Universidad de Valparaíso
CHILE

Ingeniería en Estadística

Modelo de Análisis de Varianza con medidas repetidas de dos factores

Presentado por:
Cynthia Choque Aros

Profesor Guía:
Phd. Claudia Navarro Villarroel
Profesor Co-Guía:
Enrique Cabrera Vicencio

Valparaíso, Chile, 2017

“Je planmäßiger ein Mensch vorgeht, desto wirksamer vermag ihn der Zufall zu treffen.”
“Cuando más metódico un ser humano avanza, más fuerte el azar le puede afectar.”
Friedrich Dürrenmatt

Agradecimientos

Quisiera aprovechar estas líneas para expresar mis más sinceros agradecimientos a quienes están y estuvieron acompañándome en mis desafíos, tanto personales como en este proyecto de titulación; en especial, a los profesores Enrique Cabrera y Dra. Claudia Navarro, quienes con paciencia, motivación y apoyo, me ayudaron a cumplir un sueño que compartía con mi madre desde pequeña y que sin ellos no hubiera sido posible. Realmente gracias por el seguimiento y la orientación, ya que nada de esto sería posible sin ustedes.
A mi madre, ya que sola nos sacó adelante a mi hermana y a mí, sin tu guía en la vida quien sabe dónde estaría ahora y no solo te debo la vida, te debo el cariño incondicional, gracias por el apoyo que siempre me has dado.
A Javiera y a mi hermana Karina, gracias por la ayuda, el apoyo y el cariño en estos años, no ha sido fácil para ninguna, gracias por enseñarme y por aprender en este camino universitario lleno de altos y bajos, ha sido duro, y ya estamos en el final, espero que el futuro sea brillante para todas.
Miguel, mi más profundo agradecimiento por ser mi guía y pareja en estos años, por ayudarme y apoyarme, por no dejar que me rindiera cuando ya no quería seguir, muchas gracias.
Mi más sincero reconocimiento y gratitud a todos ustedes.

Resumen

El análisis de varianza, más conocido como ANOVA (analysis of variance), en el caso univariante, mientras que en el caso multivariante se denomina MANOVA (multivariate analysis of variance), se utiliza para determinar si diferentes tratamientos muestran diferencias significativas, permitiendo compararlas bajo diversas situaciones. Este tipo de análisis permite contrastar más de dos grupos o tratamientos en forma simultánea. A partir de esto, se han desarrollado diseños para el análisis de varianza, como en el caso del análisis de varianza de un factor. Seber (2003), explica que un investigador compara el efecto de cuatro anti-inflamatorios en pacientes con artritis, tomando una muestra aleatoria de pacientes, separándolos al azar en cuatro grupos de igual tamaño, cada grupo recibe uno de los medicamentos, obteniendo cuatro muestras independientes que reciben tratamientos diferentes, este experimento completo se le denomina experimento de un factor (medicamento) o de una clasificación, el que permite comparar medias de tres o más muestras con respecto a la variable de interés. Otro de los diseños para el análisis de varianza está el de dos factores, que a diferencia del caso de un factor, éste permite estudiar simultáneamente los efectos de dos factores o fuentes de variación, por ejemplo, siguiendo el caso anterior, además de que cada grupo reciba uno de los medicamentos, se apliquen cuatro dosis distintas.

El análisis de varianza con medidas repetidas sirve para estudiar el efecto de uno o más factores, el que se caracteriza principalmente porque a un grupo de sujetos se les hace pasar por todos los niveles del o los tratamientos. Su finalidad es facilitar la recolección de datos, minimizar costos, facilitar el análisis, entre otros. A continuación se estudia el análisis de varianza con medidas repetidas en dos factores, tanto su análisis univariante como multivariante, y de esta forma comparar estructura, hipótesis, test, y presentar la posibilidad del desarrollo de un código para el software *R-project*.

Objetivos

Objetivo General

Presentar el análisis de varianza a dos factores con medidas repetidas en forma univariada y multivariada, a través del uso de R-Project.

Objetivos Específicos

1. Presentar el modelo a dos criterios de clasificación con medidas repetidas a dos factores, con sus estimadores y test de hipótesis respectivos, usando un modelo univariado y multivariado.

2. Programar un algoritmo en R-Project que permita el análisis de varianza con medidas repetidas a dos factores.

3. Aplicar el algoritmo a datos obtenidos desde la bibliografía.

Índice general

Índice de cuadros

Capítulo 1

Análisis de Varianza con Medidas Repetidas de dos Factores

En este capítulo se expone el análisis de varianza con medidas repetidas de dos factores de un modelo univariado y uno multivariado, el que consiste en aplicar al mismo grupo de sujetos más de un tratamiento, y de esta forma evaluar su efecto sobre la variable dependiente, la finalidad es comparar ambos modelos.

1.1 Análisis de varianza univariante con medidas repetidas de dos factores

Ximénes & San Martín (2000) plantea que existen dos modelos fundamentales para este tipo de análisis, el aditivo y el no aditivo, su principal diferencia consiste en que este último considera la interacción entre los n sujetos y los jk-tratamientos. A continuación, se presenta un resumen con lo más importante que plantea Ximénes & San Martín (2000). Comenzando con lo más simple, la estructura de los datos, o el orden de los mismos, lo que ayudará a comprender tanto los modelos que se plantean, sus estimadores y resultados.

1.1.1 Estructura de los datos y su modelo

Como se explica al comienzo, Ximénes & San Martín (2000), expone dos modelos para trabajar con los datos, los cuales deben tener el siguiente orden o estructura:

Cuadro 1.1: Estructura de los datos y notación

	A_1					$\dots$	A_j					$\dots$	A_a				
Sujetos	B_1	$\dots$	B_k	$\dots$	B_b		B_1	$\dots$	B_k	$\dots$	B_b		B_1	$\dots$	B_k	$\dots$	B_b
P_1	y_{111}	$\dots$	y_{11k}	$\dots$	y_{11b}		y_{1j1}	$\dots$	y_{1jk}	$\dots$	y_{1jb}		y_{1a1}	$\dots$	y_{1ak}	$\dots$	y_{1ab}
P_2	y_{211}	$\dots$	y_{21k}	$\dots$	y_{21b}		y_{2j1}	$\dots$	y_{2jk}	$\dots$	y_{2jb}		y_{2a1}	$\dots$	y_{2ak}	$\dots$	y_{2ab}
		$\dots$		$\dots$				$\dots$		$\dots$				$\dots$		$\dots$	
P_i	y_{i11}	$\dots$	y_{i1k}	$\dots$	y_{i1b}		y_{ij1}	$\dots$	y_{ijk}	$\dots$	y_{ijb}		y_{ia1}	$\dots$	y_{iak}	$\dots$	y_{iab}
		$\dots$		$\dots$				$\dots$		$\dots$				$\dots$		$\dots$	
P_n	y_{n11}	$\dots$	y_{n1k}	$\dots$	y_{n1b}		y_{nj1}	$\dots$	y_{njk}	$\dots$	y_{njb}		y_{na1}	$\dots$	y_{nak}	$\dots$	y_{nab}

A partir del cuadro 1.1, se puede apreciar que a un grupo de n sujetos, se les aplica a niveles de un factor (A) y b niveles de otro factor (B), las cuales son fijadas por el investigador, el cual debe asegurar que los n sujetos pasen por todos y cada uno de los niveles de ambos factores. Los niveles de A están representados por $A_1, \dots, A_j, \dots, A_a$, de esta forma se puede deducir que j es el subíndice utilizado para referirse a los niveles del factor A, de la misma forma, los niveles del factor B, están representados por $B_1, \dots, B_k, \dots, B_b$, siendo en este caso k el subíndice para referirse a los niveles de dicho factor, mientras que los sujetos son representados como $P_1, \dots, P_i, \dots, P_n$. Con esto último, se pueden proponer los siguientes modelos:
Modelo aditivo, incluyendo la interacción entre los tratamientos:

$$y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + P_i + \varepsilon_{ijk}$$

Modelo no aditivo:
$y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + P_i + (P\alpha)_{ij} + (P\beta)_{ik} + (P\alpha\beta)_{ijk} + \varepsilon_{ijk}$;

{i=1,...,n},{j=1,...,a} y {k=1,...,b}.

cuya gran diferencia, es que el primero no considera la interacción entre los n sujetos y los tratamientos.

