ANÁLISIS DE COVARIANZA

Dr. Cs.[1] Ruiz Aranibar, Gustavo

✉ *gustavoruiz432@hotmail.com*

RESUMEN

El análisis de covarianza es una estrategia estadística útil en la evaluación de la relación entre una variable independiente y una o más variables dependientes, dándose a conocer su determinación, facilitando su aplicación numérica por medio de los programas computacionales desarrollados, resultados que permiten comparar el análisis de varianza con el análisis de covarianza.

PALABRAS CLAVE

Concomitantes, covariable, error experimental, heterogeneidad, nivel de significación.

ABSTRACT

Analysis of covariance is a statistical strategy useful in the evaluation of the relationship between an independent variable and a dependent variable, giving to know his determination, facilitating their numerical implementation by means of computer programs developed, results that allow you to compare the analysis of variance with the analysis of covariance.

KEYWORDS

Concomitants, covariates, experimental error, heterogeneity, level of significance.

1. INTRODUCCIÓN.

Se define el análisis de covarianza (ANACO), como una prueba estadística que analiza la relación entre una variable dependiente y las independientes, eliminando y controlando el efecto de por lo menos una de estas variables independientes, procedimiento estadístico que permite eliminar la heterogeneidad causada en la variable de interés (variable dependiente) por la influencia de una o más variables cuantitativas (covariables). La inclusión de covariables puede aumentar la potencia estadística porque con frecuencia reduce la variabilidad y los sesgos provocados por las diferencias en unidades experimentales (covariables) entre los grupos. Uno de los problemas con el que frecuentemente se enfrenta el investigador, es el de controlar aquellos factores que no son posibles de medir, cuyo efecto no puede justificarse, los cuales constituyen el error experimental.

Existen dos métodos generales para controlar la variabilidad debido al error experimental: directo y estadístico. El control directo incluye tales métodos como unidades de grupos experimentales dentro estratos o bloques homogéneos, incrementando la uniformidad de la condición bajo la cual el experimento es ejecutado e incrementando la exactitud de las mediciones. El método estadístico, controla el incremento de la precisión de un experimento y saca de la fuente potencial variables extrañas, en cambio el método denominado ANACO, es aplicable cuando tales variables extrañas o concomitantes no pueden mantenerse constantes, pero que, sin embargo, pueden medirse.

[1] Se agradece a la Universidad Autónoma Gabriel René Moreno (UAGRM) por la beca otorgada con fondos del IDH, para cursar y culminar exitosamente el Doctorado en Ciencias en Educación Superior. Especializado en Estadística e Informática.

2. ECUACIÓN MODELO PARA EL ANACO

El modelo subyacente está dado por:

$$Y_{ij}= \mu+\alpha_i+\delta x_{ij}+\varepsilon_{ij} \;\; ; \;\; i=1,2,.....k \;\; ; \;\; j=1,2,.....,n$$

μ = gran media.

α_i =efecto del iésimo tratamiento.

δ =pendiente de la ecuación de regresión lineal

x_{ij} = valores de los coeficientes de regresión

ε_{ij} = variables independientes, normalmente distribuidos con medias cero y varianza común σ^2.

En el análisis de tales datos, los valores x_{ij} de la variable concomitante son eliminados por métodos de regresión, esto es, estimando δ por el método de los mínimos cuadrados y luego mediante un análisis de varianza sobre las y ajustadas, es decir, sobre las cantidades $y_{ij} = y_{ij} - \hat{\delta}x_{ij}$. A este procedimiento se le llama análisis de covarianza, que implica una subdivisión de la suma de productos total:

$$SPT = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y})(x_{ij} - \bar{x})$$

De la misma manera que un análisis ordinario de análisis de varianza implica la subdivisión de la suma de los cuadrados. En la práctica los cálculos se efectúan como sigue:

1. Las sumas totales de los cuadrados, de tratamientos y de errores se calculan para las *x* por medio de las fórmulas para una clasificación sencilla, denotadas con SST_x, $SS(Tr)_x$, y SSE_x.

$$SST_x = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - C$$

$$C = \frac{T^2}{kn} \qquad SS(Tr)_x = \frac{\sum_{i=1}^{k} T_i^2}{n} - C$$

2. Las sumas totales de los cuadrados, de tratamiento *y* errores se calculan para las y por medio de las fórmulas para una clasificación SST_y, $SS(Tr)_y$, y SSE_y.

3. Las sumas totales de los productos, de tratamientos y de errores se calculan por medio de:

$$SPT = \sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}\, y_{ij} - C \qquad SP(Tr) = \frac{\sum_{i=1}^{k} T_{x_i} T_{y_i}}{n} - C$$

$$C = \frac{T_x T_y}{kn}$$

$$SPE = SPT - SP(Tr)$$

T_{xi} = Total de las *x* para el iésimo tratamiento
T_{yi} = Total de las *y* para el iésimo tratamiento
T_x = Total de todas las *x*
T_y = Total de todas las *y*

4. Las sumas totales de los cuadrados, de tratamientos *y* errores se calculan para las y ajustadas por medio de:

$$SST_{y'} = SST_y - \frac{SPT^2}{SST_x}$$

$$SSE_{y'} = SSE_y - \frac{SPE^2}{SSE_x}$$

$$SS(Tr)_{y'} = SST_{y'} - SSE_{y'}$$

$$MS(Tr)_{y'} = \frac{SS(T_r)_{y'}}{k-1}$$

$$MSE_{y'} = \frac{SSE_{y'}}{nk-k-1}$$

$$F_{cal} = \frac{MS(Tr)_{y'}}{MSE_{y'}}$$

Los resultados obtenidos con estos cálculos se resumen convenientemente por medio de:

Tabla 1
Tabla del ANACO

Fuente de variación	**Suma de cuadrados para x**	**Suma de cuadrados para y**	**Suma de productos**	**Suma de cuadrados para y´**	**Grados de libertad**	**Cuadrado medio**	**Razón F**
Tratamiento	$SS(Tr)_x$	$SS(Tr)_y$	$SP(Tr)$	$SS(Tr)_{y'}$	$k-1$	$MS(Tr)_{y'}$	F_{cal}
Error	SSE_x	SSE_y	SPE	$SSE_{y'}$	$nk-k-1$	$MSE_{y'}$	
Total	SCT_x	SCT_y	SPT	$SCT_{y'}$	$nk-2$		

Fuente: Elaboración propia.

