

CONSTRUCCION DE INTERVALOS DE CONFIANZA
SIMULTANEOS Y METODOS DE COMPARACION
MULTIPLE

J. CONCEPCION LOREDO OSTI

T E S I S

PRESENTADA COMO REQUISITO PARCIAL PARA
OBTENER EL GRADO DE
MAESTRO EN CIENCIAS
ESPECIALIDAD DE ESTADISTICA EXPERIMENTAL



Universidad Autónoma Agraria
Antonio Narro

PROGRAMA DE GRADUADOS

Buenavista, Saltillo, Coah.

AGOSTO DE 1987

CONSTRUCCION DE INTERVALOS DE CONFIANZA SIMULTANEOS
Y METODOS DE COMPARACION MULTIPLE

J. CONCEPCION LOREDO OSTI

TESIS

Presentada como requisito parcial
para obtener el grado de
Maestro en Ciencias
Especialidad de Estadística Experimental

Universidad Autónoma Agraria
Antonio Narro
Programa de Graduados
Buenavista, Saltillo, Coah.
Agosto de 1987

Tesis elaborada bajo la supervisión del comité particular de asesoría y aprobada como requisito parcial, para optar al grado de

MAESTRO EN CIENCIAS ESPECIALIDAD
DE ESTADISTICA EXPERIMENTAL

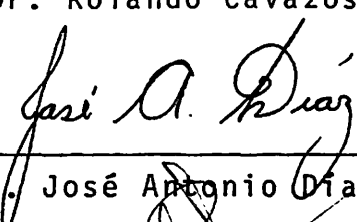
COMITE PARTICULAR

Asesor principal:



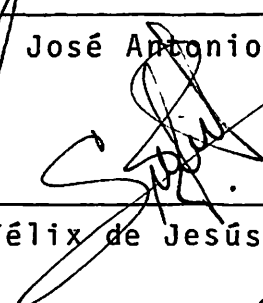
Dr. Rolando Cavazos Cadena

Asesor:

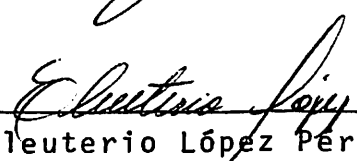


M.C. José Antonio Díaz García

Asesor:



M.C. Félix de Jesús Sánchez Pérez



Dr. Eleuterio López Pérez

Subdirector de Asuntos de Postgrado



BIBLIOTECA
EGIDIO G. REBONATO
BANCO DE TESIS
U.A.A.N.

Buenavista, Saltillo, Coahuila. Agosto de 1987

AGRADECIMIENTO

Deseo expresar mi agradecimiento

- A la UAAAN por la oportunidad de superación que me brindó.
- Al CONACYT por su apoyo económico.
- Al Dr. Rolando Cavazos Cadena por la ayuda y orientación que de él recibí durante mis estudios y en la realización del presente trabajo.
- Al M.C. José Antonio Díaz García y al M.C. Félix de Jesús Sánchez Pérez por el apoyo que me brindaron.
- A los maestros que me impartieron clase.
- A mis compañeros Octavio Martínez de la Vega, Roberto Canales Cruz, César Augusto Treviño y Javier Espinosa A.
- A Lety por su paciencia en el trabajo de mecanografía
- A todas las personas que de alguna manera contribuyeron en la realización del presente trabajo, en especial al Sr. Pedro Medina, por su desinteresado apoyo logístico.

A mis padres

A mi familia

A mis amigos

COMPENDIO

Construcción de intervalos de confianza simultáneos
y métodos de comparación múltiple

Por

J. CONCEPCION LOREDO OSTI

MAESTRIA EN CIENCIAS
ESTADISTICA EXPERIMENTAL

BUENAVISTA, SALTILLO, COAHUILA. AGOSTO 1987

Dr. Rolando Cavazos Cadena -Asesor-

Palabras clave: estimación simultánea; inferencia simultánea; métodos de pruebas simultáneas; métodos protegidos; rango múltiple; método de Scheffé; método de Tukey.

Este trabajo contiene una exposición de los aspectos más importantes de la teoría y métodos de construcción de intervalos de confianza simultáneos así como de pruebas de comparaciones múltiples. Estamos principalmente interesados en los resultados clásicos en este campo, pero también presentamos algunos desarrollos recientes.

Aunque esencialmente no obtenemos nuevas técnicas, hemos una formulación general de los problemas de mayor interés. Los principales resultados se ilustran con ejemplos.

ABSTRACT

CONSTRUCTION OF SIMULTANEOUS CONFIDENCE INTERVALS
AND MULTIPLE COMPARISON METHODS

By

J. CONCEPCION LOREDO OSTI

MASTER OF SCIENCE
EXPERIMENTAL STATISTICS

BUENAVISTA, SALTILLO, COAHUILA. AUGUST 1987

D.Sc. Rolando Cavazos Cadena -Advisor-

Key words: simultaneous estimation, simultaneous inference; simultaneous test methods; stepwise methods protected methods; multiple range; Scheffé's method, Tukey's method.

This work contains an exposition of the most important aspects of the theory and methods on the construction of simultaneous confidence intervals as well as on multiple comparisons tests. We are mainly concerned with classical results, but some recent developments in this field are presented.

Although essentially no new techniques were obtained, we provide a general formulation to the problems we are concerned with. Examples are given to illustrate the main results.

INDICE

	Página
INTRODUCCION	1
1. CONSTRUCCION DE INTERVALOS DE CONFIANZA SIMULTANEOS	5
1.1. DESCRIPCION GENERAL	5
1.1.1. OBSERVACIONES INTRODUCTORIAS ACERCA DE LA ESTIMACION SIMULTANEA	5
1.1.2. ALGUNAS DESIGUALDADES UTILES EN LA CON <u>S</u> TRUCCION DE ICS	10
1.2. ESTIMACION DE ICS PARA VARIABLES CON DISTRIBU <u>-</u> CION NORMAL	19
1.2.1. CONCEPTOS PRELIMINARES	19
1.2.2. EL METODO DE SCHEFFE	21
1.2.3. INTERPRETACION GEOMETRICA DEL METODO <u>-</u> DE SCHEFFE	27
1.2.4. EL METODO DE TUKEY	29
1.2.5. EXTENSION DEL METODO DE TUKEY A TODOS <u>-</u> LOS CONTRASTES	37
1.2.6. COMPARACION ENTRE EL METODO DE TUKEY Y EL METODO DE SCHEFFE	39
1.2.7. EXTENSION DEL METODO DE TUKEY A TODOS <u>-</u> LOS CONTRASTES EN EL CASO NO-BALANCEADO	42
1.2.8. EL METODO DE TUKEY PARA TODAS LAS FUN <u>-</u> CIONES LINEALES	44
1.2.9. EL METODO DE TUKEY-ROY-BOSE	47
1.2.10. EL METODO DE HOCHBERG	50
1.2.11. EL METODO DE TUKEY MODIFICADO	52
1.2.12. UN RESULTADO GENERAL	56
2. METODOS DE COMPARACIONES MULTIPLES	62
2.1. CONCEPTOS PRELIMINARES	62
2.2. METODOS DE PRUEBAS SIMULTANEAS	67
2.3. METODOS DE PRUEBAS PROTEGIDAS	86
2.4. METODOS DE COMPARACIONES POLIETAPICAS	89
EPILOGO	104

LITERATURA CITADA	109
APENDICE	114

INTRODUCCION

Consideremos el problema estadístico que se presenta cuando pensamos en la estimación del valor de una función paramétrica y deseamos que la estimación esté acompañada por alguna medida del posible error en la estimación; por ejemplo, supongamos que la función paramétrica es un escalar, entonces, la estimación puede consistir en un intervalo al que se le asocia una cantidad que, de algún modo, expresa la "certidumbre" o "grado de confianza" con el que, suponemos, el valor de la función cae dentro del intervalo; en este caso el problema es la construcción de un intervalo de confianza en el que su coeficiente de confianza es la magnitud del posible error en la estimación. Ahora bien, cuando tenemos no una, como en el caso anterior, sino un conjunto de funciones paramétricas y deseamos que la estimación de este conjunto esté asociada a alguna medida global del posible error en la estimación de todas las funciones en el conjunto; parece natural en este caso, asignar una región a cada función y hacer que la estimación consista en un conjunto de regiones que, con una probabilidad determinada, incluyan simultáneamente todos los valores de las funciones parametrales consideradas en sus respectivas

regiones, es decir, ahora el problema es la construcción de un conjunto de regiones de confianza simultáneas. En la mayoría de las situaciones prácticas, este conjunto de regiones está constituido por un conjunto de intervalos de confianza.

Un problema estrechamente relacionado con el de la construcción de regiones de confianza es el problema de las comparaciones múltiples.

La situación más común en que nos enfrentamos a un problema de comparaciones múltiples, se presenta en los diseños experimentales: Cuando se prueba la igualdad de medias en los tratamientos en el ANOVA con una prueba de F, raramente encontramos respuestas que sean suficientemente específicas para más de dos grupos de tratamientos, pues un resultado significativo en la prueba implica solamente que no todos los efectos de los tratamientos son iguales; la pregunta de qué tratamientos difieren no ha sido contestada. Por otro lado, hay situaciones en las que estamos interesados, por ejemplo, en detectar diferencias específicas entre parejas de tratamientos; en estos casos, la rutina del ANOVA puede no detectarlas, aunque al menos una pareja sea diferente, si las otras diferencias son pequeñas o no existen. Una gran cantidad de procedimientos han sido desarrollados en los últimos treinta años para ofrecer soluciones más específicas a problemas de este tipo. Las técnicas y métodos de comparaciones múltiples pueden diferir grandemente en muchos aspectos y algunos criterios adecuados para compararlos, han sido desarrollados muy recientemente; tal vez por esta razón, aún

persiste cierto tipo de confusión entre los usuarios de estos métodos, sobre cómo hacer una caracterización adecuada de algún problema particular de acuerdo a sus objetivos y cuáles son las potencialidades y limitantes de los métodos existentes.

En general, cuando nos enfrentamos a un problema de comparaciones múltiples contamos con un conjunto de hipótesis y el problema consiste en construir un conjunto de pruebas para todas y cada una de las hipótesis. Hay fundamentalmente tres puntos de vista acerca de cómo deben ser los métodos de comparaciones múltiples y estos puntos de vista nos permiten definir tres tipos de métodos, a saber: los métodos de pruebas simultáneas, los métodos de pruebas protegidas y los métodos de comparaciones polietápicas.

El objetivo de este trabajo es presentar algunos de los aspectos más relevantes de la teoría en que se fundamentan los métodos de construcción de intervalos de confianza simultáneos y los métodos de comparaciones múltiples.

El trabajo se divide en dos capítulos. En el primer capítulo se trata el problema de la construcción de intervalos de confianza simultáneos y se divide en dos secciones; en la primera de ellas se intenta presentar una definición formal del problema y se presentan tres desigualdades particularmente útiles en la obtención de un método para la construcción de intervalos de confianza simultáneos; las desigualdades son: la desigualdad de Bonferroni, la desigualdad de

Kimball y la desigualdad de Sidák; también se ilustra con algunos ejemplos el uso de estos resultados; en la segunda sección se trata la construcción de intervalos de confianza simultáneos para un conjunto de funciones lineales de las medias cuando las variables en estudio, se asume, están normalmente distribuidas; en esta parte se enfatiza en tres métodos, correspondientes a tres distintos conjuntos de funciones parametrales, a saber: el método de Scheffé, el método de Tukey y el método de Tukey-Roy-Bose; además, se estudia un conjunto de resultados relacionados con estos métodos; el capítulo finaliza con una formulación general de la solución al problema. Se pretende también que los resultados presentados en este capítulo sirvan como preámbulo al siguiente, en el que se trata el problema de las comparaciones múltiples. El segundo capítulo se divide en cuatro secciones. En la primera, se introducen algunos conceptos básicos que nos permiten caracterizar las familias de hipótesis y los diferentes métodos de comparaciones múltiples; en la segunda sección se definen y analizan los métodos de pruebas simultáneas; en la tercera sección se presentan los métodos de pruebas protegidas y en la última sección de este capítulo se estudian las propiedades de los métodos de comparación polietápicos.

CAPITULO 1
CONSTRUCCION DE INTERVALOS DE CONFIANZA SIMULTANEOS

1.1. Descripción General

1.1.1. Observaciones introductorias acerca de la Estimación Simultánea

Supongamos que se cuenta con una observación de un conjunto de variables aleatorias $Y = \{Y_1, Y_2, \dots, Y_k\}$ cuya distribución depende de un conjunto de parámetros desconocidos $\theta = \{\theta_1, \theta_2, \dots, \theta_m\}$. Consideremos una familia $\Psi = \{\psi_i, i \in I\}$ de funciones parametrales de interés primario; \mathbb{I} denota el conjunto de índices, posiblemente infinito. Denotemos por Λ_i el conjunto de valores posibles de ψ_i y a un valor típico de $\psi_i(\theta)$ por λ_i , esto es, $\lambda_i \in \Lambda_i$ para todo valor de θ en el espacio de parámetros Θ .

Estamos interesados en determinar, a partir de la observación y de Y , un conjunto de regiones

$$\Lambda_i(y, \alpha) \subset \Lambda_i; i \in I \quad (1.1)$$

tales que las inclusiones

$$\lambda_i \in \Lambda_i(y; \alpha) \quad (1.2.)$$

se satisfagan simultáneamente para toda $i \in I$ con una probabilidad de $1 - \alpha$, independientemente del valor de θ . En otras -

palabras, queremos, a partir de la observación de Y , obtener un conjunto $\{\Lambda_i(y; \alpha); i \in I\}$ de regiones tales que

$$P_{\theta}[\psi_i(\theta) \in \Lambda_i(y; \alpha) \text{ para toda } i \in I] = 1 - \alpha \text{ para toda } \theta \in \bar{\Omega} \quad (1.3)$$

El valor $1 - \alpha$ que satisface (1.3) se denomina coeficiente de confianza simultáneo (CCS) para las regiones definidas por (1.1) y (1.2). Así mismo, el conjunto

$$\{\Lambda_i(y; \alpha); i \in I\} \quad (1.4)$$

define una familia de regiones de confianza simultáneas para Ψ con un CCS de $1 - \alpha$.

El siguiente lema, presentado por Roy y Bose (1953), proporciona condiciones suficientes para que el problema de estimación planteado arriba sea solucionado.

Lema 1. Para que la obtención de la familia de regiones de confianza (1.4) sea posible, es suficiente que las siguientes condiciones se verifiquen:

(i) exista una familia de funciones

$$Z = \{Z_i(y; \lambda_i); i \in I\} \quad (1.5)$$

tal que para cada $i \in I$

$$\xi_1 \leq Z_i(y; \lambda_i) \leq \xi_2; \quad \xi_1 \text{ y } \xi_2 \text{ constantes independientes } i \in I \quad (1.6)$$

implique que

$$\lambda_i \in \Lambda_i(y; \alpha)$$

e inversamente, y

(ii) si, para un θ dado, $A_{i,\theta}$ denota el conjunto de puntos para los que lo anterior se satisface, es decir,

$$A_{i,\theta} = \{y: \xi_1 \leq Z_i(y; \psi_i(\theta)) \leq \xi_2\}; i \in I \quad (1.7)$$

entonces, para cada θ admisible

$$P_\theta \left[y \in \bigcap_I A_{i,\theta} \right] = 1 - \alpha \quad (1.8)$$

(y tal probabilidad no depende de θ)

Demostración. Por (i) $\xi_1 \leq Z_i \leq \xi_2$ es equivalente a $\lambda_i \in \Lambda_i(y; \alpha)$ para cualquier $i \in I$. De (ii) se tiene que $y \in A_{i,\theta}$ implica que $\xi_1 \leq Z_i \leq \xi_2$ para toda $i \in I$ y, recíprocamente, $\xi_1 \leq Z_i \leq \xi_2$ para toda $i \in I$ implica que $y \in \bigcap_I A_{i,\theta}$

Entonces, $\lambda_i \in \Lambda_i(y; \alpha)$ para toda $i \in I$, cuando y sólo cuando $y \in \bigcap_I A_{i,\theta}$; la probabilidad de este evento es por hipótesis $1 - \alpha$ y no depende de los parámetros. Luego, el conjunto (1.4) es una familia de regiones de confianza con un CCS de $1 - \alpha$. //

Note que $\bigcap_I A_{i,\theta}$ es el conjunto de los puntos y que satisfacen

$$\xi_1 \leq \inf_i \{Z_i(y; \lambda_i); i \in I\} \text{ y } \xi_2 \geq \sup_i \{Z_i(y; \lambda_i); i \in I\} \quad (1.9)$$

También note que pueden existir muchas posibilidades para elegir la familia Z definida en (1.5) y de hecho hay una familia de regiones de confianza para cada elección.

La familia de funciones Z definida en el lema, aquí - será referida como la familia de estimación. Esta familia es a la estimación simultánea lo que las funciones llamadas cantidades pivote son a la estimación ordinaria, no simultánea, de intervalos de confianza para funciones parametrales; en este sentido Z es una familia de cantidades pivote, aunque parece mejor decir que es una familia de funciones útiles en la - estimación de intervalos de confianza simultáneos, donde las propiedades de las funciones en la familia, mimetizan las propiedades relevantes de las cantidades pivote usadas en la es- timación no-simultánea. Un tratamiento más extensivo de estas últimas puede verse en Mood *et al.* (1985).

En adelante supondremos que cada función paramétrica es un escalar y su región de confianza respectiva puede ser - expresada como un intervalo en la recta.

Un procedimiento para construir un conjunto de inter- valos de confianza simultáneos (ICS) para un conjunto de fun- ciones parametrales con un CCS de $1 - \alpha$, podría ser como si - gue: primero consideremos cada función paramétrica separada - mente y asociemos un intervalo de confianza ordinario a ésta, con un coeficiente de confianza de $1 - \alpha_0$ ($\alpha_0 < \alpha$); el siguiente paso es considerar la intersección del conjunto de intervalos asociados a las funciones paramétricas individuales y - usar esta intersección para determinar la longitud de los in- tervalos de modo que éstos se satisfagan simultáneamente con un CCS de $1 - \alpha$, naturalmente $1 - \alpha < 1 - \alpha_0$; dado α puede -

determinarse α_0 y viceversa.

Cuando la intersección $\prod_I A_{i,\theta}$, usando la notación del lema, puede ser definida de manera simple y podemos encontrar los valores ξ_1 y ξ_2 para los que (1.8) se satisface, el procedimiento anterior nos permite estimar una familia de ICS con un CCS de exactamente $1 - \alpha$; queda entonces por decidir si son en algún sentido los "mejores" para el problema particular considerado. El tratamiento de algunos de estos casos será pospuesto hasta la sección (1.2) y ahí trataremos de aclarar en que sentido una familia de intervalos puede considerarse como mejor que otra. Por otro lado, es necesario aclarar que, en general, el procedimiento anterior no es el único posible.

Hay situaciones en las que la intersección $\prod_I A_{i,\theta}$ no puede ser fácilmente expresada o es técnicamente difícil de determinar la probabilidad contenida en ella. En muchos de estos casos una posibilidad tendiente a tornar en factible la estimación simultánea es la de eliminar las restricciones impuestas por el lema de que las constantes ξ_1 y ξ_2 sean las mismas para toda $i \in I$; en tales casos (1.6) es reemplazada por:

$$\xi_{i_1} \leq Z'_i(y; \lambda_i) \leq \xi_{i_2} \quad \xi_{i_1} \text{ y } \xi_{i_2} \text{ constantes, } i \in I \quad (1.10)$$

y entonces (1.7) debe cambiarse por:

$$A'_{i,\theta} = \{y: \xi_{i_1} \leq Z'_i(y; \psi_i(\theta)) \leq \xi_{i_2}\}; i \in I \quad (1.11)$$

de modo que el evento $y \in A'_{i,\theta}$ sea equivalente a $\lambda_i \in \Lambda_i(y; \alpha)$ para toda $i \in I$; sin embargo, es común que en estos casos sea

muy difícil determinar las ξ 's para tener que:

$$P_{\theta} \left[Y \in \bigcap_I A_{i,\theta} \right] = 1 - \alpha; \theta \in \mathcal{T}$$

no obstante, casi siempre es posible estimar un conjunto de intervalos de confianza simultáneos para los que la cota mínima del coeficiente de confianza conjunto sea $1 - \alpha$, es decir, que el CCS sea no menor a $1 - \alpha$. En algunos casos de este tipo $Z_i^1 = \{Z_i^1; i \in I\}$, la familia de funciones que satisfacen (1.10), puede verse como una familia de cantidades pivote, es decir, cada Z_i^1 es una cantidad pivote en el sentido ordinario; así cada cantidad pivote Z_i^1 tiene asociado un intervalo de confianza ordinario $\Lambda_i^1(y; \alpha_i) \subset \Lambda_i$ para la función ψ_i , con un coeficiente de confianza de $1 - \alpha_i$, de modo que si

$$P_{\theta} \left[\psi_i(\theta) \in \Lambda_i^1(y; \alpha_i) \text{ para toda } i \in I \right] \geq 1 - \alpha \quad (1.12)$$

el conjunto $\{\Lambda_i^1(y; \alpha_i); i \in I\}$ forman una familia de ICS con un CCS no menor a $1 - \alpha$.

A continuación veremos algunos resultados conocidos relacionados con lo anterior.

1.1.2. Algunas desigualdades útiles en la construcción de ICS.

Teorema 1. (Desigualdad de Bonferroni). Sean B_1, B_2, \dots, B_n eventos en un espacio de probabilidad y denotemos por \bar{B}_i al evento complementario a B_i , entonces

$$P \left[\bigcap_{i=1}^n B_i \right] \geq 1 - \sum_{i=1}^n P \left[\bar{B}_i \right] \quad (1.13)$$

Demostración. Por el teorema de De Morgan

$$\overline{\bigcap_{i=1}^n B_i} = \bigcup_{i=1}^n \overline{B_i}$$

y entonces

$$P\left[\bigcap_{i=1}^n B_i\right] = 1 - P\left[\bigcup_{i=1}^n \overline{B_i}\right]$$

y el resultado se sigue de la desigualdad de Boole:

$$P\left[\bigcup_{i=1}^n \overline{B_i}\right] \leq \sum_{i=1}^n P\left[\overline{B_i}\right] \quad /$$

La desigualdad de Bonferroni se estudia en muchos textos, como por ejemplo, Feller (1968), y su importancia se puede ilustrar con la siguiente explicación: supongamos que la familia $\Psi = \{\psi_i; i \in I\}$ de funciones paramétricas consta de un número finito h de tales funciones y supongamos que para cada λ_i se cuenta con un intervalo $\Lambda_i(y; \alpha_i)$ con un coeficiente de confianza $1 - \alpha_i$; entonces el teorema anterior garantiza que la probabilidad de que todos los h intervalos contengan el verdadero valor de sus respectivas funciones paramétricas es al menos $1 - \sum_{i=1}^n \alpha_i$. Una elección particular de α_i es $\alpha_i = \alpha/h$ y entonces

$$\{\Lambda_i(y; \alpha/h); i \in I\}$$

constituyen una familia de ICS con un CCS igual o mayor que $1 - \alpha$.

