = 1, \dots, k\}$. En el siguiente resultado proporcionamos los estimadores de estas f.l.e.

Proposición 1 *El estimador insesgado de mínima varianza de $\mu + \tau_i$ es $\widehat{\mu + \tau_i} = \bar{Y}_{i.}$, $i = 1, \dots, k$. Además se verifica que*

$$\text{Var}(\bar{Y}_{i.}) = \frac{\sigma^2}{n_i}, \quad \text{Cov}(\bar{Y}_{i_1.}, \bar{Y}_{i_2.}) = 0, \quad i_1 \neq i_2$$

Demostración. Utilizando la solución de las ecuaciones normales calculada anteriormente, el estimador insesgado de mínima varianza de $\mu + \tau_i$ es

$$\widehat{\mu + \tau_i} = \widehat{\mu} + \widehat{\tau}_i = \bar{Y}_{..} + \bar{Y}_{i.} - \bar{Y}_{..} = \bar{Y}_{i.}$$

El cálculo de la varianza y covarianza de los estimadores, aunque se puede realizar utilizando la teoría del Tema 3, es más sencillo si tenemos en cuenta que Y_{11}, \dots, Y_{kn_k} son variables aleatorias independientes e idénticamente distribuidas. De este modo

$$\text{Var}[\bar{Y}_{i.}] = \frac{1}{n_i^2} \sum_{j=1}^{n_i} \text{Var}[Y_{ij}] = \frac{n_i\sigma^2}{n_i^2} = \frac{\sigma^2}{n_i}$$

Por último, si $i_1 \neq i_2$, como $\bar{Y}_{i_1.}$ es función de $Y_{i_11}, \dots, Y_{i_1n_{i_1}}$ e $\bar{Y}_{i_2.}$ es función de $Y_{i_21}, \dots, Y_{i_2n_{i_2}}$ y ambos grupos de variables son independientes, también lo serán $\bar{Y}_{i_1.}$ e $\bar{Y}_{i_2.}$ y por tanto $\text{Cov}[\bar{Y}_{i_1.}, \bar{Y}_{i_2.}] = 0$.

A continuación nos centraremos en un conjunto de f.l.e., los **contrastos**, que nos serán de gran utilidad para comparar la influencia de los distintos niveles del factor sobre la variable respuesta.

Definición 1 (Contraste) *Diremos que $\psi = \sum_{i=1}^k c_i\tau_i$ es un **contraste** si $\sum_{i=1}^k c_i = 0$.*

Proposición 2 Una combinación lineal de τ_1, \dots, τ_k , $\psi = \sum_{i=1}^k c_i \tau_i$ es una función lineal estimable si y sólo si es un contraste.

El estimador insesgado de mínima varianza de ψ es $\widehat{\psi} = \sum_{i=1}^k c_i \bar{Y}_i$ y además $\text{Var}[\widehat{\psi}] = \sigma^2 \sum_{i=1}^k c_i^2 / n_i$.

Demostración.

Si $\psi = \sum_{i=1}^k c_i \tau_i$ es una función lineal estimable entonces tiene que ser combinación lineal de $\mu + \tau_1, \dots, \mu + \tau_k$:

$$\psi = \sum_{i=1}^k c_i \tau_i = \sum_{i=1}^k \lambda_i (\mu + \tau_i) = \left(\sum_{i=1}^k \lambda_i \right) \mu + \sum_{i=1}^k \lambda_i \tau_i$$

De aquí se deduce que $\lambda_i = c_i$ y que $\sum_{i=1}^k c_i = \sum_{i=1}^k \lambda_i = 0$, luego ψ es un contraste.

Recíprocamente, si $\psi = \sum_{i=1}^k c_i \tau_i$ es un contraste, entonces $\sum_{i=1}^k c_i = 0$ y por tanto

$$\psi = \sum_{i=1}^k c_i \tau_i = \sum_{i=1}^k c_i (\mu + \tau_i)$$

luego es combinación lineal de f.l.e. y por tanto es f.l.e.

La demostración de la segunda parte queda como ejercicio.

Ejemplo. Las funciones $\tau_{i_1} - \tau_{i_2}$ $i_1 \neq i_2$ son contrastes. Su estimador insesgado de mínima varianza es $\bar{Y}_{i_1} - \bar{Y}_{i_2}$ y la varianza de este estimador es $\sigma^2(1/n_{i_1} + 1/n_{i_2})$.

Por último proporcionamos el estimador insesgado de σ^2 .

Proposición 3 El estimador insesgado de σ^2 es

$$\widehat{\sigma}^2 = \frac{1}{N - k} \left(\sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}^2 - \sum_{i=1}^k n_i \bar{Y}_i^2 \right)$$

Demostración. Utilizando la Proposición 6 del Tema 3, sabemos que el estimador insesgado de σ^2 es

$$\widehat{\sigma}^2 = \frac{1}{N - k} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^t (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \frac{1}{N - k} (\mathbf{Y}^t \mathbf{Y} - \widehat{\boldsymbol{\beta}}^t \mathbf{X}^t \mathbf{Y})$$

siendo $\widehat{\boldsymbol{\beta}}$ cualquier solución de las ecuaciones normales.

En nuestro caso tenemos

$$\widehat{\boldsymbol{\beta}}^t = (\widehat{\mu}, \widehat{\tau}_1, \dots, \widehat{\tau}_k) = (\bar{Y}_{..}, \bar{Y}_1 - \bar{Y}_{..}, \dots, \bar{Y}_k - \bar{Y}_{..})$$

y

$$\mathbf{X}^t \mathbf{Y} = (Y_{..}, Y_1, \dots, Y_k)^t = (N\bar{Y}_{..}, n_1 \bar{Y}_1, \dots, n_k \bar{Y}_k)^t$$

De este modo es sencillo obtener que $\widehat{\boldsymbol{\beta}}^t \mathbf{X}^t \mathbf{Y} = \sum_{i=1}^k n_i \bar{Y}_i^2$.

Ejercicio propuesto. Demostrar la segunda parte de la Proposición 2.

Propiedades de los estimadores

Haciendo uso de la Proposición 7 del Tema 3, obtenemos las siguientes propiedades de los estimadores obtenidos anteriormente:

- $\widehat{\mu + \tau_i} = \bar{Y}_i \sim N(\mu + \tau_i, \sigma^2/n_i), i = 1, \dots, k.$
- Si $\psi = \sum_{i=1}^k c_i \tau_i$ es un contraste, entonces $\widehat{\psi} \sim N(\psi, \sigma^2 \sum_{i=1}^k c_i^2/n_i).$
- $(N - k)\widehat{\sigma}^2/\sigma^2 \sim \chi^2(N - k).$
- $\widehat{\mu + \tau_i}$ y $\widehat{\sigma}^2$ son independientes, $i = 1, \dots, k.$
- Si $\psi = \sum_{i=1}^k c_i \tau_i$ es un contraste, entonces $\widehat{\psi}$ y $\widehat{\sigma}^2$ son independientes.

Intervalos de confianza al nivel de confianza $1 - \alpha$

Particularizando los intervalos obtenidos en el Tema 3 para f.l.e. obtenemos las siguientes expresiones:

- Intervalo de confianza para $\mu + \tau_i, i = 1, \dots, k,$ al nivel $1 - \alpha$ ($0 < \alpha < 1$):

$$\left[\bar{Y}_i - \frac{\widehat{\sigma}}{\sqrt{n_i}} t_{N-k, \alpha/2}, \bar{Y}_i + \frac{\widehat{\sigma}}{\sqrt{n_i}} t_{N-k, \alpha/2} \right]$$

- Intervalo de confianza para un contraste $\sum_{i=1}^k c_i \tau_i,$ al nivel $1 - \alpha$ ($0 < \alpha < 1$):

$$\left[\sum_{i=1}^k c_i \bar{Y}_i - \widehat{\sigma} \sqrt{\sum_{i=1}^k \frac{c_i^2}{n_i}} t_{N-k, \alpha/2}, \sum_{i=1}^k c_i \bar{Y}_i + \widehat{\sigma} \sqrt{\sum_{i=1}^k \frac{c_i^2}{n_i}} t_{N-k, \alpha/2} \right]$$

siendo $t_{N-k, \alpha/2}$ el cuantil por la derecha de orden $\alpha/2$ de una distribución $t(N - k).$

Contraste de hipótesis

En este modelo, el contraste más importante a realizar es

$$H_0 : \tau_1 = \dots = \tau_k$$

Caso de rechazar $H_0,$ quedará probado que existen i_1, i_2 tales que $\tau_{i_1} \neq \tau_{i_2},$ y por tanto el factor presenta influencia significativa sobre la variable respuesta. Habrá que investigar qué niveles del factor presentan efectos diferenciados, para lo cual se llevará a cabo un procedimiento de comparaciones múltiples que se estudiará más adelante. Por el contrario, si aceptamos $H_0,$ los efectos de todos los niveles del factor sobre la variable respuesta se considerarán iguales y no se podrá afirmar que existe una influencia del factor sobre la dicha variable.

La hipótesis H_0 es estimable puesto que se puede expresar como $H_0 : \tau_1 - \tau_k = \dots = \tau_{k-1} - \tau_k = 0.$ Las $k - 1$ funciones $\tau_i - \tau_k, i = 1, \dots, k - 1$ son contrastes y, aplicando la Proposición 2, funciones lineales estimables linealmente independientes.

Para calcular el test de hipótesis, seguiremos los pasos indicados en el apartado 3.3 del Tema 3.

1er paso: modelo reducido por H_0 reparametrizado.

El modelo reducido por H_0 es

$$Y_{ij} = \mu + \tau_1 + \mathcal{E}_{ij}, \quad i = 1, \dots, k; \quad j = 1, \dots, n_i$$

Si denotamos $\gamma = \mu + \tau_1$, el modelo reducido reparametrizado es:

$$Y_{ij} = \gamma + \mathcal{E}_{ij}, \quad i = 1, \dots, k; \quad j = 1, \dots, n_i$$

2º paso: formas cuadráticas asociadas al test. La suma de cuadrados residual Q_0 , se obtuvo al calcular el estimador de σ^2 y tiene la siguiente expresión:

$$Q_0 = (N - k)\widehat{\sigma}^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}^2 - \sum_{i=1}^k n_i \bar{Y}_i^2$$

Por otro lado, con las notaciones del tema 3, se verifica que

$$Q_1 = \mathbf{Y}'(\mathbf{H} - \mathbf{H}_2)\mathbf{Y} = \mathbf{Y}'(\mathbf{H} - \mathbf{I}_n + \mathbf{I}_n - \mathbf{H}_2)\mathbf{Y} = \mathbf{Y}'(\mathbf{I}_n - \mathbf{H}_2)\mathbf{Y} - \mathbf{Y}'(\mathbf{I}_n - \mathbf{H})\mathbf{Y} = Q'_0 - Q_0$$

siendo Q'_0 la suma de cuadrados residual en el modelo reducido. Se verifica que (queda propuesto como ejercicio):

$$Q'_0 = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}^2 - N\bar{Y}_{..}^2$$

y por tanto

$$Q_1 = Q'_0 - Q_0 = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}^2 - N\bar{Y}_{..}^2 - \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}^2 + \sum_{i=1}^k n_i \bar{Y}_i^2 = \sum_{i=1}^k n_i \bar{Y}_i^2 - N\bar{Y}_{..}^2$$

3er paso: Estadístico F

A partir de las formas cuadráticas calculadas anteriormente se obtiene el estadístico de contraste

$$F = \frac{Q_1/(k-1)}{Q_0/(N-k)} = \frac{N-k}{k-1} \frac{Q_1}{Q_0}$$

cuya distribución, asumiendo H_0 , es $F \sim F(k-1, N-k)$, de modo que rechazaremos H_0 a un nivel α si $F \geq F_{k-1, N-k, \alpha}$. El cálculo del estadístico F se resume en la siguiente tabla de análisis de la varianza:

Fuentes de Variación	grados de libertad	Suma de cuadrados	Media Cuadrática	F
Debida a H_0	$k - 1$	$Q_1 = \sum_{i=1}^k n_i \bar{Y}_i^2 - N\bar{Y}_{..}^2$	$\frac{Q_1}{k-1}$	$F = \frac{(N-k)Q_1}{(k-1)Q_0}$
Error	$N - k$	$Q_0 = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}^2 - \sum_{i=1}^k n_i \bar{Y}_i^2$	$\frac{Q_0}{N-k}$	

Ejercicio propuesto: Calcular la suma de cuadrados residual, Q'_0 del modelo reducido por H_0 .

Comparaciones Múltiples.

Cuando el test de hipótesis estudiado anteriormente nos lleva a rechazar H_0 , hemos de investigar qué niveles del factor presentan efectos diferenciados. Para ello se utilizan los procedimientos de comparaciones múltiples, que consisten en contrastar, de modo simultáneo, las $\binom{k}{2}$ hipótesis

$$H_0^{(i_1, i_2)} : \tau_{i_1} = \tau_{i_2} \quad , \quad i_1, i_2 = 1, \dots, k, \quad i_1 < i_2.$$

Aunque existen multitud de procedimientos de comparaciones múltiples, solo estudiaremos algunos de ellos:

Método LSD de Fisher. Se basa en los intervalos de confianza al nivel de confianza $1 - \alpha$ para los contrastes $\tau_{i_1} - \tau_{i_2}$, $i_1, i_2 = 1, \dots, k$, $i_1 < i_2$. Particularizando los intervalos que vimos anteriormente para los contrastes, obtenemos la expresión

$$I_{(i_1, i_2)} = \left[\bar{Y}_{i_1} - \bar{Y}_{i_2} - \widehat{\sigma} \sqrt{\frac{1}{n_{i_1}} + \frac{1}{n_{i_2}}} t_{N-k, \alpha/2}, \bar{Y}_{i_1} - \bar{Y}_{i_2} + \widehat{\sigma} \sqrt{\frac{1}{n_{i_1}} + \frac{1}{n_{i_2}}} t_{N-k, \alpha/2} \right]$$

El método LSD (Least Significant Difference) propone la familia de test

$$\Phi_{i_1, i_2}(\mathbf{Y}) = \begin{cases} 1 & \text{si } 0 \notin I_{(i_1, i_2)} \\ 0 & \text{si } 0 \in I_{(i_1, i_2)} \end{cases} \quad i_1, i_2 = 1, \dots, k, \quad i_1 < i_2$$

es decir, se rechaza $H_0^{(i_1, i_2)}$ si $0 \notin I_{(i_1, i_2)}$. Como ya vimos en los temas anteriores, cada uno de estos test tiene extensión α .

Sin embargo, al realizar todas las comparaciones simultáneamente, la probabilidad de que el método cometa un error de tipo I puede aumentar. En efecto, asumiendo $\tau_{i_1} - \tau_{i_2} = 0$, $i_1, i_2 = 1, \dots, k$, $i_1 < i_2$:

$$P(\cup_{i_1 < i_2} \{\Phi_{i_1, i_2} = 1\}) \leq \sum_{i_1 < i_2} P_{\tau_{i_1} = \tau_{i_2}}(\Phi_{i_1, i_2} = 1) = \sum_{i_1 < i_2} P_{\tau_{i_1} = \tau_{i_2}}(0 \notin I_{(i_1, i_2)}) = \binom{k}{2} \alpha$$

Por tanto, la probabilidad de error puede llegar a ser de $\binom{k}{2} \alpha$ y por tanto, mucho mayor que α .