Hipótesis:

$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_i = 0$

H_1 : *no son todas iguales a cero.*

Se rechaza al nivel de significación α, si el valor obtenido para F_{cal} excede a F_α con k-1 y nk-k-1 grados de libertad.

3. CLASIFICACIÓN DEL ANÁLISIS DE COVARIANZA

Cada método del ANACO está asociado a un modelo matemático específico. Los modelos son:

3.1. Modelo de clasificación simple del experimento.

Se utiliza la tabla siguiente:

Tabla 2
Notación para el ANACO

Tratamiento 1		**…**	**Tratamiento j**		**…**	**Tratamiento k**		**Total Y**	**Total X**
Y_{11} Y_{21} . Y_{n1}	X_{11} X_{21} . X_{n1}		Y_{1j} Y_{2j} . Y_{nj}	X_{1j} X_{2j} . X_{nj}		Y_{1k} Y_{2k} . Y_{nk}	X_{1k} X_{2k} . X_{nk}		
T_{y1} $\overline{Y}_1$	T_{x1} $\overline{X}_1$		T_{yj} $\overline{Y}_j$	T_{xj} $\overline{X}_j$	…	T_{yk} $\overline{Y}_k$	T_{xk} $\overline{X}_k$	$\sum T_{yj} = G_y \ \overline{Y}$	$\sum T_{xj} = G_x \ \overline{X}$

Fuente: Elaboración propia.

$$T_{xx} = n\sum(\overline{X}_j - \overline{X})^2$$
$$E_{xx_j} = n\sum_i (X_{ij} - \overline{X}_j)^2$$
$$E_{xx} = \sum_j E_{xx_j}$$
$$S_{xx} = T_{xx} + E_{xx} = \sum\sum(X_{ij} - \overline{X})^2$$

$$T_{xy} = n\sum(\overline{X}_j - \overline{X})(\overline{Y}_j - \overline{Y})$$
$$E_{xy_j} = \sum_i (X_{ij} - \overline{X}_j)(Y_{ij} - \overline{Y}_j)$$
$$E_{xy} = \sum_j E_{xy_j}$$
$$S_{xy} = T_{xy} + E_{xy} = \sum\sum(X_{ij} - \overline{X})(Y_{ij} - \overline{Y})$$

$$T_{yy} = n\sum(\overline{Y}_j - \overline{Y})^2$$
$$E_{xY_j} = n\sum_i (Y_{ij} - \overline{Y}_j)^2$$
$$E_{yy} = \sum_j E_{yy_j}$$
$$S_{yy} = T_{yy} + E_{yy} = \sum\sum(Y_{ij} - \overline{Y})^2$$

Fuente: Elaboración propia.

Tabla 3
Fórmulas de cálculo para el ANACO para el modelo de clasificación simple

$2x_j = \sum_i X_{ij}^2$ $3x_j = T_{xj}^2 / n$	$2xy_j = \sum_i X_{ij} Y_{ij}$ $3xy_j = T_{xy_j} T_{y_j} / n$	$2y_j = \sum_i Y_{ij}^2$ $3y_j = T_{y_j}^2 / n$
$E_{xx_1} = 2x_1 - 3x_1$ $\ldots$ $E_{xx_k} = 2x_k - 3x_k$	$E_{xy_1} = 2xy_1 - 3xy_1$ $\ldots$ $E_{xy_k} = 2xy_k - 3xy_k$	$E_{xy_1} = 2y_1 - 3y_1$ $\ldots$ $E_{yy_k} = 2y_k - 3y_k$
$1x = G_x^2 / kn$ $2x = \sum 2x_j = \sum X^2$ $3x = \sum 3x_j = \sum T_{x_j}^2 / n$	$1xy = G_x G_y / kn$ $2xy = \sum 2xy_j = \sum Xy$ $3xy = \sum 3xy_j = \sum T_{x_j} T_{y_j} / n$	$1y = G_y^2 / kn$ $2y = \sum 2y_j = \sum Y^2$ $3y = \sum 3y_j = \sum T_{y_j}^2 / n$
$S_{xx} = 2x - 1x$ $E_{xx} = 2x - 3x$ $T_{xx} = 3x - 1x$	$S_{xy} = 2xy - 1xy$ $E_{xy} = 2xy - 3xy$ $T_{xy} = 3xy - 1xy$	$S_{xy} = 2y - 1y$ $E_{yy} = 2y - 3y$ $T_{yy} = 3y - 1y$
	$S_{yy}^p = S_{yy} - S_{xy}^2 / S_{xx}$ $E_{yy}^p = E_{yy} - E_{xy}^2 / E_{xx}$ $T_{yyR} = S_{yy}^p - E_{yy}^p$	

Fuente: Elaboración propia.

Tabla 4
ANAVA

Fuente de variación	SC	GL	CM	F
Tratamientos Error	T_{yy} E_{yy}	$k-1$ $k(n-1)$	CM_{trat} CM_{error}	F_{cal}
Total	S_{yy}	$kn-1$		

Fuente: Elaboración propia.

Tabla 5
ANACO

Fuente de variación	SC	GL	CM	F
Tratamientos Error	T_{yyR} E^{p}_{yy}	$k-1$ $k(n-1)-1$	CM_{tratR} CM^{p}_{error}	F_{cal}
Total	S^{p}_{yy}	$kn-2$		

Fuente: Elaboración propia.