Ejemplo 1.1. Sea X_1, X_2, \dots, X_k un conjunto de variables aleatorias con $X_i \sim N(\theta_i, \sigma_i^2)$; supongamos que $S_1^2, S_2^2, \dots, S_k^2$ son estimadores de $\sigma_1^2, \sigma_2^2, \dots, \sigma^2$ con $\nu_1, \nu_2, \dots, \nu_k$ grados de

libertad respectivamente y que $v_i S_i^2 / \sigma_i^2$ tiene distribución χ^2 con v_i grados de libertad. Un conjunto de intervalos de confianza simultáneos para las θ_i 's con un CCS de al menos $1 - \alpha$, podría ser, entre otras posibilidades, construido de las dos maneras siguientes:

(i) Si

$$P\left[\frac{|X_i - \theta_i|}{S_i} \leq \xi_i\right] = 1 - \frac{\alpha}{k} \quad i = 1, 2, \dots, k$$

donde ξ_i es el punto correspondiente al $(1 - \alpha/2k)$ -percentil de la distribución t con v_i grados de libertad para cada $i \in I$; entonces los ICS deseados los constituyen el conjunto de - desigualdades

$$x_i - s_i \xi_i \leq \theta_i \leq x_i + s_i \xi_i \quad i = 1, 2, \dots, k$$

es decir,

$$z_i^1 = \frac{|X_i - \theta_i|}{S_i} \quad \text{y} \quad \Lambda_i^1(x; \alpha/k) = \{\theta: x_i - s_i \xi_i \leq \theta_i \leq x_i + s_i \xi_i\}$$

(ii) Si

$$P\left[\frac{|X_i - \theta_i|}{S_i} < \xi_0\right] = 1 - \alpha_i \quad i = 1, 2, \dots, k$$

donde ξ_0 es el punto correspondiente al $(1 - \alpha_i)$ -percentil de la distribución t con v_i grados de libertad para cada $i \in I$, y, adicionalmente, $\alpha = \alpha_1 + \alpha_2 + \dots + \alpha_k$, entonces $\Lambda_i^1(x; \alpha_i)$ es ta constituido por los puntos θ_i que satisfacen

$$x_i - s_i \xi_0 \leq \theta_i \leq x_i + s_i \xi_0 \quad i = 1, 2, \dots, k$$

y el conjunto

$$\{\Lambda_i^1(x_i; \alpha_i); i \in I\}$$

es la familia de ICS buscada □

La extrema generalidad de la desigualdad de Bonferroni permite su aplicación en una gran variedad de situaciones, ya que para utilizarla sólo se requiere que conozcamos los intervalos de confianza ordinarios de cada una de las funciones parametrales de interés sin necesidad de hacer supuestos sobre independencia, distribución conjunta, etc. de tales funciones. La cota proporcionada por la desigualdad de Bonferroni puede ser adecuada en muchos problemas particulares, inclusive puede ser bastante "buena" cuando el número de funciones para las que se desea el conjunto de ICS sea "moderado".

En cierto tipo de problemas estadísticos la desigualdad de Bonferroni puede ser estrictamente mejorada, como se desprende del siguiente teorema.

Teorema 2. (Desigualdad de Kimball). Sea X cualquier variable aleatoria. Sea también, $\{g_i\}_{i=1}^n$ un conjunto de funciones no negativas, monótonas crecientes. Supongamos ahora que $g_i(X)$ tiene esperanza finita. Entonces

$$E\left[\prod_{i=1}^n g_i(X)\right] \geq \prod_{i=1}^n E[g_i(X)] \quad (1.14)$$

Demostración. Observe que es suficiente demostrar la desigualdad para cualquier par de funciones, g y h , que satisfagan las condiciones del teorema, es decir, es suficiente mostrar que

$$E[g(X)h(X)] \geq E[g(X)] E[h(X)] \quad (*)$$

pues note que si esta desigualdad es cierta, podemos definir las funciones

$$h_i = \prod_{j \geq i} g_j; \quad i, j = 1, 2, \dots, n$$

y las funciones del conjunto $\{h_i\}$ son también monótonas crecientes; entonces si en (*) hacemos que $g = g_1$ y $h = h_2$ se tendrá que

$$E[h_1(X)] \geq E[g_1(X)] E[h_2(X)]$$

y la aplicación repetida del mismo argumento para h_2, h_3, \dots, h_n establece el resultado.

Para ver la desigualdad (*) hagamos que

$$\begin{aligned} E[g(X)h(X)] - E[g(X)] E[h(X)] &= \int_{-\infty}^{\infty} g(x) \{h(x) - E[h(X)]\} dF(x) \\ &= \zeta \end{aligned}$$

donde F denota la función de distribución de X ; observe que $\zeta \geq 0$ implica la desigualdad (*).

Ahora bien, como $h(x)$ es monótona, debe existir una cantidad x_0 tal que

$$h(x) \leq E[h(X)] \quad \text{si} \quad x \leq x_0$$

$$h(x) > E[h(X)] \quad \text{si} \quad x > x_0$$

a saber

$$x_0 = \max \{x: h(x) \leq E[h(X)]\} .$$

Entonces

$$\begin{aligned} \zeta &= - \int_{-\infty}^{x_0} g(x) \{ E[h(X)] - h(x) \} dF(x) + \int_{x_0}^{\infty} g(x) \{ h(x) - E[h(X)] \} dF(x) \\ &= -\zeta_1 + \zeta_2 \end{aligned}$$

Por otro lado, como

$$\int_{-\infty}^{\infty} \{h(x) - E[h(X)]\} dF(x) = 0$$

debe tenerse que

$$\begin{aligned} 0 &= \int_{-\infty}^{x_0} \{h(x) - E[h(X)]\} dF(x) + \int_{x_0}^{\infty} \{h(x) - E[h(X)]\} dF(x) \\ &= \Delta_1 + \Delta_2 \end{aligned}$$

más aún, ya que $g(x)$ es una función monótona creciente y positiva, entonces

$$\zeta_1 \leq -g(x_0)\Delta_1 \quad \text{y} \quad \zeta_2 \geq g(x_0)\Delta_2$$

de modo que

$$\begin{aligned} \zeta &= -\zeta_1 + \zeta_2 \\ &\geq g(x_0)\Delta_1 + g(x_0)\Delta_2 \\ &= g(x_0)(\Delta_1 + \Delta_2) \\ &= 0 \end{aligned}$$

//

La desigualdad de Kimball es interesante por sí misma y su utilización no se restringe a la construcción de ICS - pues, al igual que la desigualdad de Bonferroni, también puede ser empleada en problemas de comparaciones múltiples para estimar la máxima tasa de error en las pruebas de comparación polietápicas y, en ambos casos, la desigualdad de Kimball supera estrictamente a la desigualdad de Bonferroni cuando las dos son aplicables.

A continuación veremos un teorema en el que se utiliza la desigualdad de Kimball para la construcción de ICS en los casos en que se tiene una familia de cantidades pivote -

Z' y cada Z'_i en Z' puede expresarse como la razón de dos variables aleatorias independientes, con numeradores estadísticamente independientes y el denominador positivo común.

Teorema 3. Supongamos que contamos con un conjunto de variables aleatorias estadísticamente independientes Y, X_1, X_2, \dots, X_k donde $Y > 0$ y $X_i \in \mathbb{R}$, $i = 1, 2, \dots, k$. Si definimos Z'_i como $Z'_i = X_i/Y$, $i = 1, 2, \dots, k$, y a cada Z'_i asociamos un intervalo (ξ_{i_1}, ξ_{i_2}) que contenga al origen, es decir, $\xi_{i_1} \leq 0$, $\xi_{i_2} \geq 0$, entonces

$$P[\xi_{i_1} \leq Z'_i \leq \xi_{i_2} \text{ para toda } i = 1, 2, \dots, k] \geq \prod_{i=1}^k P[\xi_{i_1} \leq Z'_i \leq \xi_{i_2}] \quad (1.15)$$

Demostración. Denotemos por F la función de distribución de Y . Entonces

$$\begin{aligned} P[\xi_{i_1} \leq Z'_i \leq \xi_{i_2} \text{ para toda } i=1, 2, \dots, k] &= \int_0^{\infty} P[\xi_{i_1} \leq Z'_i \leq \xi_{i_2} \text{ para toda } i=1, 2, \dots, k | Y=y] dF(y) \\ &= \int_0^{\infty} \prod_{i=1}^k P[y\xi_{i_1} \leq X_i \leq y\xi_{i_2}] dF(y) \\ &= \int_0^{\infty} \prod_{i=1}^k g_i(y) dF(y) \end{aligned}$$

donde $g_i(y) = P[y\xi_{i_1} \leq X_i \leq y\xi_{i_2}]$.

Ahora bien, las $g_i(y)$ son funciones de y no-negativas acotadas y crecientes, con lo que las condiciones del Teorema 2 se satisfacen y, por lo tanto

$$\begin{aligned} P[\xi_{i_1} \leq Z'_i \leq \xi_{i_2} \text{ para toda } i=1, 2, \dots, k] &\geq \prod_{i=1}^k \int_0^{\infty} g_i(y) dF(y) \\ &= \prod_{i=1}^k P[\xi_{i_1} \leq Z'_i \leq \xi_{i_2}] \quad // \end{aligned}$$

Observe que bajo los supuestos de este teorema, la desigualdad de Bonferroni también es aplicable; sin embargo, la desigualdad (1.15) supera estrictamente a la desigualdad de Bonferroni, ya que si

$$P[\xi_{i_1} \leq z_i \leq \xi_{i_2}] = 1 - \alpha_i \quad i = 1, 2, \dots, k$$

es fácil verificar, usando inducción por ejemplo, que

$$\prod_{i=1}^k (1 - \alpha_i) > 1 - \sum_{i=1}^k \alpha_i, \quad 0 < \alpha_i < 1, \quad i = 1, 2, \dots, k \quad (k > 1).$$

Es interesante notar que el establecer la desigualdad (1.15), no hicimos referencia a ninguna función de distribución en particular, pero asumimos que los numeradores son estadísticamente independientes. El siguiente teorema, debido a Sidák (1967), nos muestra que este requerimiento de independencia puede ser eliminado si suponemos que los numeradores tienen una distribución normal y que el denominador común es una variable aleatoria con distribución χ .

Teorema 4. (Desigualdad de Sidák). Sea $X \in \mathbb{R}^k$ un vector aleatorio con $X \sim N(0, \sigma^2 R)$, $\sigma^2 > 0$ y R una matriz de correlaciones arbitraria. Sea además S^2 una variable aleatoria estadísticamente independiente de X tal que $\nu S^2 \sim \sigma^2 \chi^2(\nu)$. Entonces

$$P\left[\frac{|X_1|}{S} < \xi_1, \frac{|X_2|}{S} < \xi_2, \dots, \frac{|X_k|}{S} < \xi_k \right] \geq \prod_{i=1}^k P\left[\frac{|X_i|}{S} < \xi_i \right] \quad (1.16)$$

La demostración puede verse en Sidák (1967). /

En palabras, este teorema nos dice que la probabilidad de que un vector aleatorio normal, con un divisor común S

y su matriz de correlaciones arbitraria, esté contenido en un rectángulo centrado en su media, es siempre igual o mayor que la probabilidad correspondiente al caso en que las covarianzas son cero, es decir, cuando R es reemplazada por la matriz identidad.

La desigualdad de Sidák y algunas de sus aplicaciones son presentadas, entre otros autores, por Sidák (1967, 1971), Miller (1977) y Games (1977). En nuestro contexto, la desigualdad tiene especial relevancia, pues una gran cantidad de problemas estadísticos relacionados con las comparaciones múltiples y la construcción de ICS, involucran la distribución normal multivariada.

Ejemplo 1.2. Supongamos que tenemos una observación del vector aleatorio $x \in \mathbb{R}^k$ y de s^2 , una variable aleatoria estadísticamente independiente de x , con $\nu s^2 \sim \sigma^2 \chi^2(\nu)$ y $x \sim N(\theta, \sigma^2 V)$, donde V son constantes conocidas, $v_{ii} = \frac{1}{\eta_i}$ $i = 1, 2, \dots, k$. Queremos construir una familia de ICS para $\Psi = \{\theta_i; i = 1, 2, \dots, k\}$ con un CCS de al menos $1 - \alpha$, usando la siguiente familia de cantidades pivote

$$\{ z_i^! = \frac{\sqrt{\eta_i} |x_i - \theta_i|}{s}; i = 1, 2, \dots, k \}$$

es decir, cada $z_i^!$ tiene distribución t con ν grados de libertad. Dos posibilidades para conseguir los ICS son:

(i) utilizando la desigualdad de Sidák, si

$$P[Z_i^! \leq \xi_i] = (1 - \alpha)^{1/k}, i = 1, 2, \dots, k$$

entonces el i -ésimo intervalo del conjunto de ICS podría

expresarse como

$$\Lambda_i^B(x; (1-\alpha)^{1/k}) = \{ \theta : x - \frac{s}{\sqrt{\eta_i}} \xi_1 \leq \theta_i \leq x + \frac{s}{\sqrt{\eta_i}} \xi_1 \}, i=1,2,\dots,k$$

(ii) Utilizando la desigualdad de Bonferroni, si

$$P[Z_i^* \leq \xi_2] = 1 - \frac{\alpha}{k} \quad i = 1, 2, \dots, k$$

entonces el i -ésimo intervalo del conjunto de ICS podría expresarse como

$$\Lambda_i^S(x; \alpha/k) = \{ \theta_i : x_i - \frac{s}{\sqrt{\eta_i}} \xi_2 \leq \theta_i \leq x_i + \frac{s}{\sqrt{\eta_i}} \xi_2 \}$$

De este modo, tanto $\{ \Lambda_i^B(x; (1-\alpha)^{1/k}); i = 1, 2, \dots, k \}$ como $\{ \Lambda_i^S(x; \alpha/k); i = 1, 2, \dots, k \}$, constituyen una familia de ICS para Ψ , con un CCS no menor que $1 - \alpha$. Una revisión rápida a las tablas de la distribución t , nos muestra que $\xi_1 \leq \xi_2$, y por lo tanto

$$\Lambda_i^B(x; (1-\alpha)^{1/k}) \subseteq \Lambda_i^S(x; \alpha/k) \text{ para toda } i = 1, 2, \dots, k$$

es decir, los intervalos obtenidos con la desigualdad de Sidák son tanto o más cortos que los obtenidos con la desigualdad de Bonferroni, lo cual era de esperarse. \square

1.2. Estimación de ICS para Variables con Distribución Normal

1.2.1. Conceptos Preliminares

La mayoría de los problemas de construcción de ICS tienen la siguiente forma: se observa un vector X de dimensión k que tiene distribución $N(\theta, \sigma^2 V)$, donde la media θ es un punto desconocido en el espacio k -dimensional, σ^2 es número real positivo y V es una matriz $k \times k$ positiva definida (p.d.) o semi definida de constantes conocidas; se cuenta también con una -

observación de la variable aleatoria S^2 que es estadísticamente independiente de x y $\nu S^2 \sim \sigma^2 \chi^2(\nu)$ para algún entero ν , entonces decimos que S^2 es un estimador insesgado de σ^2 . Se está interesado en estimar un conjunto de ICS con un CCS de $1 - \alpha$ para una familia $\Psi = \{\psi_i; i \in I\}$ de funciones parametrales - con, posiblemente, infinito número de funciones, y cada ψ en Ψ es de la forma $\psi(\theta) = c'\theta$, donde c es un vector k -dimensional, es decir, ψ_i es una función lineal de los parámetros. Si definimos a C como el conjunto de los vectores de coeficientes que definen todas y cada una de las funciones ψ en Ψ , podemos caracterizar alternativamente la familia Ψ como

$$\Psi = \{c'\theta; c \in C\}$$

y el problema consiste en construir ICS para todas las funciones $c'\theta$ con c en C .

Este es, en general, el conjunto de supuestos con el que trabajaremos el resto del capítulo y que particularizaremos en cada inciso.

En la anterior sección hemos visto un conjunto de condiciones suficientes bajo las que la estimación es posible; - también de los teoremas ahí presentados se desprende que si - el conjunto C es finito, siempre es posible construir un conjunto de ICS para los que la cota mínima del CCS sea $1 - \alpha$. Ahora veremos algunos resultados que nos permitirán obtener - un método para la construcción de ICS con un CCS de exactamente $1 - \alpha$, bajo el conjunto de supuestos descrito en el párrafo anterior; además, veremos algunos resultados específicos -

relacionados con la construcción de ICS aproximados (aproximados en el mismo sentido que en los teoremas anteriores), para los que la longitud de los intervalos es posiblemente menor que la de los obtenidos usando los resultados del inciso (1.1.2).

Comenzaremos con un resultado debido a Scheffé relacionado con el caso en que el número de funciones en C es infinito, esto es, Ψ es un espacio o subespacio lineal arbitrario; luego veremos un grupo de resultados de algún modo asociados al nombre de Tukey relacionados con algunos casos en que el número de funciones de interés primario en Ψ es finito y, finalmente, veremos un resultado general debido a Uusipai-kka (1985).

1.2.2. El Método de Scheffé.

En muchos problemas estadísticos, para obtener un análisis completo, es usualmente necesario y deseable construir un conjunto de intervalos de confianza para algunas funciones lineales de los parámetros estimados simultáneamente. Uno de los procedimientos disponibles es el método de Scheffé. Una amplia discusión del método puede encontrarse en Scheffé (1953; 1959), Roy y Bose (1953) y Miller (1966).

Para motivar el método de Scheffé consideremos, a manera de ejemplo, el modelo lineal general $y = x\beta + e$ donde $y \in \mathbb{R}^n$ es el vector de observaciones, X la matriz diseño de orden $n \times m$ y rango $k \leq m \leq n$, β un vector de parámetros y e

una variable aleatoria con distribución $N(0, \sigma^2 I)$. Sea C un subespacio vectorial de \mathbb{R}^m de rango $q \leq k$ tal que $c'\beta$ es estimable para todo $c \in C$, es decir $E(c'\beta^0) = c'\beta$, donde β^0 es cualquier solución a las ecuaciones normales. El método de Scheffé establece que para este modelo

$$P[|c'(\beta^0 - \beta)| \leq [qS^2 F_\alpha c'(X'X)^{-1}c]^{1/2} \text{ para todo } c \in C] = 1 - \alpha \quad (1.17)$$

donde $(n - k)S^2 = Y'[I - X(X'X)^{-1}X']Y$ y F_α es el $(1 - \alpha)$ -percentil de la distribución F de Snedecor con q y $(n - k)$ grados de libertad. Este resultado nos permite construir un conjunto de ICS para todas las funciones estimables con c en C . La prueba será presentada en detalle después del teorema que da origen al método.

Para hacer más evidente la relación entre el método y el Lema 1, el teorema de Scheffé será desarrollado como caso especial de un procedimiento de maximización; para ello, antes del teorema presentaremos dos lemas; el primero es meramente técnico, y el segundo nos proporciona el procedimiento de maximización y es la generalización de un resultado presentado por Rao (1973).

Lema 2. (i) Si A y B son dos matrices cuadradas de orden n . Entonces los eigenvalores de AB coinciden con los de BA .

(ii) Si A y B son dos matrices de orden $m \times n$ y $n \times m$, respectivamente, donde $m \leq n$. Entonces los eigenvalores de BA son $(n - m)$ ceros y los m valores propios de AB .

La demostración se presenta en el apéndice. //

En lo que sigue, si A es una matriz cuadrada, con $C_m[A]$ denotaremos el mayor de los eigenvalores de A ; con $L[B]$ denotaremos el espacio lineal engendrado por las columnas de B , donde B es cualquier matriz.

Lema 3. Sea A una matriz simétrica y V una matriz p.d., ambas de orden n ; sea además B una matriz de orden $n \times k$ de rango completo para columnas. Entonces

$$(i) \quad \max_x \frac{x'Ax}{x'Vx} = C_m[B(B'VB)^{-1}B'A] \quad (1.18)$$

donde el máximo se toma sobre los vectores $x \neq 0$ tales que $x \in L[B]$; y

(ii) El máximo se alcanza cuando x es cualquier eigen vector de $B(B'VB)^{-1}B'A$ correspondiente a $C_m[B'(B'VB)^{-1}B'A]$

La demostración se presenta en el apéndice. //

Teorema 5. Supóngase que X es un vector aleatorio con distribución $N(\theta, \sigma^2 V)$ donde $\theta \in \mathbb{R}^k$ y $\sigma^2 > 0$ son parámetros desconocidos y V es una matriz p.d. de constantes conocidas. Supóngase además que S^2 es una variable aleatoria estadísticamente independiente de X y $VS^2 \sim \sigma^2 \chi^2(q)$. Sea B una matriz de orden $k \times q$ con rango completo por columnas. Entonces, para vectores c no nulos

$$(i) \quad \max_c \frac{|c'(X - \theta)|^2}{c'Vc} \sim \sigma^2 \chi^2(q) \quad y \quad (1.19)$$

donde el máximo se toma sobre los vectores $c \in L[B]$; y

$$(ii) P \left[|c'(X - \theta)| \leq [qS^2 F_{\alpha} c'Vc]^{1/2} \text{ para toda } c \in L[B] \right] = 1 - \alpha \quad (1.20)$$

donde F_{α} es el punto correspondiente al $(1 - \alpha)$ -percentil de la distribución F de Snedecor con q y v grados de libertad.

Demostración. $\frac{c'(X - \theta)}{(c'Vc)^{1/2}} \sim N(0, \sigma^2)$ para $c \neq 0$, $c \in \mathbb{R}^k$

Ahora observe que

$$|c'(X - \theta)|^2 = c'(X - \theta)(X - \theta)'c$$

y aplicando el lema anterior con $A = (X - \theta)(X - \theta)'$ vemos que

$$\max_{c \neq 0} \frac{|c'(X - \theta)|^2}{c'Vc} = C_m [B(B'VB)^{-1}B'(X - \theta)(X - \theta)']$$

donde el máximo se toma sobre los vectores no nulos $c \in L[B]$.

Por el Lema 2, tenemos que

$$\begin{aligned} \max_{c \neq 0} \frac{|c'(X - \theta)|^2}{c'Vc} &= C_m [(X - \theta)'B(B'VB)^{-1}B'(X - \theta)] \\ &= (X - \theta)'B(B'VB)^{-1}B'(X - \theta) \\ &= \gamma \end{aligned}$$

Como $B(B'VB)^{-1}B'V$ es idempotente de rango q , entonces $\gamma \sim \sigma^2 \chi^2(q)$ (Searle, 1971). Esto prueba (i).

Para probar (ii) note que $vS^2 \sim \sigma^2 \chi^2(v)$ y es, por hipótesis, independiente de γ . Se sigue entonces que

$$P [\gamma \leq qS^2 F_{\alpha}] = 1 - \alpha$$

de donde el resultado deseado se obtiene observando que

$$\frac{|c'(X - \theta)|^2}{c'Vc} \leq \gamma \text{ para todo } c \neq 0, c \in L[B] \quad //$$

Note que el resultado de este teorema no es exactamente el presentado en la motivación del método de Scheffé relacionado con el modelo $Y = X\beta + e$; de hecho si

$$\beta^0 = (X'X)^{-1}X'Y$$

entonces $\beta^0 \sim N(\beta, \sigma^2(X'X)^{-1})$ y $(X'X)^{-1}$ es singular y este caso no está contemplado en la demostración del teorema; sin embargo, el resultado puede obtenerse fácilmente observando lo siguiente:

(i) Si C es un subespacio vectorial para el que las funciones $c'\beta$ son estimables y si las columnas de B , una matriz $m \times q$ forman una base para C , entonces B puede escribirse como $B = (X'X)K$ para alguna matriz K de rango completo por columnas (Searle, 1971);

(ii) Si definimos $G = (X'X)^{-1}X'$, entonces

$$\begin{aligned} |c'(\beta^0 - \beta)| &= |c'GY - c'\beta| \text{ con } c \in L[B] \\ &= |a'XGY - c'\beta| \text{ con } a \in L[G'B] \end{aligned}$$

Para ver ésto, note que cualquier $c \in L[B]$ puede escribirse como $c = X'a$, donde $a \in L[G'B]$, pues

$$\begin{aligned} c &= Bh \text{ para algún } h \in \mathbb{R}^q \\ &= (X'X)Kh \end{aligned}$$

y

$$\begin{aligned}
 a &= G'Bh \\
 &= X(X'X)^{-1}X'XKh \\
 &= XKh
 \end{aligned}$$

más aún, $a \in L[G'B]$ implica que $a'XG = a'$ de modo que

$$\max_c \frac{|c'(\beta^0 - \beta)|^2}{c'(X'X)^{-1}c} = \max_a \frac{|a'(Y - X\beta)|^2}{a'a}$$

donde $c \in L[B]$ y $a \in L[G'B]$

(iii) $G'B$ es una matriz de rango completo por columnas, y

(iv) $G[I - XG] = 0$, de modo que S^2 y Y son estadísticamente independientes.

De esta manera, se satisfacen las hipótesis hechas en el teorema y entonces (1.17) es inmediata. \square

El teorema anterior proporciona un método para construir ICS para todas las funciones en la familia $\Psi = \{c'\theta; c \in C\}$, donde C es un subespacio lineal arbitrario; cada uno de estos intervalos puede escribirse como

$$I_{C, \alpha}(X; \alpha) = \{c'\theta: c'X - S\xi(c'Vc)^{1/2} \leq c'\theta \leq c'X + S\xi(c'Vc)^{1/2}\}, c \in C \quad (1.21)$$

donde $\xi^2 = qF_\alpha$, F_α es como en el teorema y q es la dimensión del subespacio.