Para solucionar este problema, los siguientes métodos proponen calcular intervalos de confianza simultáneos a un nivel dado $1 - \alpha$ para los parámetros $\tau_{i_1} - \tau_{i_2}$ con $i_1, i_2 = 1, \dots, k$, $i_1 < i_2$, es decir, construir intervalos de tal manera que la probabilidad de que todos los intervalos contengan a sus respectivos parámetros simultáneamente sea de $1 - \alpha$. Veamos la definición:

Definición 2 La familia de intervalos de $\{I_j\}_{j \in J}$ se dice que es una familia de intervalos de confianza simultáneos, al nivel de confianza $1 - \alpha$, para la familia de parámetros $\{\theta_j\}_{j \in J}$ si

$$P(\theta_j \in I_j \text{ para todo } j \in J) \geq 1 - \alpha.$$

Sea $\{I_{(i_1, i_2)} : i_1, i_2 = 1, \dots, k, i_1 < i_2\}$ una familia de intervalos de confianza simultáneos, al nivel de confianza $1 - \alpha$, para la familia de parámetros $\{\tau_{i_1} - \tau_{i_2} : i_1, i_2 = 1, \dots, k, i_1 < i_2\}$, y construimos la familia de test

$\{\Phi_{i_1, i_2} : i_1, i_2 = 1, \dots, k, i_1 < i_2\}$ del mismo modo que lo hemos hecho en el método LSD. Asumiendo de nuevo $\tau_{i_1} - \tau_{i_2} = 0, i_1, i_2 = 1, \dots, k, i_1 < i_2$, la probabilidad de que el procedimiento basado en dichos intervalos cometa un error de tipo I será:

$$P(\cup_{i_1 < i_2} \{\Phi_{i_1, i_2} = 1\}) = 1 - P(0 \in I_{(i_1, i_2)} \text{ para todo } i_1, i_2 = 1, \dots, k, i_1 < i_2) \leq \alpha$$

Por tanto, los procedimientos basados en intervalos de confianza simultáneos sí garantizan que la probabilidad máxima de error en el procedimiento es el nivel de significación α prefijado.

Método de Bonferroni. Se basa en el siguiente resultado.

Proposición 4 Si $\delta = \alpha / \binom{k}{2}$, entonces la familia de intervalos

$$I_{(i_1, i_2)} = \left[\bar{Y}_{i_1 \cdot} - \bar{Y}_{i_2 \cdot} - \widehat{\sigma} \sqrt{\frac{1}{n_{i_1}} + \frac{1}{n_{i_2}}} t_{N-k, \delta/2}, \bar{Y}_{i_1 \cdot} - \bar{Y}_{i_2 \cdot} + \widehat{\sigma} \sqrt{\frac{1}{n_{i_1}} + \frac{1}{n_{i_2}}} t_{N-k, \delta/2} \right]$$

es una familia de intervalos de confianza simultáneos, al nivel de confianza $1 - \alpha$ para la familia de parámetros $\{\tau_{i_1} - \tau_{i_2} : i_1, i_2 = 1, \dots, k, i_1 < i_2\}$.

Demostración. Queda como ejercicio.

En base a esta familia de intervalos, el método de Bonferroni propone rechazar la hipótesis $H_0^{(i_1, i_2)} : \tau_{i_1} = \tau_{i_2}$ a un nivel de significación α si $0 \notin I_{(i_1, i_2)}$.

Método de Tukey. Este método se basa en la distribución de probabilidad de Tukey, que se define del siguiente modo:

Definición 3 Si Z_1, \dots, Z_k y U v.a. independientes con distribuciones: $Z_i \sim N(0, 1), i = 1, \dots, k$ y $U \sim \chi^2(s)$, entonces a la distribución de la variable aleatoria

$$q = \max_{i < j} \frac{|Z_i - Z_j|}{\sqrt{U/s}}.$$

la llamaremos distribución de Tukey de parámetros k y s y la denotaremos por $q(k, s)$.

Proposición 5 Si el modelo es balanceado y denotamos $n = n_1 = \dots = n_k$, entonces el estadístico

$$q = \max_{i_1 < i_2} \frac{|\bar{Y}_{i_1 \cdot} - \bar{Y}_{i_2 \cdot} - (\tau_{i_1} - \tau_{i_2})|}{\widehat{\sigma} / \sqrt{n}}$$

tiene distribución de Tukey de parámetros k y $N - k$.

En consecuencia, si denotamos $q_{k, N-k, \alpha}$ al cuantil de orden α de la distribución de Tukey de parámetros k y $N - k$, es fácil probar que la familia de intervalos

$$I_{(i_1, i_2)} = \left[\bar{Y}_{i_1 \cdot} - \bar{Y}_{i_2 \cdot} - \frac{\widehat{\sigma}}{\sqrt{n}} q_{k, N-k, \alpha}, \bar{Y}_{i_1 \cdot} - \bar{Y}_{i_2 \cdot} + \frac{\widehat{\sigma}}{\sqrt{n}} q_{k, N-k, \alpha} \right] \quad i_1, i_2 = 1, \dots, k$$

es una familia de intervalos de confianza simultáneos, al nivel de confianza $1 - \alpha$ para la familia de parámetros $\{\tau_{i_1} - \tau_{i_2} : i_1, i_2 = 1, \dots, k, i_1 < i_2\}$.

Demostración: Sabemos que $\bar{Y}_i \sim N(\mu + \tau_i, \sigma^2/n)$ y tipificando,

$$Z_i = \frac{\bar{Y}_i - \mu - \tau_i}{\sigma/\sqrt{n}} \sim N(0, 1), \quad i = 1, \dots, k$$

Además como $\bar{Y}_1, \dots, \bar{Y}_k$ son independientes, también lo son Z_1, \dots, Z_k . Por otro lado, también sabemos que

$$U = \frac{(N-k)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2(N-k)$$

y además, como $\widehat{\sigma}^2$ es independiente de $\bar{Y}_1, \dots, \bar{Y}_k$, tenemos que U es independiente de Z_1, \dots, Z_k . Estamos por tanto en condiciones de aplicar la Definición 1 y obtenemos que

$$q = \max_{i_1 < i_2} \frac{|Z_{i_1} - Z_{i_2}|}{\sqrt{U/(N-k)}} = \frac{|Y_{i_1 \cdot} - \bar{Y}_{i_2 \cdot} - (\tau_{i_1} - \tau_{i_2})|}{\frac{\sigma/\sqrt{n}}{\sqrt{\frac{(N-k)\widehat{\sigma}^2}{(N-k)\sigma^2}}}} = \max_{i_1 < i_2} \frac{|Y_{i_1 \cdot} - \bar{Y}_{i_2 \cdot} - (\tau_{i_1} - \tau_{i_2})|}{\widehat{\sigma}/\sqrt{n}}$$

tiene distribución de Tukey de parámetros k y $N-k$.

Para demostrar la segunda parte, sabemos que

$$\begin{aligned} 1 - \alpha &= P(q \leq q_{k, N-k, \alpha}) = P\left(|Y_{i_1 \cdot} - \bar{Y}_{i_2 \cdot} - (\tau_{i_1} - \tau_{i_2})| \leq \frac{\widehat{\sigma}}{\sqrt{n}} q_{k, N-k, \alpha} \text{ para todo } i_1, i_2 = 1, \dots, k, i_1 < i_2\right) \\ &= P\left(-\frac{\widehat{\sigma}}{\sqrt{n}} q_{k, N-k, \alpha} \leq Y_{i_1 \cdot} - \bar{Y}_{i_2 \cdot} - (\tau_{i_1} - \tau_{i_2}) \leq \frac{\widehat{\sigma}}{\sqrt{n}} q_{k, N-k, \alpha} \text{ para todo } i_1, i_2 = 1, \dots, k, i_1 < i_2\right) \\ &= P(\tau_{i_1} - \tau_{i_2} \in I_{(i_1, i_2)}) \text{ para todo } i_1, i_2 = 1, \dots, k, i_1 < i_2 \end{aligned}$$

lo cual concluye la demostración.

En base a este resultado, el método de Tukey propone rechazar $H_0^{(i_1, i_2)} : \tau_{i_1} = \tau_{i_2}$ al nivel de significación α si $0 \notin I_{(i_1, i_2)}$.

Método de Scheffé. Proporciona intervalos de confianza simultáneos al nivel $1 - \alpha$ para todos los contrastes $\sum_{i=1}^k c_i \tau_i$. Concretamente, si $\mathbf{c} = (c_1, \dots, c_k)^t \in \mathbb{R}^k$ verificando $\sum_{i=1}^k c_i = 0$, los intervalos

$$I_{\mathbf{c}} = \left[\sum_{i=1}^k c_i Y_i - \widehat{\sigma} \sqrt{(k-1)F_{k-1, N-k, \alpha} \sum_{i=1}^k (c_i^2/n_i)}, \sum_{i=1}^k c_i Y_i + \widehat{\sigma} \sqrt{(k-1)F_{k-1, N-k, \alpha} \sum_{i=1}^k (c_i^2/n_i)} \right]$$

constituyen una familia de intervalos simultáneos al nivel $1 - \alpha$ para el conjunto de todos los contrastes (**no probaremos este enunciado ni lo dejaremos como ejercicio**). En particular para los intervalos para los contrastes de la forma $\tau_{i_1} - \tau_{i_2}$ tienen la siguiente expresión:

$$I_{(i_1, i_2)} = \left[Y_{i_1 \cdot} - \bar{Y}_{i_2 \cdot} \pm \widehat{\sigma} \sqrt{(k-1)F_{k-1, N-k, \alpha} \left(\frac{1}{n_{i_1}} + \frac{1}{n_{i_2}} \right)} \right], \quad i_1, i_2 = 1, \dots, k, \quad i_1 < i_2$$

El método de Scheffé propone rechazar $H_0^{(i_1, i_2)} : \tau_{i_1} = \tau_{i_2}$ al nivel de significación α si $0 \notin I_{(i_1, i_2)}$.

Ejercicio propuesto. Demostrar la Proposición 4 (Indicación: utilizar la desigualdad de Bonferroni, según la cual $P(\cap_n A_n) \geq 1 - \sum_n P(\bar{A}_n)$).

4.2.2 Diseño de bloques al azar.

Llamaremos *diseño de bloques al azar* a aquel en el que las unidades experimentales están divididas en grupos, denominados *bloques* y la asignación de los niveles del factor a cada una de las unidades experimentales de que consta cada bloque se hace totalmente al azar. Además, dentro de cada bloque se aplicarán todos los niveles del factor.

Ejemplo. Supongamos que queremos comparar 4 catalizadores para el petróleo crudo. Una forma de diseñar el experimento es tomar al azar 5 barriles de petróleo crudo, cada uno de ellos dividirlo en 4 partes y asignar un catalizador a cada una de esas partes para cada barril. La asignación de los catalizadores a las partes de cada barril la haremos al azar, utilizándose en cada barril los 4 catalizadores.

El Modelo Lineal que se corresponde con esta definición es:

$$Y_{ij} = \mu + \tau_i + \delta_j + \mathcal{E}_{ij}, \quad i = 1, \dots, r; \quad j = 1, \dots, s$$

donde μ y τ_i , $i = 1, \dots, r$, y δ_j , $j = 1, \dots, s$ son parámetros desconocidos.

En este modelo, los vectores de observaciones y parámetros son, respectivamente, $\mathbf{Y} = (Y_{11}, \dots, Y_{rs})'$ y $\boldsymbol{\beta} = (\mu, \tau_1, \dots, \tau_r, \delta_1, \dots, \delta_s)'$. Por tanto $p = r + s + 1$, y la matriz del modelo es:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 1 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 \end{pmatrix} \quad \mathbf{X}'\mathbf{Y} = \begin{pmatrix} Y_{..} \\ Y_{1.} \\ \vdots \\ Y_{r.} \\ Y_{.1} \\ \dots \\ Y_{.s} \end{pmatrix}$$

siendo $Y_{i.} = \sum_{j=1}^s Y_{ij}$, $i = 1, \dots, r$, $Y_{.j} = \sum_{i=1}^r Y_{ij}$, $j = 1, \dots, s$, e $Y_{..} = \sum_{i=1}^r \sum_{j=1}^s Y_{ij}$. La matriz \mathbf{X} tiene rs filas. En las columnas de la 2 a la $r + 1$ hay s unos. Las columnas de la $r + 2$ a la $r + s + 1$ (las últimas s columnas) están formadas por r copias apiladas de la matriz \mathbf{I}_s . Se verifica que $r(\mathbf{X}) = r + s - 1 < p = r + s + 1$, por tanto es un **modelo de rango no completo**.

Asumiremos, en general, que $\mathcal{E} \sim \mathcal{N}_{rs}(\mathbf{0}, \sigma^2 \mathbf{I}_{rs})$, aunque para algún resultado la normalidad no sea necesaria.

Ecuaciones Normales del Modelo. Teniendo en cuenta que para $i = 1, \dots, r$; $j = 1, \dots, s$,

$$E[Y_{..}] = rs\mu + \sum_{k=1}^r s\tau_k + \sum_{l=1}^s r\delta_l, \quad E[Y_{i.}] = s\mu + s\tau_i + \sum_{l=1}^s \delta_l, \quad E[Y_{.j}] = r\mu + \sum_{k=1}^r \tau_k + r\delta_j,$$

obtenemos que las ecuaciones normales de este modelo ($\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}$) son:

$$\begin{aligned} rs\widehat{\mu} + \sum_{k=1}^r s\widehat{\tau}_k + \sum_{l=1}^s r\widehat{\delta}_l &= Y_{..} \\ \widehat{\mu} + \widehat{\tau}_i + \sum_{l=1}^s \widehat{\delta}_l &= Y_{i.}, \quad i = 1, \dots, r \\ \widehat{\mu} + \sum_{k=1}^r \widehat{\tau}_k + r\widehat{\delta}_j &= Y_{.j}, \quad j = 1, \dots, s, \end{aligned}$$

Imponiendo las condiciones $\sum_{k=1}^r \widehat{\tau}_k = 0$, $\sum_{l=1}^s \widehat{\delta}_l = 0$, se obtiene la siguiente solución particular de las ecuaciones normales:

$$\widehat{\boldsymbol{\beta}} = (\bar{Y}_{..}, \bar{Y}_{1.} - \bar{Y}_{..}, \dots, \bar{Y}_{r.} - \bar{Y}_{..}, \bar{Y}_{.1} - \bar{Y}_{..}, \dots, \bar{Y}_{.s} - \bar{Y}_{..})'$$

donde $\bar{Y}_{..} = Y_{..}/rs$, $\bar{Y}_{i.} = Y_{i.}/s$, $i = 1, \dots, r$, $\bar{Y}_{.j} = Y_{.j}/r$, $j = 1, \dots, s$. Estas variables aleatorias tienen interesantes propiedades que exponemos en la siguiente proposición:

Proposición 6

- a) $\text{Var}[\bar{Y}_{i.}] = \sigma^2/s$, $\text{Var}[\bar{Y}_{.j}] = \sigma^2/r$, $i = 1, \dots, r$, $j = 1, \dots, s$.
- b) $\text{Cov}[\bar{Y}_{i_1.}, \bar{Y}_{i_2.}] = 0$, $i_1, i_2 = 1, \dots, r$, $i_1 \neq i_2$.
- c) $\text{Cov}[\bar{Y}_{.j_1}, \bar{Y}_{.j_2}] = 0$, $j_1, j_2 = 1, \dots, s$, $j_1 \neq j_2$.

Demostración.

Apartado a) Puesto que los errores $\boldsymbol{\mathcal{E}} \sim \mathcal{N}_{rs}(\mathbf{0}, \sigma^2 \mathbf{I}_{rs})$ se verifica que tanto errores como observaciones son incorrelados y como hay normalidad son independientes, es decir, Y_{11}, \dots, Y_{rs} son independientes con varianza común σ^2 . Por tanto

$$\text{Var}[\bar{Y}_{i.}] = \frac{1}{s^2} \sum_{j=1}^s \text{Var}[Y_{ij}] = \frac{s\sigma^2}{s^2} = \frac{\sigma^2}{s}$$

Análogamente se prueba que $\text{Var}[\bar{Y}_{.j}] = \sigma^2/r$.

Apartado b) Las medias $\bar{Y}_{i_1.}$ e $\bar{Y}_{i_2.}$ son función de $Y_{i_11}, \dots, Y_{i_1s}$ e $Y_{i_21}, \dots, Y_{i_2s}$, respectivamente y como estos grupos de variables son independientes, también lo son las medias.

Apartado c) Se hace igual que el apartado b).

Estimación puntual.