Desde las puntuaciones de las casillas que se encuentran en la cuadro 1.1, se pueden obtener los siguientes resultados:

Cuadro 1.2: Totales y Promedios

Totales	Promedios
$T = \sum_i \sum_j \sum_k y_{ijk}$	$\bar{y} = \frac{T}{nab}$
$T_{.j.} = \sum_i \sum_k y_{ijk}$	$\bar{y}_{.j.} = \frac{T_{.j.}}{nb}$
$T_{..k} = \sum_i \sum_j y_{ijk}$	$\bar{y}_{..k} = \frac{T_{..k}}{na}$
$T_{i..} = \sum_j \sum_k y_{ijk}$	$\bar{y}_{i..} = \frac{T_{i..}}{ab}$
$T_{ij.} = \sum_k y_{ijk}$	$\bar{y}_{ij.} = \frac{T_{ij.}}{b}$
$T_{i.k} = \sum_j y_{ijk}$	$\bar{y}_{i.k} = \frac{T_{i.k}}{a}$
$T_{.jk} = \sum_i y_{ijk}$	$\bar{y}_{.jk} = \frac{T_{.jk}}{n}$

donde, a partir de las puntuaciones de las casillas, cada una representa lo siguiente:

μ: representa la media total.
α_j: representa el efecto del j-ésimo nivel de A.
β_k: representa el efecto del k-ésimo nivel de B.
$(\alpha\beta)_{jk}$: representa el efecto de la interacción entre el nivel j del factor A y el nivel k del factor B.
P_i: es el efecto debido a las diferencias entre los sujetos.
$(P\alpha)_{ij}$: representa el efecto de la interacción entre el sujeto i y el nivel j del factor A.
$(P\beta)_{ik}$: representa el efecto de la interacción entre el sujeto i y el nivel k del factor B.
$(P\alpha\beta)_{ijk}$: representa el efecto de la interacción entre el sujeto i y el efecto de la combinación de los jk tratamientos.
ε_{ijk}: representa el error aleatorio o las desviaciones de las puntuaciones del sujeto i

1.1.2 Estimadores

El modelo con medidas repetidas en dos factores es una adaptación del modelo lineal general. El modelo aditivo no es el más adecuado ya que no considera que exista la interacción entre los i sujetos y los jk niveles de los tratamientos. Además, se asume que las puntuaciones y_{ijk} son normales, independientes y existe homogeneidad en las varianzas de las variables y sus diferencias. A partir de los modelos expuestos, los estimadores correspondientes y utilizando las medias del cuadro 1.2, a los términos de los modelos, se tienen los siguientes estimadores correspondientes a cada uno de los términos de cada modelo:

$$\hat{\mu} = \bar{y} \tag{1.1}$$

$$\hat{\alpha}_j = \bar{y}_{.j.} - \bar{y} \tag{1.2}$$

$$\hat{\beta}_k = \bar{y}_{..k} - \bar{y} \tag{1.3}$$

$$(\hat{\alpha\beta})_{jk} = \bar{y}_{.jk} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y} \tag{1.4}$$

$$\hat{P}_i = \bar{y}_{i..} - \bar{y} \tag{1.5}$$

$$\hat{P\alpha}_{ij} = \bar{y}_{ij.} - \bar{y}_{.j.} - \bar{y}_{i..} + \bar{y} \tag{1.6}$$

$$\hat{P\beta}_{ik} = \bar{y}_{i.k} - \bar{y}_{..k} - \bar{y}_{i..} + \bar{y} \tag{1.7}$$

$$\hat{P\alpha\beta}_{ijk} = y_{ijk} - \bar{y}_{.jk} - \bar{y}_{ij.} - \bar{y}_{i.k} + \bar{y}_{.j.} + \bar{y}_{..k} + \bar{y}_{i..} - \bar{y} \tag{1.8}$$

$$\hat{\varepsilon}_{ijk} = y_{ijk} - \bar{y}_{.j.} - \bar{y}_{..k} - \bar{y}_{i..} + \bar{y} \tag{1.9}$$

cuyas demostraciones se presentan en el anexo A.
En el primer modelo, ya que no considera la interacción entre los sujetos y los tratamientos, puede asumirse que $\sigma^2_{p\alpha} = \sigma^2_{p\beta} = \sigma^2_{p\alpha\beta}$.Cada varianza de la interacción parcial constituye un estimador de la varianza de la interacción total y el promedio ponderado de sus medias cuadráticas, contribuye al error cuadrático medio para la prueba F del ANOVA-MR (análisis de varianza con medidas repetidas).

1.1.3 ANOVA-MR

Para analizar el efecto que ejerce cada factor respecto a la variable dependiente, se deben definir las hipótesis de interés como sigue:

para el factor A:

$H_{0A} : \mu_{+1+} = \mu_{+2+} = \ldots = \mu_{+j+} = \ldots = \mu_{+a+}$ o bien $\alpha_j = 0$ para todo valor de j o j' $(j \neq j')$.

para el factor B:

$H_{0B} : \mu_{++1} = \mu_{++2} = \ldots = \mu_{++k} = \ldots = \mu_{++b}$ o bien $\beta_k = 0$ para todo valor de k o k' $(k \neq k')$.

mientras que para la interacción sería:

$H_{0AB}: \mu_{+jk} - \mu_{+j\prime k} = \mu_{+j+} - \mu_{+j\prime +}$ o bien $(\alpha\beta)_{jk} = 0$ para todo valor de j o k.

Una vez definidas las hipótesis, y a partir de los estimadores, se puede calcular la suma de cuadrados, los que serán utilizados para definir el ANOVA-MR y el test F correspondientes a cada una de las fuentes de variación, para el caso de ambos modelos explicados con anterioridad.

$$SCT = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y})^2 \tag{1.10}$$

$$SCI = n \sum_j \sum_k (\bar{y}_{.jk} - \bar{y})^2 \tag{1.11}$$

$$SCA = nb \sum_j (\bar{y}_{.j.} - \bar{y})^2 \tag{1.12}$$

$$SCB = na \sum_k (\bar{y}_{..k} - \bar{y})^2 \tag{1.13}$$

$$SCAB = n \sum_j \sum_k (\bar{y}_{.jk} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y})^2 \tag{1.14}$$

$$SCP = ab \sum_i (\bar{y}_{i..} - \bar{y})^2 \tag{1.15}$$

$$SCE = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{.j.} - \bar{y}_{..k} - \bar{y}_{i..} + \bar{y})^2 \tag{1.16}$$

$$SCPA = b \sum_i \sum_j (\bar{y}_{ij.} - \bar{y}_{.j.} - \bar{y}_{i..} + \bar{y})^2 \tag{1.17}$$

$$SCPB = a \sum_i \sum_k (\bar{y}_{i.k} - \bar{y}_{..k} - \bar{y}_{i..} + \bar{y})^2 \tag{1.18}$$

$$SCPAB = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.} - \bar{y}_{i.k} - \bar{y}_{.jk} + \bar{y}_{.j.} + \bar{y}_{..k} + \bar{y}_{i..} + \bar{y})^2 \tag{1.19}$$

de las sumas de cuadrados se obtiene la siguiente tabla de resumen del ANOVA-MR:

Cuadro 1.3: ANOVA-MR

F.V.	S.C.	g.l	M.C.	F en el modelo aditivo	F en el modelo no aditivo
I	SCI	$(ab-1)$	$\frac{SCI}{(ab-1)}$		
A	SCA	$(a-1)$	$\frac{SCA}{(a-1)}$	$F_A = \frac{MCA}{MCE}$	$F'_A = \frac{MCA}{MCPA}$
B	SCB	$(b-1)$	$\frac{SCB}{(b-1)}$	$F_B = \frac{MCB}{MCE}$	$F'_B = \frac{MCB}{MCPB}$
AB	SCAB	$(a-1)(b-1)$	$\frac{SCAB}{(a-1)(b-1)}$	$F_{AB} = \frac{MCAB}{MCE}$	$F'_{AB} = \frac{MCAB}{MCPAB}$
P	SCP	$(n-1)$			
E	SCE	$(ab-1)(n-1)$	$\frac{SCE}{(ab-1)(n-1)}$		
PA	SCPA	$(a-1)(n-1)$	$\frac{SCPA}{(a-1)(n-1)}$		
PB	SCPB	$(b-1)(n-1)$	$\frac{SCPB}{(b-1)(n-1)}$		
PAB	SCPAB	$(a-1)(b-1)(n-1)$	$\frac{SCPAB}{(a-1)(b-1)(n-1)}$		
T	SCT	$(nab-1)$			

donde cada estadístico F, con sus respectivos grados de libertad, se distribuye, bajo H_0, de la siguiente forma:

$$F_A \cong F_{a-1,(ab-1)(n-1)} \tag{1.20}$$

$$F_B \cong F_{b-1,(ab-1)(n-1)} \tag{1.21}$$

$$F_{AB} \cong F_{(a-1)(b-1),(ab-1)(n-1)} \tag{1.22}$$

$$F'_{PA} \cong F_{a-1,(a-1)(n-1)} \tag{1.23}$$

$$F'_{PB} \cong F_{b-1,(a-1)(n-1)} \tag{1.24}$$

$$F'_{PAB} \cong F_{(a-1)(b-1),(a-1)(b-1)(n-1)} \tag{1.25}$$

La regla de decisión es rechazar H_0 si su estadístico cae en las siguientes zonas críticas:

$$
\begin{aligned}
F_A & \geq F_{a-1,(ab-1)(n-1)} \\
F_B & \geq F_{b-1,(ab-1)(n-1)} \\
F_{AB} & \geq F_{(a-1)(b-1),(ab-1)(n-1)} \\
F_A^{'} & \geq F_{a-1,(a-1)(n-1)} \\
F_B^{'} & \geq F_{b-1,(b-1)(n-1)} \\
F_{AB}^{'} & \geq F_{(a-1)(b-1),(a-1)(b-1)(n-1)}
\end{aligned}
$$

A partir de lo presentado, se puede calcular o estimar los efectos de los tratamientos considerando las medidas repetidas, ya que como se ha mencionado, este modelo fue diseñado para aplicar dos factores, con diferentes niveles de tratamiento, a un grupo de prueba.

1.2 Análisis de varianza multivariante con medidas repetidas de dos factores

En la sección siguiente, se presenta un análisis de medidas repetidas de dos factores desde un punto de vista del MANOVA, considerando que las medidas (tratamientos) están correlacionadas, ya que fueron tomadas a un mismo sujeto.

1.2.1 Estructura de los datos

Para comprender mejor este análisis de varianza es necesario entender la estructura de los datos:

Cuadro 1.4: Estructura de los datos y notación

	A_1					…	A_j					…	A_a				
Sujetos	B_1	…	B_k	…	B_b		B_1	…	B_k	…	B_b		B_1	…	B_k	…	B_b
P_1	$(y_{111}$		y_{11k}		y_{11b}		y_{1j1}		y_{1jk}		y_{1jb}		y_{1a1}		y_{1ak}		$y_{1ab}) = Y_1^t$
…			…						…						…		
P_i	$(y_{i11}$		y_{i1k}		y_{i1b}		y_{ij1}		y_{ijk}		y_{ijb}		y_{ia1}		y_{iak}		$y_{iab}) = Y_i^t$
…			…						…						…		
P_n	$(y_{n11}$		y_{n1k}		y_{n1b}		y_{nj1}		y_{njk}		y_{njb}		y_{na1}		y_{nak}		$y_{nab}) = Y_n^t$

a partir de la cuadro 1.4, se puede deducir que a diferencia de los modelos univariantes, este se reduce al más simple de los modelos multivariados, y tal como se menciona en Rencher & Christensen (2012), se asemeja superficialmente a un diseño de bloques aleatorizados univariados, sin embargo, las medias están correlacionadas debido a que las medidas fueron tomadas del mismo sujeto, mientras que en el diseño de bloques estas serían tomadas en diferentes unidades experimentales. Resultando así el siguiente modelo:

$Y_i = \mu + \varepsilon_i$, i=1,...,n.

El vector Y_i tiene dimensión $(ab \times 1)$, donde $Y \sim N_{ab}(\mu; \Sigma)$, cuyos parámetros son Σ que es la matriz de covarianzas de dimensión $(ab \times ab)$ y μ es un vector de medias de dimension $(ab \times 1)$, donde los estimadores de estos parámetros son:

$$\hat{\Sigma} = S = \frac{1}{n-1} \sum_i (Y_i - \bar{Y})(Y_i - \bar{Y})^t, \tag{1.26}$$

y

$$\hat{\mu} = \bar{Y} = \frac{\sum_i Y_i}{n}. \tag{1.27}$$

La ecuación (1.26) representa la matriz de covarianza muestral la que se ajusta a una distribución Wishart cuando $(n-1)S \sim W_{ab}(n-1; \Sigma)$ y la ecuación (1.27) representa el vector de medias muestral que se distribuye $N_{ab}(\mu; \Sigma/n)$.

1.2.2 Manova-MR

En Rencher (2002) se puede encontrar que para probar la significancia de los tratamientos, es necesario comparar todas las medias, esto bajo la siguiente hipótesis:

$$H_0 : \mu_1 = \ldots = \mu_i = \ldots = \mu_n$$

la que a su vez, puede ser expresada como:
$H_0 : \mu_1 - \mu_2 = \mu_2 - \mu_3 = \ldots = \mu_{i-1} - \mu_i = \ldots = \mu_{n-1} - \mu_n = 0$
o matricialmente
$H_0 : C\mu = 0$
donde C es una matriz de contraste con rango completo, siendo sus dimensiones $(ab-1) \times ab$ y la hipótesis es similar o equivalente al test de análisis de perfiles "planos". Utilizando la matriz C descrita, la matriz de covarianza y vector de medias, descritas en las ecuaciones (1.26) y (1.27), respectivamente, se puede obtener el estadístico de prueba con su distribución, bajo H_0, es:

$$T^2 = N(C\bar{Y})^t (CSC^t)^{-1} (C\bar{Y}) \sim T^2_{\alpha, ab-1, n-1}. \tag{1.28}$$

La región critica o de rechazo de la hipótesis $H_0 : C\mu = 0$ es $T^2 \geq T^2_{\alpha, ab-1, n-1}$, note que la dimensión es $ab - 1$ debido a que $C\bar{Y}$ es $(ab - 1 \times 1)$. También, si la hipótesis es verdadera $(H_0 : C\mu = 0)$ $C\bar{Y}$ se ajusta a una distribución $N_{ab-1}(0, C\Sigma C/n)$. Cabe destacar, que dependiendo del tipo de comparación o evaluación que se quiera realizar, la matriz de contrastes puede variar, pero debe cumplir con los requerimientos de ser de rango completo y, linealmente independiente u ortogonal. Pero, para evaluar cada factor o tratamiento, en este caso, C pasa a ser un vector o matriz $\underset{\sim}{c}$, según sea el caso, se explica también en Rencher & Christensen (2012) como un caso particular que puede ser deducido a partir del modelo de medidas repetidas con dos factores intra-sujetos y un factor entre-sujetos, se obtiene lo siguiente:

$$T^2 = n(\underset{\sim}{c}_A\bar{Y})^t(\underset{\sim}{c}_A S\underset{\sim}{c}_A^t)^{-1}(\underset{\sim}{c}_A\bar{Y}) \sim T^2_{\alpha,a-1,n-1}; \tag{1.29}$$

donde α representa el nivel de confianza, ab son la totalidad de tratamientos por los que atravesaron los n sujetos, y $\underset{\sim}{c}_A$ es un vector o matriz de contraste correspondiente a los niveles del tratamiento A de dimensión $((a-1)\times ab)$, de forma similar para B, mientras que para la interacción se obtiene a partir de los productos de los respectivos elementos del vector $\underset{\sim}{c}_A$ y los elementos de $\underset{\sim}{c}_B$. Ya que lo que único que cambia entre el test general y el análisis de los tratamientos, son los vectores o matrices, también cambia su valor de tabla, pero se utiliza el mismo método de rechazo.