3.2. Experimento factorial para el ANACO.

Es un procedimiento para un experimento factorial *p*q* teniendo *n* observaciones por celda, donde se tienen muestras de tamaños iguales, en el cual se existen *n* pares de observaciones en cada una de las *p*q* celdas. Para su determinación se tiene:

Tabla 6
Fórmulas de cálculo para el ANACO en un experimento factorial

$1x = G_x^2/npq$ $2x = \sum X^2$ $3x = (\sum A_x^2)/nq$ $4x = (\sum B_x^2)/np$ $5x = (\sum AB_x^2)/n$	$1xy = G_x G_y/npq$ $2xy = \sum XY$ $3xy = (\sum A_x A_y)/nq$ $4xy = (\sum B_x B_y)/np$ $5xy = (\sum AB_x AB_y)/n$	$1y = G_y^2/npq$ $2y = \sum Y^2$ $3y = (\sum A_y^2)/nq$ $4y = (\sum B_y^2)/np$ $5y = (\sum AB_y^2)/n$
$A_{xx} = 3x - 1x$ $B_{xx} = 4x - 1x$ $AB_{xx} = 5x - 3x - 4x + 1x$ $E_{xx} = 2x - 5x$	$A_{xy} = 3xy - 1xy$ $B_{xy} = 4xy - 1xy$ $AB_{xy} = 5xy - 3xy - 4xy + 1xy$ $E_{xy} = 2xy - 5xy$	$A_{yy} = 3y - 1y$ $B_{yy} = 4y - 1y$ $AB_{yy} = 5y - 3y - 4y + 1y$ $E_{yy} = 2y - 5y$
$E^{p}_{yy} = E_{yy} - E^2_{xy}/E_{xx}$		
$(A+E)^{p}_{yy} = A_{yy} + E_{yy} - \dfrac{(A_{xy}+E_{xy})^2}{A_{xx}+E_{xx}}$ $(B+E)^{p}_{yy} = B_{yy} + E_{yy} - \dfrac{(B_{xy}+E_{xy})^2}{B_{xx}+E_{xx}}$ $(AB+E)^{p}_{yy} = AB_{yy} + E_{yy} - \dfrac{(AB_{xy}+E_{xy})^2}{AB_{xx}+E_{xx}}$	$A^{p}_{yy} = (A+E)^{p}_{yy} - E^{p}_{yy}$ $B^{p}_{yy} = (B+E)^{p}_{yy} - E^{p}_{yy}$ $AB^{p}_{yy} = (AB+E)^{p}_{yy} - E^{p}_{yy}$	

Fuente: Elaboración propia.

Tabla 7
ANAVA

Fuente de variación	SC	GL	CM	F
A	A_{yy}	$p-1$	CM_a	F_a
B	B_{yy}	$q-1$	CM_b	F_b
AB	AB_{yy}	$(p-1)(q-1)$	CM_{ab}	
Error	E_{yy}	$pq(n-1)$	CM_{error}	
Total	S_{yy}	$pqn-1$		

Fuente: Elaboración propia.

Tabla 8
ANACO

Fuente de variación	SC	GL	CM	F
A^p	A^p_{yy}	$p-1$	CM^p_a	F^p_a
B^p	B^p_{yy}	$q-1$	CM^p_b	F^p_b
AB^p	AB^p_{yy}	$(p-1)(q-1)$	CM^p_{ab}	
$Error^p$	E^p_{yy}	$pq(n-1)-1$	CM^p_{error}	
Total	S^p_{yy}	$pqn-2$		

Fuente: Elaboración propia.

3.3. Experimento factorial para el ANACO.

Teniéndose pares de datos para frecuencias de celdas diferentes, el procedimiento para la determinación del ANAVA y ANACO se muestra a través de un experimento factorial en forma generalizada teniéndose una covariación.

El símbolo $\tilde{n}$ representa la media armónica de las frecuencias de las celdas, definida como:

$$\tilde{n} = \frac{pq}{\sum (1/n_{ij})}$$

Tabla 6
Fórmulas de cálculo para el ANACO en un experimento factorial

$1x = \tilde{n}G_x^2/pq$ $2x = \sum X^2$ $3x = \tilde{n}\sum(A_x^2)/q$ $4x = \tilde{n}\sum(B_x^2)/p$ $5x = \tilde{n}\sum \overline{AB}_x^2$ $5'x = \sum(AB_{xjk}^2)/n_{jk}$	$1xy = \tilde{n}G_xG_y/pq$ $2xy = \sum XY$ $3xy = \tilde{n}\sum(A_xA_y)/q$ $4x = \tilde{n}\sum(B_xB_y)/p$ $5xy = \tilde{n}\sum \overline{AB}_x\overline{AB}_y$ $5'x = \sum(AB_{xjk}AB_{yjk})/n_{jk}$	$1y = \tilde{n}G_y^2/pq$ $2y = \sum Y^2$ $3y = \tilde{n}\sum(A_y^2)/q$ $4y = \tilde{n}\sum(B_y^2)/p$ $5y = \tilde{n}\sum \overline{AB}_y^2$ $5'y = \sum(AB_{yjk}^2)/n_{jkk}$
$A_{xx} = 3x - 1x$ $B_{xx} = 4x - 1x$ $AB_{xx} = 5x - 3x - 4x + 1x$ $E_{xx} = 2x - 5'x$	$A_{xy} = 3xy - 1xy$ $B_{xy} = 4xy - 1xy$ $AB_{xy} = 5xy - 3xy - 4xy + 1xy$ $E_{xy} = 2xy - 5'xy$	$A_{yy} = 3y - 1y$ $B_{yy} = 4y - 1y$ $AB_{yy} = 5y - 3y - 4y + 1y$ $E_{yy} = 2y - 5'y$