Las filas de la matriz \mathbf{X} nos proporcionan un conjunto de rs funciones lineales estimables (f.l.e.), $\{\mu + \tau_i + \delta_j, i = 1, \dots, r, j = 1, \dots, s\}$. Cualquier otra f.l.e. es combinación lineal de las mismas. Sin embargo, no son linealmente independientes, ya que el máximo número de f.l.e. linealmente independientes es $r + s - 1$. Podemos extraer el siguiente conjunto de $r + s - 1$ f.l.e. que sí son linealmente independientes:

$$\{\mu + \tau_1 + \delta_1, \mu + \tau_1 + \delta_2, \dots, \mu + \tau_1 + \delta_s, \mu + \tau_2 + \delta_s, \dots, \mu + \tau_r + \delta_s\}$$

Tomando la solución de las ecuaciones normales calculada anteriormente, se deduce que el estimador insesgado de mínima varianza de $\mu + \tau_i + \delta_j$ $i = 1, \dots, r$, $j = 1, \dots, s$ es

$$\mu + \widehat{\tau}_i + \widehat{\delta}_j = \widehat{\mu} + \widehat{\tau}_i + \widehat{\delta}_i = \bar{Y}_{..} + \bar{Y}_{i.} - \bar{Y}_{..} + \bar{Y}_{.j} - \bar{Y}_{..} = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..}$$

Sin embargo, para el estudio de este modelo son más interesantes otras f.l.e., los contrastes:

Definición 4 Un contraste de los parámetros τ_1, \dots, τ_r es una función $\psi = \sum_{i=1}^r c_i \tau_i$ tal que $\sum_{i=1}^r c_i = 0$. Un contraste de los parámetros $\delta_1, \dots, \delta_s$ es una función $\Phi = \sum_{j=1}^s c_j \delta_j$ tal que $\sum_{j=1}^s c_j = 0$.

Ejemplo. Las funciones $\tau_{i_1} - \tau_{i_2}$ $i_1 \neq i_2$ y $\delta_{j_1} - \delta_{j_2}$ $j_1 \neq j_2$ son ejemplos de contrastes de τ_1, \dots, τ_r y $\delta_1, \dots, \delta_s$, respectivamente.

Proposición 7 $\psi = \sum_{i=1}^r c_i \tau_i$ y $\Phi = \sum_{j=1}^s c_j \delta_j$ son f.l.e. si y sólo si son contrastes.

Demostración. Se deja como ejercicio. Para hacer ambas implicaciones es conveniente utilizar que ψ y Φ son f.l.e. si y sólo si son combinación lineal de las funciones

$$\{\mu + \tau_1 + \delta_1, \mu + \tau_1 + \delta_2, \dots, \mu + \tau_1 + \delta_s, \mu + \tau_2 + \delta_s, \dots, \mu + \tau_r + \delta_s\}.$$

Proposición 8

- a) Si $\psi = \sum_{i=1}^r c_i \tau_i$ es un contraste de τ_1, \dots, τ_r , entonces el estimador insesgado de mínima varianza de ψ es $\widehat{\psi} = \sum_{i=1}^r c_i \bar{Y}_i$ y su varianza $\text{Var}[\widehat{\psi}] = \sigma^2 \sum_{i=1}^r c_i^2 / s$.
- b) Si $\Phi = \sum_{j=1}^s c_j \delta_j$ es un contraste de $\delta_1, \dots, \delta_s$, entonces el estimador insesgado de mínima varianza de Φ es $\widehat{\Phi} = \sum_{j=1}^s c_j \bar{Y}_{.j}$ y su correspondiente varianza $\text{Var}[\widehat{\Phi}] = \sigma^2 \sum_{j=1}^s c_j^2 / r$.

Demostración. Es análoga a la segunda parte de la Proposición 2, que fue propuesta como ejercicio. Como indicación, utilizar la Proposición 6.

Finalizamos este apartado con la estimación de σ^2 .

Proposición 9 El estimador insesgado de σ^2 es

$$\widehat{\sigma}^2 = \frac{1}{(r-1)(s-1)} \left(\sum_{i=1}^r \sum_{j=1}^s Y_{ij}^2 - \sum_{i=1}^r s \bar{Y}_i^2 - \sum_{j=1}^s r \bar{Y}_{.j}^2 + rs \bar{Y}_{..}^2 \right)$$

Demostración. Se deja como ejercicio. Como indicación, utilizar que si $\widehat{\beta}$ es una solución de las ecuaciones normales, entonces:

$$\widehat{\sigma}^2 = \frac{1}{rs - (r + s - 1)} (\mathbf{Y}'\mathbf{Y} - \widehat{\beta}'\mathbf{X}'\mathbf{Y})$$

Ejercicios propuestos. Demostrar las Proposiciones 7 y 9.

Propiedades de los estimadores

Haciendo uso de nuevo de la Proposición 7 del Tema 3, obtenemos las siguientes propiedades para los estimadores anteriores:

- Si $\psi = \sum_{i=1}^r c_i \tau_i$ es un contraste de τ_1, \dots, τ_r , entonces $\widehat{\psi} \sim N(\psi, \sigma^2 \sum_{i=1}^r c_i^2 / s)$.
- Si $\Phi = \sum_{j=1}^s c_j \delta_j$ es un contraste de $\delta_1, \dots, \delta_s$, entonces $\widehat{\Phi} \sim N(\Phi, \sigma^2 \sum_{j=1}^s c_j^2 / r)$.
- $(r-1)(s-1)\widehat{\sigma}^2 / \sigma^2 \sim \chi^2((r-1)(s-1))$.
- Si ψ es un contraste de τ_1, \dots, τ_r , entonces $\widehat{\psi}$ y $\widehat{\sigma}^2$ son independientes.
- Si Φ es un contraste de $\delta_1, \dots, \delta_s$, entonces $\widehat{\Phi}$ y $\widehat{\sigma}^2$ son independientes.

Intervalos de confianza al nivel de confianza $1 - \alpha$

Particularizando los intervalos calculados en el Tema 3 para f.l.e. obtenemos las siguientes expresiones:

- Para un contraste de τ_1, \dots, τ_r , $\sum_{i=1}^r c_i \tau_i$,

$$\left[\sum_{i=1}^r c_i \bar{Y}_{.i} - \widehat{\sigma} \sqrt{\sum_{i=1}^r c_i^2 / s} t_{(r-1)(s-1), \alpha/2}, \sum_{i=1}^r c_i \bar{Y}_{.i} + \widehat{\sigma} \sqrt{\sum_{i=1}^r c_i^2 / s} t_{(r-1)(s-1), \alpha/2} \right]$$

- Para un contraste de $\delta_1, \dots, \delta_s$, $\sum_{j=1}^s c_j \delta_j$,

$$\left[\sum_{j=1}^s c_j \bar{Y}_{.j} - \widehat{\sigma} \sqrt{\sum_{j=1}^s c_j^2 / r} t_{(r-1)(s-1), \alpha/2}, \sum_{j=1}^s c_j \bar{Y}_{.j} + \widehat{\sigma} \sqrt{\sum_{j=1}^s c_j^2 / r} t_{(r-1)(s-1), \alpha/2} \right]$$

siendo $t_{(r-1)(s-1), \alpha/2}$ el cuantil por la derecha de orden $\alpha/2$ de una distribución $t((r-1)(s-1))$.

Contraste de hipótesis

En este modelo el contraste más importante a realizar es

$$H_0 : \tau_1 = \dots = \tau_r$$

es decir, se contrasta la influencia (si rechazamos H_0) del factor en la variable respuesta. Desde el punto de vista matemático, también se puede contrastar

$$H'_0 : \delta_1 = \dots = \delta_s$$

Este contraste carece de interés desde el punto de vista aplicado, ya que investiga una posible influencia de los bloques en la variable respuesta. Además, a la hora de interpretarlo, hay que tener un poco de cuidado, ya

que en la división en bloques no hay la componente de aleatoriedad requerida, por lo que puede llevarnos a resultados engañosos.

En el ejemplo propuesto al principio de este apartado, el contraste H_0 nos permite detectar diferencias entre los catalizadores, que es el objetivo del experimento. Por su parte, el contraste de H'_0 nos permitiría detectar diferencias entre los distintos barriles, que seguramente han sido elegidos con distinta procedencia para incorporarlos al experimento como bloques.

Ambas hipótesis son estimables puesto que se pueden expresar como $H_0 : \tau_1 - \tau_r = \dots = \tau_{r-1} - \tau_r = 0$ y $H'_0 : \delta_1 - \delta_s = \dots = \delta_{s-1} - \delta_s = 0$ y las funciones $\tau_i - \tau_r, i = 1, \dots, r-1$ y $\delta_j - \delta_s, j = 1, \dots, s-1$ son contrastes y, aplicando la Proposición 7, funciones lineales estimables linealmente independientes.

Consideramos en primer lugar el contraste $H_0 : \tau_1 = \dots = \tau_r$. Para calcular el test de hipótesis debemos seguir los pasos indicados en el apartado 3.3 del Tema 3.

1er paso: modelo reducido por H_0 .

El modelo reducido por H_0 es

$$Y_{ij} = \mu + \tau_1 + \delta_j + \mathcal{E}_{ij} = \mu^* + \delta_j + \mathcal{E}_{ij}, \quad i = 1, \dots, r; \quad j = 1, \dots, s$$

siendo $\mu^* = \mu + \tau_1$. No hace falta reparametrizar este modelo porque se corresponde con un diseño completamente donde los bloques actuarían como factor. Según lo visto en el apartado anterior, la suma de cuadrados de residuos de este modelo, a la que en este caso denotaremos Q'_0 sería:

$$Q'_0 = \sum_{i=1}^s \sum_{j=1}^s Y_{ij}^2 - \sum_{j=1}^s r \bar{Y}_{.j}^2$$

Para entender esta expresión, hay que tener en cuenta que tenemos s bloques y por cada bloque hay r observaciones, una por cada nivel del factor. La media de las observaciones para el bloque j es $\bar{Y}_{.j}$

El 2º paso (**formas cuadráticas asociadas al test**) y el 3er paso (**estadístico F**) quedan como ejercicio.

Análogamente se realizarían los cálculos para contrastar H'_0 (también queda. Los resultados aparecen resumidos en la siguiente tabla de análisis de la varianza:

F.V.	g.l.	Suma de cuadrados	M.C.	F
Tr. (τ)	$r - 1$	$Q_1 = \sum_{i=1}^r s \bar{Y}_i^2 - r s \bar{Y}_{..}^2$	$\frac{Q_1}{r - 1}$	$F_T = \frac{(s - 1) Q_1}{Q_0}$
Bl. (δ)	$s - 1$	$Q'_1 = \sum_{j=1}^s r \bar{Y}_{.j}^2 - r s \bar{Y}_{..}^2$	$\frac{Q'_1}{s - 1}$	$F_B = \frac{(r - 1) Q'_1}{Q_0}$
Error	$(r - 1)(s - 1)$	$Q_0 = \sum_{i=1}^r \sum_{j=1}^s Y_{ij}^2 - \sum_{i=1}^r s \bar{Y}_i^2 - \sum_{j=1}^s r \bar{Y}_{.j}^2 + r s \bar{Y}_{..}^2$	$\frac{Q_0}{(r - 1)(s - 1)}$	

Ejercicio propuesto: Deducir los test de hipótesis para contrastar las hipótesis H_0 y H'_0 . Como indicación, seguir los pasos que se vieron en el Diseño Completamente Aleatorizado, página 7.

Comparaciones Múltiples.

Al igual que ocurría en el diseño completamente aleatorizado, cuando los test de hipótesis estudiados anteriormente nos llevan a rechazar H_0 o H'_0 , hemos de investigar qué niveles del factor o qué bloques presentan efectos diferenciados. Para ello se utilizan los procedimientos de comparaciones múltiples, que consisten en contrastar, de modo simultáneo, las $\binom{r}{2}$ hipótesis

$$H_0^{(i_1, i_2)} : \tau_{i_1} = \tau_{i_2} \quad , \quad i_1, i_2 = 1, \dots, r, \quad i_1 < i_2$$

o, aunque esto tiene menos sentido como hemos visto antes, las $\binom{s}{2}$ hipótesis

$$H_0^{(j_1, j_2)} : \delta_{j_1} = \delta_{j_2} \quad , \quad j_1, j_2 = 1, \dots, s, \quad j_1 < j_2.$$

Sólo veremos aquí comparaciones múltiples para los parámetros τ_1, \dots, τ_r . Las comparaciones múltiples para los parámetros $\delta_1, \dots, \delta_s$ se realizan de modo análogo. Los métodos que veremos, excepto el método LSD de Fisher, se basan en intervalos de confianza simultáneos a nivel $1 - \alpha$ ($0 < \alpha < 1$)

Método LSD de Fisher. Se basa en los intervalos de confianza al nivel de confianza $1 - \alpha$ para los contrastes $\tau_{i_1} - \tau_{i_2}$, $i_1, i_2 = 1, \dots, r$, $i_1 < i_2$. Particularizando los intervalos que vimos anteriormente para los contrastes, obtenemos la expresión

$$I_{(i_1, i_2)} \left[\bar{Y}_{i_1} - \bar{Y}_{i_2} - \widehat{\sigma} \sqrt{\frac{2}{s}} t_{(r-1)(s-1), \alpha/2}, \bar{Y}_{i_1} - \bar{Y}_{i_2} + \widehat{\sigma} \sqrt{\frac{2}{s}} t_{(r-1)(s-1), \alpha/2} \right]$$

El método LSD de Fisher propone rechazar $H_0^{(i_1, i_2)} : \tau_{i_1} = \tau_{i_2}$ al nivel de significación α si $0 \notin I_{(i_1, i_2)}$.

Método de Bonferroni. Si tomamos $\gamma = \alpha / \binom{r}{2}$, la familia de intervalos $\{I_{(i_1, i_2)}, i_1, i_2 = 1, \dots, r, i_1 < i_2\}$ con

$$I_{(i_1, i_2)} = \left[\bar{Y}_{i_1} - \bar{Y}_{i_2} - \widehat{\sigma} \sqrt{\frac{2}{s}} t_{(r-1)(s-1), \gamma/2}, \bar{Y}_{i_1} - \bar{Y}_{i_2} + \widehat{\sigma} \sqrt{\frac{2}{s}} t_{(r-1)(s-1), \gamma/2} \right]$$

es una familia de intervalos simultáneos al nivel de confianza $1 - \alpha$ para los parámetros $\{\tau_{i_1} - \tau_{i_2}, i_1, i_2 = 1, \dots, r, i_1 < i_2\}$. En base a esta familia, el método de Bonferroni propone rechazar $H_0^{(i_1, i_2)} : \tau_{i_1} = \tau_{i_2}$ al nivel de significación α si $0 \notin I_{(i_1, i_2)}$.

Método de Tukey. Se utiliza el siguiente resultado:

Proposición 10 *El estadístico*

$$q = \max_{i_1 < i_2} \frac{|Y_{i_1 \cdot} - \bar{Y}_{i_2 \cdot} - (\tau_{i_1} - \tau_{i_2})|}{\widehat{\sigma} / \sqrt{s}}$$

tiene distribución de Tukey de parámetros r y $(r - 1)(s - 1)$.

Demostración: Queda como ejercicio. Hay que seguir los mismos pasos que en la demostración de la Proposición 5.

Como consecuencia del resultado anterior, si denotamos $q_{r,(r-1)(s-1),\alpha}$ al cuantil por la derecha de orden α de la distribución de Tukey de parámetros r y $(r-1)(s-1)$, es fácil probar que la familia de intervalos

$$I_{(i_1, i_2)} = \left[\bar{Y}_{i_1} - \bar{Y}_{i_2} - \frac{\hat{\sigma}}{\sqrt{s}} q_{r,(r-1)(s-1),\alpha}, \bar{Y}_{i_1} - \bar{Y}_{i_2} + \frac{\hat{\sigma}}{\sqrt{s}} q_{r,(r-1)(s-1),\alpha} \right] \quad i_1, i_2 = 1, \dots, r \quad i_1 < i_2$$

es una familia de intervalos de confianza simultáneos, al nivel de confianza $1 - \alpha$ para la familia de parámetros $\{\tau_{i_1} - \tau_{i_2} : i_1, i_2 = 1, \dots, r, i_1 < i_2\}$. En base a esta familia, el método de Tukey propone rechazar $H_0^{(i,j)} : \tau_i = \tau_j$ al nivel de significación α si $0 \notin I_{(i_1, i_2)}$.