$$T^2 = n(\underset{\sim}{c}_B\bar{Y})^t(\underset{\sim}{c}_B S\underset{\sim}{c}_B^t)^{-1}(\underset{\sim}{c}_B\bar{Y}) \sim T_{\alpha,b-1,n-1}, \tag{1.30}$$

y

$$T^2 = n(\underset{\sim}{c}_{AB}\bar{Y})^t(\underset{\sim}{c}_{AB} S\underset{\sim}{c}_{AB}^t)^{-1}(\underset{\sim}{c}_{AB}\bar{Y}) \sim T_{\alpha,(a-1)(b-1),n-1}. \tag{1.31}$$

H_0 se rechaza si $T^2 \geq T^2_{\alpha,ab-1,n-1}$, y para cada tratamiento es similar. Note que para buscar el valor de tabla $p = ab-1$ en el caso general, $p = a-1$ para el tratamiento A, $p = b-1$ para el tratamiento B y $p = (a-1)(b-1)$ para la interacción. Las tablas con $\alpha = 0{,}05$ y $\alpha = 0{,}01$, se encuentran disponibles en Rencher & Christensen (2012).

Con la finalidad de una mejor comprensión de la teoría explicada en este capítulo, se presentan en el siguiente capítulo, dos ejemplos obtenidos desde la bibliografía, a los cuales se les aplicara el análisis de varianza univariada y multivariada con medidas repetidas en dos factores.

Capítulo 2

Resultados del análisis de varianza univariada y multivariada de un modelo con medidas repetidas en dos factores

En este capítulo se presentan ejemplos obtenidos desde Ximénes & San Martín (2000) y Rencher & Christensen (2012), para analizar las posibles diferencias o semejanzas entre el análisis de varianza univariada y multivariada con medidas repetidas en dos factores.

2.1 Ejemplos

A continuación se presentan ejemplos obtenidos desde Ximénes & San Martín (2000) y Rencher & Christensen (2012).

Ejemplo 1 : Un investigador está interesado en estudiar el tipo de aroma del que está compuesto un perfume. Ha considerado tres tipos de aroma: B_1:lavanda (el mas suave); B_2: espliego; y B_3: rosas (el más fuerte). En cuanto al factor A, en este caso solamente decide considerar dos tipos de concentración del perfume: A_1: la mínima (500 ppm) y A_2: la máxima (2000 ppm). Ambas variables se miden en una nueva muestra que consta de 10 sujetos, los cuales pasan por todos los niveles de los factores considerados.

Por tanto, el objetivo ahora es estudiar el efecto del grado de concentración del perfume y el tipo de aroma, y el de ambos factores en interacción sobre la respuesta afectiva hacia el perfume. La respuesta afectiva de los sujetos se mide en una escala de 1 a 7, como se muestra a continuación:

	Concentración de 500 ppm			concentración de 2000 ppm		
Sujetos	Lavanda	Espliego	Rosas	Lavanda	Espliego	Rosas
1	3	4	5	5	6	5
2	4	4	4	6	7	4
3	4	6	7	4	6	7
4	4	4	5	5	7	7
5	2	4	5	5	4	7
6	2	3	7	5	6	7
7	4	3	5	4	7	6
8	6	6	5	6	6	6
9	4	6	4	7	3	7
10	4	5	7	4	7	6

Ejemplo 2 : Fue obtenido desde Rencher & Christensen (2012), las observaciones constituyen una muestra de medidas repetidas en dos factores. El factor A es una comparación de dos tipos de ejercicios matemáticos, el factor B es una comparación de dos tipos de calculadoras. Las medidas son la velocidad del calculo.

	A_1		A_2	
Sujetos	B_1	B_2	B_1	B_2
S_1	30	21	21	14
S_2	22	13	22	5
S_3	29	13	18	17
S_4	12	7	16	14
S_5	23	24	23	8

2.2 Resultados Anova-MR

Utilizando la teoría y fórmulas expuestas en el capítulo anterior y en Ximénes & San Martín (2000), se generó un algoritmo en R que permite el cálculo del ANOVA-MR, Se puede ver, en los resultados obtenidos mediante el algoritmo creado en R, que los factores A y B ejercen un efecto significativo sobre la variable dependiente, mientras que la interacción no ejerce ningún tipo de efecto. El valor de tabla para el factor A es $F_{(1,45)} = 4{,}08$, mientras que para el factor B y la interacción es $F_{(2,45)} = 3{,}23$, ambos valores considerando un $\alpha = 0{,}05$.

Cuadro 2.1: Resultados Anova-MR en R.

Factor	S.C.	g.l.	M.C.	F	F'
Intergrupos (I)	40,60	5	8,12		
Factor A	21,60	1	21,60	12,67	24,50
Factor B	18,30	2	9,15	5,37	6,33
Interacción AB	0,70	2	0,35	0,21	0,26
Intersujetos (P)	11,07	9	1,23		
Error (E)	76,73	45	1,71		
A x Sujetos	7,93	9	0,88		
B x Sujetos	26,03	18	1,45		
AB x Sujetos	24,30	18	1,35		
Total (T)	109,40	59			

Ximénes & San Martín (2000) exponen los siguientes resultados para el caso de análisis de varianza univariante con medidas repetidas en dos factores aplicados al primer ejemplo, expuesto en la sección anterior:

Cuadro 2.2: Resultados Anova-MR.

FV	SC	g. l.	MC	$F_{(aditivo)}$	$F'_{(no-aditivo)}$
Intergrupos (I)	40,60	5	8,01		
Factor A	21,60	1	21,60	16,87	26,34
Factor B	18,30	2	9,15	7,15	6,327
Interacción AB	0,70	2	0,35	0,27	0,259
Intersujetos (P)	11,07	9	1,23		
Error (E)	57,73	45	1,28		
A x Sujetos	7,40	9	0,82		
B x Sujetos	26,03	18	1,45		
AB x Sujetos	24,30	18	1,35		
Total (T)	109,40	59			

Basándose en la tabla anterior, en los resultados obtenidos por Ximénes & San Martín (2000), los factores A y B ejercen un efecto significativo sobre la variable dependiente, mientras que la interacción entre los factores no ejerce ningún tipo de efecto, es decir, que el grado de concentración del perfume con el tipo de aroma afectan a la respuesta afectiva de los sujetos, al igual que los resultados obtenidos en el software R. Sin embargo, como se puede apreciar en el siguiente gráfico, la interacción entre ambos factores no modifica el efecto separado de cada uno de los factores.

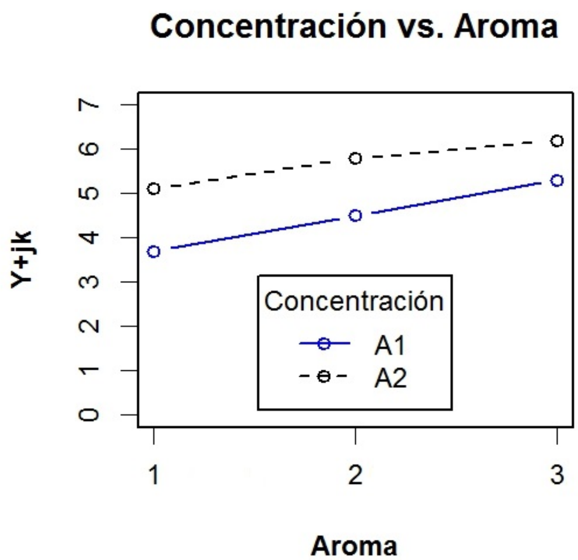

Figura 2.1: Representación gráfica de la relación entre la concentración y el tipo de aroma del perfume

Cabe destacar que en Ximénes & San Martín (2000) utilizan el software *SPSS* 7,5.