$E^{p}_{yy} = E_{yy} - E^{2}_{xy} / E_{xx}$	
$(A+E)^{p}_{yy} = A_{yy} + E_{yy} - \dfrac{(A_{xy}+E_{xy})^2}{A_{xx}+E_{xx}}$	$A^{p}_{yy} = (A+E)^{p}_{yy} - E^{p}_{yy}$
$(B+E)^{p}_{yy} = B_{yy} + E_{yy} - \dfrac{(B_{xy}+E_{xy})^2}{B_{xx}+E_{xx}}$	$B^{p}_{yy} = (B+E)^{p}_{yy} - E^{p}_{yy}$
$(AB+E)^{p}_{yy} = AB_{yy} + E_{yy} - \dfrac{(AB_{xy}+E_{xy})^2}{AB_{xx}+E_{xx}}$	$AB^{p}_{yy} = (AB+E)^{p}_{yy} - E^{p}_{yy}$

Fuente: Elaboración propia.

Las tablas de ANAVA y ANACO son similares al experimento factorial anterior.

*3.4. Experimento factorial de p*q, para el ANACO cuando los factores son repetidos*

El diseño puede ser representado esquemáticamente, en el experimento se tiene *n* sujetos.

Tabla 10
Experimento factorial para factores repetidos

	b_1 b_2 .... b_q	Fuente	SC	GL
a_1	G_1 G_1 .... G_1	**Entre sujetos A**	A_{yy}	*(np-1)/(p-1)*
a_2	G_2 G_2 .... G_2	Suj. dentro el grupo	P_{yy}	*p(n-1)*
.	. . .			
.	. . .	**Dentro los sujetos B**	B_{yy}	*np(q-1)/(q-1)*
.	. . .	AB	AB_{yy}	*(p-1)(q-1)*
a_p	G_p G_p .... G_p	Residual	E_{yy}	*p(q-1)(n-1)/(npq-1)*

Fuente: Elaboración propia.

Tabla 11
Fórmulas del ANACO para el diseño de la tabla 10

Fuente	X^2	XY	Y^2	Variación ajustada	GL
A	A_{xx}	A_{xy}	A_{yy}	$A^{'}_{yy} = A_{yy} + P_{yy} - \dfrac{(A_{xy}+P_{xy})^2}{A_{xx}+P_{xx}} - P^{'}_{yy}$	$p-1$
Suj. det. gp	P_{xx}	P_{xy}	P_{yy}	$P^{'}_{yy} = P_{yy} - \dfrac{P^2_{xy}}{P_{xx}}$	$p(p-1)-1$
B	B_{xx}	B_{xy}	B_{yy}	$B^{'}_{yy} = B_{yy} + E_{yy} - \dfrac{(B_{xy}+E_{xy})^2}{B_{xx}+E_{xx}} - E^{'}_{yy}$	$q-1$
AB	AB_{xx}	AB_{xy}	AB_{yy}	$AB^{'}_{yy} = AB_{yy} + E_{yy} - \dfrac{(AB_{xy}+E_{xy})^2}{AB_{xx}+E_{xx}} - E^{'}_{yy}$	$(p-1)(q-1)$
Residual	E_{xx}	E_{xy}	E_{yy}	$E^{'}_{yy} = E_{yy} - \dfrac{E^2_{xy}}{E_{xx}}$	$p(p-1)(q-1)-1$

Fuente: Elaboración propia.

Tabla 12
Análisis de covarianza para el diseño de la tabla 10

Fuente	X^2	XY	Y^2	Variación ajustada	g. l.
A	A_{xx}	A_{xy}	A_{yy}	$A'_{yy} = A_{yy} + P_{yy} - \frac{(A_{xy} + P_{xy})^2}{A_{xx} + P_{xx}} - P'_{yy}$	$p-1$
Suj. d. g.	P_{xx}	P_{xy}	P_{yy}	$P'_{yy} = P_{yy} - P_{xy}^2 / P_{xx}$	$p(n-1)-1$
B	B_{xx}	B_{xy}	B_{yy}	$B'_{yy} = B_{yy} + E_{yy} - \frac{(B_{xy} + E_{xy})^2}{B_{xx} + E_{xx}} - E'_{yy}$	$q-1$
AB	AB_{xx}	AB_{xy}	AB_{yy}	$AB'_{yy} = AB_{yy} + E_{yy} - \frac{(AB_{xy} + E_{xy})^2}{AB_{xx} + E_{xx}} - E'_{yy}$	$(p-1)(q-1)$
Residual	E_{xx}	E_{xy}	E_{yy}	$E'_{yy} = E_{yy} - E_{xy}^2 / E_{xx}$	$p(q-1)(n-1)-1$

Fuente: Elaboración propia.

4. TRABAJO COMPUTACIONAL

Por todo lo anterior mencionado, se observa que en la determinación de los ANACOS, se constata que son numerosos los cálculos a realizar, por esta razón, se facilita su determinación de acuerdo a las necesidades del usuario, por medio de la ejecución del menú, en el cual para cada uno de ellos se tiene un ejemplo numérico mostrado en el inciso 6, para mayor comprensión e interpretación de la teoría desarrollada.

	UNIVERSIDADES UTO[2], UMSA[3], UAGRM ANÁLISIS DE COVARIANZA Dr. Cs. Gustavo Ruiz Aranibar
1	Anaco, modelo para muestras de igual tamaño
2	Anaco, modelo de clasificación simple del experimento
3	Anaco, experimento factorial, para tamaños iguales de X y Y
4	Anaco, experimento factorial, para tamaños diferentes de X y Y
5	Anaco, experimento factorial de p*q, con factores repetidos.
0	Salir del sistema

5. APLICACIONES DEL ANACO

Mención de algunas aplicaciones del ANACO:

a) La efectividad de varios programas de capacitación industrial y los resultados dependerán de los cocientes intelectuales de los participantes.
b) La duración de varias clases de suelas para zapatos y los resultados dependieran del peso de las personas que lo usaran.
c) Si los méritos de varios productos de limpieza y los resultados dependieran de la condición original de las superficies por limpiarse.