La generalidad de este resultado radica en el hecho de que el subespacio lineal puede incluir casi cualquier cosa; esto lo hace aplicable a una amplia gama de situaciones prácticas, sin embargo, en muchos casos, se está interesado -

solamente en un conjunto finito de combinaciones lineales y, en tales casos, la longitud de los intervalos es posiblemente mayor que la de los intervalos obtenidos mediante algún otro porcedimiento; esta afirmación parece evidente, no obstante trataremos de verla explícitamente.

1.2.3. Interpretación Geométrica del Método de Scheffé.

Sabemos que si $c \in L[B]$ y si definimos $\lambda = B'\theta$, podemos escribir $c'\theta = h'\lambda$ con $h \in \mathbb{R}^q$, donde q es el rango de B , denotemos por $\hat{\lambda}$ el estimador insesgado de λ . De la discusión anterior se desprende que el mecanismo que permite la construcción de los ICS, lleva consigo la construcción de un elipsoide E , centrado en $\hat{\lambda}$, de la forma

$$E = \{\lambda: (\lambda - \hat{\lambda})' M^{-1}(\lambda - \hat{\lambda}) \leq \gamma\} \quad (1.22)$$

en el espacio de λ , de dimensión q . Además, por construcción, se tiene que

$$P[\gamma \leq S^2 \xi^2] = 1 - \alpha$$

donde ξ es como en (1.21), implica que el elipsoide E contiene al verdadero valor de λ con una probabilidad de $1 - \alpha$. Ahora bien, un punto pertenece al elipsoide, si y sólo si está situado en la intersección de todas las franjas formadas entre todas las parejas de hiperplanos paralelos, de dimensión $(q - 1)$, soporte de E , ortogonales a $h \in \mathbb{R}^q$, esto es

$$|h'(\lambda - \hat{\lambda})| \leq (\gamma h'Mh)^{1/2}$$

para todo h no-nulo. Luego, si hay una probabilidad de $1 - \alpha$ de que el elipsoide E contenga al verdadero valor de λ , habrá una probabilidad de $1 - \alpha$ de que lo contenga la intersección de todas las franjas. En otras palabras, podemos afirmar que la probabilidad de que, para todos los vectores no-nulos $h \in \mathbb{R}^q$, los intervalos

$$(h'\hat{\lambda} - s^2 \xi^2 \sqrt{h'Mh}, h'\hat{\lambda} + s^2 \xi^2 \sqrt{h'Mh})$$

contengan simultáneamente el verdadero valor de $h'\lambda$, es $1 - \alpha$.

Ahora, supongamos que solamente tenemos un conjunto finito de vectores h ; entonces, el conjunto de los puntos, en el espacio de λ , contenidos en la intersección de todas las franjas formadas entre las parejas de hiperplanos paralelos soporte de E , correspondientes a ese conjunto finito de vectores, es un poliedro de dimensión q que circunscribe al elipsoide; luego, la probabilidad asociada al poliedro excede a la asociada al elipsoide.

Del párrafo anterior se desprende que si estamos interesados en un conjunto finito de funciones, es probable que podamos encontrar algún procedimiento para la construcción de ICS, alternativo al método de Scheffé, para el que la longitud de los intervalos sea no mayor que la de los obtenidos con este último, cuando el CCS sea el mismo para ambos.

1.2.4. El Método de Tukey

Uno de los casos más comunes en que se está interesado en obtener un conjunto de ICS para una familia Ψ de funciones parametrales con un número finito de funciones, se presenta cuando

$$\Psi = \{\theta_i - \theta_j; i \neq j, i, j = 1, 2, \dots, k\}.$$

Un método para la construcción de ICS, diseñado específicamente para algunas situaciones en que este caso se presenta, es el método de Tukey.

A diferencia del método de Scheffé, que utiliza los puntos porcentuales de la distribución F de Snedecor, el método de Tukey involucra la distribución del rango studentizado; a continuación presentamos la definición de esta variable aleatoria.

Definición 1. Sea X_1, X_2, \dots, X_k una muestra aleatoria de $N(\theta, \sigma^2)$, $\theta \in \mathbb{R}$ y $\sigma^2 > 0$; sea S^2 una variable aleatoria independiente de $\{X_i\}$ tal que $\nu S^2 \sim \sigma^2 \chi^2(\nu)$. El rango studentizado $Q(k, \nu)$ se define como

$$Q(k, \nu) = (\text{máx} \{X_1, X_2, \dots, X_k\} - \text{mín} \{X_1, X_2, \dots, X_k\})/S$$

La distribución del rango studentizado puede encontrarse en Miller (1966). Tablas para algunos percentiles de esta distribución se presentan en muchos textos y artículos, como por ejemplo, en Scheffé (1959), Harter (1960) y Miller (1966).

En lo que sigue denotaremos por C_p al conjunto

$$\{e_i - e_j; i \neq j, i, j = 1, 2, \dots, k\}$$

donde e_i es el i -ésimo vector unitario de la base canónica para \mathbb{R}^k .

En el siguiente teorema se alude un caso particular, el más conocido quizás, del método de Tukey.

Teorema 6. Sea $X \in \mathbb{R}^k$ un vector aleatorio con distribución $N(\theta, a^2 \sigma^2 I)$, donde $\theta \in \mathbb{R}^k$ y $\sigma^2 > 0$ son parámetros desconocidos y $a > 0$ es una constante conocida. Entonces

$$P[|c'(X - \theta)| \leq a s q_\alpha \text{ para toda } c \in C_p] = 1 - \alpha \quad (1.23)$$

donde q_α es el punto correspondiente al $(1 - \alpha)$ percentil de la distribución del rango studentizado $Q_{(k, \nu)}$ y S es como antes

Demostración. Definamos $Y = (X - \theta)$, de este modo se tiene que

$$|c'(X - \theta)| \leq a s \xi \text{ para toda } c \in C_p$$

es equivalente a

$$|Y_i - Y_j| \leq a s \xi \text{ para toda } i \neq j, i, j = 1, 2, \dots, k$$

Ahora observe que las desigualdades de este conjunto se satisfacen simultáneamente si y sólo si

$$(\max\{X_1, X_2, \dots, X_k\} - \min\{X_1, X_2, \dots, X_k\}) / a s \leq \xi$$

Como $Y \sim N(0, a^2 \sigma^2 I)$ y, además, Y es independiente de s^2 , el resultado es una consecuencia inmediata de la definición de

rango studentizado. /

Nos es bastante familiar la aplicación de este resultado en problemas de comparación de medias o en la construcción de ICS para todas las diferencias entre efectos de tratamientos, cuando ensayamos un conjunto de k tratamientos en un diseño completamente al azar con igual número de repeticiones. Hay abundante literatura sobre este y algunos otros ejemplos de problemas estadísticos que se solucionan mediante la aplicación directa del teorema anterior, sin embargo, como mencionamos previamente, éste es sólo un caso particular del método de Tukey.

Si observamos detenidamente la demostración del teorema anterior, veremos que ésta se basa fundamentalmente en la maximización del valor absoluto de las diferencias entre parejas de variables aleatorias independientes e idénticamente distribuidas, es decir, se basa en el hecho de que el valor absoluto de la diferencia entre cualquier par de variables es igual o menor que el rango, lo cual, aunado al requerimiento de distribución, nos permite utilizar la definición de rango studentizado. Ahora bien, como el rango de un conjunto $\{X_i\}$ de variables es lo mismo que el máximo de $\{|X_i - X_j|; i \neq j\}$, parece intuitivamente claro que la distribución del rango está determinada por la estructura de la matriz de varianzas-covarianzas de $\{X_i - X_j; i \neq j\}$. Por esta razón, es posible encontrar una generalización del teorema anterior para el caso en que $X \sim N(\theta, \sigma^2 V)$, que nos

permita seguir utilizando la distribución del rango studentizado en la construcción de ICS para la familia de funciones $\Psi = \{c'\theta; c \in C_p\}$, cuando los elementos de V satisfacen cierta restricción, a saber, cuando $v_{ii} - 2v_{ij} + v_{jj} = 2a^2$ para toda $i \neq j$, donde $a > 0$ es una constante.

Definición 2. Una matriz simétrica V de orden k para la que se satisface que $v_{ii} - 2v_{ij} + v_{jj} = 2a^2$ para toda $i \neq j$, $i = 1, 2, \dots, k$, donde a es una constante positiva, se dice que posee la propiedad de ser balanceada; para abreviar diremos simplemente que es balanceada.

Para establecer el teorema en el que se fundamenta el método de Tukey para el caso balanceado, y en posteriores extensiones del método, haremos uso de algunas de las propiedades de las matrices balanceadas. Estas propiedades se presentan en los dos lemas siguientes.

Denotemos por L_c el espacio lineal de contrastes, es decir, $L_c \subset \mathbb{R}^k$ y toda c en L_c cumple con $\sum_{i=1}^k c_i = 0$, donde $c' = (c_1, c_2, \dots, c_k)$

Lema 4. Si una matriz V de orden k satisface que $v_{ii} + 2v_{ij} + v_{jj} = 2a^2$ para toda $i \neq j$, $i, j = 1, 2, \dots, k$, entonces

$$c'Vc = a^2c'c \text{ para toda } c \text{ en } L_c$$

Demostración. Hagamos que $B' = [I_{k-1}, -\mathbf{1}]$ y note que cualquier $c \in L_c$ puede escribirse como $c = Bh$, $h \in \mathbb{R}^{k-1}$, y además $c'c = h'[I_{k-1} + \mathbf{1}\mathbf{1}']h$.

Denotemos por b_i la i -ésima columna de B , entonces

$$\begin{aligned} b_i' V b_i &= v_{ii} - 2v_{ik} + v_{kk}, \quad i = 1, 2, \dots, k-1 \\ &= 2a^2 \end{aligned}$$

$$\begin{aligned} b_i' V b_j &= v_{ij} - v_{ik} - v_{jk} + v_{kk}, \quad i \neq j, \quad i, j = 1, 2, \dots, k-1 \\ &= a^2 \end{aligned}$$

ya que $2a^2 - v_{ii} = v_{kk} - 2v_{ik}$, $i = 1, 2, \dots, k-1$

Entonces

$$B'VB = a^2 [I_{k-1} + \mathbb{1}\mathbb{1}']$$

de donde se sigue que

$$\begin{aligned} c'Vc &= h'B'VBh \\ &= a^2 h' [I_{k-1} + \mathbb{1}\mathbb{1}'] h \\ &= a^2 c'c \end{aligned}$$

//

El siguiente lema muestra una característica general de las matrices balanceadas que puede ser usada como definición alternativa.

Lema 5. Una matriz simétrica V , de orden k , satisface que $c'Vc = a^2 c'c$ para toda $c \in L_c$, si y sólo si puede encontrarse una matriz Q tal que $QQ' = V$ y $Q'c = ac$ para toda $c \in L_c$.

La demostración se presenta en el apéndice. //

Teorema 7. Supongamos que $x \in \mathbb{R}^k$ es un vector aleatorio con distribución $N(\theta, \sigma^2 V)$, donde $\theta \in \mathbb{R}^k$ y $\sigma^2 > 0$ son parámetros desconocidos y V es una matriz p.d. balanceada, es decir, -

$c'Vc = a^2c'c$ para toda $c \in L_c$, donde $a > 0$ es una constante conocida. Supóngase además, que S^2 es una variable aleatoria estadísticamente independiente de X tal que $vS^2 \sim \sigma^2\chi^2(v)$. Entonces.

$$P\{|c'(X - \theta)| \leq aSq_\alpha \text{ para toda } c \in C_p\} = 1 - \alpha \quad (1.24)$$

donde q_α es el punto que corresponde al $(1 - \alpha)$ -percentil de la distribución del rango studentizado para k variables con v grados de libertad.

Demostración. Como V es p.d. y balanceada, del Lema 5 se sigue que existe una matriz inversible Q tal que $QQ' = V$ y $Q'c = ac$ para toda $c \in L_c$ y en particular $Q'c = ac$ para toda $c \in C_p$.

De este modo si $c \in C_p$

$$\begin{aligned} |c'(X - \theta)| &= |c'QQ^{-1}(X - \theta)| \\ &= a|c'y| \quad \text{donde } y \sim N(0, \sigma^2I) \\ &\leq aR_y \text{ para toda } c \in C_p \end{aligned}$$

donde $R_y = \text{máx}\{Y_1, Y_2, \dots, Y_k\} - \text{mín}\{Y_1, Y_2, \dots, Y_k\}$

Ahora, empleando un argumento similar al de la demostración del teorema anterior, observe que el evento $R_y \leq \xi$ ocurre, si y sólo si, el evento $|c'(X - \theta)| \leq a\xi$ para toda $c \in C_p$ ocurre, es decir, los dos eventos son equivalentes, de modo que

$$P[R_y \leq \xi] = P\{|c'(X - \theta)| \leq a\xi \text{ para toda } c \in C_p\}$$

Finalmente, observe que $R_y/S \sim Q(k, \nu)$, es decir, tiene la distribución del rango studentizado, entonces

$$P[\underline{R}_y \leq S q_\alpha] = 1 - \alpha$$

lo cual prueba el resultado del teorema. //

En resumen, de lo hasta aquí expuesto acerca del método de Tukey, sabemos que este método nos permite obtener un conjunto de ICS para la familia $\Psi = \{c'\theta; c \in C_p\}$; sabemos además que es aplicable cuando la matriz de varianzas-covarianzas de X es balanceada. Cuando X , S^2 y V satisfacen los supuestos del teorema anterior, cada uno de los intervalos proporcionados por el método puede escribirse como

$$I_c(X; \alpha) = \{c'\theta: c'X - aSq_\alpha \leq c'\theta \leq c'X + aSq_\alpha\}; c \in C_p \quad (1.25)$$

donde a y q_α son como en el teorema. Estos intervalos son óptimos en el sentido de que, para esta familia Ψ y bajo los mismos supuestos, ningún otro procedimiento proporciona intervalos con una longitud promedio más corta con el mismo CCS.

Es conveniente remarcar que en el caso balanceado el CCS de la familia de ICS, obtenida por el método de Tukey, tiene un valor de exactamente $1 - \alpha$; más adelante veremos una versión modificada del método para el caso no-balanceado con la que esto no ocurre.

Ejemplo 1.3. Existen problemas de interés en los que la matriz de varianzas-covarianzas cumple con los supuestos del teorema anterior, con lo que se posibilita la aplicación del método de Tukey. A continuación citaremos tres casos en los que la matriz p.d. V es balanceada, es decir, $c'Vc = a^2 c'c$ para toda $c \in L_c$.

$$(i) \text{ Cuando } \text{var}(X_i) = \frac{\sigma^2}{n} \quad i = 1, 2, \dots, k$$

$$\text{cov}(X_i, X_j) = 0 \quad i \neq j, \quad i, j = 1, 2, \dots, k$$

Es decir, cuando $V = \frac{1}{n} I$ y entonces $a = \left(\frac{1}{n}\right)^{1/2}$

Un ejemplo de este caso se presenta cuando x es el vector solución a las ecuaciones normales en un diseño completamente al azar con igual número de repeticiones por tratamiento.

$$(ii) \text{ Cuando } \text{var}(X_i) = \frac{\sigma^2}{n}; \quad i = 1, 2, \dots, k$$

$$\text{cov}(X_i, X_j) = \frac{\rho}{n} \sigma^2; \quad i \neq j; \quad i, j = 1, 2, \dots, k \quad -\frac{1}{k-1} < \rho < 1$$

Es decir $V = \frac{1}{n} [(1 - \rho)I + \rho \mathbb{1}\mathbb{1}']$ y entonces $a = \left(\frac{1 - \rho}{n}\right)^{1/2}$

Un ejemplo de este caso se presenta cuando x es el vector de estimaciones de los efectos de tratamientos de un diseño en bloques incompletos balanceados.

(iii) En general, cuando V puede escribirse como

$$V = a^2 [I - \mathbb{1}\mathbb{1}'] + \frac{1}{2} [b\mathbb{1}' + \mathbb{1}b']$$

donde b es el vector formado por los elementos diagonales de V .

Note que este caso incluye a los dos anteriores. Hay diseños experimentales para los que los elementos en el vector solución a las ecuaciones normales no son ni homocedásticos ni están equicorrelacionados, sin embargo, su matriz de varianzas-covarianzas tiene esta estructura y, entonces, para estos diseños la aplicación del método de Tukey también es posible. Algunos ejemplos de tales diseños pueden encontrarse en Pearce (1964).

1.2.5. Extensión del método de Tukey a todos los contrastes.

En el teorema anterior la familia de funciones involucrada es finita. Ahora veremos como extender el método al caso en que se consideran todos los contrastes posibles. Primero estableceremos el siguiente resultado auxiliar.

Lema 6. Sea $y \in \mathbb{R}^k$ y $R_y = \max\{Y_1, Y_2, \dots, Y_k\} - \min\{Y_1, Y_2, \dots, Y_k\}$.

Entonces, para todo $c \in L_c$ se tiene que

$$|c'y| \leq R_y \frac{1}{2} \sum_{i=1}^k |c_i| \quad (1.25)$$

donde $c' = (c_1, c_2, \dots, c_k)$.

Demostración. $\sum c_i = 0$ y la desigualdad del triángulo implican que

$$\begin{aligned} |c'y| &= |c'(y - a \mathbb{1})| \text{ para cualquier } a \in \mathbb{R} \\ &\leq \sum_{i=1}^k |c_i (Y_i - a)| \end{aligned}$$

eligiendo $a = \frac{1}{2} [\max\{Y_1, Y_2, \dots, Y_k\} + \min\{Y_1, Y_2, \dots, Y_k\}]$
y observando que para esta elección

$$|z_i - a| \leq \frac{1}{2} R_y \quad i = 1, 2, \dots, k$$

se obtiene el resultado deseado. //

La extensión al método de Tukey se desprende del siguiente teorema.

Teorema 8. Bajo las condiciones del teorema anterior.

$$P[|c'(X - \theta)| \leq a q_\alpha \left(\frac{1}{2} \sum_{i=1}^k |c_i|\right) \text{ para toda } c \in L_c] = 1 - \alpha \quad (1.26)$$

donde las constantes a y q_α son constantes.

Demostración. Aplicando los Lemas 5 y 6 se tiene que

$$\begin{aligned} |c'(X - \theta)| &= a |c'Y| \text{ donde } Y \sim N(0, \sigma^2 I) \\ &\leq a R_y \frac{1}{2} \sum_{i=1}^k |c_i| \end{aligned}$$

Luego $R_y \leq \xi$ implica que $|c'(X - \theta)| \leq a \xi \frac{1}{2} \sum_{i=1}^k |c_i|$ para toda $c \in L_c$. Inversamente, si la última desigualdad es válida para toda $c \in L_c$, debe tenerse que $R_y \leq \xi$ de modo que

$$P[R_y \leq \xi] = P[|c'(X - \theta)| \leq a \xi \frac{1}{2} \sum_{i=1}^k |c_i| \text{ para toda } c \in L_c]$$

el resto de la prueba es similar a la del teorema anterior. /

Esta extensión es de gran utilidad cuando se tiene interés primario en construir ICS para el conjunto $\{c'\theta; c \in L_c\}$ de funciones parametrales pero, adicionalmente, se desea que la familia de ICS incluya también intervalos para todos o algunos contrastes específicos entre medias, sin que por ello se incremente la longitud de los intervalos para las funciones de mayor interés. Ahora bien, si comparamos la longitud de

los intervalos obtenidos por este método con los de los obtenidos mediante el método de Scheffé, tenemos que para algunos patrones de configuración de los contrastes, el método de Tukey proporciona intervalos con menor longitud y, en este sentido, es más eficiente que el método de Scheffé para esos contrastes.

1.2.6. Comparación entre el Método de Tukey y el Método de Scheffé.

A continuación veremos una comparación entre el método de Tukey y el método de Scheffé en el caso balanceado. Hasta aquí sabemos que el primer método es mejor cuando de construir un conjunto de ICS para todas las diferencias entre medias se trata y que el segundo lo es cuando los ICS son para todos los contrastes posibles entre medias; sin embargo, la comparación puede ser de utilidad para saber que pasa en casos "intermedios".

En el caso que nos ocupa, la longitud de los intervalos con el método de Scheffé es

$$2\sqrt{(k-1)F_{\alpha}} \text{ aS } \|c\| \quad c \in L_c$$

y con el método de Tukey es

$$q_{\alpha} \text{ aS } \sum_{i=1}^k |c_i| \quad c \in L_c$$

o equivalentemente

$$q_{\alpha} \text{ aS } \|c\| \sum_{i=1}^k |w_i| \quad \text{donde } c = \|c\| w$$

donde F_{α} es el punto correspondiente al $(1-\alpha)$ -percentil de la

distribución F de Snedecor con $k - 1$ y ν grados de libertad y q_α es el punto que corresponde al $(1-\alpha)$ -percentil de la distribución del rango studentizado con k variables y ν grados de libertad.

Un posible índice de comparación es la precisión o eficiencia relativa de los intervalos obtenidos por el método de Tukey con respecto a los obtenidos por el método de Scheffé, expresada como R la razón de las longitudes. De este modo, para algún contraste en particular, $R = 1$ indica que los métodos son equivalentes, $R > 1$ indica que el método de Tukey es más eficiente para ese contraste y viceversa, $R < 1$ indica que el de Scheffé es mejor.

Ahora bien, como los contrastes pueden involucrar un número r de variables, esto es, r coeficientes del contraste son diferentes de cero, podemos definir R_r como la precisión relativa de un contraste con un número fijo r de variables involucradas en el contraste. Por otro lado, cualquier R_r siempre puede escribirse en función de un contraste normalizado con un número fijo r de coeficientes diferentes de cero. Denotemos por $\{w|r\}$ el conjunto de tales contrastes; entonces

$$R_r = \frac{\sqrt{F_\alpha}}{q_\alpha} \frac{\sqrt{k-1}}{\sum |w_i|}; \quad w \in \{w|r\}, \quad r=2,3,\dots,k$$

y como para cualquier r fijo se tiene que

$$\min \sum |w_i| = 2 \sqrt{1 - \frac{1}{r}} \quad r=2,3,\dots,k$$

$$\max \sum |w_i| = \begin{cases} \sqrt{r} & \text{si } r \text{ es par} \\ \sqrt{r - \frac{1}{r}} & \text{si } r \text{ es impar} \end{cases}$$

entonces tenemos que

$$\begin{aligned} \text{máx } R_r &= \frac{\sqrt{F_\alpha}}{q_\alpha} \sqrt{\frac{r(k-1)}{r-1}} \quad r = 2, 3, \dots, k \\ \text{mín } R_r &< \begin{cases} 2 \frac{\sqrt{F_\alpha}}{q_\alpha} \sqrt{\frac{k-1}{r}} & \text{si } r \text{ es par} \\ 2 \frac{\sqrt{F_\alpha}}{q_\alpha} \sqrt{\frac{r(k-1)}{(r+1)(r-1)}} & \text{si } r \text{ es impar} \end{cases} \end{aligned}$$

donde máx y mín se toman sobre $\{w|r\}$.

Es fácil verificar que $\text{máx } R_r = \text{mín } R_r$ para $r = 2, 3$; también es fácil ver que el máximo sobre todos los contrastes se alcanza cuando $r = 2$ y el mínimo cuando $r = k$ y el grupo de coeficientes con signo positivo es de igual o casi igual tamaño que el de los coeficientes con signo negativo. Por otra parte, observemos que para k fija, R es una función decreciente de los grados de libertad para un contraste dado, por lo que valores grandes de v pueden favorecer al método de Scheffé; lo contrario sucede si fijamos los grados de libertad, ya que así R es una función creciente en k , entonces el número de configuraciones que favorecen al método de Tukey crece con k . Ahora bien, si $k > 3$ tenemos que máx R y mín R sobre todos los contrastes son siempre mayor y menor que la unidad respectivamente, de modo que habrá patrones de configuración que favorezcan a uno u otro método.

De lo anterior se desprende que para los casos "intermedios", un método será más eficiente que el otro para algunos patrones de configuración de los contrastes y menos

eficiente para otros, de manera que, salvo en situaciones muy particulares, la elección de uno u otro método deberá hacerse en base al problema particular que estemos considerando.