Método de Scheffé. Si $F_{r-1,(r-1)(s-1),\alpha}$ es el cuantil por la derecha de orden α de una distribución $F(r-1, (r-1)(s-1))$, entonces la familia de intervalos $\{I_{(i_1, i_2)}, i_1, i_2 = 1, \dots, r, i_1 < i_2\}$ con

$$I_{(i_1, i_2)} = \left[\bar{Y}_{i_1} - \bar{Y}_{i_2} - \hat{\sigma} \sqrt{\frac{2(r-1)}{s} F_{r-1,(r-1)(s-1),\alpha}}, \bar{Y}_{i_1} - \bar{Y}_{i_2} + \hat{\sigma} \sqrt{\frac{2(r-1)}{s} F_{r-1,(r-1)(s-1),\alpha}} \right]$$

es una familia de intervalos simultáneos al nivel de confianza $1 - \alpha$ para los parámetros $\{\tau_{i_1} - \tau_{i_2}, i_1, i_2 = 1, \dots, r, i_1 < i_2\}$. En base a esta familia, el método de Scheffé propone rechazar $H_0^{(i,j)} : \tau_i = \tau_j$ al nivel de significación α si $0 \notin I_{(i_1, i_2)}$.

Ejercicio propuesto. Demostrar la Proposición 10.

4.3 Experimentos con dos factores

En este apartado se consideran experimentos con dos factores de efectos fijos, que serán tratados como variables cualitativas y una variable cuantitativa continua, la variable respuesta. El experimentador introduce cambios simultáneos en los valores de ambos factores y posteriormente observa cómo dichos cambios afectan a la variable respuesta en varias repeticiones del experimento. Vamos a desarrollar modelos que nos permitan determinar si los dos factores tienen influencia en la variable respuesta. De acuerdo a la disposición de los dos factores en el diseño del experimento podemos distinguir dos tipos de modelos:

Modelos de Clasificación cruzada: Suponiendo que el primer factor tiene r niveles y el segundo factor tiene s niveles, se consideran todas las combinaciones posibles de los r niveles del primer factor con los s niveles del segundo factor. Para la combinación del i del primer factor con el nivel j del segundo factor se toman n_{ij} observaciones. Asumiremos $n_{ij} > 0$ para todo $i = 1, \dots, r, j = 1, \dots, s$, en caso contrario se dice que el diseño es incompleto. Si $n_{ij} = n$ para todo $i = 1, \dots, r, j = 1, \dots, s$, se dice que el diseño es **balanceado** y en caso contrario se dice **no balanceado**.

Denotaremos Y_{ijk} a la k -ésima observación de la variable respuesta cuando se ha aplicado el nivel i del primer factor y el nivel j del segundo factor, $i = 1, \dots, r, j = 1, \dots, s, k = 1, \dots, n_{ij}$.

Ejemplo 1. Se investiga si la variedad de trigo y la marca de abono utilizado tienen influencia sobre la producción de trigo (en kg. por metro cuadrado, Y). Se consideran 3 variedades de trigo (V1, V2, V3) y cuatro marcas de abono (A1, A2, A3, A4). Se utiliza cada abono con cada variedad de trigo en tres parcelas. Los datos obtenidos en este experimento se representan en la siguiente tabla:

	A1	A2	A3	A4
V1	Y_{111}	Y_{121}	Y_{131}	Y_{141}
	Y_{112}	Y_{122}	Y_{132}	Y_{142}
	Y_{113}	Y_{123}	Y_{133}	Y_{143}
V2	Y_{211}	Y_{221}	Y_{231}	Y_{241}
	Y_{212}	Y_{222}	Y_{232}	Y_{242}
	Y_{213}	Y_{223}	Y_{233}	Y_{243}
V3	Y_{311}	Y_{321}	Y_{331}	Y_{341}
	Y_{312}	Y_{322}	Y_{332}	Y_{342}
	Y_{313}	Y_{323}	Y_{333}	Y_{343}

En este caso, $r = 3, s = 4$ y $n_{ij} = 3$ para todo $i = 1, 2, 3, j = 1, 2, 3, 4$, es decir, se ha realizado un diseño balanceado. Además, Y_{ijk} denotará en este caso la producción de trigo en la k -ésima parcela donde se ha sembrado la i -ésima variedad de trigo y se ha abonado con la j -ésima marca de abono, $i = 1, 2, 3, j = 1, 2, 3, 4, k = 1, 2, 3$.

Cada combinación de una variedad de trigo con una marca de abono está representada en una celda de la tabla anterior. Estamos suponiendo que tenemos observaciones en todas las celdas. Esta denominación de **celda** se utiliza habitualmente en los modelos de clasificación cruzada para denotar a cada combinación de un nivel del primer factor con un nivel del segundo factor. El modelo que hemos de utilizar en el ejemplo es un modelo de clasificación cruzada, balanceado con 3 observaciones por celda.

Modelos Anidados: En estos modelos, cada nivel del segundo factor está supeditado a un nivel del primer factor. Si el primer factor (factor principal) tiene r niveles, el nivel i de este factor tiene asignados s_i niveles del segundo factor (factor anidado).

Para j -ésimo nivel del segundo factor anidado al nivel i factor principal se toman n_{ij} observaciones. Asumiremos $n_{ij} > 0$ para todo $i = 1, \dots, r, j = 1, \dots, s_i$, en caso contrario se dice que el diseño es incompleto. Si $n_{ij} = n$ para todo $i = 1, \dots, r, j = 1, \dots, s_i$, se dice que el diseño es **balanceado** y en caso contrario se dice **no balanceado**.

Denotaremos Y_{ijk} a la k -ésima observación de la variable respuesta cuando se ha aplicado el nivel j del segundo factor, anidado al nivel i del factor principal, $i = 1, \dots, r, j = 1, \dots, s_i, k = 1, \dots, n_{ij}$.

Ejemplo 2. En un instituto de educación secundaria, los alumnos de 2° de Bachillerato pueden elegir entre 3 optativas: Biología (B), Física (F) y Química (Q). Como hay muchos alumnos en este curso, se forman 3 grupos de Biología, 2 grupos de Física y 2 grupos de Química. Se investiga si la optativa elegida y el grupo al que se adscribió al alumno han influido en su nota de selectividad (variable respuesta). Para ello se seleccionaron al azar 3 alumnos de cada grupo.

Biología	{	B1	$Y_{111}, Y_{112}, Y_{113}$
		B2	$Y_{121}, Y_{122}, Y_{123}$
		B3	$Y_{131}, Y_{132}, Y_{133}$
Física	{	F1	$Y_{211}, Y_{212}, Y_{213}$
		F2	$Y_{221}, Y_{222}, Y_{223}$
Química	{	Q1	$Y_{311}, Y_{312}, Y_{313}$
		Q2	$Y_{321}, Y_{322}, Y_{323}$

En este caso $r = 3, s_1 = 3, s_2 = s_3 = 2$ y $n_{ij} = 3$ para todo i, j , es decir, se ha realizado un diseño balanceado. Además, Y_{ijk} denota la nota de selectividad del k -ésimo alumno elegido en el grupo j de la asignatura i .

4.3.1 Modelos de clasificación cruzada

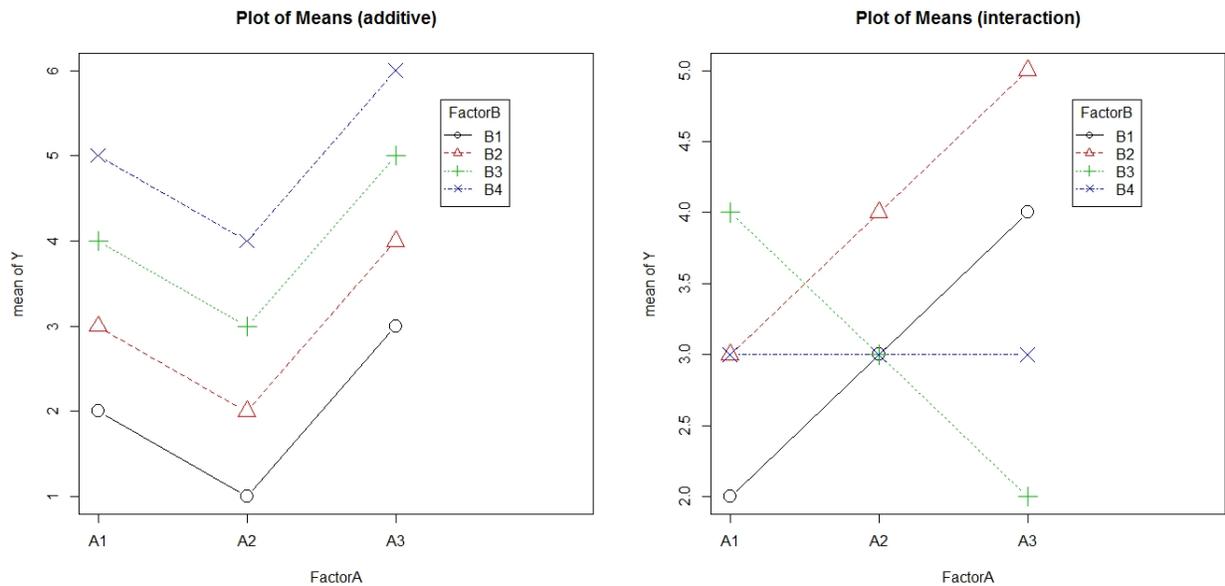
Supongamos que tenemos un primer factor con r niveles y un segundo factor con s niveles. Supondremos que el diseño es balanceado se han tomado n observaciones por celda. Sea Y_{ijk} la variable aleatoria que representa la observación de la variable respuesta en la k -ésima observación para la combinación del nivel i del primer factor con el nivel j del segundo factor. Asumiremos que estas variables son idénticamente distribuidas dentro de una celda, es decir, supondremos que Y_{ij1}, \dots, Y_{ijn} son idénticamente distribuidas y sea $\mu_{ij} = E[Y_{ijk}], k = 1, \dots, n$. El concepto clave en estos modelos es el de *interacción*.

Definición 1 Diremos que un modelo es **aditivo** si

$$\mu_{ij} - \mu_{ij'} = \mu_{i'j} - \mu_{i'j'} \quad \text{para todo } i, i' = 1, \dots, r \quad j, j' = 2, \dots, s.$$

Si no se verifica alguna de esas igualdades se dice que el modelo presenta **interacción**.

Gráficamente:



En el primer gráfico se puede ver como las diferencias entre $\mu_{i1}, \mu_{i2}, \mu_{i3}$ y μ_{i4} se mantienen constantes para todo $i = 1, 2, 3$, dando lugar a líneas paralelas, lo que corresponde con un modelo aditivo.

Sin embargo, en el segundo gráfico se aprecia que aunque las diferencias entre μ_{i1} y μ_{i2} sí se mantienen constantes para $i = 1, 2, 3$, tenemos, por ejemplo, que la media μ_{13} es mayor que μ_{11} , pero la media μ_{23} es igual a μ_{21} y además la media μ_{33} es menor que μ_{31} . El resultado es que μ_{13}, μ_{23} y μ_{33} forman una línea decreciente, mientras que μ_{11}, μ_{21} y μ_{31} forman una línea creciente. Si seguimos observando vemos que μ_{14}, μ_{24} y μ_{34} forman una línea horizontal. Es decir, las diferencias entre las líneas dependen de si estamos en A1, A2 o A3. Esto corresponde a un modelo con interacción.

Desde un punto de vista intuitivo, diremos que existe interacción en el modelo, si la relación que puede haber entre la variable respuesta y uno de los factores puede variar para cada uno de los niveles del otro factor. Si hubiera interacción los factores presentarían un efecto conjunto en la media de la variable respuesta. En caso contrario, los efectos de los factores se sumarían y se diría que el modelo es aditivo.

En el Ejemplo 1, la existencia de interacción significaría que alguno de los abonos produce un mejor (o peor) efecto en la producción de trigo si lo utilizamos en una variedad determinada que si lo utilizamos en las otras. Si el efecto de los abonos en la producción es el mismo con todas las variedades, el modelo es aditivo.

Ejercicio propuesto. Poner un ejemplo, ajustado a la realidad, de experimento de dos factores con un diseño de clasificación cruzada y otro ejemplo con un diseño anidado.

4.3.1.1. Diseño de dos factores sin interacción.

Si contamos con n observaciones por celda y no hay evidencias de interacción, el Modelo Lineal que se corresponde con este diseño es:

$$Y_{ijk} = \mu + \tau_i + \delta_j + \mathcal{E}_{ijk}, \quad i = 1, \dots, r; \quad j = 1, \dots, s; \quad k = 1, \dots, n$$

donde μ , τ_i y δ_j son constantes desconocidas y \mathcal{E}_{ijk} son variables aleatorias tales que $E[\mathcal{E}_{ijk}] = 0$, $i = 1, \dots, r$, $j = 1, \dots, s$, $k = 1, \dots, n$. Por tanto, se verifica que

$$\mu_{ij} = E[Y_{ijk}] = \mu + \tau_i + \delta_j$$

El parámetro τ_i (resp. δ_j) representa el efecto que el nivel i del primer factor (resp. el nivel j del segundo factor) produce en la media de la variable respuesta. El número de observaciones en este modelo es rsn y el número de parámetros $p = r + s + 1$.

Este modelo verifica que

$$\mu_{ij} - \mu_{i'j} = \delta_j - \delta_{j'} = \mu_{ij'} - \mu_{i'j'} \quad \text{para todo } i, i' = 1, \dots, r \quad j, j' = 2, \dots, s.$$

luego, de acuerdo a la Definición 1, es aditivo.

La matriz del modelo, \mathbf{X} , tiene las mismas filas que la matriz del modelo de bloques al azar, aunque cada fila repetida n veces. Por tanto, el rango de ambas matrices es el mismo, $r(\mathbf{X}) = r + s - 1 < p$ y en consecuencia es también un modelo lineal de rango no completo. De hecho, el modelo del diseño de bloques al azar, matemáticamente hablando, se corresponde al caso particular de este modelo, donde hay una única observación por celda, $n = 1$. Por ello, toda la teoría que se expone a continuación es similar a la estudiada en dicho modelo.

Asumiremos, en general, que $\mathcal{E} \sim \mathcal{N}_{rsn}(\mathbf{0}, \sigma^2 \mathbf{I}_{rsn})$, aunque para algún resultado la normalidad no sea necesaria.

Ecuaciones Normales. Es sencillo obtener que

$$\mathbf{X}^t \mathbf{Y} = (Y_{...}, Y_{1..}, \dots, Y_{r..}, Y_{.1.}, \dots, Y_{.s.})^t$$

siendo $Y_{...} = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^n Y_{ijk}$, $Y_{i..} = \sum_{j=1}^s \sum_{k=1}^n Y_{ijk}$, $i = 1, \dots, r$, $Y_{.j.} = \sum_{i=1}^r \sum_{k=1}^n Y_{ijk}$, $j = 1, \dots, s$.

Procediendo como en el diseño de bloques al azar se obtienen las siguientes ecuaciones normales (se deja como ejercicio su desarrollo):

$$\begin{aligned} rsn\widehat{\mu} + \sum_{h=1}^r sn\widehat{\tau}_h + \sum_{l=1}^s rn\widehat{\delta}_l &= Y_{...} \\ ns\widehat{\mu} + sn\widehat{\tau}_i + \sum_{l=1}^s n\widehat{\delta}_l &= Y_{i..}, \quad i = 1, \dots, r \\ nr\widehat{\mu} + \sum_{h=1}^r n\widehat{\tau}_h + rn\widehat{\delta}_j &= Y_{.j.}, \quad j = 1, \dots, s \end{aligned}$$

Una solución particular de estas ecuaciones, con las restricciones $\sum_{i=1}^r \widehat{\tau}_i = 0$, $\sum_{j=1}^s \widehat{\delta}_j = 0$, es

$$\widehat{\boldsymbol{\beta}} = (\bar{Y}_{...}, \bar{Y}_{1..} - \bar{Y}_{...}, \dots, \bar{Y}_{r..} - \bar{Y}_{...}, \bar{Y}_{.1.} - \bar{Y}_{...}, \dots, \bar{Y}_{.s.} - \bar{Y}_{...})^t$$

donde $\bar{Y}_{...} = Y_{...}/rsn$, $\bar{Y}_{i..} = Y_{i..}/sn$, $i = 1, \dots, r$, $\bar{Y}_{.j.} = Y_{.j.}/rn$, $j = 1, \dots, s$.

Estimación puntual e intervalos de confianza.