[2] Universidad Técnica de Oruro
[3] Universidad Mayor de San Andrés

Aplicación desarrollada para comprensión del lector.

Un investigador tiene tres diferentes productos de limpieza A_1, A_2 y A_3 y desea seleccionar el más eficiente para limpiar una superficie metálica. La limpieza de una superficie se mide por su reflectividad, expresada en unidades arbitrarias como la razón de reflectividad observada a la superficie de un espejo estándar. El investigador obtuvo los siguientes resultados:

Datos.

A_1		A_2		A_3	
Reflectividad original, x	Reflectividad final, y	Reflectividad original, x	Reflectividad final, y	Reflectividad original, x	Reflectividad final, y
0,50	1,00	0,75	0,75	0,60	1,00
0,55	1,20	0,65	0,60	0,90	0,70
0,60	0,80	1,00	0,55	0,80	0,80
0,35	1,40	1,10	0,50	0,70	0,90

Fuente: Elaboración propia.

Efectuar un análisis de covarianza para determinar si hay diferencias en las mejoras de reflectividad producidas por los tres productos de limpieza, en el nivel de significación de 0,05.

Solución.

El análisis de covarianza debe usarse porque el efecto de un producto de limpieza sobre la reflectividad depende de la limpieza original, esto es, de la reflectividad original de la superficie. Se realiza los cálculos, de manera, que el lector fácilmente comprenda lo realizado.

Solución.

Hipótesis:

$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_i = 0$

H_1 : *no son todas iguales a cero.*

Nivel de significación: α = 0.05

Criterio: Rechazar H_0 si $F_{cal} > F_t$

$F_t = F_{\alpha\,,\,k-1\,,\,mk-k-1} = F_{0.95\,,\,3-1\,,\,4*3-3-1} = F_{0.95\,,\,2\,,\,8} = 4,46$

Los totales son:

A_1		A_2		A_3	
R.O. , x	**R.F. , y**	**R.O. , x**	**R.F. , y**	**R.O. , x**	**R.F. , y**
0,50	1,00	0,75	0,75	0,60	1,00
0,55	1,20	0,65	0,60	0,90	0,70
0,60	0,80	1,00	0,55	0,80	0,80
0,35	1,40	1,10	0,50	0,70	0,90
Tx_1=2,00	Ty_1=4,40	Tx_2=4,50	Ty_2=2,40	Tx_3=3,00	Ty_3=3,40

Fuente: Elaboración propia.

Cálculos para X:

$$T_x = \sum_{i=1}^{k} T_{x_i} = 9{,}50 \qquad C = \frac{T_{\cdot}^2}{kn} = \frac{9.50^2}{3*4} = 7{,}52$$

$$SST_x = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - C = 0{,}50^2 + 0{,}55^2 + \ldots + 0{,}70^2 - 7{,}52 = 1{,}31$$

$$SS(Tr)_x = \frac{\sum_{i=1}^{k} T_i^2}{n} - C = \frac{2{,}00^2 + 4{,}50^2 + 3{,}00^2}{4} - 7{,}52 = 0{,}79$$

$$SSE_x = SST_x - SS(Tr)_x = 1{,}31 - 0{,}79 = 0{,}52$$

Cálculos para *y*:

$$T_y = \sum_{i=1}^{k} T_{y_i} = 10{,}20 \qquad C = \frac{T_{\cdot}^2}{kn} = \frac{10{,}20^2}{3*4} = 8{,}67$$

$$SST_y = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - C = 1{,}00^2 + 1{,}20^2 + \ldots + 0{,}90^2 - 8{,}67 = 0{,}79$$

$$SS(Tr)_y = \frac{\sum_{i=1}^{k} T_i^2}{n} - C = \frac{4{,}40^2 + 2{,}40^2 + 3{,}40^2}{4} - 8{,}67 = 0{,}50$$

$$SSE_y = SST_y - SS(Tr)_y = 0{,}79 - 0{,}50 = 0{,}29$$

Cáculos para las sumas de productos:

$$C = \frac{T_x T_y}{kn} = \frac{9{,}50*10{,}20}{3*4} = 8{,}08$$

$$SPT = \sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij} y_{ij} - C = 0{,}50*1{,}00 + 0{,}55*1{,}20 + \ldots + 0{,}70*0{,}90 - 8{,}08 = -0{,}80$$

$$SP(Tr) = \frac{\sum_{i=1}^{k} T_{x_i} T_{y_i}}{n} - C = \frac{2{,}00*4{,}40 + 4{,}50*2{,}40 + 3{,}00*3{,}40}{4} - 8{,}08 = -0{,}63$$

$$SPE = SPT - SP(Tr) = -0{,}80 - (0{,}63) = -0{,}17$$

Cálculos para las *y* ajustadas se obtiene:

$$SST_{y'} = SST_y - \frac{SPT^2}{SST_x} = 0{,}79 - \frac{-0{,}80^2}{1{,}31} = 0{,}30$$

$$SSE_{y'} = SSE_y - \frac{SPE^2}{SSE_x} = 0{,}29 - \frac{-0{,}17^2}{0{,}52} = 0{,}23$$