Para el ejemplo 2, se obtienen los siguientes resultados al utilizar el software R:

Cuadro 2.3: Resultados Anova-MR, ejemplo 2

Factor	S.C.	g.l.	M.C.	F	F'
Intergrupos (I)	385,6	3	128,53		
Factor A	64,8	1	64,80	1,24	2,53
Factor B	320,0	1	320	6,12	27,53
Interacción AB	0,8	1	0,80	0,02	0,02
Intersujetos (P)	218,3	4	54,58		
Error (E)	627,7	12	52,31		
A x Sujetos	102,5	4	25,62		
B x Sujetos	46,5	4	11,62		
AB x Sujetos	138,7	4	34,68		
Total (T)	910,8	19			

Para este caso, el valor de tabla con un $\alpha = 0{,}05$ para el factor A es $F_{(1,12)} = 4{,}75$, al igual que para el factor B y la interacción. Al considerar la zona de rechazo, explicada en el capitulo anterior, se puede apreciar que en este caso, el factor B ejerce un efecto significativo sobre la variable dependiente. En donde se puede apreciar, que el tipo de calculadora ejerce

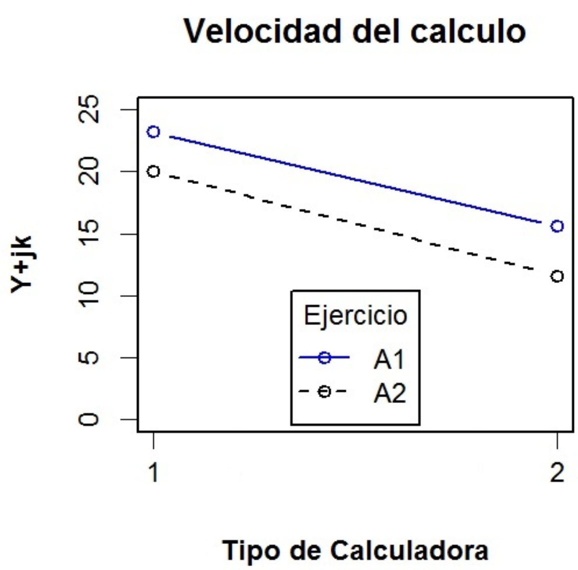

Figura 2.2: Representación gráfica de la relación entre el ejercicio (matemático) y el tipo de calculadora

un efecto sobre la variable dependiente.

2.3 Resultados Manova-MR

A partir del primer ejemplo, y utilizando la teoría expuesta en el capítulo anterior, se desarrolló un código que permitiera obtener los resultados. Con ello, se pretende explicar cómo funciona la metodología, así también, ver posibles semejanzas o diferencias entre ambos métodos.
Para testear las hipótesis explicadas en el capítulo anterior, con la finalidad de estudiar la significancia de los niveles de cada tratamiento, se utilizaron los siguientes vectores y matrices de contrastes:

$$\underset{\sim}{c}_A = \begin{pmatrix} 1 & 1 & 1 & -1 & -1 & -1 \end{pmatrix}$$

$$\underset{\sim}{c}_B = \begin{pmatrix} 2 & -1 & -1 & 2 & -1 & -1 \\ 0 & 1 & -1 & 0 & 1 & -1 \end{pmatrix}$$

$$\underset{\sim}{c}_{AB} = \begin{pmatrix} 2 & -1 & -1 & -2 & 1 & 1 \\ 0 & 1 & -1 & 0 & -1 & 1 \end{pmatrix}$$

$$C = \begin{pmatrix} 1 & 1 & 1 & -1 & -1 & -1 \\ 2 & -1 & -1 & 2 & -1 & -1 \\ 0 & 1 & -1 & 0 & 1 & -1 \\ 2 & -1 & -1 & -2 & 1 & 1 \\ 0 & 1 & -1 & 0 & -1 & 1 \end{pmatrix}$$

Donde $\underset{\sim}{c}_A$, $\underset{\sim}{c}_B$ y $\underset{\sim}{c}_{AB}$representan los contrastes entre los niveles de las matrices de A, B y la interacción AB respectivamente, tal como se puede apreciar en el anexo. Similarmente, en la matriz C, la primera fila representa los dos contrastes de A, la segunda y tercera fila representa los contrastes de los 3 niveles de B y la cuarta y quinta fila representa los contrastes de la interacción, la cual, tal como se explicó con anterioridad, es la multiplicación de los valores que se encuentran en las primeras filas, que representan los contrastes entre los niveles de A y los niveles de B.
Las dos estimaciones más importantes descritas en el capítulo anterior son el vector de medias y la matriz de covarianza, en las ecuaciones ((1.27) y (1.26)), dando como resultado lo siguiente:

$$\bar{Y} = \begin{pmatrix} 3.7 \\ 4.5 \\ 5.3 \\ 5.1 \\ 5.8 \\ 6.2 \end{pmatrix}$$

y

$$S = \begin{pmatrix} 1.344 & 0.833 & -0.344 & 0.255 & 0.377 & -0.266 \\ & 1.388 & 0.166 & 0.388 & -0.666 & 0.222 \\ & & 1.122 & -0.033 & -0.377 & 0.711 \\ & & & 0.988 & -0.644 & -0.133 \\ & & & & 1.733 & -0.622 \\ & & & & & 1.066 \end{pmatrix}$$

estas medidas fueron calculadas con el código que se puede encontrar en el anexo C, a partir de los cuales se obtuvieron los siguientes resultados:

Factor	T^2-Hotelling
General	69,47
A	26,27
B	12,07
AB	1,18

Como se dijo en el capítulo anterior, en el caso del test T^2 de Hotelling, la hipótesis se rechaza si el valor resultante es mayor o igual que el valor de tabla, en cualquier otro caso se acepta. Para este caso en particular, $T^2_{0 \cdot 05,5,9} = 45{,}453$, y tal como se puede apreciar en la tabla, el test T^2 que llamamos General, se considera significativa, es decir, a nivel general, cuando se comparan todos los niveles y tratamientos, ejercen un efecto significativo respecto de la variable dependiente. Al analizar por cada tratamiento, y con la finalidad de poder realizar comparaciones, se aplica el test en los factores para corroborar si ocurre lo mismo, para el factor A el valor de tabla es $T^2_{0 \cdot 05,1,9} = 5{,}117$, para el factor B y la interacción es $T^2_{0 \cdot 05,2,9} = 10{,}033$, donde se puede apreciar que los tratamientos A y B son los que ejercen un efecto significativo sobre la variable dependiente, no así, la interacción entre ellos. Tal como se puede apreciar, se obtuvieron los mismos resultados que en el caso univariante.

A continuación, se puede apreciar gráficamente los resultados obtenidos del primer ejemplo:

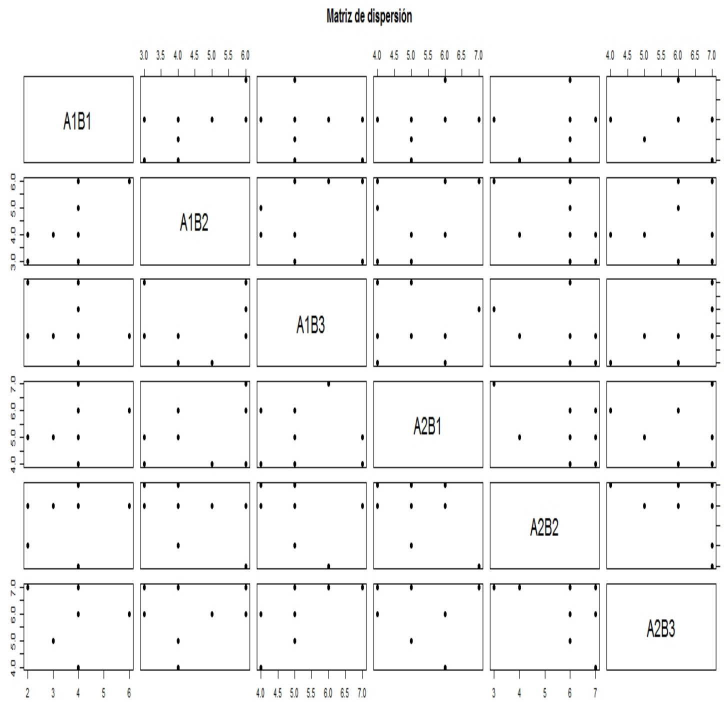

Figura 2.3: Representación gráfica que corresponde a los cruces de la relación entre la concentración y el tipo de aroma del perfume

Donde se puede apreciar que no hay correlación entre las variables, es decir, no hay relación entre los niveles concentración y los tipos de aromas del perfume en estudio.