La teoría de estimación en este modelo es similar a la del diseño de bloques al azar. Comenzamos dando el estimador de σ^2

Estimación de $\widehat{\sigma}^2$: Partiendo de la solución de las ecuaciones normales, podemos obtener el siguiente estimador insesgado de σ^2 :

$$\widehat{\sigma}^2 = \frac{1}{nrs - r - s + 1} \left(\sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^n Y_{ijk}^2 - \sum_{i=1}^r sn\bar{Y}_{i..}^2 - \sum_{j=1}^s rn\bar{Y}_{.j.}^2 + rsn\bar{Y}_{...}^2 \right)$$

Los cálculos para obtener este estimador son los mismos que los realizados en la Proposición 9 del apartado de diseños con un factor (diseño de bloques al azar).

Funciones Lineales Estimables: Las f.l.e. $\{\mu + \tau_i + \delta_j, i = 1, \dots, r, j = 1, \dots, s\}$ generan el espacio de f.l.e., al igual que ocurría en el modelo de bloques al azar. El estimador lineal insesgado de mínima varianza de $\mu + \tau_i + \delta_j$ es $\bar{Y}_{i..} + \bar{Y}_{.j.} - \bar{Y}_{...}$, $i = 1, \dots, r, j = 1, \dots, s$.

Contrastes: Llamaremos *contraste* de los parámetros τ_1, \dots, τ_r a una función $\psi = \sum_{i=1}^r c_i \tau_i$ tal que $\sum_{i=1}^r c_i = 0$. Llamaremos *contraste* de los parámetros $\delta_1, \dots, \delta_s$ a una función $\Phi = \sum_{j=1}^s c_j \delta_j$ tal que $\sum_{j=1}^s c_j = 0$.

La definición anterior es idéntica a la que dimos en el diseño de bloques al azar, por tanto, la teoría relativa a la estimación de estos contrastes, se puede extender con pequeños cambios, como se observa en las siguientes propiedades:

- $\psi = \sum_{i=1}^r c_i \tau_i$ y $\Phi = \sum_{j=1}^s c_j \delta_j$ son f.l.e. si y sólo si son contrastes.
- Si $\psi = \sum_{i=1}^r c_i \tau_i$ es un contraste τ_1, \dots, τ_r , entonces su estimador insesgado de mínima varianza es $\widehat{\psi} = \sum_{i=1}^r c_i \bar{Y}_{i..}$ y se verifica que $\widehat{\psi} \sim N(\psi, \sigma^2 \sum_{i=1}^r c_i^2 / sn)$.
- Si $\Phi = \sum_{j=1}^s c_j \delta_j$ es un contraste de $\delta_1, \dots, \delta_s$, entonces su estimador insesgado de mínima varianza es $\widehat{\Phi} = \sum_{j=1}^s c_j \bar{Y}_{.j.}$ y se verifica que $\widehat{\Phi} \sim N(\Phi, \sigma^2 \sum_{j=1}^s c_j^2 / rn)$.
- $(rsn - r - s + 1)\widehat{\sigma}^2 / \sigma^2 \sim \chi^2(rs n - r - s + 1)$.
- Si ψ es un contraste de τ_1, \dots, τ_r , entonces $\widehat{\psi}$ y $\widehat{\sigma}^2$ son independientes.
- Si Φ es un contraste de $\delta_1, \dots, \delta_s$, entonces $\widehat{\Phi}$ y $\widehat{\sigma}^2$ son independientes.

Aplicando las propiedades anteriores, obtenemos los siguientes intervalos de confianza al nivel $1 - \alpha$:

- Para un contraste $\sum_{i=1}^r c_i \tau_i$,

$$\left[\sum_{i=1}^r c_i \bar{Y}_{i..} \pm \widehat{\sigma} \sqrt{\sum_{i=1}^r \frac{c_i^2}{sn}} t_{rsn-r-s+1, \alpha/2} \right]$$

- Para un contraste $\sum_{j=1}^s c_j \delta_j$,

$$\left[\sum_{j=1}^s c_j \bar{Y}_{.j.} \pm \widehat{\sigma} \sqrt{\sum_{j=1}^s \frac{c_j^2}{rn}} t_{rsn-r-s+1, \alpha/2} \right]$$

Contraste de hipótesis

Contrastaremos las hipótesis $H_0 : \tau_1 = \dots = \tau_r$ y $H'_0 : \delta_1 = \dots = \delta_s$.

Para ambos contraste obtenemos que

$$Q_0 = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^n Y_{ijk}^2 - \sum_{i=1}^r sn\bar{Y}_{i..}^2 - \sum_{j=1}^s rn\bar{Y}_{.j.}^2 + rsn\bar{Y}_{...}^2$$

El modelo reducido por H_0 corresponden a modelo de un factor, diseño completamente aleatorizados. En consecuencia, la suma de cuadrados de los residuos del modelo reducido y la forma cuadrática Q_1 para el contraste de esta hipótesis tienen las siguientes expresiones:

$$Q_0^{H_0} = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^n Y_{ijk}^2 - \sum_{j=1}^s rn\bar{Y}_{.j.}^2 \quad Q_1 = Q_0^{H_0} - Q_0 = \sum_{i=1}^r sn\bar{Y}_{i..}^2 - rsn\bar{Y}_{...}^2$$

Análogamente, la suma de cuadrados de los residuos en el modelo reducido por H'_0 y la forma cuadrática Q'_1 para el contraste de esta hipótesis tienen las siguientes expresiones:

$$Q_0^{H'_0} = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^n Y_{ijk}^2 - \sum_{i=1}^r sn\bar{Y}_{i..}^2 \quad Q'_1 = Q_0^{H'_0} - Q_0 = \sum_{j=1}^s rn\bar{Y}_{.j.}^2 - rsn\bar{Y}_{...}^2$$

Los test de hipótesis para contrastar H_0 y H'_0 quedan resumidos en la siguiente tabla de análisis de la varianza:

F.V.	g.l.	Suma de cuadrados	M.C.	F
Ft. A	$r - 1$	$Q_1 = \sum_{i=1}^r sn\bar{Y}_{i..}^2 - rsn\bar{Y}_{...}^2$	$\frac{Q_1}{r - 1}$	$F_A = \frac{(rsn - r - s + 1)Q_1}{(r - 1)Q_0}$
Ft. B	$s - 1$	$Q'_1 = \sum_{j=1}^s rn\bar{Y}_{.j.}^2 - rsn\bar{Y}_{...}^2$	$\frac{Q'_1}{s - 1}$	$F_B = \frac{(rsn - r - s + 1)Q'_1}{(s - 1)Q_0}$
Error	$rsn - r - s + 1$	$Q_0 = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^n Y_{ijk}^2 - \sum_{i=1}^r sn\bar{Y}_{i..}^2 - \sum_{j=1}^s rn\bar{Y}_{.j.}^2 + rsn\bar{Y}_{...}^2$	$\frac{Q_0}{rsn - r - s + 1}$	

Bajo H_0 el estadístico F_A tiene distribución $F(r - 1, rsn - r - s + 1)$. Rechazaremos H_0 a un nivel de significación α si $F_A \geq F_{r-1, rsn-r-s+1, \alpha}$, probando que existen diferencias entre los efectos de los tratamientos del factor A.

Por otra parte, bajo H'_0 , el estadístico F_B tiene distribución $F(s - 1, rsn - r - s + 1)$. Rechazaremos H'_0 a un nivel de significación α si $F_B \geq F_{s-1, rsn-r-s+1, \alpha}$, probando que existen diferencias entre los efectos de los tratamientos del factor B.

En caso de rechazar cualquiera de las hipótesis anteriores, deberíamos llevar a cabo un procedimiento de comparaciones múltiples. Los métodos que podemos emplear son los mismos que en el modelo de Bloques al azar, sustituyendo $(r - 1)(s - 1)$ por $rsn - r - s - 1$ donde proceda.

Ejercicio propuesto: Deducir las ecuaciones normales de este modelo.

Ecuaciones Normales. Es sencillo obtener que

$$\mathbf{X}'\mathbf{Y} = (Y_{...}, Y_{1..}, \dots, Y_{r..}, Y_{.1.}, \dots, Y_{.s.}, Y_{11.}, \dots, Y_{rs.})^t$$

siendo $Y_{...} = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^n Y_{ijk}$, $Y_{i..} = \sum_{j=1}^s \sum_{k=1}^n Y_{ijk}$, $Y_{.j.} = \sum_{i=1}^r \sum_{k=1}^n Y_{ijk}$, $Y_{ij.} = \sum_{k=1}^n Y_{ijk}$, $i = 1, \dots, r$, $j = 1, \dots, s$.

Teniendo en cuenta que $E[\mathbf{X}'\mathbf{Y}] = \mathbf{X}'\mathbf{X}\boldsymbol{\beta}$, podemos obtener fácilmente la parte izquierda de las ecuaciones normales calculando las esperanzas de las variables aleatorias anteriormente definidas:

$$E[Y_{...}] = rsn\mu + \sum_{h=1}^r sn\tau_h + \sum_{l=1}^s rn\delta_l + \sum_{h=1}^r \sum_{l=1}^s n(\tau\delta)_{hl} \quad , \quad E[Y_{i..}] = sn\mu + sn\tau_i + \sum_{l=1}^s n\delta_l + \sum_{l=1}^s n(\tau\delta)_{il}$$

$$E[Y_{.j.}] = rn\mu + \sum_{h=1}^r n\tau_h + rn\delta_j + \sum_{h=1}^r n(\tau\delta)_{hj} \quad , \quad E[Y_{ij.}] = n\mu + n\tau_i + n\delta_j + (\tau\delta)_{ij}$$

De este modo, las ecuaciones normales $\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}$ son las siguientes:

$$\begin{aligned} rsn\widehat{\mu} + \sum_{h=1}^r sn\widehat{\tau}_h + \sum_{l=1}^s rn\widehat{\delta}_l + \sum_{h=1}^r \sum_{l=1}^s n(\widehat{\tau\delta})_{hl} &= Y_{...} \\ sn\widehat{\mu} + sn\widehat{\tau}_i + \sum_{l=1}^s n\widehat{\delta}_l + \sum_{l=1}^s n(\widehat{\tau\delta})_{il} &= Y_{i..}, \quad i = 1, \dots, r \\ rn\widehat{\mu} + \sum_{h=1}^r n\widehat{\tau}_h + rn\widehat{\delta}_j + \sum_{h=1}^r n(\widehat{\tau\delta})_{hj} &= Y_{.j.}, \quad j = 1, \dots, s \\ n\widehat{\mu} + n\widehat{\tau}_i + n\widehat{\delta}_j + n(\widehat{\tau\delta})_{ij} &= Y_{ij.}, \quad i = 1, \dots, r, \\ & \quad j = 1, \dots, s \end{aligned}$$

Una solución $\widehat{\boldsymbol{\beta}} = (\widehat{\mu}, \widehat{\tau}_1, \dots, \widehat{\tau}_r, \widehat{\delta}_1, \dots, \widehat{\delta}_s, (\widehat{\tau\delta})_{11}, \dots, (\widehat{\tau\delta})_{rs})^t$ verificando las restricciones

$$\sum_{h=1}^r \widehat{\tau}_h = 0, \quad \sum_{l=1}^s \widehat{\delta}_l = 0, \quad \sum_{h=1}^r (\widehat{\tau\delta})_{hj} = 0, \quad \sum_{l=1}^s (\widehat{\tau\delta})_{il} = 0, \quad i = 1, \dots, r, \quad j = 1, \dots, s$$

viene dada por

$$\widehat{\mu} = \bar{Y}_{...} \quad , \quad \widehat{\tau}_i = \bar{Y}_{i..} - \bar{Y}_{...} \quad , \quad \widehat{\delta}_j = \bar{Y}_{.j.} - \bar{Y}_{...} \quad , \quad (\widehat{\tau\delta})_{ij} = \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...} \quad , \quad \begin{matrix} i = 1, \dots, r \\ j = 1, \dots, s \end{matrix}$$

siendo $\bar{Y}_{...} = Y_{...}/rsn$, $\bar{Y}_{i..} = Y_{i..}/sn$, $\bar{Y}_{.j.} = Y_{.j.}/rn$ e $\bar{Y}_{ij.} = Y_{ij.}/n$.

Estimación puntual

La estimación puntual no tendrá gran importancia en este modelo. A partir de las filas de \mathbf{X} podemos obtener rs funciones lineales estimables linealmente independientes: $\mu_{ij} = \mu + \tau_i + \delta_j + (\tau\delta)_{ij}$, $i = 1, \dots, r$, $j = 1, \dots, s$. Se verifica el siguiente resultado:

Proposición 1 *El estimador insesgado de mínima varianza de μ_{ij} es $\widehat{\mu}_{ij} = \bar{Y}_{ij.}$ y además se verifica que $\widehat{\mu}_{ij} \sim N(\mu_{ij}, \sigma^2/n)$.*

Demostración. Queda como ejercicio.

Respecto a la estimación de σ^2 , utilizando la solución de las ecuaciones normales, $\widehat{\beta}$, indicada anteriormente podemos establecer el siguiente resultado

Proposición 2 *Se verifica que el estimador insesgado de σ^2 es*

$$\widehat{\sigma}^2 = \frac{1}{rs(n-1)}(\mathbf{Y}'\mathbf{Y} - \widehat{\beta}'\mathbf{X}'\mathbf{Y}) = \frac{1}{rs(n-1)} \left(\sum_{i=1}^r \sum_{j=1}^s \sum_{n=1}^k Y_{ijk}^2 - \sum_{i=1}^r \sum_{j=1}^s n\bar{Y}_{ij}^2 \right)$$

Además $Q_0 = rs(n-1)\widehat{\sigma}^2/\sigma^2 \sim \chi^2(rs(n-1))$ y $\widehat{\sigma}^2$ es independiente de $\widehat{\mu}_{ij}$ para todo $i = 1, \dots, r, j = 1, \dots, n$.

Demostración. Queda como ejercicio. Los cálculos de la segunda parte son consecuencia directa de los resultados del Tema 3, con lo cual sólo hay que indicar el resultado que se utiliza.

Ejercicio propuesto. Demostrar las Proposiciones 1 y 2.

Contraste de hipótesis. Formulación del problema.

El principal contraste que se realiza en este modelo es el denominado **contraste de interacción**. En un lenguaje ordinario, este contraste se podría expresar del siguiente modo:

H_0 : el modelo es aditivo, no presenta interacción

H_1 : existe interacción entre los dos factores del modelo

Aplicando la Definición 1, podemos reescribir H_0 como sigue

$$H_0 : (\mu_{ij} - \mu_{i'j}) - (\mu_{ij'} - \mu_{i'j'}) = ((\tau\delta)_{ij} - (\tau\delta)_{i'j}) - ((\tau\delta)_{ij'} - (\tau\delta)_{i'j'}) = 0, \quad i, i' = 1, \dots, r; \quad j, j' = 1, \dots, s$$

Sin embargo, esta formulación de H_0 no resulta conveniente a la hora de aplicar la metodología que hemos visto en el Tema 3 para calcular un test de hipótesis que resuelva este contraste. El cálculo del modelo reducido por H_0 parece complicado y en consecuencia una de las formas cuadráticas que interviene en el estadístico del test sería difícil de determinar.

Para resolver este problema vamos a introducir nuevos parámetros que nos permitirán reescribir el modelo. No se trata de hacer una reparametrización en el sentido que vimos en el Tema 3, simplemente de un cambio de parámetros de modo que los distintos términos en que se descompone el modelo sean funciones lineales estimables.

Para introducir estos nuevos parámetros definimos en primer lugar los “efectos medios” de τ_i , δ_j y $(\tau\delta)_{ij}$ como sigue

$$\bar{\tau}_. = \frac{1}{r} \sum_{i=1}^r \tau_i, \quad \bar{\delta}_. = \frac{1}{s} \sum_{j=1}^s \delta_j, \quad \overline{(\tau\delta)}_{.i} = \frac{1}{s} \sum_{j=1}^s (\tau\delta)_{ij}, \quad \overline{(\tau\delta)}_{.j} = \frac{1}{r} \sum_{i=1}^r (\tau\delta)_{ij}, \quad \overline{(\tau\delta)}_{..} = \frac{1}{rs} \sum_{i=1}^r \sum_{j=1}^s (\tau\delta)_{ij}$$

con $i = 1, \dots, r, j = 1, \dots, s$. Utilizando estos valores podemos definir los nuevos parámetros:

$$\mu^* = \mu + \bar{\tau}_. + \bar{\delta}_. + \overline{(\tau\delta)}_{..} \quad ; \quad \tau_i^* = \tau_i - \bar{\tau}_. + \overline{(\tau\delta)}_{.i} - \overline{(\tau\delta)}_{..} \quad ; \quad \delta_j^* = \delta_j - \bar{\delta}_. + \overline{(\tau\delta)}_{.j} - \overline{(\tau\delta)}_{..}$$

$$(\tau\delta)_{ij}^* = (\tau\delta)_{ij} - \overline{(\tau\delta)}_{.i} - \overline{(\tau\delta)}_{.j} + \overline{(\tau\delta)}_{..}, \quad i = 1, \dots, r; \quad j = 1, \dots, s,$$

Es sencillo verificar que, para $i = 1, \dots, r, j = 1, \dots, s$

$$\mu_{ij} = \mu^* + \tau_i^* + \delta_j^* + (\tau\delta)_{ij}^*$$

y además que $\sum_{i=1}^r \tau_i^* = \sum_{j=1}^s \delta_j^* = 0$, y también $\sum_{i=1}^r (\tau\delta)_{ij}^* = 0$ para todo $j = 1, \dots, s$ y $\sum_{j=1}^s (\tau\delta)_{ij}^* = 0$, para todo $i = 1, \dots, r$.