1.2.7. Extensión del método de Tukey a todos los contrastes en el caso no-balanceado.

Hasta aquí hemos visto que el método de Tukey nos permite construir un conjunto de ICS para la familia $\{c'\theta; c \in L_c\}$ de funciones parametrales en el caso balanceado. A continuación veremos un resultado que nos permite, para esta misma familia de funciones parametrales, aplicar el método de Tukey en el caso no-balanceado. El resultado se basa en la existencia de cierto tipo de matrices Q que satisfacen $QQ' = V$ y, además, $Q'c \in L_c$ siempre que $c \in L_c$.

Teorema 9. Supóngase que X es un vector aleatorio con distribución $N(\theta, \sigma^2 V)$, donde $\theta \in \mathbb{R}^k$ y $\sigma^2 > 0$ son parámetros desconocidos y V es una matriz p.d. de constantes conocidas. Supóngase además que S^2 es una variable aleatoria estadísticamente independiente de X y que $v^2 S^2 \sim \sigma^2 \chi^2(v)$.

(i) Si Q es una matriz que satisface $QQ' = V$ y, además $Q'c \in L_c$ siempre que $c \in L_c$, entonces

$$P[|c'(X - \theta)| \leq S q_\alpha \left(\frac{1}{2} \sum |(Q'c)_i|\right) \text{ para toda } c \in L_c] = 1 - \alpha \quad (1.27)$$

donde q_α es como antes y $(Q'c)_i$ es el i -ésimo elemento del vector $Q'c$.

(ii) Existe al menos una matriz Q que satisface lo anterior.

Demostración. Si Q es como en el teorema, podemos definir $\bar{d} = Q'c$ y $Y = Q^{-1}(X - \theta)$ de modo que $\bar{d} \in L_c$ y $Y \sim N(0, \sigma^2 I)$ y como $|c'(X - \theta)| = |d'Y|$ la prueba de la primera parte se reduce a una aplicación del Teorema 8.

Para ver la parte (ii) observe que es necesario y suficiente mostrar que siempre podemos encontrar una matriz Q tal que $Q\mathbb{1} = a\mathbb{1}$ para alguna constante positiva a y $QQ' = V$.

Supongamos que Q_0 es cualquier matriz para que $Q_0 Q_0' = V$. Si hacemos que $Q = Q_0 H$ para alguna matriz ortogonal H , entonces $QQ' = V$. Ahora bien, $Q\mathbb{1} = a\mathbb{1}$ implica que $H\mathbb{1} = aQ_0^{-1}\mathbb{1}$, es decir, H debe ser una matriz ortogonal tal que la suma de sus filas es proporcional a la suma de las filas de Q_0^{-1} ; esto puede solucionarse de muchas maneras, como por ejemplo, haciendo que $H = FG'$ donde F y G son matrices ortogonales tales que sus primeras columnas son $\frac{a}{\sqrt{k}}Q_0^{-1}\mathbb{1}$ y $\frac{1}{\sqrt{k}}\mathbb{1}$ respectivamente, pues

$$H = \frac{a}{k} Q_0^{-1} \mathbb{1} \mathbb{1}' + f_2 g_2' + \dots + f_k g_k'$$

de modo que $H\mathbb{1} = aQ_0^{-1}\mathbb{1}$ ya que $g_i'\mathbb{1} = 0$ para $i = 2, 3, \dots, k$ //

Este teorema es interesante pues en él se trata el caso no-balanceado, sin embargo, cuando aplicamos en la construcción de ICS, tenemos que la longitud de los intervalos depende en mucho de la arbitrariedad con la que Q sea elegida. El lector interesado en este tema puede consultar el

trabajo de Genizi y Hochberg (1978).

1.2.8. El Método de Tukey para todas las Funciones Lineales

Sabemos ya que el método de Tukey puede ser utilizado para construir un conjunto de ICS para todas las diferencias entre medias cuando la matriz de varianzas-covarianzas es balanceada; también hemos visto la generalización del método al subespacio lineal de contrastes tanto al caso balanceado como al no-balanceado. Ahora consideraremos otra extensión del método, también debida a Tukey, que nos permite efectuar la construcción de ICS para todas las funciones lineales, sin necesidad de suponer que la matriz de varianzas-covarianzas es balanceada. Esta extensión utiliza la distribución del rango aumentado studentizado que a continuación definimos.

Definición 3. Sea $X \in \mathbb{R}^k$ un vector aleatorio con distribución $N(0, \sigma^2 I)$ donde $\sigma^2 > 0$. Sea S^2 una variable aleatoria estadísticamente independiente de X y tal que $\nu S^2 \sim \sigma^2 \chi^2(\nu)$. Si

$$R_X = \max\{X_1, X_2, \dots, X_k\} - \min\{X_1, X_2, \dots, X_k\}$$

el rango aumentado studentizado $Q_{(k, \nu)}$ se define como

$$Q_{(k, \nu)} = [\max\{R_X, |X_1|, |X_2|, \dots, |X_k|\}] / S$$

Tablas para la distribución de esta variable pueden encontrarse en Stoline (1978).

Para cualquier $c \in \mathbb{R}^k$ denotaremos por $\bar{m}(c)$ el valor

$$\max \left\{ \sum_{i=1}^k c_i^+, \sum_{i=1}^k c_i^- \right\}$$

donde $c_i^+ = \max \{c_i, 0\}$ y $c_i^- = \max \{-c_i, 0\}$, $c' = (c_1, c_2, \dots, c_k)$

Teorema 10. Supóngase que $X \in \mathbb{R}^k$ es un vector aleatorio con distribución $N(\theta, \sigma^2 V)$ donde θ y $\sigma^2 > 0$ son parámetros desconocidos. Adicionalmente, supóngase que S^2 es una variable aleatoria estadísticamente independiente de X y que $\nu S^2 \sim \sigma^2 \chi^2(\nu)$. Sea Q una matriz tal que $QQ' = V$, entonces

$$P[|c'(X - \theta)|^2 \leq S q_{\alpha}^1 \bar{m}(Q'c) \text{ para todo } c \in \mathbb{R}^k] = 1 - \alpha \quad (1.28)$$

donde q_{α}^1 es el punto correspondiente al $(1 - \alpha)$ percentil de la distribución del rango aumentado studentizado con k variables y ν grados de libertad.

Demostración. Definamos $d = Q'c$, entonces

$$\begin{aligned} |c'(X - \theta)| &= |c'QQ^{-1}(X - \theta)| \\ &= |d'Y| \quad Y \sim N(0, \sigma^2 I) \end{aligned}$$

Ahora observemos que

$$\begin{aligned} d'Y &= [d', -d'\Pi] \begin{bmatrix} Y \\ 0 \end{bmatrix} \\ &= \tilde{d}'\tilde{Y} \end{aligned}$$

y además $\tilde{d} \in L_c$. Entonces por el Lema 6

$$\begin{aligned} |d'Y| &= |\tilde{d}'\tilde{Y}| \\ &\leq R_{\tilde{Y}} \frac{1}{2} \sum_{i=1}^{k+1} |\tilde{d}_i| \end{aligned}$$

$$\text{pero } \sum_{i=1}^{k+1} |\tilde{d}_i| = \sum_{i=1}^k d_i^+ + \sum_{i=1}^k d_i^- + \left| \sum_{i=1}^k d_i^+ - \sum_{i=1}^k d_i^- \right| = 2\bar{m}(d)$$

y como $R\tilde{y} = \text{máx} \{R, |Y_1|, |Y_2|, \dots, |Y_k|\}$ entonces $\frac{R\tilde{y}}{S}$ es por definición el rango aumentado studentizado.

Así tenemos que $P[R\tilde{y} \leq S q_\alpha] = 1 - \alpha$ implica el resultado del teorema.

Observemos que la matriz Q en el teorema no es única, por tanto las propiedades de algún procedimiento para la construcción de ICS para cualquier conjunto específico de funciones lineales, dependerá en gran medida de la elección de Q .

Información adicional de este método puede encontrarse en Scheffé (1959), Miller (1966), Genizi y Hochberg (1978) entre otros trabajos.

Un caso particular de este teorema se presenta cuando V es una matriz diagonal y tomamos Q como

$$Q = \text{diag} \{ \sqrt{v_{11}}, \sqrt{v_{22}}, \dots, \sqrt{v_{kk}} \} .$$

A este caso, principalmente asociado a la construcción de ICS para todas las diferencias entre medias, se le conoce como método de Spjøtvoll-Stoline; en este caso (1.28) se transforma en

$$P[|c'(X-\theta)| \leq S q_\alpha \text{ máx}\{\sqrt{v_{ii}}, \sqrt{v_{jj}}\} \text{ para toda } c = e_i - e_j, i \neq j, i, j = 1, 2, \dots, k] \geq 1 - \alpha \quad (1.29)$$

Este método fue propuesto como una solución en la construcción de ICS para todas las diferencias entre medias para el caso en que las variables son estadísticamente independientes, cuando se pensaba que la corrección consistente en reemplazar la constante a en el Teorema 7 por $[(v_{ii} + v_{jj})/2]^{1/2}$,

no tenía control sobre la cota mínima del CCS, más adelante trataremos este tema.

1.2.9. El Método de Tukey-Roy-Bose

Otro método para la construcción de ICS, basado en la distribución el máximo módulo studentizado, fue propuesto independientemente por Tukey (1951) y por Roy y Bose (1953). El método, aplicable en el caso en que las variables son estadísticamente independientes, originalmente fue diseñado para construir ICS para la familia $\Psi = \{\theta_i; i = 1, 2, \dots, k\}$ de funciones parametrales, aunque también se conoce una extensión al método que nos permite construir ICS para todas las funciones lineales. A continuación veremos el teorema en que se basa el método de Tukey-Roy-Bose y una de sus posibles aplicaciones.

Definición 4. Sea $X \in \mathbb{R}^k$ un vector aleatorio con distribución $N(0, \sigma^2 I)$ y sea S^2 una variable aleatoria estadísticamente independiente de X tal que $\nu S^2 \sim \sigma^2 \chi^2(\nu)$. El máximo módulo studentizado $M(k, \nu)$ se define como

$$M(k, \nu) = [\text{máx}\{|X_1|, |X_2|, \dots, |X_k|\}] / S$$

Tablas de la distribución de esta variable aleatoria pueden encontrarse en Stoline y Uri (1979).

Teorema 11. Supóngase que $X \in \mathbb{R}^k$ es un vector aleatorio con distribución $N(\theta, \sigma^2 V)$, donde θ y $\sigma^2 > 0$ son parámetros desconocidos y V es una matriz diagonal p.d.; supóngase, además, que S es una variable aleatoria estadísticamente

independiente de X y que $vS^2 \sim \sigma^2 \chi^2(v)$. Entonces

$$(i) \quad P[|X_i - \theta_i| \leq Sm_\alpha \sqrt{v_{ii}} \text{ para toda } i=1,2,\dots,k] = 1 - \alpha \quad (1.30)$$

$$(ii) \quad P[|c'(X - \theta)| \leq Sm_\alpha \sum_{i=1}^k |c_i \sqrt{v_{ii}}| \text{ para toda } c \in \mathbb{R}^k] = 1 - \alpha \quad (1.31)$$

donde m_α es el punto que corresponde al $(1 - \alpha)$ percentil de la distribución del máximo módulo studentizado con k variables y v grados de libertad.

Demostración. Observe que (ii) implica (i), por lo tanto, probaremos solamente la segunda parte.

$$\text{Definamos } Q = \text{diag}\{\sqrt{v_{11}}, \sqrt{v_{22}}, \dots, \sqrt{v_{kk}}\}$$

entonces, para cualquier $c \in \mathbb{R}^k$

$$\begin{aligned} |c'(X - \theta)| &= |c'QQ^{-1}(X - \theta)| \\ &= |c'QY| \end{aligned}$$

donde $Y \sim N(0, \sigma^2 I)$.

Ahora, si $M = \max\{|Y_1|, |Y_2|, \dots, |Y_k|\}$, entonces

$$|c'(X - \theta)| \leq M \sum_{i=1}^k |c_i \sqrt{v_{ii}}|$$

Luego, $M \leq \xi$ implica que $|c'(X - \theta)| \leq \xi \sum_{i=1}^k |c_i \sqrt{v_{ii}}|$ para toda $c \in \mathbb{R}^k$. Inversamente, si la última desigualdad es válida para toda c no nula, debe tenerse que $M \leq \xi$ y, por lo tanto

$$P[M \leq \xi] = P[|c'(X - \theta)| \leq \xi \sum_{i=1}^k |c_i \sqrt{v_{ii}}| \text{ para toda } c \text{ no nula } c \in \mathbb{R}^k]$$

y como M/S es por definición el máximo módulo studentizado -

$$P[M \leq Sm_\alpha] = 1 - \alpha$$

implica el resultado del teorema. //

Este resultado o alguno similar se presenta en muchos lugares, por ejemplo, en Roy y Bose (1953), Scheffé (1959), y Miller (1966), por mencionar algunos.

Ejemplo 1.4. Una posible aplicación de este teorema se presenta en los experimentos factoriales, cuando se está interesado en intervalos para un conjunto finito de funciones lineales de los efectos de tratamientos cuyas estimaciones son independientes y normalmente distribuidas y la varianza de cada función se puede estimar como un múltiplo apropiado del cuadrado medio del error en el ANOVA. Supongamos, por ejemplo, que tenemos las observaciones de un experimento factorial 2^k y que estamos interesados en la construcción de ICS para los efectos principales y las interacciones de dos factores solamente; supongamos además, que ninguna de estas funciones está confundida con las repeticiones. Denotemos por τ_{ii} y τ_{ij} , $i \neq j$, $i, j = 1, 2, \dots, k$ en el verdadero valor de los efectos principales y las interacciones de dos factores respectivamente. Sea $\hat{\tau}_{ij}$, $i, j = 1, 2, \dots, k$ la estimación insesgada de τ_{ij} . Entonces el teorema nos permite afirmar que el conjunto $k(k+1)/2$ desigualdades.

$$\hat{\tau}_{ij} - \sqrt{\text{Var}(\hat{\tau}_{ij})} m_\alpha \leq \tau_{ij} \leq \hat{\tau}_{ij} + \sqrt{\text{Var}(\hat{\tau}_{ij})} m_\alpha ; i, j = 1, 2, \dots, k$$

se satisfacen simultáneamente con una probabilidad de $1 - \alpha$, cuando m_α es el punto correspondiente al $(1 - \alpha)$ percentil de la distribución del máximo módulo studentizado para $k(k+1)/2$ variables con el número de grados de libertad disponibles para la estimación del error.

1.2.10. El Método de Hochberg.

El método de Hochberg consiste básicamente en una aplicación del teorema anterior en combinación con la desigualdad de Sidák. Este método fue propuesto fundamentalmente para la construcción de ICS para la familia $\Psi = \{c'\theta; c \in C_p\}$ de funciones parametrales en el caso no-balanceado, pero también puede extenderse para permitir la construcción de ICS para todos los contrastes. Veamos.

Teorema 12. Sea $X \in \mathbb{R}^k$ un vector aleatorio con distribución $N(\theta, \sigma^2 V)$, donde θ y $\sigma^2 > 0$ son parámetros desconocidos y una matriz p.d. de constantes conocidas. Sea también, S^2 una variable aleatoria estadísticamente independiente de X tal que $\nu S^2 \sim \sigma^2 \chi^2(\nu)$

$$(i) P[|c'(X - \theta)| \leq S m_\alpha (v_{ii} - 2v_{ij} + v_{jj})^{1/2} \text{ para toda } c = e_i - e_j, i \neq j, i, j = 1, 2, \dots, k] \geq 1 - \alpha \quad (1.32)$$

$$(ii) P[|c'(X - \theta)| \leq S m_\alpha \frac{2}{\sum_{i=1}^k |c_i|} \sum_{i=1}^k \sum_{j=1}^k c_i^+ c_j^- (v_{ii} - 2v_{ij} + v_{jj}) \text{ para toda } c \in L_c] \geq 1 - \alpha \quad (1.33)$$

donde m_α es el punto correspondiente al $(1 - \alpha)$ -percentil de la distribución del máximo módulo studentizado para $k(k-1)/2$ variables con ν grados de libertad.

Demostración. Sea $g = k(k-1)/2$ y sea Y un vector en el espacio g -dimensional cuyos elementos Y_r , $r = 1, 2, \dots, g$, son

$$Y_\delta(i, j) = \frac{(X_i - X_j) - (\theta_i - \theta_j)}{(v_{ii} - 2v_{ij} + v_{jj})^{1/2}} \quad 1 \leq i < j \leq k$$

donde

$$\delta(i, j) = (i-1)k + j - i(i+1)/2$$

Entonces $Y \sim N(0, \sigma^2 R)$ donde R es una matriz de correlación de orden g .

Por la desigualdad de Sidák se tiene que

$$P\left[\frac{|Y_1|}{S} \leq \xi, \frac{|Y_2|}{S} \leq \xi, \dots, \frac{|Y_g|}{S} \leq \xi\right] \geq \prod_{r=1}^g P\left[\frac{|Y_r|}{S} \leq \xi\right]$$

y de acuerdo a los comentarios que siguen al Teorema 4, el lado derecho de la desigualdad es equivalente al caso en que reemplazamos R por la matriz identidad. Ahora bien, si hacemos esto, es decir, si pensamos en Y como $Y \sim N(0, \sigma^2 I)$, por un argumento similar al del teorema anterior tenemos que

$$\prod_{r=1}^g P\left[\frac{|Y_r|}{S} \leq \xi\right] = P\left[\max\{|Y_1|, |Y_2|, \dots, |Y_g|\} \leq S\xi\right]$$

y tal probabilidad es igual a $1 - \alpha$ si elegimos ξ como $\xi = m_\alpha$ donde m_α es como en el enunciado del presente teorema.

Sea ahora $Z = (X - \theta)$ y $a_{ij} = S m_\alpha (v_{ii} - 2v_{ij} + v_{jj})^{1/2}$, $i \neq j$, $i, j = 1, 2, \dots, k$, entonces $\frac{|Y_r|}{S} \leq m_\alpha$ para todo $r = 1, 2, \dots, g$ implica que $|Z_i - Z_j| \leq a_{ij}$ para toda $i \neq j$, $i, j = 1, 2, \dots, k$ de modo que la desigualdad de Sidák y el teorema anterior implican que

$$P\left[|Z_i - Z_j| \leq a_{ij} \text{ para toda } i \neq j, 1, 2, \dots, k\right] \geq 1 - \alpha$$

con lo que se demuestra la primera parte del teorema.

Veremos ahora que (ii) se sigue de (i)

Si $z = (x - \theta)$, tomemos ahora cualquier $c \in L_c$, entonces

$$\begin{aligned} |c'(x - \theta)| &= |c'z| \\ &= \left| \sum_{i=1}^k c_i^+ z_i - \sum_{i=1}^k c_i^- z_i \right| \\ &= \left(\frac{1}{2} \sum_{i=1}^k |c_i| \right)^{-1} \left| \sum_{i=1}^k \sum_{j=1}^k c_i^+ c_j^- z_i - \sum_{i=1}^k \sum_{j=1}^k c_i^+ c_j^- z_j \right| \\ &\leq 2 \left(\sum_{i=1}^k |c_i| \right)^{-1} \sum_{i=1}^k \sum_{j=1}^k c_i^+ c_j^- |z_i - z_j| \end{aligned}$$

entonces si $|z_i - z_j| \leq a_{ij}$ para todo $i \neq j$, $i, j = 1, 2, \dots, k$, se tiene que

$$c'(x - \theta) \leq 2 \left(\sum_{i=1}^k |c_i| \right)^{-1} \sum_{i=1}^k \sum_{i=1}^k c_i^+ c_j^- a_{ij}$$

de modo que (i) implica (ii) //

El método de Hochberg es comentado en Miller (1977) y Gabriel (1978b).

1.2.11. El Método de Tukey Modificado

La construcción de ICS para todas las diferencias entre medias, es el más común y quizás el más importante de los problemas de estimación simultánea. De los que ya hemos visto, los resultados en los que se hace referencia a este problema son

(i) El método de Tukey para el caso balanceado que dice:

$$P[|c'(x - \theta)| \leq a_{SQ\alpha} \text{ para toda } c = e_i - e_j, i \neq j, i, j = 1, 2, \dots, k] = 1 - \alpha$$

donde las constantes a y q_α son como en el Teorema 6.

(ii) El método de Spjøtvoll-Stoline para muestras estadísticamente independientes, esto es cuando V es una matriz diagonal, que dice

$$P \left[|c'(X - \theta)| \leq S q_\alpha' \max\{\sqrt{v_{ii}}, \sqrt{v_{jj}}\} \text{ para toda } c = e_i - e_j, i \neq j, i, j = 1, 2, \dots, k \right] \geq 1 - \alpha$$

donde q_α' es como en el Teorema 10.

(iii) El método de Hochberg para el caso no-balanceado, es decir, cuando V es cualquier matriz p.d., que dice

$$P \left[|c'(X - \theta)| < S m_\alpha (v_{ii} - 2v_{ij} + v_{jj})^{1/2} \text{ para toda } c = e_i - e_j, i \neq j, i, j = 1, 2, \dots, k \right] \geq 1 - \alpha$$

donde m_α es como en el Teorema 12.

Antes de que estos dos últimos métodos se conocieran, Tukey propuso efectuar una modificación al método que lleva su nombre para afrontar el caso no-balanceado, reemplazando a en el Teorema 6 por $\left[\frac{1}{2}(v_{ii} - 2v_{ij} + v_{jj}) \right]^{1/2}$ y manifestó que posiblemente

$$P \left[|c'(X - \theta)| \leq S q_\alpha \frac{1}{2}(v_{ii} - 2v_{ij} + v_{jj})^{1/2} \text{ para toda } c = e_i - e_j, i \neq j, i, j = 1, 2, \dots, k \right] \geq 1 - \alpha$$

(1.34)

Durante mucho tiempo se ha intentado probar o desaprobar esta conjetura sin que se haya encontrado una solución general; sin embargo, Brown (1983) demostró la validez de (1.34) cuando $k = 3$ y, recientemente Hayter (1984) demostró un teorema del cual se desprende la validez de (1.34) para cualquier $v_{ij} = 0$, $i \neq j$, $i, j = 1, 2, \dots, k$. El teorema de Hayter en cuestión, será presentado en seguida sin demostración, el -

lector interesado puede consultar el trabajo de Hayter (1984).

Teorema 13. Sean $X_i \sim N(0, \sigma_i^2)$, $i = 1, 2, \dots, k$ un conjunto de variables aleatorias independientes con $\sigma_i^2 > 0$ y sea $\xi \geq 0$ alguna constante. Entonces, como función de σ_i , $i = 1, 2, \dots, k$

$$P[|X_i - X_j| \leq \xi(\sigma_i^2 + \sigma_j^2)^{1/2} \text{ para toda } i \neq j, i, j = 1, 2, \dots, k] \quad (1.35)$$

es una función estrictamente minimizada cuando los valores de σ_i^2 , $i = 1, 2, \dots, k$, son iguales //

Nótese que el método de Tukey modificado puede extenderse para construir ICS para todos los contrastes entre medias, haciendo exactamente el mismo razonamiento utilizado en la extensión del método de Hochberg.

El teorema de Hayter garantiza que el método de Tukey modificado preserva el CCS cuando las variables son independientes; más aún, para este caso, los resultados de una simulación hecha por Dunnett (1980) y los cálculos presentados por Uusipaikka (1985) han demostrado que, a menos que el grado de imbalance sea muy severo, la igualdad en (1.34) casi se alcanza.

La modificación al método de Tukey para el caso no-balanceado es conocida en algunos contextos como corrección de Kramer, debido a que Kramer (1956, 1957), trabajando con pruebas de comparación polietápicas, propuso un cambio similar al propuesto por Tukey.

Una comparación, en el caso no-balanceado con variables aleatorias independientes, de los ICS para todas las diferencias entre medias obtenidas mediante el método de Tukey modificado contra los obtenidos por el método de Spjotvoll-Stoline y el método de Hochberg, nos muestra que el primer método proporciona los intervalos de menor longitud; veamos esta comparación.