Para el segundo ejemplo, se utilizó el mismo algoritmo generado en R, y utilizando las mismas matrices y vectores de contraste disponibles en Rencher & Christensen (2012), similarmente al ejemplo anterior, se busca evaluar la significancia de los niveles de cada tratamiento, es decir, las siguientes matrices y vectores de contraste:

$$\underset{\sim}{c}_A = \begin{pmatrix} 1 & 1 & -1 & -1 \end{pmatrix}$$

$$\underset{\sim}{c}_B = \begin{pmatrix} 1 & -1 & 1 & -1 \end{pmatrix}$$

$$\underset{\sim}{c}_{AB} = \begin{pmatrix} 1 & -1 & -1 & 1 \end{pmatrix}$$

$$C = \begin{pmatrix} 1 & 1 & \text{-1} & \text{-1} \\ 1 & \text{-1} & 1 & \text{-1} \\ 1 & \text{-1} & \text{-1} & 1 \end{pmatrix}$$

se obtuvieron los siguientes resultados:

	T^2 Hotelling
General	29,736
A	2,13
B	27,527
AB	0,023

A diferencia del caso anterior, los resultados obtenidos para este ejemplo, ninguno de los factores ejerce un efecto significativo sobre la variable dependiente, esto considerando que el valor general que arroja el algoritmo y considerando el valor de tabla $T^2_{0\cdot 05,4,3} = 114{,}986$. Pese a ello, si se analiza para cada tratamiento, se tiene que el valor de tabla para ellos es $T^2_{0\cdot 05,4,1} = 7{,}709$, en donde se puede apreciar que el tratamiento B, ejerce un efecto significativo sobre la variable dependiente, lo que coincide con los resultados obtenidos en la sección anterior.
El valor del vector de medias es:

$$\bar{Y} = \begin{pmatrix} 23.2 \\ 15.6 \\ 20 \\ 11.6 \end{pmatrix}$$

Y el valor de la matriz de covarianza:

$$S = \begin{pmatrix} 51.70 & 29.85 & 9.25 & 7.35 \\ & 46.80 & 16.25 & -8.70 \\ & & 8.50 & -10.50 \\ & & & 24.30 \end{pmatrix}$$

Su representación gráfica es:

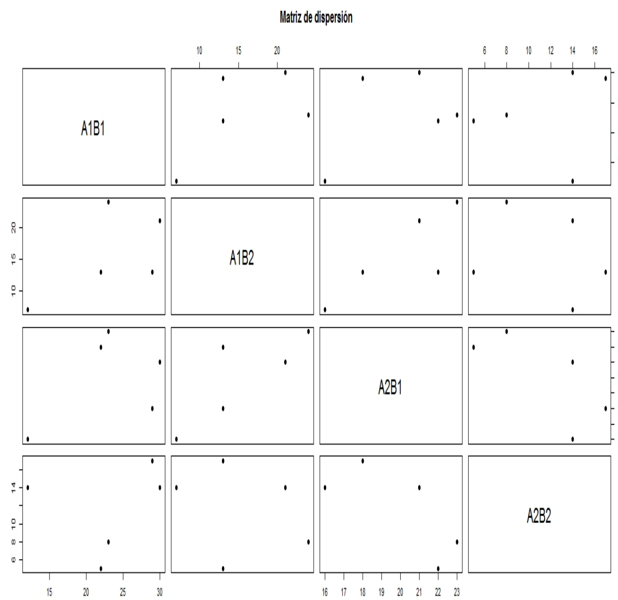

Figura 2.4: Representación gráfica que corresponde a los cruces entre el ejercicio (matemático) y el tipo de calculadora

De la gráfica correspondiente a los cruces entre los ejercicios matemáticos y los tipos de calculadoras, se puede apreciar una baja relación entre ambos.

Conclusión

El análisis de varianza con medidas repetidas, si bien, se puede resolver sin la necesidad de saber multivariado, no implica que pueda resultar más eficiente, ya que el modelo univariante contiene más elementos que el multivariado, por ejemplo; en el caso univariante es necesario calcular la correlación de manera apartada, mientras que en el caso multivariante, estos resultados se pueden apreciar en la matriz de covarianzas. Otra diferencia que se puede apreciar, es que mientras el modelo univariante toma uno a uno los datos, el modelo multivariante toma el conjunto de datos, es decir, en este último caso, los trabaja de forma matricial, obteniendo los mismos resultados en ambos casos, pero de forma más rápida. Se debe tener en cuenta que en ambos casos se mide el efecto que ejercen los dos factores y su interacción sobre la variable dependiente. Sin embargo, mientras en el caso univariado se mide el efecto de la interacción entre los n sujetos y los tratamientos, incluyendo su interacción, no así para el caso multivariante, ya que cada sujeto debe atravesar por todos los niveles de los factores, en ambos casos se debe tener en cuenta la independencia entre los sujetos.

Observando los resultados expuestos en el capitulo dos, ambos ejemplos provenientes desde la bibliografía, se tienen los mismos resultados en ambos análisis de varianza (modelo univariante y multivariante), la principal diferencia, comentada en el párrafo anterior, es el caso multivariante, ya que debido a que se trabaja de forma matricial, es más eficiente, ya que no solamente se obtienen los mismos resultados, el cálculo es rápido.

Si bien, para el caso del análisis multivariante de medidas repetidas en dos factores no existe un paquete o códigos en el programa R-project, los códigos desarrollados en este proyecto de titulación quedarán disponibles a los usuarios del programa. Los vectores ó matrices que reconoce el código son vectores fila, para el caso $\underset{\sim}{c}_A$, $\underset{\sim}{c}_B$ y $\underset{\sim}{c}_{AB}$, en el caso de la matriz de contrastes, no se estipuló nada, ya que puede quedar abierta a los tipos de contrastes que se quieran realizar por parte del investigador. Se recomienda, pese a lo anterior, considerar una matriz de contraste que considere los vectores de $\underset{\sim}{c}_A$, $\underset{\sim}{c}_B$ y $\underset{\sim}{c}_{AB}$, para realizar el análisis general de los datos, ya que da un indicio del comportamiento de estos, y tal como se explicó con anterioridad el vector $\underset{\sim}{c}_{AB}$ es la multiplicación de los valores de los vectores

$\underline{c}_A$ y $\underline{c}_B$, tal como se mostró en ambos ejemplos. Cabe destacar que los códigos se pueden modificar para trabajar el caso del análisis de varianza con una medida repetida.
El principal defecto del análisis de medidas repetidas en dos factores es principalmente que es sensible a datos faltantes, imputación y violación de la esfericidad, lo que puede resultar en sesgo de muestreo y tasas infladas de error tipo I.