Proposición 3 *Se verifica que*

a) $\mu^*, \tau_i^*, \delta_j^*$ y $(\tau\delta)_{ij}^*, i = 1, \dots, r, j = 1, \dots, s$, son f.l.e.

b) El modelo es aditivo si y sólo si $(\tau\delta)_{ij}^* = 0, i = 1, \dots, r, j = 1, \dots, s$.

Demostración.

Apartado a). Probemos que $(\tau\delta)_{ij}^*$ es f.l.e. Sabemos que

$$((\tau\delta)_{ij} - (\tau\delta)_{i'j'}) - ((\tau\delta)_{i'j} - (\tau\delta)_{i'j'}) = (\mu_{ij} - \mu_{i'j'}) - (\mu_{i'j} - \mu_{i'j'}) \quad (1)$$

Sumando en i' y en j' en el lado izquierdo de (1), obtenemos que

$$\begin{aligned} \sum_{i'=1}^r \sum_{j'=1}^s [((\tau\delta)_{ij} - (\tau\delta)_{i'j'}) - ((\tau\delta)_{i'j} - (\tau\delta)_{i'j'})] &= rs(\tau\delta)_{ij} - r \sum_{j'=1}^s (\tau\delta)_{i'j'} - s \sum_{i'=1}^r ((\tau\delta)_{i'j} + \sum_{j'=1}^s (\tau\delta)_{i'j'}) \\ &= rs(\tau\delta)_{ij} - rs\overline{(\tau\delta)}_{i.} - rs\overline{(\tau\delta)}_{.j} + rs\overline{(\tau\delta)}_{..} = rs(\tau\delta)_{ij}^* \end{aligned}$$

y por tanto, igualando a la suma en i' y en j' en el lado derecho de (1), tenemos que

$$(\tau\delta)_{ij}^* = \frac{1}{rs} \sum_{i'=1}^r \sum_{j'=1}^s (\mu_{ij} - \mu_{i'j'}) - (\mu_{i'j} - \mu_{i'j'}) \quad (2)$$

es decir, $(\tau\delta)_{ij}^*$ es combinación lineal de las f.l.e. $\mu_{i'j'}$, $i' = 1, \dots, r$, $j' = 1, \dots, s$ y por tanto es f.l.e. Análogamente, se prueba que

$$\mu^* = \frac{1}{rs} \sum_{i'=1}^r \sum_{j'=1}^s \mu_{i'j'}, \quad \tau_i^* = \frac{1}{rs} \sum_{i'=1}^r \sum_{j'=1}^s (\mu_{ij'} - \mu_{i'j'}), \quad \delta_j^* = \frac{1}{rs} \sum_{i'=1}^r \sum_{j'=1}^s (\mu_{i'j} - \mu_{i'j'})$$

y por tanto μ^* , τ_i^* , $i = 1, \dots, r$, y δ_j^* , $j = 1, \dots, s$, son f.l.e.

Apartado b). Hay que probar que $(\tau\delta)_{ij}^* = 0$, $i = 1, \dots, r$, $j = 1, \dots, s$ si y sólo si $(\mu_{ij} - \mu_{i'j}) - (\mu_{ij'} - \mu_{i'j'}) = 0$, $i, i' = 1, \dots, r$, $j, j' = 1, \dots, s$.

Si $(\tau\delta)_{ij}^* = 0$, para todo $i = 1, \dots, r$, $j = 1, \dots, s$, como se verifica que $\mu_{ij} = \mu^* + \tau_i^* + \delta_j^* + (\tau\delta)_{ij}^*$, obtenemos que

$$(\mu_{ij} - \mu_{i'j}) - (\mu_{ij'} - \mu_{i'j'}) = ((\tau\delta)_{ij}^* - (\tau\delta)_{i'j}^*) - ((\tau\delta)_{ij'}^* - (\tau\delta)_{i'j'}^*) = 0, \quad i, i' = 1, \dots, r; \quad j, j' = 1, \dots, s$$

y por tanto el modelo es aditivo.

Recíprocamente, si $(\mu_{ij} - \mu_{i'j}) - (\mu_{ij'} - \mu_{i'j'}) = 0$, $i, i' = 1, \dots, r$, $j, j' = 1, \dots, s$, teniendo en cuenta (2), se obtiene que $(\tau\delta)_{ij}^* = 0$, para todo $i = 1, \dots, r$, $j = 1, \dots, s$.

De este modo, el contraste de interacción se puede formular también

$$\begin{aligned} H_0 : (\tau\delta)_{ij}^* &= 0 \quad \text{para todo } i = 1, \dots, r, \quad j = 1, \dots, s \\ H_1 : (\tau\delta)_{ij}^* &\neq 0 \quad \text{para algún } i = 1, \dots, r, \quad j = 1, \dots, s \end{aligned}$$

Ejercicio propuesto. Probar que μ^* , τ_i^* , $i = 1, \dots, r$, y δ_j^* , $j = 1, \dots, s$, son f.l.e.

Contraste de hipótesis. Cálculo del test.

Tenemos que contrastar

$$\begin{aligned} H_0 &: (\tau\delta)_{ij}^* = 0 \text{ para todo } i = 1, \dots, r, j = 1, \dots, s \\ H_1 &: (\tau\delta)_{ij}^* \neq 0 \text{ para algún } i = 1, \dots, r, j = 1, \dots, s \end{aligned}$$

en el modelo $Y_{ijk} = \mu + \tau_i + \delta_j + (\tau\delta)_{ij} + \mathcal{E}_{ijk}$, $i = 1, \dots, r$; $j = 1, \dots, s$; $k = 1, \dots, n$, que puede reescribirse en términos de los nuevos parámetros como

$$Y_{ijk} = \mu^* + \tau_i^* + \delta_j^* + (\tau\delta)_{ij}^* + \mathcal{E}_{ijk}, \quad i = 1, \dots, r; \quad j = 1, \dots, s; \quad k = 1, \dots, n$$

Esta hipótesis es estimable porque $(\tau\delta)_{ij}^*$, $i = 1, \dots, r$, $j = 1, \dots, s$ son f.l.e. Sin embargo, no son linealmente independientes, ya que estos parámetros están ligados por las $r + s$ restricciones $\sum_{i=1}^r (\tau\delta)_{ij}^* = 0$ para todo $j = 1, \dots, s$ y $\sum_{j=1}^s (\tau\delta)_{ij}^* = 0$, para todo $i = 1, \dots, r$. Teniendo en cuenta que una de estas restricciones es a su vez redundante porque $\sum_{j=1}^s \sum_{i=1}^r (\tau\delta)_{ij}^* = \sum_{i=1}^r \sum_{j=1}^s (\tau\delta)_{ij}^*$, el número de f.l.e. linealmente independientes que definen H_0 es $rs - (r + s - 1) = (r - 1)(s - 1)$.

Para calcular el test de hipótesis que resuelva este contraste, vamos a seguir los pasos que vimos en el tema 3.

1er paso: Modelo reducido por H_0 .

Si H_0 es cierta, el modelo

$$Y_{ijk} = \mu^* + \tau_i^* + \delta_j^* + \mathcal{E}_{ijk}, \quad i = 1, \dots, r; \quad j = 1, \dots, s; \quad k = 1, \dots, n$$

que es un modelo de dos factores sin interacción.

La suma de cuadrados de los residuos en este modelo es

$$Q'_0 = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^n Y_{ijk}^2 - \sum_{i=1}^r sn\bar{Y}_{i..}^2 - \sum_{j=1}^s rn\bar{Y}_{.j.}^2 + rsn\bar{Y}_{...}^2$$

2º paso: Formas cuadráticas asociadas al test.

La suma de cuadrados de los residuos del modelo con interacción, calculada cuando estimamos σ^2 , es

$$Q_0 = \sum_{i=1}^r \sum_{j=1}^s \sum_{n=1}^k Y_{ijk}^2 - \sum_{i=1}^r \sum_{j=1}^s n\bar{Y}_{ij.}^2$$

Ya hemos visto en la Proposición 2 que Q_0/σ^2 tiene distribución chi cuadrado con $rsn - rs = rs(n - 1)$ grados de libertad. Por otro lado, a partir de Q_0 y Q'_0 obtenemos la otra forma cuadrática, Q_1 , que necesitamos para el cálculo del estadístico F :

$$\begin{aligned} Q_1 = Q'_0 - Q_0 &= \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^n Y_{ijk}^2 - \sum_{i=1}^r sn\bar{Y}_{i..}^2 - \sum_{j=1}^s rn\bar{Y}_{.j.}^2 + rsn\bar{Y}_{...}^2 - \sum_{i=1}^r \sum_{j=1}^s \sum_{n=1}^k Y_{ijk}^2 + \sum_{i=1}^r \sum_{j=1}^s n\bar{Y}_{ij.}^2 \\ &= \sum_{i=1}^r \sum_{j=1}^s n\bar{Y}_{ij.}^2 - \sum_{i=1}^r sn\bar{Y}_{i..}^2 - \sum_{j=1}^s rn\bar{Y}_{.j.}^2 + rsn\bar{Y}_{...}^2 \end{aligned}$$

Según lo estudiado en el Tema 3, si H_0 es cierta Q_1/σ^2 tiene distribución chi cuadrado. Los grados de libertad de esta distribución vienen determinados por el número de f.l.e. linealmente independientes que definen H_0 , y que ya vimos que es igual a $(r-1)(s-1)$. Además Q_0 y Q_1 son variables aleatorias independientes.

3er paso: Estadístico F .

A partir de las formas cuadráticas calculadas anteriormente se obtiene el estadístico de contraste

$$F = \frac{Q_1/(r-1)(s-1)}{Q_0/rs(n-1)} = \frac{rs(n-1)Q_1}{(r-1)(s-1)Q_0}$$

cuya distribución, asumiendo H_0 , es $F \sim F((r-1)(s-1), rs(n-1))$, de modo que rechazaremos H_0 a un nivel α si $F \geq F_{(r-1)(s-1), rs(n-1), \alpha}$.

Los cálculos anteriores se resumen en la siguiente tabla de Análisis de la Varianza:

F.V.	g.l.	Suma de Cuadrados	M.C.	F
Interacción	$(r-1)(s-1)$	$Q_1 = \sum_{i,j} n\bar{Y}_{ij}^2 - \sum_i sn\bar{Y}_{i.}^2 - \sum_j rn\bar{Y}_{.j}^2 + rs n\bar{Y}_{...}^2$	$\frac{Q_1}{(r-1)(s-1)}$	$F = \frac{rs(n-1)Q_1}{(r-1)(s-1)Q_0}$
Error	$rs(n-1)$	$Q_0 = \sum_{i,j,k} Y_{ijk}^2 - \sum_{i,j} n\bar{Y}_{ij}^2$	$\frac{Q_0}{rs(n-1)}$	

Si aceptamos H_0 y concluimos que no hay interacción, entonces debemos ajustar un modelo aditivo

$$Y_{ijk} = \mu + \tau_i + \delta_j + \mathcal{E}_{ijk}, \quad i = 1, \dots, r; \quad j = 1, \dots, s; \quad k = 1, \dots, n.$$

Si probamos la existencia de interacción entre ambos factores, entonces el efecto del factor A en la respuesta depende del nivel del factor B en el que nos encontremos, y viceversa. Así, parece lógico que fijemos cada nivel del factor B (A) y estudiemos el efecto del factor A (B) en la respuesta, por ejemplo mediante un procedimiento de comparaciones múltiples como el que ya estudiamos para un modelo con un factor.

Otra posible actuación en este último caso consistiría en contrastar los “efectos principales” de cada factor, esto es, las hipótesis

$$H'_0 : \tau_i^* = 0, \quad i = 1, \dots, r, \quad H''_0 : \delta_j^* = 0, \quad j = 1, \dots, s.$$

No estudiaremos estos contrastes desde un punto de vista teórico.

4.3.2 Modelos anidados

Supongamos que el factor principal tiene r niveles. Asumiremos por simplicidad que todos los niveles del factor principal tienen asociados el mismo número de niveles del factor anidado, es decir, $s_1 = \dots = s_r = s$. Asumiremos también que el modelo es balanceado, es decir, que se ha tomado el mismo número de observaciones, n , para todos los niveles del factor anidado. El Modelo Lineal que utilizaremos para esta situación es

$$Y_{ijk} = \mu + \tau_i + (\tau\delta)_{ij} + \mathcal{E}_{ijk} \quad i = 1, \dots, r; \quad j = 1, \dots, s; \quad k = 1, \dots, n.$$

donde μ , τ_i , y $(\tau\delta)_{ij}$ son constantes desconocidas y \mathcal{E}_{ijk} son variables aleatorias tales que $E[\mathcal{E}_{ijk}] = 0$, $i = 1, \dots, r$, $j = 1, \dots, s$, $k = 1, \dots, n$. Por tanto, se verifica que

$$\mu_{ij} = E[Y_{ijk}] = \mu + \tau_i + (\tau\delta)_{ij}$$

El parámetro τ_i representa ahora el efecto “principal” que el nivel i del factor principal produce en la media de la variable respuesta. El parámetro $(\tau\delta)_{ij}$ representa el efecto en la media de la variable respuesta del nivel j del segundo factor, anidado al nivel i del factor principal. El número de observaciones en este modelo es rsn y el número de parámetros $p = rs + r + 1$, por tanto para que haya más observaciones que parámetros necesitamos (en el caso balanceado) que $n \geq 2$. Si esto no ocurre, es decir, si tenemos una única observación por cada nivel del factor anidado, no podemos ajustar este modelo.

La matriz de este Modelo es la que aparece a continuación¹

$$\mathbf{X} = \left(\begin{array}{cccc|cccccccc} 1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & \dots & \dots & \dots & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 & 1 & \dots & \dots & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & & & & & \vdots \\ 1 & 1 & 0 & \dots & 0 & & & & & & 1 & & \vdots \\ 1 & 0 & 1 & \dots & 0 & \vdots & \vdots & & & & 1 & & \vdots \\ 1 & 0 & 1 & \dots & 0 & & & & & & 1 & & \vdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & & & \ddots & & \vdots \\ 1 & 0 & 1 & \dots & 0 & & & & & & 1 & & \vdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & & & \ddots & & \vdots \\ 1 & 0 & 0 & \dots & 1 & & & & & & 1 & & \vdots \\ 1 & 0 & 0 & \dots & 1 & \vdots & \vdots & & & & & 1 & \vdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & & & & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & \dots & \dots & \dots & \dots & 1 \end{array} \right)$$

Se verifica que $r(\mathbf{X}) = rs < p$. Si $\mathcal{E} = (\mathcal{E}_{111}, \dots, \mathcal{E}_{rsn})^t$, asumiremos, en general, que $\mathcal{E} \sim \mathcal{N}_{rsn}(\mathbf{0}, \sigma^2 \mathbf{I}_{rsn})$, aunque para algún resultado la normalidad no sea necesaria.

¹Hay que tener en cuenta que cada fila se repite n veces, lo cual no afecta al rango.