En el caso que nos ocupa, sabemos que cuando la cota mínima para el CCS es la misma para los tres métodos, la mitad de la longitud de los intervalos para cada método es:

(i) Método de Tukey modificado

$$sq_{\alpha} \left[\frac{1}{2} (v_{ii} + v_{ij}) \right]^{1/2}$$

(ii) Método de Spjotvoll-Stoline

$$sq_{\alpha}^* [\max(v_{ii}, v_{ij})]^{1/2}$$

(iii) Método de Hochberg

$$sm_{\alpha} (v_{ii} + v_{ij})^{1/2}$$

donde q_{α} , q_{α}^* y m_{α} son los $(1 - \alpha)$ -cuantiles de las distribuciones de $Q(k, \nu)$, $Q^*(k, \nu)$ y $M(g, \nu)$, $g = k(k - 1)/2$, respectivamente.

Observemos ahora que

(a) El método de Tukey modificado es mejor que el de Spjotvoll-Stoline pues

$$q_{\alpha} \leq q_{\alpha}^* \quad \text{y} \quad v_{ii} + v_{ij} \leq 2 \max \{v_{ii}, v_{ij}\}$$

(b) El método de Tukey modificado es mejor que el de Hochberg ya que

$$q_{\alpha} \leq \sqrt{2} m_{\alpha}$$

(c) El método de Hochberg puede ser más o menos eficiente que el método de Spjøtvoll-Stoline, ya que la razón de las longitudes de los intervalos

$$\frac{m_{\alpha}}{q_{\alpha}} \left(1 + \frac{\min\{v_{ii}, v_{jj}\}}{\max\{v_{ii}, v_{jj}\}}\right)^{1/2}$$

puede ser mayor o menor que la unidad, dependiendo de α , k , v y del grado de imbalance, esto es, del grado de discrepancia entre los valores de v_{ii} y v_{jj} .

1.2.12. Un resultado general

De los métodos para la construcción de ICS aquí presentados, básicamente sólo tres de ellos poseen propiedades óptimas, esto quiere decir que cuando se cumplen los supuestos bajo los que son aplicables, cada uno de estos métodos proporciona intervalos con la mínima longitud posible y además el CCS tiene exactamente el valor nominal; estos métodos corresponden a tres diferentes familias de funciones parametrales y son: el método de Scheffé para el conjunto de todas las combinaciones lineales de $c'\theta$ con c en un espacio o subespacio vectorial; el método de Tukey para todas las diferencias entre medias $\theta_i - \theta_j$, $i \neq j$, $i, j = 1, 2, \dots, k$, en el caso balanceado y; el método de Tukey-Roy-Bose para las coordenadas individuales del vector de medias, es decir para θ_i , $i = 1, 2, \dots, k$,

cuando las componentes del vector de observaciones son estadísticamente independientes. Hemos visto también que cuando se alteran las especificaciones bajo las cuales cada método tiene propiedades óptimas, por ejemplo, cuando no se cumplen los supuestos acerca de la distribución o cuando se modifica el conjunto de funciones parametrales consideradas, es posible, en muchos casos, efectuar alguna modificación de modo que el resultado central del método sea aplicable y CCS o el valor mínimo de este se preserve; sin embargo, casi siempre los cambios sacrifican las propiedades óptimas del método. Para finalizar esta sección veremos una formulación general, debida a Uusipaikka (1985), que en teoría ofrece una solución al problema de estimación de ICS exactos para un conjunto específico de funciones parametrales de la forma $c'\theta$ con c en C donde C es algún subconjunto finito o infinito de un subespacio vectorial de dimensión q , cuando $x \sim N(\theta, \sigma^2 v)$ y sólo se supone que v es una matriz positiva definida de constantes conocidas y, además, se cuenta con una estimación de σ^2 estadísticamente independiente de x , es decir, bajo el conjunto de supuestos hechos en (1.2:1.)

El resultado de Uusipaikka considera la familia de estimación o familia de cantidades pivote compuesta por

$$z(x; c'\theta) = \frac{|c'(x - \theta)|}{(qS^2 c'vc)^{1/2}} \quad c \in C \quad (1.36)$$

donde q es la dimensión del subespacio vectorial al que pertenece C . De este modo, los ICS son de la forma

$$c'x - \xi_\alpha [qS^2 c'vc]^{1/2} \leq c'\theta \leq c'x + \xi_\alpha [qS^2 c'vc]^{1/2} \quad (1.37)$$

para toda c en C ; donde ξ_α es el punto correspondiente al $(1 - \alpha)$ -percentil de cierta función de distribución H , que depende de C , V y v . El siguiente teorema contiene una representación de esta función de distribución.

Teorema 14. Los ICS mostrados en (1.37) tienen exactamente un CCS de $1 - \alpha$ para toda $c' \theta$ con c en C si el número real ξ_α es el punto correspondiente al $(1 - \alpha)$ -percentil de la siguiente función de distribución:

$$\begin{aligned} H(w) &= F(w^2) + 2 \int_w^{w/\tau} G(w/r) f(r^2) r dr \\ &= \int_\tau^1 F(w^2/t^2) g(t) dt \end{aligned} \quad (1.38)$$

para todo $w > 0$; donde F y f son la función de distribución y función de densidad de la distribución F de Snedecor con q y v grados de libertad respectivamente y G y g son la función de distribución y función de densidad de la siguiente variable aleatoria

$$T = \sup_{c \in C_*} |c' \beta| \quad (1.39)$$

donde β tiene distribución uniforme sobre la esfera unitaria

$$E = \{ \eta : \eta \in \mathbb{R}^q, \|\eta\| = 1 \}$$

y C_* es algún subconjunto de esa esfera tal que la siguiente correspondencia se verifica

$$c \downarrow C_* = \frac{c' V \tilde{c}}{(c' V c \tilde{c}' V \tilde{c})^{1/2}} \quad (1.40)$$

para cualquier par de vectores en C y C_* . El número real τ es

igual a

$$\inf_{\eta \in E} \sup_{c \in C_*} |c' \eta|$$

y la variable aleatoria T satisface

$$P [\tau \leq T \leq 1] = 1$$

Demostración. Claramente si ξ_α es el punto correspondiente al $(1 - \alpha)$ -percentil de la variable aleatoria

$$M = \sup_{c \in C} \frac{|c'(X - \theta)|}{(qS'c'vc)^{1/2}}$$

los intervalos (1.37) tienen un CCS igual a $1 - \alpha$. Veamos entonces que H , la función de distribución de esta variable, puede representarse como en el teorema.

Sea Q una matriz inversible tal que $QQ' = v$, entonces

$$\begin{aligned} M &= \sup_{c \in C} \frac{|c'QQ^{-1}(X - \theta)|}{\sqrt{q} \ S \ \|Q'c\|} \\ &= \sup_{c \in C_*} \frac{c'Z}{(qY)^{1/2}} \end{aligned}$$

donde $Z = \frac{1}{\sigma} Q^{-1}(X - \theta)$, $Y = S^2/\sigma^2$ y $C_* = \{c: c = \|Q'd\|^{-1}Q'd, d \in C\}$, de modo que $Z \sim N(0, \sigma^2 I)$ y $vY \sim \chi^2(v)$ y además Z y Y son, por hipótesis, estadísticamente independientes; así mismo, C_* es algún subconjunto de un subespacio vectorial de dimensión q , - digamos L . Sea P el operador proyección sobre ese subespacio, entonces $b = PZ$ tiene distribución normal estandar en ese subespacio. Entonces

$$M = \sup_{c \in C_*} \frac{|c'b|}{(qY)^{1/2}}$$

de modo que si hacemos que $b = \|b\|\beta$, podemos escribir

$$M = \frac{\|b\|}{(qY)^{1/2}} \sup_{c \in C_*} |c'\beta|$$

Obsérvese que $\|b\|$ y β son estadísticamente independientes (Anderson, 1958) y, además $\|b\|^2 \sim \chi^2(q)$ y β tiene una distribución uniforme sobre la esfera unitaria en el subespacio L de dimensión q ; más aún, C_* es algún subconjunto de esa esfera unitaria y es fácil ver que la correspondencia (1.40) en el teorema se verifica para cualquier par de vectores en C y C_* .

Por lo anterior, podemos escribir $M = RT$, donde R^2 tiene la distribución F de Snedecor con q y v grados de libertad y T es como en el teorema.

Finalmente, si H y h , F y f , G y g son las funciones de probabilidad y de densidad de M , R^2 y T respectivamente y si

$$\tau = \inf_{\eta \in E} \sup_{c \in C_*} |c'\eta|$$

entonces $P[\tau \leq T \leq 1] = 1$

y

$$h(m) = \int_{\tau}^1 \frac{2m}{t^2} f(m^2/t^2) g(t) dt$$

Luego

$$\begin{aligned} H(w) &= \int_0^w \int_{\tau}^1 \frac{2m}{t^2} f(m^2/t^2) g(t) dt dm \\ &= \int_{\tau}^1 \int_0^{\frac{w^2}{t^2}} f(s) g(t) ds dt = \int_{\tau}^1 F(w/t) g(t) dt \\ &= F(w^2/t^2) G(t) \Big|_{\tau}^1 + 2 \int_{\tau}^1 G(t) f(w^2/t^2) \frac{w^2}{t^3} dt \end{aligned}$$

$$= F(w^2) + 2 \int_w^{w/\tau} G\left(\frac{w}{r}\right) f(r^2) r dr \quad //$$

La utilidad de este teorema depende de nuestra habilidad para determinar la distribución de la variable aleatoria τ en (1.39); en Uusipaikka (1983, 1985) se deriva esta distribución para los casos en que $q = 2$ y $q = 3$ para un conjunto finito C , sin embargo, para mayores valores de q parece ser difícil de determinar, excepto cuando se satisfacen los supuestos para los tres métodos antes mencionados, ya que en esos casos, el resultado de este teorema es equivalente al de esos métodos; dicho de otra forma, los tres métodos son casos particulares de la aplicación de este teorema.

CAPITULO 2

METODOS DE COMPARACIONES MULTIPLES

2.1. Conceptos Preliminares

Contamos con la observación de un vector aleatorio x , cuya distribución depende de un conjunto de parámetros θ ; sea Θ el espacio de parámetros. Estamos interesados en probar una familia de hipótesis H dada por

$$H = \{H_i: \theta \in w_i, i \in I\}$$

donde $w_i \subset \Theta$ es un subconjunto en el espacio de parámetros restringido por la hipótesis H_i ; como antes, I denota el conjunto de índices. Podemos caracterizar los elementos de H diciendo que la familia contiene una hipótesis global H_0 , así como un conjunto de hipótesis $H_i, i \in I - \{0\}$ implicados por H_0 ; es decir, la hipótesis $H_i: \theta \in w_i$ es tal que $w_0 \subset w_i \subset \Theta$ para toda $i \in I$. Se asume que $w_0 = \bigcap_I w_i$, de modo que podemos decir que H_0 es la intersección de todas las hipótesis. A la relación $w_i \subset w_j$ para cualquier par de hipótesis en H , la referiremos diciendo que H_j es componente de H_i , y si la relación es estricta, esto es, si $w_i \subset w_j$, diremos que H_j es componente propio de H_i ; denotaremos esta relación estricta escribiendo $i \rightarrow j$. Nótese que esta relación es transitiva, pero -

no es reflexiva, ni simétrica. Podemos dividir H en dos subfamilias, a saber, una subfamilia de hipótesis minimales, - que denotaremos por H_m e indizada por I_m , formada todas aque- llas hipótesis que no tienen componentes propios y, por ex- tensión otra subfamilia de hipótesis no-minimales, compuesta por todas las hipótesis en el complemento de H_m .

Ejemplo 2.1. Para ilustrar lo anterior, consideremos la siguiente familia de hipótesis:

$$\begin{array}{ll}
 H_0: \theta_1 = \theta_2 = \theta_3 = \theta_4 & H_5: \theta_1 = \theta_2 \\
 & H_6: \theta_1 = \theta_3 \\
 H_1: \theta_1 = \theta_2 = \theta_3 & H_7: \theta_1 = \theta_4 \\
 H_2: \theta_1 = \theta_2 = \theta_4 & H_8: \theta_2 = \theta_3 \\
 H_3: \theta_1 = \theta_3 = \theta_4 & H_9: \theta_2 = \theta_4 \\
 H_4: \theta_2 = \theta_3 = \theta_4 & H_{10}: \theta_3 = \theta_4
 \end{array}$$

En este caso H_0 es la hipótesis global; H_i , $i = 1, 2, \dots, 10$, son las hipótesis implicadas por H_0 ; H_i , $i = 0, 1, \dots, 4$, son hipótesis no-minimales y; H_i , $i = 5, 7, \dots, 10$, son hipóte- sis minimales. Vemos también que $0 \rightarrow i$, $i = 1, 2, \dots, 10$; $1 \rightarrow j$, $j = 5, 6, 8$; $2 \rightarrow j$, $j = 5, 7, 9$; etc... \square

Cuando una hipótesis es probada por una prueba de sig- nificancia y no es rechazada, generalmente agregamos que to- das las hipótesis implicadas por esta (sus componentes) deben considerarse como no-rechazadas; sin embargo, cuando la hipó- tesis es rechazada, nos enfrentamos al problema de tener que decidir qué componentes deben ser rechazados también; en -

otras palabras, cuando consideramos una familia de hipótesis deseamos saber si todas las hipótesis son ciertas, si todas son falsas o, de otro modo, saber qué hipótesis son ciertas y cuáles no lo son. Estamos ante un problema de decisiones múltiples, que puede o no admitir una solución simple y natural.

En un intento por ofrecer una respuesta a algunos problemas de este tipo, se ha propuesto una variedad de métodos de comparación múltiple basados ya sea en pruebas de significancia para cada una de las hipótesis en la familia, o en regiones de confianza simultáneas para funciones parametrales relacionadas con la familia de hipótesis.

Hay fundamentalmente tres enfoques acerca de cómo deben ser los métodos de comparación múltiple. El primero, aparentemente debido a Tukey (1951), indica que la estructura -- del método debe ser tal que permita probar simultáneamente to das y cada una de las hipótesis en la familia; la manera más común de conseguir esto es comparando todos los estadísticos de prueba contra el valor crítico obtenido para la hipótesis global, de manera que las hipótesis pueden ser probadas sin referencia de una a otra; a los métodos de este género se les conoce como métodos de pruebas simultáneas (MPS); como ejemplos de estos tenemos los conocidos métodos de Tukey y Scheffé. El segundo enfoque, sugerido por Student e implementado por Fisher (1935), parte del hecho de que si la hipótesis global es falsa, utilizando esta información es posible, más adelante lo veremos, obtener métodos, en algún sentido más poderosos que los métodos anteriores, para probar simultáneamente -

las hipótesis restantes en la familia; los métodos de este tipo generalmente constan de dos etapas: primero se hace una prueba ordinaria de nivel α para la hipótesis global H_0 , en una segunda etapa la subfamilia H_i , $i \in I - \{0\}$ es probada simultáneamente si y sólo si la hipótesis global es rechazada, de lo contrario el procedimiento termina; el valor crítico para las hipótesis en la subfamilia se obtiene bajo el supuesto de que H_0 es falsa; a este tipo de procedimientos les denominamos métodos de pruebas protegidas (MPP). En el tercer enfoque, también sugerido por Student e implementado de diferentes maneras por varios autores, los métodos tienen una estructura de naturaleza polietápica tendiente a alcanzar la mayor capacidad para detectar significancia en cada hipótesis; para los métodos más comunes de este tipo, se establece que en la primera etapa se prueba la hipótesis global y en las etapas subsiguientes una hipótesis H_j es probada si y sólo si H_j es componente propio de una hipótesis H_i y, además, H_i es rechazada en una etapa anterior; el valor crítico contra el que se compara el estadístico de prueba correspondiente a la i -ésima hipótesis; es obtenido bajo la distribución especificada por H_i considerada individualmente; una característica de estos métodos es que el nivel de significancia puede no ser el mismo para todas las hipótesis; a este tipo de procedimientos les llamaremos métodos de comparación polietápicas (MCP); aquí se incluyen los procedimientos conocidos con el nombre genérico de pruebas de rango múltiple.

En el párrafo anterior, implícitamente hemos asociado a cada hipótesis H_i , $i \in I$, un estadístico de prueba $Z_i = Z_i(X)$; entonces para la familia H tenemos una familia de estadísticos de prueba $Z = \{Z_i, i \in I\}$. Además, asumimos que la distribución de Z_i , $i \in I$, está completamente especificada bajo H_i , es decir, es la misma para toda $\theta \in w_i$. Ahora observemos que para cada $i \in I$, puede haber más de un estadístico que nos permita construir una prueba para H_i , de modo que hay varias posibilidades para elegir alguna familia Z ; más aún la elección de Z casi siempre juega un papel importante en cualquier procedimiento de prueba.

Antes de describir cada uno de los tres tipos de métodos mencionados previamente, veamos dos requerimientos, deseables en cualquier procedimiento, que se refieren a la relación entre decisiones sobre hipótesis que se implican una a otra. Estos requerimientos son la coherencia y la consonancia en las decisiones. Con un procedimiento de comparaciones múltiples que satisfaga el requerimiento de coherencia, ninguna hipótesis puede ser aceptada si cualquier hipótesis implicada por ésta es rechazada, la coherencia es esencial ya que nos previene de contradicciones en el sentido de que impide rechazar una hipótesis sin haber rechazado todas las hipótesis que la implican. La consonancia se refiere a la relación recíproca, esto es, con un procedimiento que satisfaga el requerimiento de consonancia, el rechazo de alguna hipótesis no-minimal implica que al menos una hipótesis minimal implicada por ésta es rechazada también; la consonancia evita cierto tipo de

inconsistencia que se presenta cuando una hipótesis es rechazada y no se rechaza ninguna de las hipótesis implicadas por esta; tal inconsistencia, aunque no deseable, algunas veces es inevitable en la práctica, pero esta es una deficiencia inherente no sólo a las pruebas de comparación múltiple, pues se presenta en otros tipos de pruebas estadísticas. Estos dos requerimientos son tratados en Gabriel (1964, 1969) y Einot y Gabriel (1975).

Coherencia y consonancia son relaciones entre las decisiones tomadas en base a pruebas sobre hipótesis que se implican una a otra, esto es, si $i \rightarrow j$, estos requerimientos indican cierto tipo de relación que debe haber entre las decisiones sobre H_i y H_j ; sin embargo, la relación de implicación de $i \rightarrow j$ no es la única que puede obtenerse dentro de H . Lehmann (1957) señala algunos procedimientos que preservan una amplia clase de relaciones de implicación, a los que denomina compatibles; un procedimiento se dice compatible si, sobre el conjunto de todas las posibles decisiones inducidas por H y la familia de pruebas para H , la probabilidad de tomar una decisión inconsistente, con respecto a las otras decisiones, es nula.

2.2. Métodos de Pruebas Simultáneas

Para una familia $\{H, Z\}$ y una constante ξ , un método de pruebas simultáneas (MPS) se define como la familia de pruebas para toda H_i , $i \in I$, que rechaza H_i si $(Z_i > \xi)$, usando la misma constante ξ para todo Z_i en Z . De este modo, un MPS

puede caracterizarse por la tríada $\{H, Z, \xi\}$. La probabilidad

$$\alpha = P_{W_0}(Z_0 > \xi)$$

de rechazar erróneamente la intersección de las hipótesis se rá referida como nivel del MPS y la constante ξ como el valor crítico.

El término de simultáneo es usado para enfatizar que ningún ordenamiento o secuencia debe ser impuesto al proceso de prueba de todas las hipótesis, esto es, las pruebas se ha cen sin referencia de una a otra.

Para un MPS, los requerimientos de coherencia y consonancia pueden escribirse en términos de los siguientes even tos: para toda $i \in I - I_m$

(i) coherencia

$$(X:Z_i > \xi) \supseteq U (X:Z_j > \xi) \text{ c.s.} \quad (2.1)$$

(ii) consonancia

$$(X:Z_i > \xi) \subseteq U (X:Z_j > \xi) \text{ c.s.} \quad (2.2)$$

(iii) coherencia y consonancia

$$(X:Z_i > \xi) = U (X:Z_j > \xi) \text{ c.s.} \quad (2.3)$$

donde las uniones son sobre los subíndices j tales que $i \rightarrow j$. En estas ecuaciones c.s. significa "casi siempre" o "casi en todas partes", es decir, si $A \subseteq B$ c.s., entonces $A \cap \bar{B}$ puede ser no vacío pero de medida cero; similarmente, $A = B$ c.s. significa que ambos $A \cap \bar{B}$ y $\bar{A} \cap B$, pueden ser no vacíos pero

su probabilidad debe ser cero.

Obviamente no todo MPS satisface estos requerimientos, sin embargo, si para una H dada restringimos la elección de Z , podemos encontrar condiciones necesarias y suficientes para las que sí se satisfagan, como se muestra en los siguientes teoremas.

Teorema 1. Cualquier MPS basado en la familia $\{H, Z\}$, es coherente para todo ξ si y sólo si para cualquier par $i, j \in I$, tal que $i \rightarrow j$, se satisface que

$$Z_i(X) \geq Z_j(X) \quad \text{c.s.} \quad (2.4)$$

Demostración. Consideremos cualquier par H_i, H_j en H para el que $i \rightarrow j$. La coherencia de $\{H, Z, \xi\}$ para alguna ξ , exige que se cumpla que

$$(X:Z_i > \xi) \supseteq (X:Z_j > \xi) \quad \text{c.s.}$$

ya que estos eventos determinan el rechazo de H_i y H_j respectivamente. Ahora bien

$$Z_i \geq Z_j \quad \text{c.s.}$$

es claramente equivalente a

$$(Z_i > \xi) \supseteq (Z_j > \xi) \quad \text{para toda } \xi \quad \text{c.s.}$$

que es lo que queríamos demostrar. //

Observación: bajo las hipótesis del teorema

$$Z_i = \max\{Z_r : r = i \text{ ó } i \rightarrow r\}$$

Si Z es como en el teorema, esto es, si $Z_i \geq Z_j$ c.s. siempre que $i \rightarrow j$, decimos entonces que Z es una familia monótona; lo mismo decimos de $\{H, Z\}$. Además, si para cualquier H_i , $i \in I - I_m$ y cualquier punto muestral X hay un componente propio H_j de H_i para el que la igualdad se verifica, esto es, si

$$Z_i = \max_j \{ Z_j, i \rightarrow j \} \quad \text{c.s.} \quad (2.5)$$

entonces se dice que la familia $\{H, Z\}$ es estrictamente monótona. Es fácil ver que si una familia es estrictamente monótona, también es monótona; igualmente es fácil verificar que para una familia monótona, el estadístico de prueba para H_0 , la intersección de las hipótesis, es

$$Z_0 = \max_i \{ Z_i, i \in I \} \quad (2.6)$$

Teorema 2. Cualquier MPS basado en la familia $\{H, Z\}$ es coherente y consonante si y sólo si $\{H, Z\}$ es estrictamente monótona.

Demostración. Un MPS es coherente y consonante si para toda $i \in I - I_m$

$$(Z_i > \xi) = \bigcup_{\{j: i \rightarrow j, j \in I_m\}} (Z_j > \xi) \quad \text{c.s.}$$

Ahora de la definición de monotonidad estricta, tenemos que para cualquier $i \in I - I_m$ se verifica que

$$Z_i = \max_j \{ Z_j, i \rightarrow j \} \quad \text{c.s.}$$

entonces para toda $i \in I - I_m$ se cumple que

$$Z_i = \max_j \{ Z_j, i \rightarrow j, j \in I_m \} \quad \text{c.s.}$$

de modo que coherencia y consonancia de $\{H, Z, \xi\}$ para toda ξ , es claramente equivalente a monotonía estricta de $\{H, Z\}$. //

El siguiente teorema se refiere a la probabilidad de rechazar erróneamente al menos una hipótesis de la familia H .

Teorema 3. La probabilidad de que un MPS $\{H, Z, \xi\}$ coherente de nivel α , rechace erróneamente al menos una hipótesis de H es α si H es cierta y no excede el valor de α , sea o no H_0 cierta, siempre que una o ambas de las siguientes condiciones se verifiquen:

(i) para cualquier subfamilia $H^* = \{H_i, i \in I^*\}$ donde $I^* \subset I$, la distribución conjunta de $\{Z_i, i \in I^*\}$ esté completamente especificada bajo $H_0: \theta \in \prod_{i \in I^*} w_i$

(ii) la familia H sea cerrada bajo intersección.