$$SS(Tr)_{y'} = SST_{y'} - SSE_{y'} = 0{,}30 - 0{,}23 = 0{,}07$$

$$MS(Tr)_{y'} = \frac{SS(T_r)_{y'}}{k-1} = \frac{0{,}07}{3-1} = 0{,}035$$

$$MSE_{y'} = \frac{SSE_{y'}}{nk-k-1} = \frac{0{,}23}{4*3-3-1} = 0{,}026$$

$$F_{cal} = \frac{MS(Tr)_{y'}}{MSE_{y'}} = \frac{0{,}035}{0{,}026} = 1{,}34$$

Tabla 12
Tabla de análisis de covarianza

Fuente de variación	Suma de cuadrados para x	Suma de cuadrados para y	Suma de productos	Suma de cuadrados para y´	Grados de libertad	Cuadrado Medio	Razón F
Tratamientos	$SS(Tr)_x$ 0,79	$SS(Tr)_y$ 0,50	$SP(Tr)$ -0,63	$SS(Tr)_{y'}$ 0,07	$k-1$ 2	$MS(Tr)_{y'}$ 0,035	F_{cal} 1,34
Error	SSE_x 0,52	SSE_y 0,29	SPE -0,17	$SSE_{y'}$ 0,23	$nk-k-1$ 8	$MSE_{y'}$ 0,026	
Total	SCT_x 1,31	SCT_y 0,79	SPT -0,80	$SCT_{y'}$ 0,30	$nk-2$ 10		

Fuente: Elaboración propia.

Decisión:

Rechazar al nivel de significación α si el valor obtenido para F_{cal} excede a F_α con k-1 y nk-k-1 grados de libertad.

$F_{cal} > F_t$ *se rechaza* H_0
$F_{cal} = 1{,}34 < F_t = 4{,}46$ *se acepta* H_0

Como se puede observar el F_{cal} es mayor al F_t, la hipótesis nula debe rechazarse.En otras palabras, no puede concluirse que alguno de los productos de limpieza sea más efectivo que los otros.

Nota. Las aplicaciones del ANACO, para cada una de las instrucciones del menú señalado, se muestran a continuación; los datos y programas computacionales, son elaboraciones propias.

5.1. ANACO, modelo para muestras de igual tamaño.

Tres configuraciones diferentes de tableros de instrumentos se probaron con pilotos de aerolíneas en simuladores de vuelo donde se midieron sus tiempos de reacción frente a emergencias de vuelos simulados. Cada piloto enfrento diez condiciones de emergencias en secuencia aleatorizada y se midió el tiempo total requerido para tomar acción correctiva en las diez condiciones con los siguientes resultados:

Datos.

	Tablero de instrumentos 1		**Tablero de instrumentos 2**		**Tablero de instrumentos 3**	
	x	**Y**	**x**	**y**	**x**	**y**
1	7,30	7,32	4,60	5,43	9,50	7,48
2	9,40	6,55	10,45	6,21	14,10	6,21
3	8,10	6,55	12,10	5,74	15,20	6,37
4	19,40	6,40	2,10	5,93	8,70	6,97
5	11,60	5,93	3,90	6,16	7,20	7,38
6	24,60	6,79	5,20	5,68	6,10	6,43
7	6,20	7,16	4,60	5,41	11,80	7,59
8	3,80	5,64	14,40	6,29	12,10	7,16
9	18,40	5,87	16,10	5,55	9,50	7,02
10	9,40	6,32	8,50	4,82	2,60	6,85

Fuente: Elaboración propia.

En esta tabla, x es el número de años de experiencia del piloto y y es el tiempo total de reacción en segundos. Efectúe un análisis de covarianza para probar si las configuraciones de los tableros dan resultados significativamente diferentes a un nivel de significación $\alpha = 5$ %.

Solución.

Tabla del ANACO

Fuente de variación	Suma de cuadrados para x	Suma de cuadrados para y	Suma de productos	Suma de cuadrados para y'	Grados de libertad	Cuadrado Medio	Razón F
Tratamientos	$SS(Tr)_x$ 67,5772	$SS(Tr)_y$ 7,5852	$SP(Tr)$ 11,3986	$SS(Tr)_{y'}$ 7,4940	$k-1$ 2	$MS(Tr)_{y'}$ 3,7470	F_{cal} 14,7965
Error	SSE_x 756,0730	SSE_y 6,5917	SPE -2,3834	$SSE_{y'}$ 6,5842	$nk-k-1$ 26	$MSE_{y'}$ 0,2532	
Total	SCT_x 823,6501	SCT_y 14,1769	SPT 9,0151	$SCT_{y'}$ 14,0782	$nk-2$ 28		

Fuente: Elaboración propia.

Decisión:

Rechazar al nivel de significación α si el valor obtenido para Fcal excede a Fα con k-1 y nk-k-1 grados de libertad.

$F_{cal} > F_t$ *se rechaza* H_0

$F_{cal} = 14,7965 < F_{5\%,(2,26)} = 3,37$ *se acepta* H_0

Como se puede observar el F_{cal} es mayor al F_t, la hipótesis nula debe rechazarse.

5.2. ANACO, modelo de clasificación simple del experimento para muestras de igual tamaño. Teniéndose tres tratamientos (grupos) que representan diferentes métodos de enseñanza o de entrenamiento, y siete sujetos en cada uno de los grupos, realizar un análisis de covarianza. Los grupos han sido asignados aleatoriamente para estos métodos de enseñanza.

Datos.

Métodos de enseñanza					
Tratamiento 1		**Tratamiento 2**		**Tratamiento 3**	
X	***Y***	***X***	***Y***	***X***	***Y***
4	7	5	9	4	7
2	5	6	10	4	8
4	6	6	8	3	8
2	4	5	10	4	8
3	5	4	9	5	9
2	4	2	6	2	6
5	7	3	8	5	8

Fuente: Elaboración propia.

Solución

	ANAVA				ANACO			
FV	**SC**	**GL**	**CM**	**F**	**SC**	**GL**	**CM**	**F**
Tratamientos	36.95	2	18.48	12.38	15,51	2	7,76	13,82
Error	26.86	18	1.49		9,55	17	0,56	
Total	63.81	20			25,06	19		

Fuente: Elaboración propia.