Ecuaciones normales. Las ecuaciones normales de este modelo son:

$$\begin{aligned} nrs\widehat{\mu} + \sum_{h=1}^r sn\widehat{\tau}_h + \sum_{h=1}^r \sum_{l=1}^s n(\widehat{\tau\delta})_{hl} &= Y_{\dots} \\ sn\widehat{\mu} + sn\widehat{\tau}_i + \sum_{l=1}^s n(\widehat{\tau\delta})_{il} &= Y_{i\dots}, \quad i = 1, \dots, r \\ n\widehat{\mu} + n\widehat{\tau}_i + n(\widehat{\tau\delta})_{ij} &= Y_{ij}, \quad i = 1, \dots, r, \quad j = 1, \dots, s. \end{aligned}$$

Una solución $\widehat{\beta} = (\widehat{\mu}, \widehat{\tau}_1, \dots, \widehat{\tau}_r, (\widehat{\tau\delta})_{11}, \dots, (\widehat{\tau\delta})_{rs})^t$ viene dada por

$$\widehat{\mu} = \bar{Y}_{\dots}, \quad \widehat{\tau}_i = \bar{Y}_{i\dots} - \bar{Y}_{\dots}, \quad (\widehat{\tau\delta})_{ij} = \bar{Y}_{ij} - \bar{Y}_{i\dots}, \quad i = 1, \dots, r, \quad j = 1, \dots, s$$

siendo $\bar{Y}_{\dots} = Y_{\dots}/rsn$, $\bar{Y}_{i\dots} = Y_{i\dots}/sn$ e $\bar{Y}_{ij} = Y_{ij}/n$.

Estimación puntual

A partir de las filas de \mathbf{X} podemos obtener rs funciones lineales estimables linealmente independientes: $\mu_{ij} = \mu + \tau_i + (\tau\delta)_{ij}$, $i = 1, \dots, r$, $j = 1, \dots, s$. Se verifica el siguiente resultado:

Proposición 4 *El estimador insesgado de mínima varianza de μ_{ij} es $\widehat{\mu}_{ij} = \bar{Y}_{ij}$, y se verifica que $\widehat{\mu}_{ij} \sim N(\mu_{ij}, \sigma^2/n)$.*

Demostración. Es similar a la de la Proposición 1 en el diseño de dos factores con interacción.

Los contrastes de los parámetros τ_i no tienen por qué ser funciones lineales estimables (f.l.e.), pero si i permanece fijo, sí son f.l.e. los contrastes de los parámetros $(\tau\delta)_{ij}$, es decir, las combinaciones lineales de la forma $\sum_{j=1}^s c_j(\tau\delta)_{ij}$ con $\sum_{j=1}^s c_j = 0$.

Proposición 5 *El estimador insesgado de mínima varianza de $\psi = \sum_{j=1}^s c_j(\tau\delta)_{ij}$ es $\widehat{\psi} = \sum_{j=1}^s c_j\bar{Y}_{ij}$, y además se verifica que $\widehat{\psi} \sim N(\sum_{j=1}^s c_j(\tau\delta)_{ij}, \sum_{j=1}^s c_j^2\sigma^2/n)$.*

Demostración. Queda como ejercicio.

Respecto a la estimación de σ^2 , utilizando la solución de las ecuaciones normales, $\widehat{\beta}$, indicada anteriormente podemos establecer el siguiente resultado

Proposición 6 *Se verifica que el estimador insesgado de σ^2 es*

$$\widehat{\sigma}^2 = \frac{1}{rs(n-1)} \left(\sum_{i=1}^r \sum_{j=1}^s \sum_{n=1}^k Y_{ijk}^2 - \sum_{i=1}^r \sum_{j=1}^s n\bar{Y}_{ij}^2 \right)$$

Además $Q_0 = rs(n-1)\widehat{\sigma}^2/\sigma^2 \sim \chi^2(rs(n-1))$ y $\widehat{\sigma}^2$ es independiente de $\widehat{\mu}_{ij}$ para todo $i = 1, \dots, r$, $j = 1, \dots, s$ y del estimador insesgado de mínima varianza de cualquier contraste $\sum_{j=1}^s c_j(\tau\delta)_{ij}$.

Demostración. Es análoga a la de la Proposición 2 en el diseño de dos factores con interacción.

Contraste de hipótesis

En este modelo nos planteamos el contraste de la hipótesis

$$H_0 : (\tau\delta)_{i1} = \dots = (\tau\delta)_{is} \quad \text{para todo } i = 1, \dots, r,$$

En caso de aceptar H_0 , el factor anidado no influye en la variable respuesta. Si la rechazamos, en alguno de los niveles del factor principal habrá diferencias entre los efectos de los niveles del factor anidado en la media de la variable respuesta.

El modelo reducido por H_0 queda $Y_{ijk} = \mu + \tau_i + (\tau\delta)_{i1} + \mathcal{E}_{ijk} = \mu + \tau_i^* + \mathcal{E}_{ijk}$, siendo $\tau_i^* = \tau_i + (\tau\delta)_{i1}$, $i = 1, \dots, r$, $j = 1, \dots, s$, $k = 1, \dots, n$. Este es un modelo de un único factor (el factor principal), concretamente un diseño completamente aleatorizado. Realizando los cálculos pertinentes, que quedan como ejercicio, obtenemos la siguiente tabla de análisis de la varianza:

F. V.	g.l.	Suma de Cuadrados	M.C.	F
Fct anid	$r(s-1)$	$Q_1 = \sum_{i,j} n\bar{Y}_{ij}^2 - \sum_i sn\bar{Y}_{i.}^2$	$\frac{Q_1}{r(s-1)}$	$F = \frac{rs(n-1)Q_1}{r(s-1)Q_0}$
Error	$rs(n-1)$	$Q_0 = \sum_{i,j,k} Y_{ijk}^2 - \sum_{i,j} n\bar{Y}_{ij}^2$	$\frac{Q_0}{rs(n-1)}$	

Rechazaremos H_0 al nivel de significación α si $F \geq F_{r(s-1), rs(n-1), \alpha}$.

Si aceptamos H_0 y concluimos que el factor anidado no tiene influencia en el experimento, debemos ajustar el modelo con un factor para comprobar si la variable respuesta varía con los niveles del factor principal

$$Y_{ijk} = \mu + \tau_i + \mathcal{E}_{ijk}.$$

Si rechazamos H_0 , entonces se utilizará un procedimiento de comparaciones múltiples, dentro de cada nivel del factor principal para determinar qué niveles del factor anidado presentan diferencias. Las hipótesis a contrastar de manera simultánea son (para cada i fijo) son

$$H_0^{(j,j')} : (\tau\delta)_{ij} = (\tau\delta)_{ij'}.$$

Tenemos que realizar $s(s-1)/2$ comparaciones dos a dos de modo simultáneo. Podemos adaptar a este caso cualquiera de los procedimientos que ya conocemos (LSD de Fisher, Bonferroni, Scheffé o Tukey), sin más que cambiar los grados de libertad cuando sea necesario.

También es necesario determinar la posible influencia del factor principal. El contraste que se plantea es similar al de los efectos principales en el modelo con interacción y no lo estudiaremos a nivel teórico.

Ejercicios propuestos:

1. Deducir las ecuaciones normales y calcular una solución particular de las mismas.
2. Demostrar la Proposición 5.
3. Demostrar que la hipótesis H_0 es estimable y deducir el test F para resolver el contraste de hipótesis.

Tema 5: Modelos Lineales Generalizados

5.1 Introducción a los Modelos Lineales Generalizados

Los modelos lineales generalizados extienden el modelo lineal normal en varios sentidos. En primer lugar, la distribución que se asocia al modelo no es la normal, sino que pertenece a una familia más extensa, de la cual también forma parte la normal que es la familia exponencial de distribuciones.

Sea Θ un intervalo de \mathbb{R} , y sea $\{f_\theta: \theta \in \Theta\}$ una familia de funciones de densidad (o funciones masa de probabilidad) definidas sobre \mathbb{R}^n . Supondremos que el conjunto $\{\mathbf{y} \in \mathbb{R}^n: f_\theta(\mathbf{y}) > 0\}$ es independiente de θ .

Definición 1 Diremos que la familia $\{f_\theta: \theta \in \Theta\}$ es una familia exponencial uniparamétrica, si existen funciones reales $Q(\theta)$ y $D(\theta)$ sobre Θ y funciones reales medibles $T(\mathbf{y})$ y $S(\mathbf{y})$ sobre \mathbb{R}^n tales que:

$$f_\theta(\mathbf{y}) = \exp\{Q(\theta)T(\mathbf{y}) + D(\theta) + S(\mathbf{y})\}.$$

- Para el caso $n = 1$, si $T(x) = x$ diremos que la familia viene expresada en forma canónica.
- A $Q(\theta)$ a veces se le llama parámetro natural de la familia.
- Para el caso $n = 1$, si la distribución de la variable aleatoria Y pertenece a la familia exponencial se verifica que:

$$\begin{aligned} E[T(Y)] &= -D'(\theta)/Q'(\theta) \\ \text{Var}[T(Y)] &= [Q''(\theta)D'(\theta) - D''(\theta)Q'(\theta)]/[Q'(\theta)]^3 \end{aligned} \tag{1}$$

Sea Θ un intervalo k -dimensional de \mathbb{R}^k , y sea $\{f_\theta: \theta \in \Theta\}$ una familia de funciones de densidad (o funciones masa de probabilidad) definidas sobre \mathbb{R}^n . Supondremos que el conjunto $\{\mathbf{y} \in \mathbb{R}^n: f_\theta(\mathbf{y}) > 0\}$ es independiente de θ .

Definición 2 Diremos que la familia $\{f_\theta: \theta \in \Theta\}$ es una familia exponencial k -paramétrica, si existen funciones reales $Q_1(\theta), Q_2(\theta), \dots, Q_k(\theta)$ y $D(\theta)$ sobre Θ y funciones reales medibles $T_1(\mathbf{y}), T_2(\mathbf{y}), \dots, T_k(\mathbf{y})$ y $S(\mathbf{y})$ sobre \mathbb{R}^n tales que:

$$f_\theta(\mathbf{y}) = \exp \left\{ \sum_{i=1}^k Q_i(\theta)T_i(\mathbf{y}) + D(\theta) + S(\mathbf{y}) \right\}$$

Definición 3 (Nelder and Wedderburn (1972)) Sea $\mathbf{Y} = (Y_1, \dots, Y_n)^t$ un vector aleatorio n -dimensional, \mathbf{X} una matriz $n \times p$ ($p < n$) de constantes conocidas:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1^t \\ \vdots \\ \mathbf{X}_n^t \end{pmatrix},$$

$g: \mathbb{R} \rightarrow \mathbb{R}$ una función monótona y derivable (llamada función de enlace ó link function) y $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^t$ un vector de parámetros desconocidos. Diremos que Y satisface un **modelo lineal generalizado** si

- a) Y_1, \dots, Y_n son v.a. independientes con funciones de densidad (o funciones masa de probabilidad) pertenecientes a una familia exponencial uniparamétrica expresada en forma canónica, i.e.

$$f_{\theta_i}(y) = \exp\{Q(\theta_i)y + D(\theta_i) + S(y)\}, \quad y \in \mathbb{R}; \quad i = 1, \dots, n.$$

- b) $g(E[Y_i]) = \mathbf{X}_i^t \boldsymbol{\beta}$, $i = 1, \dots, n$.

Observaciones:

1. Si definimos $G: \mathbb{R}^n \rightarrow \mathbb{R}^n$ $(x_1, \dots, x_n)^t \rightarrow G(x_1, \dots, x_n) = (g(x_1), \dots, g(x_n))^t$, entonces la segunda condición se puede escribir:

$$G(E[\mathbf{Y}]) = \mathbf{X}\boldsymbol{\beta}$$

2. Puesto que g es continua y monótona es biyectiva con su imagen y por tanto existe g^{-1} , de modo que

$$\mu_i = E[Y_i] = g^{-1}(\mathbf{X}_i^t \boldsymbol{\beta})$$

Ejemplos:

- a) Modelo lineal normal.
- b) Tendencia de la mortalidad: Para poblaciones grandes la probabilidad de que un individuo elegido al azar muera de una enfermedad dada en un instante determinado es pequeña. Si suponemos que la muerte de diferentes individuos son sucesos independientes, entonces el número de muertos, Y , durante un período de tiempo fijo puede ser modelizado por una distribución de Poisson

$$f_\lambda(y) = \frac{\lambda^y e^{-\lambda}}{y!} \quad y = 0, 1, \dots$$

siendo λ el número medio de muertos por período de tiempo.

La tendencia de la mortalidad puede ser modelizada tomando v.a. independientes Y_1, \dots, Y_n que sean el número de muertes ocurridas en intervalos de tiempos sucesivos numerados por $i = 1, \dots, n$. Sea $\lambda_i = E[Y_i]$, $i = 1, \dots, n$.

El número de muertes por Sida en Australia en períodos de tres meses entre 1983 y 1986 aparecen en la siguiente tabla y el siguiente gráfico:

i	1	2	3	4	5	6	7
y_i	0	1	2	3	1	4	9
i	8	9	10	11	12	13	14
y_i	18	23	31	20	25	37	45

Claramente el número de muertos crece con i . Para estos datos un modelo posible es el basado en la distribución de Poisson con $\lambda_i = i^\beta$, donde β es un parámetro a ser estimado. Este modelo puede ser descrito como un modelo lineal generalizado en el cual la función de enlace es $g(\lambda_i) = \log \lambda_i = \beta \log i$ y por tanto $\mathbf{X}_i^t = (\log i)$, $i = 1, \dots, n$, y $\boldsymbol{\beta} = (\beta)$.

5.2 Estimación en Modelos Lineales Generalizados

Consideremos un modelo lineal generalizado como el dado en la definición y utilicemos el método de máxima verosimilitud para estimar el vector de parámetros $\boldsymbol{\beta}$. Así la log-verosimilitud será:

$$\ell(\boldsymbol{\theta}; \mathbf{Y}) = \sum_{i=1}^n Y_i Q(\theta_i) + \sum_{i=1}^n D(\theta_i) + \sum_{i=1}^n S(Y_i) \quad \boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^t, \quad \mathbf{y} = (y_1, \dots, y_n)^t$$

En esta expresión, $\mu_i = E[Y_i] = -D'(\theta_i)/Q'(\theta_i)$ y por tanto es función de θ_i , que asumiremos invertible, de hecho, en muchas ocasiones $\theta_i = \mu_i$. Por tanto, si denotamos $\eta_i = g(\mu_i) = \mathbf{X}_i^t \boldsymbol{\beta}$, $i = 1, \dots, n$, entonces

$$\mu_i = \mu_i(\theta_i) = g^{-1}(\eta_i)$$

y en consecuencia podemos asumir que θ_i es función de $\boldsymbol{\beta}$, de modo que la log-verosimilitud también es función de $\boldsymbol{\beta}$:

$$\ell = \ell(\boldsymbol{\beta}; \mathbf{Y})$$

Una propiedad de la familia exponencial de distribuciones es que satisface suficientes condiciones de regularidad para asegurar que el máximo global de la función de log-verosimilitud, $\widehat{\boldsymbol{\beta}}$, se obtiene de manera única como solución de las ecuaciones

$$\frac{\partial \ell}{\partial \beta_j} = 0, \quad j = 1, \dots, p.$$

Denotaremos $U_j = U_j(\boldsymbol{\beta}) = \partial \ell / \partial \beta_j$ y denominaremos a estas derivadas parciales, como *scores*. El vector de scores se denotará por $\mathbf{U} = (U_1, \dots, U_p)^t$ y jugará un papel muy importante en el estudio de los Modelos Lineales Generalizados.