Demostración. Por el Teorema 1 $\{H, Z\}$ debe ser monótona y entonces $Z_0 = \max_i \{Z_i, i \in I\}$, de modo que el rechazo de H_0 , evento $(Z_0 > \xi)$, es equivalente al rechazo de cualquier H_i , $i \in I$, es decir al evento $\cup_I (Z_i > \xi)$. Dado que bajo H_0 la probabilidad del primer evento es α , es decir

$$P_{w_0}(Z_0 > \xi) = \alpha$$

cuando H_0 es cierta, entonces la probabilidad del segundo evento es la misma cuando H_0 es cierta, lo cual demuestra la primera parte del teorema.

Supongamos ahora que \bar{H} , la subfamilia de hipótesis ciertas, es no-vacía; entonces su intersección \bar{H}_0 es necesariamente cierta; para esta subfamilia el rechazo de alguna hipótesis cierta ocurre si $(\bar{Z}_0 > \xi)$, \bar{Z}_0 definida por

$$\bar{Z}_0 = \max_i \{Z_i, i \in I\}$$

donde $\bar{I} \subseteq I$ es el conjunto de índices de \bar{H} . La probabilidad asociada a este evento es

$$P_{\bar{w}_0}(\bar{Z}_0 > \xi)$$

Si podemos mostrar que

$$P_{\bar{w}_0}(\bar{Z}_0 > \xi) = P_{w_0}(\bar{Z}_0 > \xi) \quad (*)$$

entonces la segunda parte del teorema se sigue, ya que $\bar{Z}_0 \leq Z_0$ implica que

$$P_{w_0}(\bar{Z}_0 > \xi) \leq P_{w_0}(Z_0 > \xi)$$

Ahora bien, si (i) se cumple, la distribución conjunta de $\{Z_i, i \in \bar{I}\}$ está completamente especificada bajo \bar{H}_0 y entonces la distribución marginal de \bar{Z}_0 también está completamente especificada bajo \bar{H}_0 indiferentemente de si H_0 es o no cierta y, por lo tanto (*) se cumple. El mismo argumento se aplica si (ii) se verifica, más aún, como en este caso H es cerrada bajo intersección $\bar{H}_0 \in H$ y \bar{Z}_0 es el estadístico de prueba para \bar{H}_0 , y, por hipótesis, a fortiori su distribución está completamente especificada bajo \bar{H}_0 sea o no H_0 cierta, con lo que también en este caso (*) se cumple. //

En resumen, bajo las condiciones de este teorema, la probabilidad de rechazar al menos una hipótesis H que sea cierta, no excede el valor de α . Esta probabilidad es la que en muchos trabajos se conoce como tasa de error experimental.

Observemos que la condición (i) del teorema implica que todas las hipótesis ciertas $H_i, i \in \bar{I}$, están incluidas en ambas \bar{H} y H , pero no necesariamente que $\bar{H}_0 \in H$, esto se garantiza solamente si H es cerrada bajo intersección, es decir, mediante la condición (ii). Para ilustrar esto veamos el siguiente ejemplo.

Ejemplo 2.2. Consideremos un diseño completamente al azar con k tratamientos ($k > 3$). Supongamos que H_0 es la hipótesis de igualdad en todas las medias y además

(a) Si la subfamilia $H-H_0$ está compuesta por las $k(k-1)/2$ hipótesis de igualdad de todos los pares de medias. Esta familia no es cerrada bajo intersección ya que no contiene las hipótesis de igualdad de grupos de q medias si $2 < q < k$; no obstante, si para cada hipótesis empleamos estadísticos F aumentada (estadísticos F multiplicados por los grados de libertad del numerador) como estadísticos de prueba, entonces la distribución conjunta de los estadísticos de prueba de las hipótesis minimales, está completamente especificada para cualquier subconjunto de q medias iguales, independientemente de los valores de las otras $k - q$ medias. Esto demuestra que (i) se cumple.

(b) Si cada $H_i; I - \{0\}$ es la hipótesis de que el i -ésimo contraste de medias es igual a cero y, la subfamilia $H - H_0$ incluye hipótesis para todos los contrastes, entonces H es una familia cerrada bajo intersección. \square

Hasta aquí hemos visto que las familias de estadísticos de prueba monótonas o estrictamente monótonas desempeñan un papel importante para asegurar la coherencia o la coherencia y consonancia de un MPS. De manera natural surge la pregunta ¿hay algún criterio de elección para la familia Z que nos garantice la monotonidad o la monotonidad estricta?. El siguiente teorema ofrece una respuesta.

Teorema 4. (i) Si la familia de estadísticos de prueba Z se relaciona con H mediante el Principio de la Razón de Verosimilitud (RV), entonces $\{H, Z\}$ es monótona.

(ii) Una familia $\{H, Z\}$ es estrictamente monótona si y sólo si los estadísticos de prueba pueden escribirse en términos del Principio Unión-Intersección de Roy (UI), esto es, cuando y sólo cuando Z se relaciona con H mediante el principio UI, $\{H, Z\}$ es estrictamente monótona.

Demostración. Sea la familia $\{H, Z\}$. Sabemos que el estadístico de prueba para cualquier H_i en H basado en el principio RV, puede escribirse como

$$Z_i = g \left[\frac{\sup_{w_i} f}{\sup_{\bar{D}} f} \right]$$

donde g es una función monótona decreciente y f la función de densidad de probabilidad (Wilks, 1962).

Tomemos ahora dos hipótesis $H_i, H_j \in H$ tales que $i \rightarrow j$. Como $i \rightarrow j$ es equivalente a $w_i \subset w_j$, lo cual implica que

$$\sup_{w_i} f \leq \sup_{w_j} f$$

se sigue que $Z_i \geq Z_j$ ya que $g(\cdot)$ es decreciente. Esto demuestra (i).

Ahora, bajo el principio UI, la construcción de una familia de pruebas simultáneas para H de nivel α , consta de los siguientes pasos:

(a) Primero consideremos la subfamilia H_m de hipótesis minimales; construyamos una prueba separadamente para cada $H_j, j \in I_m$, de tamaño digamos α_* ($\alpha_* \leq \alpha$) tal que la región crítica de la prueba para H_j sea $(X:Z_j > \xi)$.

(b) Una prueba para cualquier $H_i, i \in I_m$, se construye a partir de sus componentes minimales, definiendo la región crítica para H_i como

$$U (X:Z_j > \xi)$$

donde la unión es sobre los subíndices j tales que $i \rightarrow j$.

(c) Consideremos la intersección $w_0 = \cap_I w_i$ para determinar ξ de manera que la prueba ordinaria para H_0 , considerada separadamente, tenga tamaño α , esto es

$$P_{w_0} = [Z_0 > \xi] = \alpha$$

Dado α puede determinarse α_* y viceversa.

De este modo por (b) tenemos que la región crítica - para $H_i, i \in I - I_m$ puede escribirse equivalentemente como

$$(X: \max_j \{Z_j, i \rightarrow j, j \in I_m\} > \xi)$$

y extendiendo la definición para todos los posibles valores de ξ , el estadístico de prueba para H_i se define como

$$Z_i = \max_j \{Z_j, i \rightarrow j, j \in I_m\}$$

lo cual es claramente equivalente a la definición de monotonicidad estricta. Con esto se demuestra (ii). //

Como corolario a este teorema tenemos que los MPS basados en la familia $\{H, Z\}$ son coherentes siempre que Z se relacione con H mediante el principio RV y, son coherentes y consonantes si y sólo si Z se relaciona con H mediante el principio UI.

A continuación veremos las familias de hipótesis más comunes en problemas de comparaciones múltiples cuando contamos con la observación de un vector aleatorio $X \in \mathbb{R}^k$, $X \sim N(\theta, \sigma^2 V)$ donde V es una matriz p.d. de constantes conocidas y, además, se tiene una observación de S^2 , una variable aleatoria estadísticamente independiente de X , tal que $vS^2 \sim \sigma^2 \chi^2(v)$. Posteriormente veremos algunos ejemplos de MPS para estas familias.

Las familias de hipótesis en cuestión son:

(i) H^P , la familia indizada por I^P , que, además de H_0 , contiene las hipótesis de igualdad de pares de medias, - esto es

$$H_i: \theta_{r_i} = \theta_{r'_i} \quad , \quad r_i \neq r'_i \quad i \in I^P - \{0\}$$

H^P contiene $1 + k(k-1)/2$ hipótesis de las cuales solamente H_0 es hipótesis no minimal.

(ii) H^Q , la familia indizada por I^Q , que contiene, - además de H_0 , las hipótesis de igualdad de todas las medias de cada subconjunto Q_i de tamaño q_i , es decir, para toda $i \in I - \{0\}$

H_i : todas las medias dentro del subconjunto Q_i son homogéneas

H^Q consta de $2^k - k - 1$ hipótesis y, además $H_M^Q = H_M^P$

(iii) La familia H^C , indizada por I^C , que además de la hipótesis global H_0 , contiene las hipótesis de que cada - contraste es igual a cero, esto es

$$H_j: c_j^! \theta = 0 \quad , \quad c_j \in L_C \quad j = 1, 2, \dots$$

El número de hipótesis en H^C es infinito y todas las hipótesis, excepto H_0 , son hipótesis minimales.

(iv) La familia H^D , indizada por I^D , que contiene - además de H_0 , las hipótesis de igualdad de la i -ésima media ($i < k$) con la media testigo θ_k , o sea

$$H_i: \theta_i = \theta_k \quad i = 1, 2, \dots, k-1$$

La familia H^D consta de k hipótesis y todas, excepto H_0 , son hipótesis minimales.

Nótese que la hipótesis global H_0 es la misma para estas cuatro familias; H_0 es la hipótesis de igualdad de todas las medias. Note además que para $k > 2$ se tiene que

$$I^D \subset I^P \subset I^Q \subset I^C \quad \square$$

Ahora presentaremos algunos MPS para estas familias de hipótesis, suponiendo que V , las constantes de la matriz de varianzas-covarianzas, satisface que $V^{-1} = \text{diag}\{n_1, n_2, \dots, n_k\}$

(1) Para H^P , si definimos Z^R como

$$Z_j^R = \frac{|X_{r_i} - X_{r'_j}|}{S \left| \frac{1}{n_r} + \frac{1}{n_{r'}} \right|^{1/2}} \quad j \in I_m^P \quad (2.7)$$

es decir, Z_j^R es el estadístico de prueba para $H_j: \theta_{r_j} = \theta_{r'_j}$, además,

$$Z_0^R = \max_j \{Z_j, j \in I_m^P\}$$

y si ξ^R es el punto correspondiente al $(1 - \alpha)$ -percentil de la distribución del rango studentizado con k variables y v grados de libertad, entonces $\{H^P, Z^R, \xi^R\}$ es un MPS coherente y consonante, si $n_r = n_{r'}$ para toda $r \neq r'$, este MPS es de nivel α , de otro modo, por el teorema de Hayter en el capítulo anterior, es de nivel a lo más α .

2) Definimos $\{H^Q, Z^F, \xi^F\}$ como un MPS de nivel α si

$$Z^F = \frac{\sum_{r=1}^k n_r (X_r - \bar{X}_0)^2}{S^2} \quad (2.8)$$

$$Z_i^F = \sum n_r (X_r - \bar{X}_i)^2 / S^2 \quad i \in I^Q - \{0\} \quad (2.9)$$

donde $\bar{X}_i = \frac{\sum n_r X_r}{\sum n_r}$

y las sumatorias corren sobre $\{r: \theta_r \in Q_i\}$ y además $Q_0 = \{\theta_1, \theta_2, \dots, \theta_k\}$, es decir Z_i , $i \in I_Q$, es el estadístico F (aumentado), y ξ^F es el punto correspondiente al $(1 - \alpha)$ -percentil de la distribución F (aumentada) con $k - 1$ y ν grados de libertad, esto es $\xi^F = (k - 1)F_\alpha$.

Este MPS es coherente ya que $\{H^Q, Z^F\}$ es monótona; más aún, es fácil mostrar (ver Searle, 1971) que cada Z_i^F , $i \in I_Q$, se relaciona con su hipótesis respectiva mediante el principio RV. Ahora bien, $\{H^Q, Z^F, \xi^F\}$ aunque es coherente no es consonante ya que

$$Z_0^F > \max_j \{Z_j^F, j \in I_M^Q\} \quad \text{c.s.}$$

y entonces $(Z_0 > \xi)$ no necesariamente implica que $(Z_j > \xi)$ para alguna $j \in I_M^Q$. Este método es ampliamente descrito en Gabriel (1964).

(3) Un MPS coherente y consonante para la familia H^Q de nivel no mayor que α es $\{H^Q, Z^R, \xi^R\}$, donde

$$Z_j = \frac{|X_{rj} - X_{r'j}|}{S \left| \frac{1}{2} \left(\frac{1}{n_r} + \frac{1}{n_{r'}} \right) \right|^{1/2}}, \quad r \neq r', j \in I_M^Q \quad (2.10)$$

$$Z_i = \max_j \{Z_j, i \rightarrow j, j \in I_M^Q\}, i \in I^Q - I_M^Q$$

y además ξ^R es como en (1), ya que por construcción Z^R y H^Q están relacionadas mediante el principio UI. Este MPS consiste en las pruebas de significancia del método de Tukey.

(4) Supongamos que tenemos la familia $\{H^C, Z^F\}$ donde la familia de estadísticos Z^F se define como

$$Z_i^F = \left[\sum_{r=1}^k c_{ri} X_r \right]^2 / \left[S^2 \sum_{r=1}^k c_{ri} / n_r \right] \quad i \in I^C - \{0\} \quad (2.11)$$

donde $c_i' = (c_{1i}, c_{2i}, \dots, c_{ki}), c_i \in L_c$

Es claro que H_0 la hipótesis de nulidad de todos los contrastes es equivalente a la hipótesis de igualdad de todas las medias y entonces Z_0^F es como en (2), por lo tanto, un MPS de nivel α se obtiene si ξ^F como en (2); también H^C y Z^F se relacionan mediante el principio RV; más aún, ya que $Z_0 = \max_j \{Z_j, j \in I_m^C\}$, también se relacionan por el principio UI. Por lo anterior $\{H^C, Z^F, \xi^F\}$ define un MPS coherente y consonante. Este MPS consiste en las pruebas de significancia del método de Scheffé.

(5) La familia de pruebas $\{H^D, Z^D, \xi^D\}$, define un MPS cuyo nivel no excede el valor de α si

$$Z_i^D = \frac{|X_i - X_k|}{\left(\frac{1}{n_i} + \frac{1}{n_k}\right)^{1/2}} \quad i \in I - \{0\} \quad (2.12)$$

$$Z_0^D = \max_j \{Z_j, i \rightarrow j, j \in I_m^D\}$$

y ξ^D es obtenido de las tablas proporcionadas por Dunnett (1955, 1964).

Este MSP es coherente y consonante y se conoce con el nombre de método de Dunnett. \square

El propósito de las comparaciones múltiples en general, y de cualquier MPS en particular, es proporcionar la mayor resolución. Por resolución entendemos la capacidad para detectar significancia en las hipótesis implicadas por H_0 ,

especialmente en las hipótesis minimales, cuando la hipótesis global es rechazada.

La resolución de un MPS cualquiera, puede ser definida en términos de la probabilidad de rechazar las hipótesis minimales siempre que la hipótesis H_0 sea rechazada. Consideremos dos MPS de nivel α con la misma hipótesis H_0 ; sean $\{H, Z, \xi\}$ y $\{H^*, Z^*, \xi^*\}$ estos MPS y supongamos $I_m \subset I^*$; entonces el primer MPS se dice más resolvente que el segundo si

$$(i) P_{W_0}(Z_0 > \xi) = P_{W_0}(Z_0^* > \xi^*) \quad (2.13)$$

es decir, tienen el mismo nivel y,

$$(ii) \text{ para todo } j \in I_m$$

$$(Z_j > \xi) \subset (Z_j^* > \xi^*) \quad \text{c.s.} \quad (2.14)$$

decimos es estrictamente más resolvente si la inclusión es estricta. Es claro que la comparación entre dos métodos fuera de las hipótesis que les son comunes no tiene sentido. La comparación puede hacerse sólo si H_0 es la misma para ambos métodos y las hipótesis minimales de una de las familias están incluidas en la otra.

Una posible medida de resolución de un método es la razón de las probabilidades de rechazar erróneamente las hipótesis minimales sobre la probabilidad de rechazar la hipótesis global. Supongamos, por ejemplo, que bajo H_0 todas las hipótesis minimales tienen la misma probabilidad de ser erróneamente rechazadas, es decir,

$$\alpha_1 = P_{W_j}(Z_j > \xi) \text{ para toda } j \in I_m$$

entonces, la resolución puede ser medida como α_1/α . Obviamente, si un método es más resolvente que otro, también tiene un valor mayor para α_1/α ; sin embargo, un mayor valor de α_1/α para alguno de ellos indica solamente más probable resolución y no necesariamente que sea estrictamente más resolvente que el otro.

Ejemplo 2.3. Consideremos la familia H^P y los métodos $\{H^P, Z^R, \xi^R\}$ y $\{H^P, Z^F, \xi^F\}$. Como para toda $j \in I_m^P$

$$(Z_j^R > \xi^R) \supset (Z_j^F > \xi^F) \quad \text{c.s.}$$

entonces el primer método es estrictamente más resolvente; - por lo tanto, para H^P , este método será preferible. \square

Ejemplo 2.4. Ahora consideremos la familia H^Q con $k = 8$ y $v = 40$. El método $\{H^Q, Z^K, 4.0990\}$ tiene una resolución de 0.06060, mientras que $\{H^Q, Z^F, 13.1077\}$ tiene resolución de 0.00818. El mismo argumento del ejemplo anterior nos lleva a concluir que el primer método es estrictamente más resolvente que el segundo.

Es interesante notar que en este caso podemos obtener una medida más general de resolución basándonos en las probabilidades de rechazar erróneamente las hipótesis en $H^Q - H_0$, pues en este caso la resolución depende de la secuencia

$$\alpha_1 : \alpha_2 : \dots : \alpha_{k-1}$$

donde $\alpha_{k-1} = \alpha$ y α_r es la probabilidad de rechazar -

erróneamente una hipótesis de rango r (ver Gabriel, 1964).

Para $\{H^Q, Z^R, 4.0990\}$ la secuencia entera satisface la relación $q_{(r+1, 40, \alpha_r)} = 4.0990$, $r = 1, 2, \dots, 7$

Para $\{H^Q, Z^F, 13.1077\}$ la secuencia satisface la relación $F_{(r, 40, \alpha_r)} = \frac{1}{r} 13.1077$, $r = 1, 2, \dots, 7$

De modo que las secuencias son:

$\{H^Q, Z^R, 4.0990\}$	$\{H^Q, Z^F, 13.1077\}$
α_7 0.100000	0.100000
α_6 0.080748	0.064650
α_5 0.062409	0.038427
α_4 0.045209	0.020491
α_3 0.029677	0.009418
α_2 0.016335	0.003432
α_1 0.006060	0.000818

lo cual nos muestra que $\{H^Q, Z^R, 4.0990\}$ es más resolvente para todas las hipótesis. Un resultado similar es válido para la comparación entre $\{H^Q, Z^R, \xi^R\}$ y $\{H^C, Z^F, \xi^F\}$. \square

Ejemplo 2.5. Supongamos que $n_r = n$, $r = 1, 2, \dots, k$ y consideremos la familia H^C . Si definimos Z^R como

$$Z_j^R = \frac{2\sqrt{n} \sum_{r=1}^k |c_{rj} X_r|}{s \sum_{i=1}^k |c_{ij}|} \quad j \in I^C - \{0\}$$

donde $c_j^i = \{c_{1j}, c_{2j}, \dots, c_{kj}\}$ y $c_j \in L_C$

$$Z_0^R = \max_j \{Z_j^R, j \in I_m^C\}$$

Entonces H^C se relaciona con Z^R por el principio UI, al igual que H^C con Z^F ; sin embargo, ahora $\{H^C, Z^R, \xi^R\}$ no es más resolvente que $\{H^C, Z^F, \xi^F\}$. \square

Ejemplo 2.6. Para las hipótesis en la familia H^Q , es interesante notar que $\{H^Q, Z^F, \xi^F\}$ y $\{H^C, Z^R, \xi^R\}$ proporcionan las mismas decisiones, ya que $I^Q \subset I^C$ y, además, cualquier hipótesis H_i en H^Q es juzgada significativamente solamente si al menos una hipótesis sobre algún contraste en H^C , relacionado con H_i , $i \in I^Q$, es juzgada significativa; por lo tanto para esta familia los métodos son equivalentes. \square

La resolución de dos métodos es comparable si éstos tienen el mismo nivel. Hay un tipo alternativo de comparación entre cualquier par de métodos que proporcionan idénticas decisiones sobre las hipótesis minimales y que se basa en el costo para detectar significancia en las hipótesis no-minimales.

Un método se dice menos exigente que otro, si es capaz de detectar la significancia en las hipótesis no-minimales, especialmente de las hipótesis global, a un costo no mayor. En otras palabras, $H \subset H^*$ y $H_0 = H_0^*$, $\{H, Z, \xi\}$ es menos exigente que $\{H^*, Z^*, \xi^*\}$ si

$$(i) \text{ para toda } j \in I \quad (Z_j > \xi) = (Z_j^* > \xi^*) \text{ c.s.} \quad (2.15)$$

$$(ii) \text{ para toda } i \in I \quad (Z_i > \xi) \subset (Z_i^* > \xi^*) \text{ c.s.} \quad (2.16)$$

$$(iii) \text{ bajo } H_0 \quad P_{W_0}(Z_0 > \xi) \leq P_{W_0}(Z_0^* > \xi^*) \quad (2.17)$$

si la desigualdad (iii) es estricta, entonces se dice que $\{H, Z, \xi\}$ es estrictamente menos exigente.

Ejemplo 2.7. Comparemos $\{H^D, Z^D, \xi^D\}$ con $\{H^P, Z^R, \xi^R\}$. Observe que si $\xi^R = \sqrt{2}\xi^D$, los métodos tienen idénticas decisiones para toda H_i , $i \in I^D$, pero el evento

$$\max_j \{Z_j^R, j \in I_m^P\} > \sqrt{2} \max \{Z_j^D, j \in I_m^D\}$$

bajo H_0 , puede ocurrir con probabilidad positiva, se sigue entonces que

$$P_{w_0}(Z_0^D > \xi^D) < P_{w_0}(Z_0^R > \xi^R)$$

ya que

$$\begin{aligned} P_{w_0}(Z_0^R > \xi^R) &= P_{w_0}(Z_0^D > \xi^D) + P_{w_0}(Z^D \leq \xi^D, Z^R > \xi^R) \\ &= P_{w_0}(Z^D > \xi^D) + P_{w_0}(\sqrt{2} Z_0^D < \xi^R < Z_0^R) \end{aligned}$$

Por lo anterior, tenemos que para H^D , $\{H^D, Z^D, \xi^D\}$ es estrictamente menos exigente que $\{H^P, Z^R, \xi^R\}$

Por otro lado tenemos que si ξ^D y ξ^R son elegidos de manera que los dos métodos tengan el mismo nivel, el método de Dunnett es más resolvente. Unos cálculos ilustrarán mejor el ejemplo. Tomemos el caso en que $k = 4$, entonces

	$\{H^D, Z^D, 2.44\}$	$\{H^P, Z^R, 3.4507\}$	$\{H^P, Z^R, 3.7910\}$
$P_{w_0}(Z_0 > \xi)$	0.050000	0.085187	0.050000
$P_{w_j}(Z_j > \xi)$	0.019214	0.019214	0.010621

Los valores de la primera y la segunda columnas nos muestran que el método de Dunnett es menos exigente, los de la primera y la tercera, nos muestran que es más resolvente. \square

Una característica importante de los MPS es su estrecha relación con los ICS. Si las hipótesis de una familia

H se relacionan con una familia de funciones parametrales Ψ , podemos construir ICS para los valores de esas funciones. La correspondencia entre los ICS para esas funciones y los MPS para las hipótesis que asignan un valor particular a cada función permite la extensión de las propiedades de los MPS a propiedades equivalentes para los ICS y viceversa. Más aún, supongamos que la hipótesis H_i en H se asocia con la función ψ_i e Ψ asignándole un valor λ_i^0 y que la aceptación de H_i por un MPS dado está determinada por el evento

$$Z_i(X; \lambda_i^0) \leq \xi$$

entonces este evento es equivalente a la inclusión del valor paramétrico λ_i^0 inducido por la hipótesis H_i , cuando la familia de cantidades pivote utilizadas por el método de construcción de ICS es

$$\{Z_i(X; \lambda_i), i \in I\}$$

Esta interpretación puede resultar de utilidad para resolver problemas de comparaciones múltiples, mediante ICS.