Decisión:

$F_{cal} = 13{,}82 < F_{5\%,(2,17)} = 3{,}59$ *se acepta* H_0

Como se puede observar el F_{cal} es mayor al F_t, la hipótesis nula debe rechazarse.

Es de interés examinar la magnitud de las correlaciones dentro las clases, obteniéndose el cuadrado de estas correlaciones como sigue:

$$r_1^2 = \frac{E_{xy_1}^2 / E_{xx_1}}{E_{yy_1}} = 0{,}85 \qquad r_2^2 = \frac{E_{xy_2}^2 / E_{xx_2}}{E_{yy_2}} = 0{,}53 \qquad r_3^2 = \frac{E_{xy_3}^2 / E_{xx_3}}{E_{yy_3}} = 0{,}59$$

El cuadrado de la correlación conjunta dentro las clases es:

$$r^2 = \frac{E_{xy}^2 \;/ E_{xx}}{E_{yy}} = 0{,}64$$

El cuadrado de la correlación entre la covariación y el criterio es:

$$r_{total}^2 = \frac{S_{xy}^2 \;/ S_{xx}}{S_{yy}} = 0{,}964$$

5.3. Experimento factorial para el ANACO.

Un experimento factorial con p = 2 niveles para el factor A que representan los niveles de instrucción en la enseñanza de una asignación y q = 3 niveles del factor B, que representan a los instructores. Ambos factores son considerados fijos. Determinar ANAVA y ANACO para el nivel de significación α = 5%.

Datos

	Métodos de instrucción					
Método	**b1 Instructor 1**		**b2 Instructor 2**		**b3 Instructor 3**	
	X	**Y**	**X**	**Y**	**X**	**Y**
a1	39	94	29	84	49	89
	36	81	41	100	39	86
	39	94	44	84	41	91
	52	107	42	92	32	82
	44	100	43	92	41	86
a2	49	100	48	100	46	96
	32	97	31	91	31	84
	36	96	41	96	24	73
	44	108	43	92	51	106
	29	87	42	96	37	88

Fuente: Elaboración propia.

Solución:

	ANAVA				ANACO			
FV	**SC**	**GL**	**CM**	**F**	**SC**	**GL**	**CM**	**F**
Métodos A	76,80	1	76,800	1,23	163,17	1	163,17	6,99
Instructores B	345,80	2	172,900	2,70	284,49	2	142,25	6,09
AB	7,41	2	3,700		5,65	2	2,83	
Error	1493,00	24	62,220		537,02	23	23,35	

Fuente: Elaboración propia.

Decisión:

$F_{cal} = 6{,}99 < F_{5\%\,,\,(1,23)} = 4{,}28$ *se rechaza* H_0

Como se puede observar el F_{cal} es mayor al F_t, la hipótesis nula debe rechazarse.

$F_{cal} = 6{,}99 < F_{5\%\,,\,(2,23)} = 3{,}42$ *se rechaza* H_0

Como se puede observar el Fcal es mayor al Ft, la hipótesis nula debe rechazarse.

La estimación del cuadrado de correlación dentro las celdas es:

5.4. Experimento factorial para el ANACO, para tamaños diferentes de X y Y.

Se tiene la información de un experimento factorial de 2*3, determinar ANAVA y ANACO, al nivel de significación α = 5%.

Datos

	Métodos de instrucción					
Método	**b_1 Docente 1**		**b_2 Docente 2**		**b_3 Docente 3**	
	X	**Y**	**X**	**Y**	**X**	**Y**
a_1	6	9	3	15	4	20
	6	17	2	12	3	15
	5	11	9	21	4	15
	10	25	8	16	5	20
			5	13	7	23
					4	18
a_2	8	19	0	9	0	11
	0	8	5	17	2	16
	5	11	9	21	10	26
	7	16	6	19	5	19
	10	24			6	17
					8	25
					9	25

Fuente: Elaboración propia

Solución:

	ANAVA				ANACO			
FV	**SC**	**GL**	**CM**	**F**	**SC**	**GL**	**CM**	**F**
A	5,41	1	5,41	0,2	5,01	1	5,01	0,84
B	78,54	2	39,27	1,42	150,07	2	75,03	12,64
AB	2,19	2	1,09	0,04	6,32	2	3,16	0,53
Dentro las celdas	690,76	25	27,63		142,49	24	5,94	

Fuente: Elaboración propia

*5.5. Experimento factorial de p*q, para el ANACO con factores repetidos.*

Datos

		b_1		b_2	
		X	**Y**	**X**	**Y**
a_1	1	5	10	6	16
	2	6	12	10	19
	3	12	17	15	23
a_2	4	5	10	6	15
	5	9	14	10	15
	6	12	10	9	11
a_3	7	8	12	6	12
	8	9	15	11	19
	9	11	17	14	24

Fuente: Elaboración propia

Solución

Totales de filas, columnas.

	23,00	39,00	31,00	58,00	54,00	97,00
	26,00	34,00	25,00	41,00	51,00	75,00
	28,00	44,00	31,00	55,00	59,00	99,00
Total	**77,00**	**117,00**	**87,00**	**154,00**	**164,00**	**271,00**

Fuente: Elaboración propia

Cálculos de acuerdo a Tabla 9

1494,22	2469,11	4080,06
1652,00	2627,00	4385,00
1499,67	2484,00	4139,17
1499,78	2489,67	4156,11
1512,00	2513,67	4227,67
1625,00	2585,00	4278,50

Fuente: Elaboración propia

Análisis de varianza y covarianza

	X2	XY	Y2
A	5,44	14,89	59,11
Suj. dentro A	125,33	101,00	139,33
B	5,56	20,56	76,06
AB	6,78	9,11	12,44
B*(Sujt. Dentro A)	14,67	12,33	18,00
Total	157,78	157,89	304,94