Scores

Si denotamos

$$\ell_i = Y_i Q(\theta_i) + D(\theta_i) + S(Y_i),$$

aplicando la regla de la cadena, se verifica que

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{d \ell_i}{d \theta_i} \frac{d \theta_i}{d \mu_i} \frac{\partial \mu_i}{\partial \beta_j}, \quad j = 1, \dots, p.$$

Ahora bien, es sencillo obtener que:

- $\frac{d \ell_i}{d \theta_i} = Y_i Q'(\theta_i) + D'(\theta_i)$
- Por la ecuación (1) tenemos que $\mu_i(\theta_i) = -D'(\theta_i)/Q'(\theta_i)$ y $\text{Var}[Y_i] = [Q''(\theta_i)D'(\theta_i) - D''(\theta_i)Q'(\theta_i)]/Q'(\theta_i)^3$, de lo cual se deduce que

$$\frac{d \mu_i}{d \theta_i} = \frac{-D''(\theta_i)Q'(\theta_i) + D'(\theta_i)Q''(\theta_i)}{Q'(\theta_i)^2} = Q'(\theta_i) \text{Var}[Y_i]$$

y en consecuencia,

$$\frac{d\theta_i}{d\mu_i} = \frac{1}{Q'(\theta_i) \text{Var}[Y_i]}$$

- Tomando $\mathbf{X}_i^t = (x_{i1}, \dots, x_{ip})$, se verifica que

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial g^{-1}(\eta_i)}{\partial \beta_j} = (g^{-1})'(\eta_i) \frac{\partial \mathbf{X}_i^t \boldsymbol{\beta}}{\partial \beta_j} = (g^{-1})'(\eta_i) x_{ij}$$

De todo lo anterior, utilizando de nuevo que $\mu_i(\theta_i) = -D'(\theta_i)/Q'(\theta_i)$, se obtiene que

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{Y_i Q'(\theta_i) + D'(\theta_i)}{Q'(\theta_i) \text{Var}[Y_i]} (g^{-1})'(\eta_i) x_{ij} = \frac{(Y_i - \mu_i) x_{ij}}{\text{Var}[Y_i]} (g^{-1})'(\eta_i)$$

y en consecuencia, se prueba que para $j = 1, \dots, p$

$$U_j = U_j(\boldsymbol{\beta}) = \frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^n \frac{(Y_i - \mu_i) x_{ij}}{\text{Var}[Y_i]} (g^{-1})'(\eta_i) \quad (2)$$

siendo x_{ij} es el j -ésimo elemento de \mathbf{X}_i^t . Se deduce fácilmente que

$$E[U_j] = \sum_{i=1}^n \frac{E[(Y_i - \mu_i)] x_{ij}}{\text{Var}[Y_i]} (g^{-1})'(\eta_i) = 0$$

A la matriz de covarianzas de \mathbf{U} se le denomina *matriz de información (de Fisher)* y se denota por $\mathcal{I} = \mathcal{I}(\boldsymbol{\beta}) = \text{Cov}[\mathbf{U}] = E[\mathbf{U}\mathbf{U}^t]$. Asumiremos que esta matriz es definida positiva, y por tanto no singular. Sus coeficientes son:

$$\mathcal{I}_{jk} = E[U_j U_k] = E \left[\sum_{i=1}^n \frac{(Y_i - \mu_i) x_{ij}}{\text{Var}[Y_i]} (g^{-1})'(\eta_i) \sum_{l=1}^n \frac{(Y_l - \mu_l) x_{lk}}{\text{Var}[Y_l]} (g^{-1})'(\eta_l) \right]$$

Teniendo en cuenta que Y_1, \dots, Y_n son independientes, los productos cruzados en la expresión anterior tienen esperanza igual a 0 y se obtiene que

$$\mathcal{I}_{jk} = E[U_j U_k] = \sum_{i=1}^n \frac{E[(Y_i - \mu_i)^2] x_{ij} x_{ik}}{\text{Var}[Y_i]^2} (g^{-1})'(\eta_i)^2 = \sum_{i=1}^n \frac{x_{ij} x_{ik}}{\text{Var}[Y_i]} \left((g^{-1})'(\eta_i) \right)^2$$

Cálculo de $\widehat{\boldsymbol{\beta}}$

Como ya hemos indicado, para calcular el estimador de máxima verosimilitud $\widehat{\boldsymbol{\beta}}$, hemos de resolver el sistema de ecuaciones

$$U_1(\widehat{\boldsymbol{\beta}}) = 0, \dots, U_p(\widehat{\boldsymbol{\beta}}) = 0$$

En general, estas ecuaciones son no lineales y deben ser resueltas por métodos numéricos.

Si utilizamos el Método de Newton-Raphson p -dimensional, denotando $\mathbf{b}^{(m)}$ la aproximación de $\widehat{\boldsymbol{\beta}}$ obtenida en la m -ésima iteración, se obtiene la siguiente ecuación:

$$\mathbf{b}^{(m)} = \mathbf{b}^{(m-1)} - \left(\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k} (\mathbf{b}^{(m-1)}) \right)_{j,k=1,\dots,p}^{-1} \mathbf{U}^{(m-1)}$$

donde $\mathbf{U}^{(m-1)} = \mathbf{U}(b^{(m-1)})$. La convergencia del Método de Newton-Raphson nos garantiza que

$$\mathbf{b}^{(m)} \xrightarrow{m \rightarrow \infty} \widehat{\boldsymbol{\beta}}$$

Un procedimiento alternativo, a veces más sencillo que el método de Newton-Raphson, es el *método de tanteo* (*method of scoring*), que consiste en reemplazar en la ecuación anterior la matriz Hessiana de ℓ por su esperanza. Ahora bien, se verifica que

$$E \left[\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k} \right] = -E[U_j U_k] = -I_{jk}$$

y por tanto, el método de tanteo, en su iteración m -ésima, queda definido por la siguiente ecuación:

$$\mathbf{b}^{(m)} = \mathbf{b}^{(m-1)} + \mathbf{I}(\mathbf{b}^{(m-1)})^{-1} \mathbf{U}^{(m-1)}$$

5.3 Distribución de los estimadores

Para obtener la distribución asociada al estimador máximo verosímil de $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}$, juegan un papel muy importante el vector de scores $\mathbf{U} = (U_1, \dots, U_p)^t$. Vamos a obtener en primer lugar la distribución aproximada de este vector aleatorio y, a partir de esta distribución, deduciremos la distribución aproximada de $\widehat{\boldsymbol{\beta}}$. En esta sección, dejaremos un poco de lado el rigor matemático, siendo nuestro objetivo justificar, que no demostrar, los distintos resultados.

Distribución del vector de scores

Utilizando (2), deducimos que el vector de scores \mathbf{U} es suma de n vectores aleatorios independientes. Como $E[\mathbf{U}] = \mathbf{0}$ y $\text{Cov}[\mathbf{U}] = \mathbf{I}$, definida positiva, el Teorema Central del Límite, en su versión multidimensional, nos garantiza que, bajo ciertas condiciones de regularidad, $\mathbf{I}^{-1/2} \mathbf{U}$ converge en distribución, cuando $n \rightarrow \infty$, a una normal p -dimensional con vector de medias $\mathbf{0}$ y matriz de covarianzas \mathbf{I}_p .

Abreviadamente escribiremos

$$\mathbf{I}^{-1/2} \mathbf{U} \overset{\text{aprox}}{\sim} N_p(\mathbf{0}, \mathbf{I}_p)$$

o, equivalentemente, $\mathbf{U} \overset{\text{aprox}}{\sim} N_p(0, \mathbf{I})$. De lo anterior se deduce que $\mathbf{U}^t \mathbf{I}^{-1} \mathbf{U} \overset{\text{aprox}}{\sim} \chi^2(p)$.

Distribución de $\widehat{\boldsymbol{\beta}}$

Aproximando la función $\mathbf{U}(\boldsymbol{\beta})$ por su desarrollo de Taylor de orden 1 en torno a $\widehat{\boldsymbol{\beta}}$, se obtiene que

$$\mathbf{U}(\boldsymbol{\beta}) \simeq \mathbf{U}(\widehat{\boldsymbol{\beta}}) + \left(\frac{\partial U_j}{\partial \beta_k} \right)_{j,k=1,\dots,p} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) = \mathbf{U}(\widehat{\boldsymbol{\beta}}) + \left(\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k} \right)_{j,k=1,\dots,p} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \simeq \mathbf{U}(\widehat{\boldsymbol{\beta}}) + \left(E \left[\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k} \right] \right)_{j,k=1,\dots,p} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})$$

Teniendo en cuenta que $\mathbf{U}(\widehat{\boldsymbol{\beta}}) = \mathbf{0}$ y que $E \left[\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k} \right] = -I_{jk}$, se deduce que $\mathbf{U}(\boldsymbol{\beta}) \simeq \mathbf{I}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ o equivalentemente

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \simeq \mathbf{I}^{-1} \mathbf{U}.$$

Como $\mathbf{U} \overset{\text{aprox}}{\sim} N_p(0, \mathbf{I})$, se concluye que:

$$\widehat{\boldsymbol{\beta}} \overset{\text{aprox}}{\sim} N_p(\boldsymbol{\beta}, \mathbf{I}^{-1})$$

y por tanto

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^t \mathbf{I} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \overset{\text{aprox}}{\sim} \chi^2(p)$$

Al estadístico $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^t \mathcal{I}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ se le denomina *estadístico de Wald*.

Observaciones:

- a) Para modelos lineales normales los resultados anteriores son exactos.
- b) \mathcal{I} dependerá en general de $\boldsymbol{\beta}$ con lo cual, en aplicaciones prácticas, la consideraremos evaluada en $\widehat{\boldsymbol{\beta}}$.
- c) El resultado anterior nos permite obtener intervalos de confianza aproximados para $\beta_i, i = 1, \dots, p$.

Intervalos y regiones de confianza

Si $\mathcal{I}^{-1} = (v_{ij})_{i,j=1,\dots,p}$, entonces un intervalo de confianza aproximado al nivel $1 - \alpha$ ($0 < \alpha < 1$) para el parámetro β_i , viene dado por la expresión:

$$[\widehat{\beta}_i - z_{\alpha/2} \sqrt{v_{ii}}; \widehat{\beta}_i + z_{\alpha/2} \sqrt{v_{ii}}]$$

siendo $z_{\alpha/2}$ un valor tal que si $Z \sim N(0, 1)$, entonces $P(Z > z_{\alpha/2}) = \alpha/2$.

Del mismo modo, el estadístico de Wald nos permite deducir la siguiente región de confianza p -dimensional al nivel $1 - \alpha$ ($0 < \alpha < 1$) para el parámetro $\boldsymbol{\beta}$:

$$\mathcal{R} = \{\mathbf{z} \in \mathbb{R}^p : (\mathbf{z} - \widehat{\boldsymbol{\beta}})^t \mathcal{I}(\mathbf{z} - \widehat{\boldsymbol{\beta}}) \leq \chi_{p,\alpha}\}$$

siendo $\chi_{p,\alpha}$ un valor tal que, si $Q \sim \chi^2(p)$, entonces $P(Q > \chi_{p,\alpha}) = \alpha$.

Se deja como ejercicio deducir el intervalo y la región de confianza anteriores.

5.4 Contraste de Hipótesis en Modelos Lineales Generalizados

Contraste de Bondad de Ajuste.

Este contraste valora lo adecuado que es un modelo para describir un conjunto de datos. Se basa en comparar la verosimilitud del modelo propuesto con la verosimilitud del modelo maximal o saturado, siendo éste el que tiene la misma distribución y función de enlace que el modelo de interés, y además el número de parámetros es igual al total de observaciones (se considera que este modelo proporciona una descripción completa de los datos, al menos para la distribución considerada).

Intuitivamente, estaríamos contrastando

- H_0 : el modelo SÍ es adecuado para describir el conjunto de datos
- H_1 : el modelo NO es adecuado para describir el conjunto de datos

El estadístico de la razón de verosimilitudes es

$$\Lambda = \frac{L(\widehat{\boldsymbol{\beta}}_{max}; \mathbf{Y})}{L(\widehat{\boldsymbol{\beta}}; \mathbf{Y})} \quad \text{o} \quad \log \Lambda = \ell(\widehat{\boldsymbol{\beta}}_{max}; \mathbf{Y}) - \ell(\widehat{\boldsymbol{\beta}}; \mathbf{Y})$$

siendo $L(\widehat{\boldsymbol{\beta}}_{max}; \mathbf{Y})$, $L(\widehat{\boldsymbol{\beta}}; \mathbf{Y})$ las funciones de verosimilitud del modelo saturado y del modelo ajustado, respectivamente. Valores grandes de $\log \Lambda$ indican que el modelo propuesto proporciona un mal ajuste. A partir $\log \Lambda$, se define un coeficiente de gran importancia en Modelos Lineales Generalizados, al que se conoce como *deviance*

$$D = 2 \log \lambda = 2(\ell(\widehat{\boldsymbol{\beta}}_{max}; \mathbf{Y}) - \ell(\widehat{\boldsymbol{\beta}}; \mathbf{Y}))$$

Para determinar el valor teórico con el que comparar D vamos a deducir la distribución aproximada de este estadístico. Para ello, aproximaremos $\ell(\boldsymbol{\beta}) = \ell(\boldsymbol{\beta}; \mathbf{Y})$ por su desarrollo de Taylor de orden 2 en torno a $\widehat{\boldsymbol{\beta}}$:

$$\ell(\boldsymbol{\beta}) \simeq \ell(\widehat{\boldsymbol{\beta}}) + \mathbf{U}(\widehat{\boldsymbol{\beta}})(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + \frac{1}{2}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^t \left(\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k} \right)_{j,k=1,\dots,p} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})$$

Utilizando de nuevo que $\mathbf{U}(\widehat{\boldsymbol{\beta}}) = \mathbf{0}$ y que $E \left[\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k} \right] = -I_{jk}$, se obtiene que

$$\ell(\boldsymbol{\beta}) - \ell(\widehat{\boldsymbol{\beta}}) \simeq -\frac{1}{2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^t \mathbf{I} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

Por tanto, obtenemos el estadístico de Wald cuya distribución aproximada conocemos:

$$2(\ell(\widehat{\boldsymbol{\beta}}) - \ell(\boldsymbol{\beta})) \simeq (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^t \mathbf{I} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \stackrel{aprox}{\sim} \chi^2(p) \quad (3)$$

Ahora, descomponemos D del siguiente modo:

$$D = 2(\ell(\widehat{\boldsymbol{\beta}}_{max}; \mathbf{Y}) - \ell(\boldsymbol{\beta}_{max}; \mathbf{Y})) - 2(\ell(\widehat{\boldsymbol{\beta}}; \mathbf{Y}) - \ell(\boldsymbol{\beta}; \mathbf{Y})) + 2(\ell(\boldsymbol{\beta}_{max}; \mathbf{Y}) - \ell(\boldsymbol{\beta}; \mathbf{Y})) \quad (4)$$

Aplicando (3) tanto a nuestro modelo como al modelo saturado, obtenemos $2(\ell(\widehat{\boldsymbol{\beta}}) - \ell(\boldsymbol{\beta})) \stackrel{aprox}{\sim} \chi^2(p)$, $2(\ell(\widehat{\boldsymbol{\beta}}_{max}) - \ell(\boldsymbol{\beta}_{max})) \stackrel{aprox}{\sim} \chi^2(n)$.

Por otro lado, si el modelo propuesto es adecuado, entonces el último sumando en (4) convergerá a 0 (al menos en distribución). Por tanto, asumiendo independencia entre los dos primeros sumandos de (4), se obtiene que si H_0 es cierta:

$$D \stackrel{aprox}{\sim} \chi^2(n - p)$$

En consecuencia, se rechaza H_0 al nivel α si $D \geq \chi_{n-p,\alpha}^2$.

La deviance está relacionada con otras medidas de la bondad de ajuste como AIC o BIC que ya vimos en el Modelo de Regresión Lineal.

Contraste de Hipótesis sobre $\boldsymbol{\beta}$.

Si consideramos $q < p$, el contraste de mayor interés es el siguiente:

$$\begin{aligned} H_0 : & \quad \beta_{q+1} = \dots = \beta_p = 0 \\ H_1 : & \quad \beta_i \neq 0 \text{ para algún } i = q + 1, \dots, p \end{aligned}$$

El test de hipótesis se basa en la comparación de los estadísticos de bondad de ajuste del modelo original y del modelo reducido por H_0 . En consecuencia, si D_0 es la deviance asociada al modelo reducido por H_0 y D_1 la deviance asociada al modelo original, un estadístico adecuado para construir un test de hipótesis es $\Delta D = D_0 - D_1$.

Para hallar la distribución aproximada de este estadística, tengamos en cuenta que, $D_1 \stackrel{aprox}{\sim} \chi^2(n - p)$. Por otro lado, si H_0 cierta, se verificaría que $D_0 \stackrel{aprox}{\sim} \chi^2(n - q)$. Bajo ciertas condiciones de independencia, de lo anterior se obtiene que,

$$\Delta D \stackrel{aprox}{\sim} \chi^2(p - q)$$

De este modo, se rechaza H_0 al nivel α si $\Delta D \geq \chi_{p-q,\alpha}^2$.