2.3. Métodos de Pruebas Protegidas

En la sección anterior hemos visto bajo que condiciones la probabilidad α_e de rechazar erróneamente al menos una hipótesis de la familia H con MPS de nivel α , es igual a α cuando la hipótesis global es cierta; hemos visto también bajo que condiciones el valor de α_e es menor que α si la hipótesis global es falsa. Luego, si partimos del supuesto de que H_0 es falsa y las condiciones se sostienen, en general es

posible modificar un MPS dado, para obtener un método más sensible para detectar las hipótesis falsas, y en este sentido con mayor poder, que pueda ser usado simultáneamente para todas las hipótesis H_i , $i \in I - \{0\}$, sin que por ello el valor de α_e exceda al de α . Surgen dos preguntas: ¿podemos de algún modo eliminar la posibilidad de que H_0 sea cierta?, si esto no es posible, ¿podemos modificar un MPS de manera que, aún y cuando H_0 sea verdadera, obtengamos un método de pruebas múltiples para el que el valor de α_e no exceda α , con mayor resolviendo que el método no modificado del nivel α ? Concerniente a la primera posibilidad, la respuesta es negativa; sin embargo, en algunas situaciones puede ser permisible asumir que H_0 es falsa y entonces podemos obtener el método modificado para probar simultáneamente todas las hipótesis implicadas por H_0 . Con respecto a la segunda posibilidad, la respuesta es afirmativa y está implícita en el siguiente "método modificado": primero se prueba H_0 por una prueba estándar de tamaño α , en la que H_0 es rechazada si $(Z_0 > \xi_0)$; si H_0 no es rechazada, entonces todas las hipótesis en H son no-rechazadas y el procedimiento termina; si H_0 se rechaza, entonces se procede a efectuar pruebas simultáneas para todas las hipótesis de H restantes, comparando sus estadísticos de prueba contra un nuevo valor crítico y rechazando cualquier H_i , $i \in I - \{0\}$, cuando $(Z_i > \xi_*)$. El valor de ξ_* se obtiene bajo el supuesto de que H_0 es falsa. Obviamente $\xi_0 > \xi_*$ y entonces este método modificado es más resolvente que el MPS no modificado, ya que la sensibilidad para detectar significancia aumenta cuando el valor crítico disminuye.

El método modificado que acabamos de describir es lo que llamaremos un método de pruebas protegidas (MPP) de nivel α y lo denotaremos por $\{H, Z, (\xi_0, \xi_*)\}$

A continuación veremos dos ejemplos de tales métodos.

Ejemplo 2.8. Encontramos un MPP de nivel α basados en la familia $\{H^C, Z^F\}$.

Para obtener los valores ξ_0^F y ξ_*^F observemos que

$$\begin{aligned} P(Z_0 > \xi) &= P[F_{(k-1, \nu)} > \xi/(k-1)] \text{ si } H_0 \text{ es cierta} \\ &= P[F_{(k-2, \nu)} > \xi/(k-2)] \text{ si } H_0 \text{ es falsa} \end{aligned}$$

donde $F_{(r, \nu)}$ denota la variable aleatoria que tiene la distribución F de Snedecor con r y ν grados de libertad.

De esta manera si elegimos a ξ_0^F y ξ_*^F como los puntos correspondientes al $(1 - \alpha)$ -percentil de las distribuciones F (aumentada) de Snedecor con $(k - 1)$ y ν grados de libertad y con $(k - 2)$ y ν grados de libertad, respectivamente, obtendremos un MPP $\{H^C, Z^F, (\xi_0^F, \xi_*^F)\}$ de nivel α .

Es fácil mostrar que $\xi_0^F \geq \xi_*^F$ lo cual confirma que este MPP tiene mayor resolviendo que su equivalente MPS. \square

Ejemplo 2.9. Un MPP de nivel no mayor que α , para con estadísticos de prueba Z^M definidos como

$$\begin{aligned} Z_0^M &= Z_0^F \\ Z_i^M &= Z_i^R \text{ para toda } i \in I \setminus \{0\} \end{aligned}$$

ha sido propuesto por Hayter (1986) y consiste en elegir -

las constantes ξ_0^M y ξ_*^M para $\{H^Q, Z^M, (\xi^M, \xi_*^M)\}$ como $\xi_0^M = \xi_0^F$ donde ξ_0^F es como en el ejemplo anterior y ξ_*^M es el punto correspondiente al $(1 - \alpha)$ -percentil de la distribución del rango - studentizado con $(k - 1)$ variables y v grados de libertad.

Observemos que este MPP es una modificación del conocido método DMS de Fisher (1935). Esta modificación tiene como objeto limitar el valor máximo de la tasa de error, que como se sabe, el DMS tradicional no controla (Hayter, 1986; Gill, 1973). \square

La principal desventaja de los MPP es que, al destruir la simultaneidad del procedimiento de prueba, la siguiente estimación de los ICS asociados a la familia de hipótesis y consistentes con las decisiones tomadas, no es posible; a diferencia de los MPS que, como vimos anteriormente, sí admiten esa interpretación.

2.4. Métodos de Comparación Polietápicas

Supongamos que contamos con un procedimiento mediante el cual podemos decidir entre rechazar ó no cada hipótesis H_i e H , usando pruebas de significancia basadas en los estadísticos de prueba Z_i ($i \in I$) de cuya distribución suponemos que está completamente especificada y es conocida bajo su respectiva hipótesis H_i , $i \in I$, y además, tal distribución es empleada para determinar la región crítica para H_i . Para caracterizar un procedimiento de este tipo, necesitamos, además de las familias H y Z , una familia $\{\alpha_i, i \in I\}$ de niveles de significancia α , equivalentemente, una familia E ,

$$E = \{\xi_i: P_{w_i}(Z_i > \xi_i) = \alpha_i, i \in I\} \quad (2.18)$$

Entonces mediante $\{H, Z, E\}$ podemos definir una familia de pruebas para H , que declara cualquier hipótesis H_i , $i \in I$, como α_i -crítica si $(Z_i > \xi_i)$. Un procedimiento arbitrario de este tipo que rechaza toda hipótesis que sea α_i -crítica, no garantiza el requerimiento de coherencia. Como vimos anteriormente, con un MPS la elección de una familia monótona de estadísticos de prueba equivale a coherencia entre las decisiones tomadas de acuerdo al método, esto se debe a que un MPS utiliza un solo valor crítico ξ contra el que se comparan todos los estadísticos de prueba. Sin embargo, ahora esto no necesariamente se cumple, pues en el caso presente, estamos contemplando una familia E de valores críticos, que a priori no tienen por que ser iguales.

Para un método $\{H, Z, E\}$ podemos escribir el requerimiento de coherencia como:

$$(X: Z_j > \xi_j) \underset{\{i:i \rightarrow j\}}{C} \cap (X: Z_i > \xi_i) \text{ c.s.} \quad (2.19)$$

para todo $j \in I - \{0\}$. En palabras, un método $\{H, Z, E\}$ se dice coherente si y sólo si cualquier hipótesis H_j es rechazada solamente si $(Z_i > \xi_i)$ para toda i tal que $H_i \subset H_j$, o, equivalentemente, el método se dice coherente si H_j es no-rechazada siempre que $(Z_i \leq \xi_i)$ para alguna i tal que $H_i \subset H_j$. En la práctica, esto puede forzarse dividiendo el procedimiento en etapas y efectuando las pruebas paso a paso, secuencialmente, partiendo de la hipótesis global que no está

implicada por ninguna otra hipótesis y continuando con la prueba de las hipótesis subsiguientes las cuales son menos restrictivas, si alguna hipótesis H_i no es rechazada, el procedimiento se interrumpe, y todas las hipótesis H_j implicadas por H_i , esto es $i \rightarrow j$, tampoco son rechazadas sin necesidad de pruebas adicionales. De esta manera, la coherencia es una consecuencia directa de la naturaleza secuencial del procedimiento (Campbell y Skillings, 1985).

Un método de comparaciones polietápico (MCP) se define como el procedimiento que utiliza la familia $\{H, Z, E\}$ de pruebas para H y es ejecutado paso a paso, para decidir la aceptación/rechazo de cualquier hipótesis H_i en H . Esta definición concuerda con la adoptada por muchos autores, por ejemplo, Einot y Gabriel (1975), Lehmann y Shaffer (1977) y Ramsey (1978), sin embargo no es la única.

Welsh (1977) define un tipo de MCP para H con la misma familia de pruebas en el que la secuencia de pruebas es ejecutada en sentido inverso. En lo que sigue, consideraremos solamente los métodos que concuerdan con la primera definición, pues además de ser los más comunes, las características principales de los dos tipos de métodos son muy similares.

Para un MCP dado, convendremos en llamar nivel del método a la máxima probabilidad de rechazar erróneamente al menos una hipótesis de la familia H y lo denotaremos por α . Los elementos de la familia $\{\alpha_i, i \in I\}$ con que definimos

a E , serán referidos como niveles de significancia nominales; nominales porque con un MCP la hipótesis H_i , $i \in I - \{0\}$, no necesariamente es rechazada cuando el evento $(Z_i > \xi_i)$ ocurre; los niveles reales son usualmente menores que los nominales, ya que el requerimiento de coherencia impuesto al MCP, es decir, el procedimiento de prueba en etapas, exige que para rechazar H_j , $j \in I - \{0\}$, los eventos $(Z_j > \xi_j)$ y $\bigcap_{\{i:i \rightarrow j\}} (Z_i > \xi_i)$ ocurran simultáneamente.

Para encontrar el valor de α_e , la probabilidad de rechazar erróneamente al menos una hipótesis de H , denotemos por \bar{I} al conjunto de índices correspondiente a la subfamilia de hipótesis que son ciertas y definamos \bar{H}_* como

$$\bar{H}_* : \theta \in \bigcap_{\bar{I}} w_i$$

es decir \bar{H}_* es la intersección de las hipótesis H_i , $i \in \bar{I}$, de modo que \bar{H}_* es necesariamente cierta, más aún las hipótesis H_i que son ciertas, son exactamente las que satisfacen $\bar{H}_* \subset H_i$. Cuando \bar{H}_* pertenece a H el valor de α_e , la probabilidad de rechazar erróneamente al menos una hipótesis de H , no excederá al valor de α_* , el nivel nominal para probar H_* . De este modo si \bar{H}_* pertenece a H se sigue que

$$\alpha_e \leq \max\{\alpha_i, i \in I\} \quad (2.21)$$

Sin embargo, esto se cumple para todos los casos solamente si H es cerrada bajo intersección y, por lo tanto, no se cumple necesariamente para cualquier MCP; así cuando la familia H no es cerrada bajo intersección el nivel de los MCP para -

estas familias, esto es, el valor máximo de α_ϵ , debe analizarse para cada caso (Petrodas y Gabriel, 1983).

Observemos que esta formulación de MCP es muy general y no depende de las propiedades de los estadísticos de prueba Z_i , $i \in I$, excepto que ξ_i , $i \in I$, pueda ser obtenido bajo su respectiva hipótesis H_i . Un MCP es, por construcción, coherente, esto, como mencionamos anteriormente, es inherente a la naturaleza secuencial del procedimiento de prueba. Por otro lado, la consonancia de cualquier MCP sí depende de las propiedades de los estadísticos de prueba y de los niveles de significancia nominales, o equivalentemente, de la familia Ξ de valores críticos (Campbell y Skillings, 1985).

A continuación veremos dos condiciones suficientes para garantizar la consonancia de un MCP.

Teorema 5. Para que un MCP (coherente por construcción) sea consonante, es suficiente que se cumpla lo siguiente:

(i) la familia de estadísticos de prueba Z sea estrictamente monótona, esto es,

$$Z_i = \max_j \{Z_j, i \rightarrow j\} \quad i \in I - I_m$$

(ii) que para cualquier par i, j tal que $i \rightarrow j$, se cumpla que

$$(Z_j > \xi_j) \supseteq (Z_i > \xi_i) \text{ c.s.} \quad (2.22)$$

Demostración. Observe que si la familia Z es monótona, (ii) implica que

$$\xi_i \geq \xi_j \text{ para toda } i \rightarrow j$$

Ahora bien, el requerimiento de coherencia indica que si H_j es aceptada, deben aceptarse todas las hipótesis implicadas por esta, y, el requerimiento de consonancia indica que si H_i es rechazada al menos una hipótesis minimal implicada por H_i debe ser rechazada.

Supongamos ahora que H_i , $i \in I - I_m$, ha sido rechazada, y consideremos la secuencia H_i', \dots, H_i'' de hipótesis implicadas por H_i en ese orden y hagamos que H_i'' sea la hipótesis minimal que tiene como estadístico de prueba a $\max_j \{Z_j; i \rightarrow j, j \in I_m\}$, entonces (i) implica que $Z_i \equiv Z_i' \equiv Z_i''$ y (i) e (ii) implican que $\xi_i \geq \xi_{i'} \geq \dots \geq \xi_{i''}$ de modo que $(Z_i > \xi_i)$ implica que $(Z_{i'} > \xi_{i'})$, ..., y $(Z_{i''} > \xi_{i''})$, lo cual muestra que estas condiciones son suficientes. //

Observación: Supongamos que (i) se cumple pero (ii) no, entonces puede tenerse que $\xi_i < \xi_j$, $i \rightarrow j$, de modo que

$$(\xi_i < \max_j \{Z_j, i \rightarrow j\} < \xi_j)$$

tiene probabilidad no nula, de donde se sigue que el MCP puede rechazar alguna hipótesis e interrumpir el procedimiento sin rechazar ninguna hipótesis minimal implicada por la hipótesis rechazada.

Supongamos ahora que (ii) se cumple pero no (i), entonces sólo se garantiza que $(Z_j < \xi_j) \cap (Z_i > \xi_i)$ para $i \rightarrow j$ tiene medida nula, pero esto no implica la coherencia del MCP; para ver esto, pensemos, por ejemplo, que Z es una familia monótona, pero no estrictamente monótona; entonces, nuevamente tendremos que $\xi_i \geq \xi_j$ para $i \rightarrow j$, sin embargo puede ocurrir que

$$(Z_j \leq \xi_j \leq \xi_i < Z_i), \quad i \rightarrow j, \quad i, j \in I$$

de modo que, con probabilidad no nula, un MCP puede rechazar alguna hipótesis y posteriormente interrumpir el procedimiento sin rechazar ninguna hipótesis implicada por la hipótesis rechazada, con lo que el MCP no sería consonante.

Si (i) y (ii) no se cumplen, es fácil construir un ejemplo de un MCP no consonante. En este sentido, podemos concluir que estas condiciones son necesarias y suficientes.

Un resultado similar al de este teorema es presentado en Lehmann y Shaffer (1977).

Note que si la condición (i) del teorema es reemplazada por la de que la familia de estadísticos de prueba sea monótona, podemos hablar de métodos que son asintóticamente consonantes.

En lo que sigue, limitaremos nuestra atención a los MCP que consideren solamente la familia de hipótesis H^Q ; además, restringiremos el estudio a los métodos cuyos estadísticos de prueba satisfacen el requerimiento de

monotonicidad. De las familias de estadísticos de prueba monótonas para H^Q , las más importantes son Z^F y Z^R definidas en la sección 2.2, por esta razón, haremos mayor énfasis en ellas.

Con las familias Z^F y Z^R , los estadísticos Z_i , $i \in I$, que nos permiten probar las hipótesis H_i de igualdad de medias dentro del subconjunto Q_i , son tales que la distribución de Z_i bajo H_i , depende de H_i solamente a través de q_i , el número de medias contenidas en Q_i . De esta manera, los estadísticos de prueba correspondientes a subconjuntos de igual tamaño, pueden ser comparados contra un mismo valor crítico y de hecho así se hace; así los $2^k - k - 1$ valores en Ξ pueden escogerse de modo que a lo más $k - 1$ de ellos sean diferentes.

Observe que con un MCP para la familia H^Q el valor de α_ϵ , la probabilidad de rechazar erróneamente al menos una hipótesis de H^Q , es equivalente a la probabilidad de rechazar al menos una hipótesis de H_m^Q , es decir, es equivalente a la probabilidad de declarar que una o más parejas de medias son diferentes cuando en realidad no lo son; tal probabilidad depende de la verdadera configuración de las medias. Así, supongamos que la configuración correspondiente a la igualdad de todas las medias es cierta, con lo que un error se cometería si H_0 , la hipótesis global, es rechazada; como $H_0 \in H^Q$, de lo discutido previamente se sigue que $\alpha_\epsilon \leq \alpha_0$. Ahora, consideremos una configuración general en la que las

medias se agrupan en n subconjuntos y éstos forman una partición de Q_0 , el conjunto de todas las medias. Note que es posible que algunos de estos subconjuntos no correspondan a ninguna hipótesis de H^Q , tales subconjuntos son aquéllos - que tienen tamaño uno, pues por ser trivialmente ciertos no necesitan ser probados. Note también que si $n = k$, entonces $\alpha_e = 0$. Supongamos que $n < k$ y que alguna configuración particular, digamos la r -ésima, es la verdadera, y que, de los subconjuntos en su respectiva partición, agrupamos en I_r - los índices que corresponden a alguna hipótesis en H^Q , de este modo se tendrá que

$$Q_0 \supseteq \bigcup_{I_r} Q_i \text{ y } Q_i \cap Q_j = \emptyset \text{ para } i, j \in I_r \quad (2.23)$$

y por lo tanto

$$\sum_{I_r} q_i = P_r, \quad P_r = 2, 3, \dots, k \quad (2.24)$$

Entonces, para rechazar erróneamente alguna hipótesis H_j en H_m^Q , debemos rechazar erróneamente la hipótesis H_i , $i \in I_r$, tal que $i \rightarrow j$ o $i = j$, es decir, para concluir que algún par de medias es diferente debemos declarar que las medias en el subconjunto que contiene ese par, también son diferentes. Se sigue que

$$\begin{aligned} \alpha_e &\leq P \left[\bigcup_{I_r} (Z_i > \xi_i) \right] \\ &= 1 - \left[P \cap_{I_r} (Z_i \leq \xi_i) \right] \end{aligned} \quad (2.25)$$

Ahora bien, si empleamos la familia Z^R o Z^F como estadísticos de prueba, note que los numeradores de los Z_i , $i \in I_r$, son estadísticamente independientes y aplicando la desigualdad de Kimball (ver el Teorema 3 del capítulo -

anterior), se tiene que

$$\begin{aligned} \alpha_e &\leq 1 - \prod_{i \in I_r} P(Z_i \leq \xi_i) \\ &= 1 - \prod_{i \in I_r} (1 - \alpha_i) \end{aligned} \quad (2.26)$$

A esto debemos agregar que la desigualdad casi siempre será estricta, como se desprende de los comentarios anteriores y de la naturaleza de la desigualdad de Kimball.

De la ecuación (2.26) se sigue que si deseamos que el conjunto de los posibles valores para α_e se encuentre acotado superiormente por un valor predeterminado, digamos α , debemos escoger la familia $\{\alpha_i, i \in I\}$ de niveles nominales de manera tal que para todas las posibles configuraciones

$$\max_{x_r} \{1 - \prod_{i \in I_r} (1 - \alpha_i)\} \leq \alpha \quad (2.27)$$

es decir, que α sea igual o mayor que el máximo sobre todas las posibles particiones del conjunto de k medias en subconjuntos de medias iguales u homogéneas.

Supongamos ahora que las condiciones bajo las que es aplicable la desigualdad de Kimball no se cumplen, entonces, a partir de (2.25) y utilizando la desigualdad de Bonferroni tenemos que (2.27) puede transformarse en

$$\max_{x_r} \{ \sum_{i \in I_r} \alpha_i \} \leq \alpha \quad (2.28)$$

Note que si la desigualdad de Kimball es aplicable y los valores de $\{\alpha_i, i \in I\}$ satisfacen (2.28), entonces también satisfacen (2.27).

Como mencionamos previamente, las familias de estadísticos de prueba más importantes en la construcción de un MCP para H^Q son Z^R y Z^F ; a continuación veremos las familias de niveles nominales comunmente usadas cuando se desea obtener un MCP de nivel α , asumiendo que los estadísticos de prueba corresponden a una de estas dos familias. Las familias de niveles en cuestión son:

(1) La familia de niveles propuesta por Newman (1939) y Keuls (1952) cuyos elementos son

$$\alpha_i = \alpha \quad \text{para toda } i \in I$$

Esta familia es inaceptable para $k > 3$ ya que de (2.24) se tiene que

$$\max_{i \in I_r} \{1 - \prod_{i \in I_r} (1 - \alpha)\} = 1 - (1 - \alpha)^k$$

y por lo tanto (2.27) no se cumple, de manera que no podemos garantizar que $\alpha_e \leq \alpha$, es decir, el MCP no es de nivel α . Si $k \leq 3$ entonces H^Q es una familia cerrada y de (2.21) se sigue que en este caso el MCP sí es de nivel α . Para ($k > 3$) la situación puede remediarse adicionando algunas pruebas de manera que α_e permanezca acotado por α . Este punto no será tratado aquí, el lector interesado puede consultar el trabajo de Begun y Gabriel (1981).

(2) La familia de niveles basada en el criterio de Duncan (1955) que consiste en elegir

$$\alpha_i = 1 - (1 - \alpha)^{d_i} \quad \text{para toda } i \in I \quad (2.29)$$

donde $d_i = (q_i - 1)/(k - 1)$. En realidad, esta es una modificación del criterio original ideado para limitar el número de rechazos erróneos por comparación y no por experimento, como en este caso.

(3) La familia de niveles basada en el criterio de Ryan (1959) cuyos elementos están dados por

$$\alpha_i = 1 - (1 - \alpha)^{t_i} \text{ para toda } i \in I \quad (2.30)$$

donde $t_i = q_i / k$

(4) La familia de niveles de Ryan modificada por Welsch (1977) y Ramsey (1978), que es básicamente la misma que la anterior, solamente que ahora

$$\alpha_i = \alpha \quad \text{si } q_i = k - 1$$

Sin embargo, con este cambio no siempre se cumple (2.22), es decir, puede suceder que $\xi_i < \xi_j$ cuando $q_i = k - 1$ y $q_j = k - 2$, pero esta anomalía puede remediarse definiendo a ξ_i como $\xi_i = \max\{\xi_i, \xi_j\}$ donde $q_i = k - 1$ y $q_j = k - 2$, sin que se afecte el nivel del método.

Es fácil ver que para las tres últimas familias (2.27) se cumple, de modo que los MCP basados en estas son de nivel α .

Ahora bien, si comparamos (2.29) y (2.30) tenemos que

$$(1 - \alpha)^{t_i} \leq (1 - \alpha)^{d_i}$$

y la desigualdad es estricta para $t_i < 1$, es decir, para

todo $q_i < k$, por lo que los métodos que utilizan las familias $\{\alpha_i, i \in I\}$ basada en el criterio de Ryan son estrictamente más resolventes y estrictamente menos exigentes que los que se basan en el criterio de Duncan; por lo anterior, no existe ninguna razón para utilizar un MCP basado en el criterio de Duncan.

Por otro lado, debido a que para cada familia de niveles nominales corresponde una familia de valores críticos, (a diferencia de los MPS que usan sólo un valor crítico), los MCP son computacionalmente muy laboriosos, ya que es muy difícil de conseguir tablas para las familias de valores críticos.

Ejemplo 2.10. Utilicemos el criterio de Duncan y el de Ryan para obtener la familia de niveles nominales y de valores críticos para un MCP de nivel 0.01, con $k = 5$, y $v = 20$ cuando las hipótesis son:

H_i : todas las medias dentro del subconjunto Q_i ,
de tamaño q_i , son iguales

(i) Duncan

q_i	α_i	ξ_i^R	ξ_i^F
5	0.0100	5.294	17.722
4	0.0075	5.161	15.888
3	0.0050	5.074	13.973
2	0.0025	4.886	11.941

(ii) Ryan:

q_i	α_i	ξ_i^R	ξ_i^F
5	0.0100	5.294	17.722
4	0.0080	5.123	15.645
3	0.0060	4.933	13.359
2	0.0040	4.598	10.569

donde:

(a) ξ_i^R es el i -ésimo elemento de Ξ para el MCP basado en $\{H^Q, Z^R\}$; además, ξ_i^R corresponde al $(1-\alpha_i)$ -percentil de la distribución del rango studentizado con q_i variables y ν grados de libertad.