Fuente: Elaboración propia

	ANAVA				ANACO			
FV	**SC**	**GL**	**CM**	**F**	**SC**	**GL**	**CM**	**F**
Entre sujetos A	59,11	2	29,56	1,27	37,81	2	18,90	1,63
Suj. dentro A	139,33	6	23,22		57,94	5	11,59	
Entre los suj. B	76,06	1	76,06	25,35	32,94	1	32,94	21,59
AB	12,44	2	6,22	2,07	1,37	2	0,69	0,45
B*Suj.dent. A	18,00	6	3,00		7,63	5	1,53	

Fuente: Elaboración propia

Las sumas de cuadrados y las sumas de productos para las variaciones y covariaciones se muestran en la primera parte del cuadro anterior. El análisis de la varianza se resume en la segunda parte El análisis de covarianza se resume en la tercera parte. La variación ajustada se define en la tabla 11, teniéndose:

$$E'_{yy} = E_{yy} - \frac{E_{xy}^2}{E_{xx}} = 18 - \frac{12{,}33^2}{14{,}67} = 7{,}64$$

$$B'_{yy} = B_{yy} + E_{yy} - \frac{(B_{xy} + E_{xy})^2}{B_{xx} + E_{xx}} - E'_{yy} = 76{,}06 + 18 - \frac{(20{,}56 + 12{,}33)^2}{5{,}56 + 14{,}67} - 7{,}64 = 32{,}95$$

Los efectos entre objetos en la tercera parte, se ajustan de acuerdo con los procedimientos descritos en la tabla 11. Las estimaciones de β_p y β_w son:

$$b_p = \frac{P_{xy}}{P_{xx}} = \frac{101{,}00}{125{,}33} = 0{,}81$$

$$b_w = \frac{E_{xy}}{E_{xx}} = \frac{12{,}33}{14{,}67} = 0{,}841$$

En este caso, la inspección indica que b_w puede usarse a lo largo del proceso de ajuste. Si el ajuste de los efectos entre sujetos se realiza de acuerdo con los procedimientos indicados en la parte inferior de la tabla 12, las sumas de cuadrados ajustadas son las siguientes:

$$A_{yy} = A_{yy} - 2b_w A_{xy} + b_w^2 A_{xx}$$

$$= 59{,}11 - 2*0{,}84*14{,}89 + 0{,}84^2*5{,}44 = 37{,}91$$

$$P_{yy} = P_{yy} - 2b_w P_{xy} + b_w^2 P_{xx}$$

$$= 139{,}33 - 2*0{,}841*101{,}00 + 0{,}841^2*125{,}33 = 58{,}09$$

La prueba F general para la hipótesis $\sigma_{\alpha}^{2} = 0$ tiene la forma:

$$F = \frac{A_{yy}^{\cdot}/(p-1)}{P_{yy}^{\cdot}/p(n-1)} = \frac{37.91/(3-1)}{58.09/3(3-1)} = 1{,}96$$

Se observa que el término de error para los efectos dentro del objeto es 3,00 en el análisis de varianza, y 1,53 en el análisis de covarianza. Para los efectos entre sujetos, los términos de error correspondientes son:

23,22 y 139,33 / 6 = 23,22.

6. CONCLUSIONES

Se ha observado que a lo largo del estudio del ANACO, aplicando las fórmulas descritas, se requiere demasiados cálculos, lo cual se evita con el programa computacional desarrollado que se muestra en el menú mencionado. La parte teórica y los ejemplos ayudan a los lectores a comprender la utilización de los modelos, que permiten su aplicación y adaptación a diferentes trabajos de investigación o de la industria.

La razón F se interpreta igual que en el ANAVA, con la diferencia que las inferencias y conclusiones, se realizan tomando en cuenta que las medias de la variable dependiente a través de las categorías de las variables independientes, han sido ajustadas, eliminando el efecto de la covariable.

Colaboración.

Prof. M. Nilda Avilés de Ruiz.
Licenciada en Idiomas. Universidad Autónoma Gabriel René Moreno
Santa Cruz – Bolivia

M.Sc. Astrid Keitel Ruiz Avilés
Lic. Adm. Emp. Universidad Católica Boliviana, Santa Cruz – Bolivia
Academia Diplomática Boliviana ¨Rafael Bustillo¨, La Paz – Bolivia

BIBLIOGRAFÍA

- Berenson L. Mark, LEVINE M. David, "*Estadística Básica en Administración*". Prentice Hall Hispanoamericana, S.A., México D.F. 1997 (6ta. edición), pp. 943 - XIX.

- Dagnelie Pierre, "*Analyse Statistique á Plusieurs Variables*". Les Presses Agronomiques de Gembloux, Gembloux, Bélgica, 1975 (2da. Edición), pp. 362 - XIV.

- Johnson A. Richard, "*Probabilidad y Estadística para Ingenieros*". Prentice Hall Hispanoamericana, S.A., México D.F. 1997 (5ta. Edición), pp. 630 - XV.

- Ruiz Aranibar Gustavo. "*Varianza. Análisis de Varianza*". Revista del IETA Nº 13. UMSA. F.C.P.N., Carrera de Estadística. La Paz – Bolivia. 2017, pp. 87 – IX.

- Ruiz Aranibar Gustavo . "*Libreria Científica de Programas Informáticos*", La Paz - Bolivia.

- Winer B. J. "*Statistical Principle in Experimental Design*", McGraw Hill, New York, USA, 1971, (2da. Edición), pp. 907 - X.

Pensamiento

Toda una vida se debe aconsejar estudiar, actualizandose en aspectos tecnológicos, humanisticos, y de servicio a la humanidad, lo que se consigue con voluntad y perseverancia, adquiriendo constantemente nuevos conocimientos.

Gustavo Ruiz Aranibar