Anexo A

A partir del modelo $y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + P_i + (P\alpha)_{ij} + (P\beta)_{ik} + (P\alpha\beta)_{ijk} + \varepsilon_{ijk}$ se sabe que la suma de sus fuentes de variación, son cero y son restricciones al modelo, es decir, $\sum_j \alpha_j = \sum_k \beta_k = \sum_j \sum_k (\alpha\beta)_{jk} = \sum_i P_i = \sum_i \sum_j (P\alpha)_{ij} = \sum_i \sum_k (P\beta)_{ik} = \sum_i \sum_j \sum_k (P\alpha\beta)_{ijk} = \sum_j (P\alpha)_{ij} = \sum_k (P\beta)_{ik} = \sum_j \sum_k (P\alpha\beta)_{ijk} = 0$, Utilizando el método de mínimos cuadrados, se pueden calcular sus estimaciones:

$$\sum_i \sum_j \sum_k \varepsilon^2 = \sum_i \sum_j \sum_k (y_{ijk} - \mu - \alpha_j - \beta_k - (\alpha\beta)_{jk} - P_i - (P\alpha)_{ij} - (P\beta)_{ik} - (P\alpha\beta)_{ijk})^2$$

$$\frac{\partial \sum_i \sum_j \sum_k \varepsilon^2}{\partial \mu} = -2 \sum_i \sum_j \sum_k (y_{ijk} - \mu - \alpha_j - \beta_k - (\alpha\beta)_{jk} - P_i - (P\alpha)_{ij} - (P\beta)_{ik} - (P\alpha\beta)_{ijk}) = 0$$

$$\sum_i \sum_j \sum_k y_{ijk} - nab\mu - nb \sum_j \alpha_j - na \sum_k \beta_k - n \sum_j \sum_k (\alpha\beta)_{jk} - ab \sum_i P_i - b \sum_i \sum_j (P\alpha)_{ij} - a \sum_i \sum_k (P\beta)_{ik} - \sum_i \sum_j \sum_k (P\alpha\beta)_{ijk} = 0$$

por la demostración inicial se obtiene:

$$\sum_i \sum_j \sum_k y_{ijk} - nab\mu = 0$$

$$nab\hat{\mu} = \sum_i \sum_j \sum_k y_{ijk}$$

$$\hat{\mu} = \frac{\sum_i \sum_j \sum_k y_{ijk}}{nab} = \bar{y}$$

de esta forma se obtiene el estimador para la media, tal como era de esperar. Para el caso de $\hat{\alpha}_j$ sería como sigue:

$$\sum_i \sum_j \sum_k \varepsilon^2 = \sum_i \sum_j \sum_k (y_{ijk} - \mu - \alpha_j - \beta_k - (\alpha\beta)_{jk} - P_i - (P\alpha)_{ij} - (P\beta)_{ik} - (P\alpha\beta)_{ijk})^2$$

$$\frac{\partial \sum_i \sum_j \sum_k \varepsilon^2}{\partial \alpha} = -2\sum_i \sum_k (y_{ijk}-\mu-\alpha_j-\beta_k-(\alpha\beta)_{jk}-P_i-(P\alpha)_{ij}-(P\beta)_{ik}-(P\alpha\beta)_{ijk}) = 0$$

$$\sum_i \sum_k y_{ijk} - nb\mu - nb\alpha_j - n\sum_k \beta_k - n\sum_k (\alpha\beta)_{jk} - b\sum_i P_i - b\sum_i (P\alpha)_{ij} - \sum_i \sum_k (P\beta)_{ik} - \sum_i \sum_k (P\alpha\beta)_{ijk} = 0$$

que de manera similar al caso anterior, al utilizar las restricciones expuestas al comienzo, se obtiene:

$$\sum_i \sum_k y_{ijk} - nb\hat{\mu} - nb\alpha_j = 0$$

$$nb\hat{\alpha}_j = \sum_i \sum_k y_{ijk} - nb\hat{\mu}$$

$$\hat{\alpha}_j = \frac{\sum_i \sum_k y_{ijk}}{nb} - \hat{\mu}$$

al reemplazar la estimación de la media se tiene:

$$\hat{\alpha}_j = \bar{y}_{.j.} - \bar{y}$$

con $j = 1, \ldots, a$ de forma similar se pueden obtener las estimaciones del resto de las puntuaciones.

Anexo B

En Timm (2002) se define Q e Y, variables aleatorias independientes donde $Y \sim N_p(\mu, \Sigma)$ y $Q \sim W_p(n, \Sigma)$, y $n > p$. Entonces el estadístico T^2 de Hotelling es:

$$T^2 = nY^tQ^{-1}Y$$

tiene una distribución proporcional a una distribución no central F

$$\frac{n-p+1}{p}\frac{T^2}{n} \sim F(p, n-p+1, \gamma)$$

donde $\gamma = \mu^t\Sigma^{-1}\mu$.

Ejemplo: Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria desde una población MVN, $Y_i \sim N_p(\mu, \Sigma)$. Entonces $\bar{Y} \sim N_p(\mu, \Sigma/n)$ y $(n-1)S \sim W_p(n-1, \Sigma)$, donde $\bar{Y}$ y S son idempotentes. Por lo tanto, para testear $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$, $T^2 = n(\bar{Y}-\mu_0)^tS^{-1}(\bar{Y}-\mu_0)$ ó

$$\frac{n-p}{p}\frac{T^2}{n-1} = \frac{n(n-p)}{p(n-1)}(\bar{Y}-\mu_0)^tS^{-1}(\bar{Y}-\mu_0) \sim F_{(p,n-p,\gamma)}$$