(b) ξ_i^R es el i -ésimo elemento de Ξ para el MCP basado en $\{H^Q, Z^F\}$; además, $\xi_i^F = (q_i - 1)F_{\alpha_i}$, donde F_{α_i} es el punto correspondiente al $(1-\alpha_i)$ -percentil de la distribución F de Snedecor con $(q_i - 1)$ y ν grados de libertad. \square

Para finalizar, veamos cuales son las ventajas y desventajas de un MCP con respecto a un MPS. La única ventaja de un MCP es que con este se tiene más capacidad para detectar significancia en las hipótesis, que con un MPS; a cambio de esto, destruye la simultaneidad del procedimiento de prueba con lo que las decisiones acerca de significancia de una hipótesis, dependen en parte de la configuración de los parámetros no involucrados por la hipótesis, es decir, dependen de las decisiones tomadas para otras hipótesis; esta dependencia es intuitivamente indeseable (Miller, 1966; Einot y Gabriel, 1975); así mismo, con los MCP la construcción de -

ICS para las funciones paramétricas asociadas a las hipótesis y consistentes con las decisiones tomadas, no es posible. A diferencia de los MCP, lo MPS, aunque con un poco menos de capacidad para detectar significancia, tienen la ventaja de ser computacionalmente simples, las hipótesis son probadas sin referencia de una a otra y los ICS para las funciones paramétricas asociadas a la familia de hipótesis son consistentes con las decisiones tomadas y, además, no requieren de tablas extensivas para los valores críticos y éstas pueden encontrarse en muchas partes. Una discusión más extensiva sobre estos aspectos puede encontrarse en Gabriel (1964, 1978a) Einot y Gabriel (1975) y Begun y Gabriel (1981).

EPILOGO

En el presente trabajo hemos estudiado los métodos de construcción de ICS y los métodos de comparaciones múltiples. En la primera parte del trabajo expusimos un grupo de conceptos y resultados relacionados con el problema de la construcción de ICS; aquí hicimos énfasis en el hecho de que la elección de un método depende fundamentalmente de la familia de funciones parametrales para los que se desea el conjunto de ICS. En la parte técnica, hemos presentado un resultado, el Teorema de Uusipaikka, que ofrece, al menos en teoría, una solución general del problema para el caso en que las variables en estudio están normalmente distribuidas; se desarrollaron algunos métodos en el caso normal, principalmente para tres diferentes familias de funciones parametrales, con un diferente conjunto de supuestos acerca de la distribución de las variables, para ser más precisos, con diferentes supuestos sobre la estructura de la matriz de varianzas-covarianzas; estas familias de funciones y los métodos recomendados para ellas son:

(1) La familia de todas las funciones de la forma $c'\theta$ con c en un subespacio lineal arbitrario; para esta familia

se recomienda la aplicación del método de Scheffé.

(2) La familia de todas las diferencias entre pares de medias. Para esta familia de funciones se recomienda el método de Tukey si la matriz de varianzas-covarianzas de las variables en estudio es balanceada; cuando esto no se cumple, podemos aplicar el método de Tukey modificado si las variables son independientes o el método de Hochberg en el otro caso; estos dos últimos métodos garantizan que la cota mínima del valor para el CCS será igual o mayor al valor nominal.

(3) La familia de todas las coordenadas individuales del vector de medias, es decir, los intervalos son para cada una de las medias; para esta familia se recomienda el empleo del método de Tukey-Roy-Bose siempre que las variables sean independientes.

En la segunda parte del trabajo hemos estudiado el problema de las comparaciones múltiples. Ahí, vimos la importancia de considerar las relaciones de implicación existente entre las hipótesis pertenecientes a la familia de hipótesis; utilizamos esto para ver que cuando deseamos tener cierta estructura lógica en las decisiones inducidas por la familia de hipótesis y el método de comparaciones múltiples para esta familia, podemos obtenerla imponiendo ciertos requerimientos a los métodos; en nuestro caso, vimos que las condiciones de coherencia y consonancia nos permiten que las relaciones de implicación entre las hipótesis se preserven en las decisiones obtenidas. En este sentido, mostramos que las

familias de estadísticos de pureba que poseen la propiedad de monotonicidad reúnen atributos que se relacionan con ésto y encontramos algunas ventajas en los estadísticos de prueba que satisfacen el principio UI, es decir, en las familias de estadísticos que cumplen con la condición de monotonicidad estricta.

Con respecto a los diferentes tipos de métodos de comparaciones múltiples, hemos visto que la analogía existente entre MPS y métodos de construcción de ICS es de especial importancia, pues esta relación amplía el espectro de métodos posibles con los que podemos trabajar ya que, como vimos, los MPS pueden ser formulados en términos de regiones de confianza; recíprocamente, con un MPS podemos encontrar un conjunto de ICS para todas las funciones parametrales asociadas a la familia de hipótesis consistentes con las decisiones obtenidas en base al MPS; esto no sucede con los MPP ni con los MCP, aunque estos últimos superan a los MPS en capacidad para detectar significancia en las hipótesis. Muchos investigadores consideran que los MCP son más apropiados para la explotación de datos, aunque en muchas situaciones prácticas la capacidad de los MCP para detectar significancia no es contundentemente superior a la de los MPS y, en cambio, sí son computacionalmente más complicados.

Antes de finalizar, es necesario tocar tres puntos importantes que apenas si fueron mencionados en el escrito. Primero, cuando nos referimos al nivel de un método de comparaciones múltiples, hemos considerado al experimento como

unidad conceptual; por esta razón podemos interpretar el valor del nivel como la tasa de error experimental; cuando los métodos son coherentes, este nivel parece ser igual a la probabilidad de error de tipo I en las decisiones para todas las hipótesis. Esto es importante tomarlo en cuenta ya que existen algunos enfoques en los que la unidad conceptual básica, en función de la cual se determina el nivel, es cada comparación; este es, por ejemplo, el caso del método de Duncan tradicional. Segundo, en general, solamente hemos considerado el error de tipo I, pues aunque definimos algunos conceptos relacionados al error de tipo II, esto es, el error que se comete al no rechazar una hipótesis que en realidad es falsa, este tipo de error no fué tratado explícitamente. Tampoco tratamos un tipo de error especialmente importante en comparaciones múltiples, a saber: el error de tipo III, que se define como el error que se comete al rechazar una hipótesis en una dirección equivocada, como por ejemplo, cuando se declara que $\theta_i < \theta_j$ cuando en realidad $\theta_j < \theta_i$; se sabe que el riesgo de cometer este tipo de error es mayor con un MCP. Tercero, al desarrollar los métodos de comparaciones múltiples, hemos asumido que el experimento está diseñado específicamente para investigar ciertas hipótesis que de alguna manera reflejan nuestros objetivos, es decir, que contamos con la familia de hipótesis antes de la inspección de los datos; por esta razón, es erróneo pensar, como muchas veces se hace, que los métodos de comparaciones múltiples son procedimientos para hacer inferencia acerca de todas las hipótesis sugeridas por los datos; esta interpretación daña las propiedades distribucionales de los

estadísticos bajo los que los métodos fueron desarrollados.

Es así como hemos intentado presentar los aspectos más relevantes de la teoría en que se fundamentan los métodos de construcción de ICS y los métodos de comparaciones múltiples.

LITERATURA CITADA

- Anderson, T.W. 1958. An introduction to multivariate statistical analysis. Wiley. New York, U.S.A. 375 p.
- Begun, J.M. and K.R. Gabriel. 1981. Closure of the Newman-Keuls multiple comparisons procedure. J. Amer. Statist. Assoc. 76:241-245. U.S.A.
- Brown, H. 1983. A proof that Tukey-Kramer multiple comparison procedure for difference between means is level- α for 3, 4 or 5 means. Amer. Statist. 27:74-81. U.S.A.
- Campbell, G. and J.H. Skillings. 1985. Nonparametric stepwise multiple comparison procedures. J. Amer. Statist. Assoc. 80:998-1003. U.S.A.
- Duncan, D.B. 1955. Multiple range and multiple F tests. Biometrics. 11:1-42. U.S.A.
- Dunnnett, C.W. 1955. A multiple comparison procedure for comparing several treatments with a control. J. Amer. Statist. Assoc. 50:1096-1121. U.S.A.
- _____. 1964. New tables for multiple comparisons with a control. Biometrics 20:482-491. U.S.A.
- _____. 1980. Pairwise multiple comparisons in the homogeneous variance, unequal sample size case. J. Amer. Statist. Assoc. 75:789-795. U.S.A.

- Einot, I. and K.R. Gabriel. 1975. A study of the powers of several methods of multiple comparisons. J. Amer. Statist. Assoc. 70:574-583. U.S.A.
- Feller, W. 1968. An introduction to probability theory and applications. Vol. 1. Wiley. New York. U.S.A. 419 p.
- Fisher, R.A. 1935. The design of experiments. Oliver and Boyd Ltd. London, U.K. 358 p.
- Gabriel, K.R. 1964. A procedure for testing the homogeneity of all sets of means in analysis of variance. Biometrics 20:459-477. U.S.A.
- _____. 1969. Simultaneous test procedures-Some theory of multiple comparisons. Ann. Math. Stat. 40:224-250. U.S.A.
- _____. 1978a. Comments on Ramsey's paper. J. Amer. Statist. Assoc. 73:485-487. U.S.A.
- _____. 1978b. A simple method of multiple comparisons of means. J. Amer. Statist. Assoc. 73:724-729. U.S.A.
- Games, P.A. 1977. An improved t table for simultaneous control on g contrasts. J. Amer. Statist. Assoc. 72:531-534. U.S.A.
- Genizi, A. and Y. Hochberg. 1978. On improved extensions of the T-method of multiple comparisons for unbalanced designs. J. Amer. Statist. Assoc. 73:879-884. U.S.A.
- Gill, J.L. 1973. Current status of multiple comparisons of means in designed experiments. J. of Daily Sci. 56: 973-977. U.S.A.
- Graybill, F.A. 1969. Introduction to matrices with applications in Statistics. Wadsworth Publishing Co. Belmont, Cal. U.S.A. 372 p.

- Harter, H.L. 1960. Tables of range and studentized range. *Ann. Math. Stat.* 31:1122-1147. U.S.A.
- Hayter, A.J. 1984. A proof of the conjecture that the Tukey-Kramer multiple comparisons procedure is conservative. *Ann. Stat.* 12:61-75. U.S.A.
- _____. 1986. The maximum family wise error rate of Fisher's least significant difference test. *J. Amer. Stat. Assoc.* 81:1000-1004. U.S.A.
- Keuls, M. 1952. The use of the 'studentized range' in connection with an analysis of variance. *Euphytica* 1:112-122. Wageningen, Ned.
- Kramer, C.V. 1956. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics* 12:307-310. U.S.A.
- _____. 1957. Extension of multiple range tests to group correlated adjusted means. *Biometrics* 13:13-18. U.S.A.
- Lehmann, E.L. 1957. A theory of some multiple decision problems. *Ann. Math. Stat.* 28:1-25, 547-572. U.S.A.
- _____ and J.P. Shaffer. 1977. On a fundamental theorem in multiple comparisons. *J. Amer. Statist. Assoc.* 72:576-578. U.S.A.
- Miller, R.G., Jr. 1966. *Simultaneous statistical inference*. McGraw-Hill. New York, U.S.A. 272 p.
- _____. 1977. Developments in multiple comparisons 1966-1976. *J. Amer. Statist. Assoc.* 72:779-788. U.S.A.
- Mood, A.M.; F.A. Graybill and D.C. Boes. 1985. *Introduction to the theory of statistics*. McGraw-Hill. Singapore. 564 p.

- Newman, D. 1939. The distribution of the range in samples - from the normal population, expressed in terms of an independent estimate of standard deviation. *Biometrika* 31:20-30. United Kingdom.
- Pearce, S.C. 1964. Experimenting with blocks of natural size. *Biometrics* 20:699-706. U.S.A.
- Petrodas, D.A. and K.R. Gabriel. 1983. Multiple comparisons - by rerandomization test. *J. Amer. Statist. Assoc.* 78: 949-957. U.S.A.
- Ramsey, H.P. 1978. Power differences between pair wise multiple comparisons. *J. Amer. Statist. Assoc.* 73:479-485. U.S.A.
- Rao, C.R. 1973. Linear statistical inference and its applications. Wiley. New York, U.S.A. 625 p.
- Roy, S.N. and R.C. Bose. 1953. Simultaneous confidence interval estimation. *Ann. Math. Stat.* 20:513-536. U.S.A.
- Ryan, T.A. 1959. Multiple comparisons in psychological - research. *Psychological Bulletin* 56:26-47. United - Kingdom.
- Sidák, Z. 1967. Rectangular confidence regions for the means of multivariate normal distributions. *J. Amer. Statist. Assoc.* 62:626-633. U.S.A.
- _____. 1971. On probabilities of rectangles in multivariate student distributions: their dependence on correlations. *Ann. Math. Stat.* 42:169-175. U.S.A.
- Scheffé, H. 1953. A method for judging all contrasts in the analysis of variance. *Biometrika* 40:87-104. United - Kingdom.
- _____. 1959. The analysis of variance. Wiley. New York. U.S.A. 477 p.

- Searle, S.R. 1971. Linear models. Wiley. New York. U.S.A. -
532 p.
- Stoline, M.R. 1978. Tables of the studentized augmented range
and applications to problems of multiple comparisons.
J. Amer. Statist. Assoc. 73:656-660. U.S.A.
- _____ and H.K. Uri. 1979. Tables of the studentized -
maximum modulus distribution and an application to -
multiple comparison among means. Technometrics 21:87-
93. U.S.A.
- Tukey, J.W. 1951. Quick and dirty methods in statistics. Part
II. Simple analyses for standard designs. Proc. Fifth
Ann. Conven. Amer. Soc. Quality Control. 189-197. -
U.S.A.
- Uusipaikka, E. 1983. Exact confidence bands for linear regre-
ssion over intervals. J. Amer. Statist. Assoc. 78:638
-644. U.S.A.
- _____. 1985. Exact simultaneous confidence intervals
for multiple comparisons among three or four mean va-
lues. J. Amer. Statist. Assoc. 80:196-201. U.S.A.
- Welsh, R.E. 1977. Stepwise multiple comparison procedures. J.
Amer. Statist. Assoc. 72:566-575. U.S.A.
- Wilks, S.S. 1962. Mathematical statistics. Wiley. New York.
U.S.A. 644 p.

APENDICE

En este apéndice presentaremos la demostración de los tres lemas expuestos en el primer capítulo cuya prueba dejamos pendiente.

Lema 2. (i) Si A y B son dos matrices cuadradas de orden n . Entonces los eigenvalores de AB coinciden con los de BA .

(ii) Si A y B son dos matrices de orden $m \times n$ y $n \times m$, respectivamente, donde $m \leq n$. Entonces los eigenvalores de BA son $(n - m)$ ceros y los m valores propios de AB .

Demostración. Si A y B son cuadradas, sabemos que los eigenvalores de AB y BA son, respectivamente, las soluciones a

$$|AB - \lambda I| = 0 \quad |BA - \lambda I| = 0$$

vamos a demostrar que

$$|AB - \lambda I| = |BA - \lambda I|$$

lo cual es equivalente a la afirmación hecha en (i).

Dada una matriz cuadrada C de orden n , consideremos la función

$$f(t) = |tC + \lambda I|$$

Obviamente $f(t)$ es un polinomio de grado n ; si desarrollamos el polinomio en un entorno de $t = 0$, obtenemos

$$f(t) = f(0) + f'(0)t + \frac{1}{2} f''(0)t^2 + \dots + \frac{1}{n!} f^{(n)}(0)t^n$$

Utilizando las reglas usuales de derivación de determinantes (ver Graybill, 1969), tenemos que

$$f(0) = \lambda^n$$

$$f'(0) = \lambda^{n-1} + \text{tr}(C)$$

y en general $f^{(j)}(0) = \lambda^{n-j} h_j(C)$ $j = 2, 3, \dots, n$
 donde $h_j(C)$ sólo depende de $t(C), t(C^2), \dots, t(C^j)$.

De acuerdo a ésto, la ecuación característica de -
 viene dada por

$$f(-1) = \lambda - \lambda^{n-1} \text{tr}(C) + \lambda^{n-2} h_2(C) - \lambda^{n-3} h_3(C) + \dots + (-1)^n h_n(C)$$

Definamos ahora $C_1 = AB$ y $C_2 = BA$ y observemos que

$$\text{tr}(C_1) = \text{tr}(C_2)$$

y en general

$$\text{tr}(C_1^j) = \text{tr}(C_2^j)$$

ya que

$$C_2^j = BC_1^{j-1} A, \quad j = 2, 3, \dots, n$$

con lo que

$$\begin{aligned} \text{tr}(C_2) &= \text{tr}(BC_1^{-1}A) \\ &= \text{tr}(C_2^{-1}AB) \\ &= \text{tr}(C_1) \end{aligned}$$

y en consecuencia, tenemos que

$$h_j(C_1) = h_j(C_2)$$

y por lo tanto

$$|\lambda I - AB| = |\lambda I - BA|$$

lo cual prueba (i). Veremos ahora que (ii) se sigue de (i).

Ahora supongamos que A y B son matrices $m \times n$ y $n \times m$
 y definamos las matrices $A_0 = (A', 0')$ y $B_0 = (B, 0)$, de modo

que A_0 y B_0 sean matrices cuadradas de orden n . Por (i) los eigenvalores de A_0B_0 coinciden con los de B_0A_0 ; pero

$$A_0B_0 = \begin{bmatrix} AB & 0 \\ 0 & 0 \end{bmatrix} \quad B_0A_0 = BA$$

y por lo tanto

$$|BA - \lambda I| = \begin{vmatrix} AB - \lambda I & 0 \\ 0 & \lambda I \end{vmatrix}$$

de modo que

$$|BA - \lambda I| = \lambda^{n-m} |AB - \lambda I| \quad //$$

Observe que para cualquier par de matrices A y B para las que los productos AB y BA están definidos, se cumple que si AB es no-negativa definida entonces $C_m(AB) = C_m(BA)$. Esto se utiliza en la siguiente demostración.

Lema 3. Sea A una matriz simétrica y V una matriz p.d., ambas de orden n ; sea además B una matriz $n \times k$ de rango completo por columnas. Entonces

$$\max \frac{x'Ax}{x'Vx} = C_m [B(B'VB)^{-1} B'A]$$

donde el máximo se toma sobre los vectores $x \neq 0$ tales que $x \in L[B]$; y

(ii) el máximo se alcanza cuando x es cualquier eigenvector de $B(B'VB)^{-1}B'A$ correspondiente a $C_m[B(B'VB)^{-1}B'A]$

Demostración. Haremos la prueba de (i) en tres etapas. Primero veremos que

(1) Si tomamos el máximo sobre todos los vectores no-nulos en \mathbb{R}^n tenemos que

$$\max \frac{x'Ax}{x'x} = C_m [A]$$

Sea D la matriz que en la diagonal tiene los eigenvalores de A y P la matriz cuyas columnas son los eigenvectores correspondientes, es decir, $AP = PD$ y $PP' = I$

Obsérvese que cualquier x puede escribirse como $x = P'c$, $c \in \mathbb{R}^n$. Entonces

$$\begin{aligned} \frac{x'Ax}{x'x} &= \frac{c'Dc}{c'c} \\ &= \bar{w}'Dw \end{aligned}$$

donde $\|w\| = 1$ y $c = \|c\| w$

Luego

$$\begin{aligned} \max_{x \in \mathbb{R}^n} \frac{x'Ax}{x'x} &= \max_{\{w: \|w\|=1\}} [w'Dw] \\ &= C_m [D] \\ &= C_m [A] \end{aligned}$$

de manera que (1) se cumple.

Ahora veremos que

(2) si tomamos el máximo como en (1), entonces

$$\max \frac{x'Ax}{x'Vx} = C_m [V^{-1}A]$$

Para ver ésto, observe que como V es p.d. podemos suponer que es simétrica ya que

$$x'Vx = \frac{1}{2} x'(V + V')x$$

y entonces existe una matriz L tal que $V = LL'$, así, si escribimos $c = L'x$, $c \in \mathbb{R}^n$

$$\frac{x'Ax}{x'Vx} = \frac{c'L^{-1}A(L^{-1})'c}{c'c}$$

de manera que usando ésto y lo establecido en (1), además del Lema 2, tenemos que

$$\begin{aligned} \max_x \frac{x'Ax}{x'Vx} &= C_m |L^{-1}A(L^{-1})'| \\ &= C_m |(L^{-1})'L^{-1}A| \\ &= C_m |V^{-1}A| \end{aligned}$$

(3) Finalmente, observemos que cualquier $x \in L[B]$ puede escribirse como $x = Bc$, $c \in \mathbb{R}^k$. De este modo tenemos que

$$\max_x \frac{x'Ax}{x'Vx} = \max_c \frac{c'B'ABC}{c'B'VBC}$$

donde $x \in L[B]$ y $c \in \mathbb{R}^k$.

Ahora utilizando lo establecido en (2) junto con el Lema 2, tenemos que

$$\begin{aligned} \max_c \frac{c'B'ABC}{c'B'VBC} &= C_m [(B'VB)^{-1} B'AB] \\ &= C_m [B(B'VB)^{-1} B'A] \end{aligned}$$

que es lo que queríamos demostrar.

(ii) Sea $\gamma = C_m [B(B'VB)^{-1} B'A]$ y x cualquier eigenvec tor de $B(B'VB)^{-1} B'A$ correspondiente a γ . Entonces

$$B(B'VB)^{-1} B'Ax = \gamma x$$

implica que

$$x'VB(B'VB)^{-1}B'Ax = \gamma x'Vx$$

que a su vez, implica que

$$\frac{x'Ax}{x'Vx} = \gamma$$

ya que $x \in L[\bar{B}]$ implica que

$$x'VB(B'VB)^{-1}B' = X'$$

//

Lema 5. Una matriz simétrica, de orden k , satisface que $c'Vc = a^2c'c$ para toda $c \in L_C$, si y sólo si puede encontrarse una matriz Q tal que $QQ' = V$ y $Q'c = ac$ para toda $c \in L_C$.

Demostración. Si $QQ' = V$ y $Q'c = ac$ para toda $c \in L_C$, es claro que $c'Vc = a^2c'c$ para toda $c \in L_C$.

Ahora veremos que $c'Vc = a^2c'c$ para toda $c \in L_C$ implica que $QQ' = V$ y $Q'c = ac$ para toda $c \in L_C$.

Sea $C = (c_1, c_2, \dots, c_{k-1})$ la matriz cuyas columnas son una base ortonormal para L_C . Entonces

$$c'Vc = a^2c'c \text{ implica que } C'VC = a^2I_{k-1} \text{ ya que } c_i'Vc_j = 0.$$

$i \neq j, i, j = 1, 2, \dots, k-1.$

$$\begin{aligned} (c_i + c_j)'V(c_i + c_j) &= a^2(c_i'c_j + 2c_i'c_j + c_j'c_j) \\ &= a^2(c_i'c_i + c_j'c_j) \end{aligned}$$

Tomemos ahora cualquier matriz Q_0 tal que $Q_0 Q_0' = V$.
Entonces $c' V c = a^2 c' c$ para toda $c \in L_C$ implica que $Q_0' c = a \bar{c}$
y las columnas de \bar{C} también son ortogonales ya que

$$\begin{aligned} C' V C &= C' Q_0 Q_0' C \\ &= a^2 \bar{C}' \bar{C} \\ &= a^2 I_{k-1} \end{aligned}$$

Es fácil ahora encontrar una matriz ortogonal H tal -
que

$$H' \bar{C} = C$$

tomando C y \bar{C} y completándolas a dos conjuntos ortonormales -
(C_0, C) y (\bar{C}_0, \bar{C}), de manera que

$$H = (\bar{C}_0, \bar{C}) (C_0, C)'$$

y entonces $Q = Q_0 H$ satisface que $Q Q' = V$ y $Q' c = a c$
ya que $Q' c = H' Q_0' c$

$$= a H' \bar{C}$$

$$= a c$$

//