donde $\gamma = n(\bar{Y}-\mu_0)^tS^{-1}(\bar{Y}-\mu_0)$

Anexo C

A continuación se presentan los códigos utilizados en el capitulo 2.

```
##caso multivariante

MR<-function(Y,C,CA,CB,CAB){

Y<-data.matrix(Y)

medY<-colSums(Y)/(length(Y[,1]))

S<-var(Y)
T<-length(Y[,1])%*%t(C%*%t(t(medY)))%*%solve(C%*%S%*%t(C))%*%(C%*%t(t(medY)))
T1<-length(Y[,1])%*%t(CA%*%t(t(medY)))%*%solve(CA%*%S%*%t(CA))%*%(CA%*%t(t(medY)))
T2<-length(Y[,1])%*%t(CB%*%t(t(medY)))%*%solve(CB%*%S%*%t(CB))%*%(CB%*%t(t(medY)))
T3<-length(Y[,1])%*%t(CAB%*%t(t(medY)))%*%solve(CAB%*%S%*%t(CAB))%*%(CAB%*%t(t(medY)))


resultados<-matrix(0,ncol=1,nrow=4)
MedY<-matrix(medY,ncol=length(Y[1,]),nrow=1)
S<-round(S,digits=3)
resultados[1,1]<-round(T1,digits=3)
resultados[2,1]<-round(T2,digits=3)
resultados[3,1]<-round(T3,digits=3)
resultados[4,1]<-round(T,digits=3)


dimnames(resultados) = list(c("A","B","AB","General"),c("T2 Hotelling"))
dimnames(MedY) = list(c("Media"),c(1:length(MedY[1,])))
dimnames(S) = list(c("Varianza",2:length(S[,1])),c(1:length(S[,1])))
```

```
print(resultados)
print(MedY)
print(S)
}
Y<-read.table(file.choose(),header=F,sep=';')

CA<-matrix(c(1,1,1,-1,-1,-1),nrow=1,ncol=6,byrow=T)
CB<-matrix(c(2,-1,-1,2,-1,-1,0,1,-1,0,1,-1),nrow=2,ncol=6,byrow=T)
CAB<-matrix(c(2,-1,-1,-2,1,1,0,1,-1,0,-1,1),nrow=2,ncol=6,byrow=T)
C<-matrix(c(1,1,1,-1,-1,-1,2,-1,-1,2,-1,-1,0,1,-1,0,1,-1,2,-1,-1,-2,
1,1,0,1,-1,0,-1,1),nrow=5,ncol=6,byrow=T)

MR(Y,C,CA,CB,CAB)

Y<-read.table(file.choose(),header=F,sep=';')
CA<-matrix(c(1,1,-1,-1),nrow=1,ncol=4,byrow=T)
CB<-matrix(c(1,-1,1,-1),nrow=1,ncol=4,byrow=T)
CAB<-matrix(c(1,-1,-1,1),nrow=1,ncol=4,byrow=T)
C<-matrix(c(1,1,-1,-1,1,-1,1,-1,1,-1,-1,1),nrow=3,ncol=4,byrow=T)
MR(Y,C,CA,CB,CAB)



##caso univariado

anovamr<-function(Y,a,b,n){
Y<-data.matrix(Y)
T<-sum(Y)
su<-sum(Y^2)

SCT<-su-((T^2)/(a*b*n))

SCI<-((sum(colSums(Y)^2))/(n))-((T^2)/(a*b*n))

su1<-sum(Y[,1:b])^2
for(i in 1:a){
l<-(b*i)+1
k<-l+(b-1)
su1<-su1+(sum(Y[,l:k])^2)
break(l>(b*a));
}
```

```
SCA<-((su1)/(n*b))-((T^2)/(a*b*n))


su2<-Y[,1:b]
for(i in 1:a){
l<-(b*i)+1
k<-l+(b-1)
su2<-su2+Y[,l:k]
break(l>(b*a));
}
m<-0
for(i in 1:b){
m<-m+(sum(su2[,i])^2)
}
SCB<-((m)/(n*a))-((T^2)/(a*b*n))

SCAB<-((sum(colSums(Y)^2))/(n))-((su1)/(n*b))-((m)/(n*a))+((T^2)/(a*b*n))

SCP<-(sum(rowSums(Y)^2)/(a*b))-((T^2)/(a*b*n))

SCE<-sum(Y^2)-((su1)/(n*b))-(sum(rowSums(Y)^2)/(a*b))+((T^2)/(a*b*n))

SCPA<-((sum(colSums(Y)^2))/(n))-((su1)/(n*b))-(sum(rowSums(Y)^2)/(a*b))+
((T^2)/(a*b*n))

SCPB<-(sum(su2^2)/a)-((m)/(n*a))-(sum(rowSums(Y)^2)/(a*b))+((T^2)/(a*b*n))


su3<-rowSums(Y[,1:b])^2
for(i in 1:a){
l<-(b*i)+1
k<-l+(b-1)
su3<-su3+(rowSums(Y[,l:k])^2)
break(l>(b*a));
}
sum(su3)/b

SCPAB<-sum(Y^2)-(sum(su3)/b)-(sum(su2^2)/a)-((sum(colSums(Y)^2))/(n))+
(sum(rowSums(Y)^2)/(a*b))+((su1)/(n*b))+((m)/(n*a))-((T^2)/(a*b*n))
```

```
MCI<-SCI/((a*b)-1)
MCA<-SCA/(a-1)
MCB<-SCB/(b-1)
MCAB<-SCAB/((a-1)*(b-1))
MCE<-SCE/(((a*b)-1)*(n-1))
MCP<-SCP/(n-1)
MCPA<-SCPA/((a-1)*(n-1))
MCPB<-SCPB/((b-1)*(n-1))
MCPAB<-SCPAB/((a-1)*(b-1)*(n-1))

FA<-MCA/MCE
FB<-MCB/MCE
FAB<-MCAB/MCE
FPA<-MCA/MCPA
FPB<-MCB/MCPB
FPAB<-MCAB/MCPAB

resultados<-matrix(0,ncol=5,nrow=10)
resultados[1,1]<-round(SCI,digits=2)
resultados[2,1]<-round(SCA,digits=2)
resultados[3,1]<-round(SCB,digits=2)
resultados[4,1]<-round(SCAB,digits=2)
resultados[5,1]<-round(SCP,digits=2)
resultados[6,1]<-round(SCE,digits=2)
resultados[7,1]<-round(SCPA,digits=2)
resultados[8,1]<-round(SCPB,digits=2)
resultados[9,1]<-round(SCPAB,digits=2)
resultados[10,1]<-round(SCT,digits=2)

resultados[1,2]<-((a*b)-1)
resultados[2,2]<-(a-1)
resultados[3,2]<-(b-1)
resultados[4,2]<-(a-1)*(b-1)
resultados[5,2]<-(n-1)
resultados[6,2]<-((a*b)-1)*(n-1)
resultados[7,2]<-(a-1)*(n-1)
resultados[8,2]<-(b-1)*(n-1)
resultados[9,2]<-(a-1)*(b-1)*(n-1)
resultados[10,2]<-((a*b*n)-1)

resultados[1,3]<-round(MCI,digits=2)
```

```
resultados[2,3]<-round(MCA,digits=2)
resultados[3,3]<-round(MCB,digits=2)
resultados[4,3]<-round(MCAB,digits=2)
resultados[5,3]<-round(MCP,digits=2)
resultados[6,3]<-round(MCE,digits=2)
resultados[7,3]<-round(MCPA,digits=2)
resultados[8,3]<-round(MCPB,digits=2)
resultados[9,3]<-round(MCPAB,digits=2)


resultados[2,4]<-round(FA,digits=2)
resultados[3,4]<-round(FB,digits=2)
resultados[4,4]<-round(FAB,digits=2)

resultados[2,5]<-round(FPA,digits=2)
resultados[3,5]<-round(FPB,digits=2)
resultados[4,5]<-round(FPAB,digits=2)

dimnames(resultados) = list(c("Intergrupos (I)","Factor A","Factor B",
"Interacción AB","Intersujetos (P)","Error (E)","A x Sujetos","B x Sujetos",
"AB x Sujetos","Total (T)"),c("SC","g.l.","MC","F","F’"))
print(resultados)
}

Y<-read.table(file.choose(),header=F,sep=’;’)
a<-2
b<-3
n<-10

anovamr(Y,a,b,n)


Y<-read.table(file.choose(),header=F,sep=’;’)
a<-2
b<-2
n<-5

anovamr(Y,a,b,n)

##ejemplo 1
B<- c(1,2,3)
```

```
A1 <- c(3.7,4.5,5.3)
A2 <- c(5.1,5.8,6.2)
par(lwd=2, cex=1.5, font.lab=2)
plot(B, A1, type="b", lty=1, col="blue", ylim=c(0,7),xlim=c(1,3),
main="Concentración vs. Aroma", xlab="Aroma", ylab="Y+jk")
lines(B,A2, type="b", lty=2, col="black")
legend("bottom", inset=.05, title="Concentración", c("A1","A2"),
lty=c(1,2),pch=c(1,1), col=c("blue", "black"))

##ejemplo 2
B<- c(1,2)
A1 <- c(23.2,15.6)
A2 <- c(20,11.6)
par(lwd=2, cex=1.5, font.lab=2)
plot(B, A1, type="b", lty=1, col="blue", ylim=c(0,25),xlim=c(1,2),
main="Velocidad del calculo", xlab="Tipo de Calculadora", ylab="Y+jk")
lines(B,A2, type="b", lty=2, col="black")
legend("bottom", inset=.02, title="Ejercicio", c("A1","A2"),
lty=c(1,2),pch=c(1,1), col=c("blue", "black"))

##ejem 1 multi
Y<-read.table(file.choose(),header=F,sep=';')
plot(Y,labels=c("A1B1","A1B2","A1B3","A2B1","A2B2","A2B3"),
main='Matriz de dispersión',cex.main=1,pch = 19, bg="light blue")

##ejem 2 multi
Y<-read.table(file.choose(),header=F,sep=';')
plot(Y,labels=c("A1B1","A1B2","A2B1","A2B2"),
main='Matriz de dispersión',cex.main=1,pch = 19, bg="light blue")
```

Bibliografía

Ximénes, C., San Martín, R. (2000) *Análisis de Varianza con Medidas Repetidas*. Madrid, España: La Muralla S.A.

Rencher, Alvin (2002) *Methods of Multivariate Analysis-2nd ed.* New York: John Wiley & Sons, Inc.

Rencher, A., Christensen, W. (2012) *Methods of Multivariate Analysis-3rd ed.* New York: John Wiley & Sons, Inc.

R Core Team (2014) *R Foundation for Statistical Computing*. Vienna, Austria. URL http://www.R-project.org/. Acceso Diciembre, 23, 2016.

Seber, G. A., & Lee, A. J.(2003) *Linear Regression Analysis-2rd ed.* John Wiley & Sons, Inc.

Timm, Neil (2002) *Applied Multivariate Analysis*. New York: Springer-Verlag.

Montgomery, D., Peck, E., Vining, G. (2006) *Introducción al Análisis de Regresión Lineal 3er. ed. en ingles (1era. ed. en español)*. Mexico: Compañía Editorial Continental.