Daena: International Journal of Good Conscience. 9(2) 12-17. Agosto 2014. ISSN 1870-557X

Aplicación de ANOVA Anidada en la Investigación Científica
(Nested ANOVA Application in Scientific Research)

Badii, M.H., A. Guillen & J.L. Abreu
UANL, San Nicolás, N.L., México

Abstract. Nested ANOVA or Nested Analysis of Variance is a statistical tool that can be utilized when there is interest in determining the differences among of various subgroups each within a group. The application of this technique along with an example is described.
Keywords: ANOVA, groups, nested, sub-groups

Resumen. El ANOVA anidada es una herramienta estadística usada cuando se desea estimar la diferencia entre los promedios de varios subgrupos cada uno localizado dentro del un grupo. Se describe junto cn un ejemplo práctico la aplicación de esta técnica estadística .
Palabras clave: Anidada, ANOVA, grupos, sub-grupos

Introducción

Se utiliza la técnica del Análisis de Varianza Anidado (ANOVA anidada) cuando se tiene una variable de medición y dos o más variables nominales (Badii et al., 2009). Las variables nominales se anidan, lo que significa que cada valor de una variable nominal (los subgrupos) se encuentra en combinación con un solo valor de la variable nominal de más alto nivel (los grupos). El nivel nominal de la variable superior puede ser Modelo I o modelo II pero el nivel nominal de las variables más bajos deben ser el Modelo II (Steel and Torrie, 1960; Zar, 1996; Pagano & Gauvreau, 2000; Casella & Berger, 2002)
El análisis de varianza anidado es una extensión del análisis de varianza de una vía en que se divide cada grupo en subgrupos. En teoría, estos subgrupos se eligen al azar de un conjunto más amplio posible de los subgrupos. Un análisis de varianza anidado tiene una hipótesis nula para cada nivel. En el ANOVA anidada, una hipótesis nula sería que los subgrupos dentro de cada grupo tienen promedios iguales, la segunda hipótesis nula sería que los grupos tienen los mismos medios.

Objetivo

Comparar varios factores que estén relacionados con una muestra, determinando su igualdad de las medias.

Requisitos

Como todos los diferentes métodos de ANOVA, se supone que las observaciones dentro de cada subgrupo se distribuyen normalmente, los residuales son independientes y con distribución normal y además, existe homogeneidad de varianzas para diferentes valores de las medias.

Ejemplo. Hay 5 laboratorios (A, B, C, D y E), cada uno con 4 trabajadores (I, II, III y IV), el objetivo es 1) el analizar y comparar si hay diferencia entre las medias de grupos o laboratorios y 2) si existe diferencias significativas entre los promedios de número de análisis realizado por cada trabajador (subgrupo) de cada laboratorio durante 4 meses consecutivos (Tabla 1).

Tabla 1. Número de análisis por trabajador de cada laboratorio*.

Mes	**Laboratorio**																			
	A				**B**				**C**				**D**				**E**			
	Trabajador				**Trabajador**				**Trabajador**				**Trabajador**				**Trabajador**			
	I	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**
1	6	13	1	7	10	2	4	0	0	10	8	7	11	5	1	0	1	6	3	3
2	2	3	10	4	9	1	1	3	0	11	5	2	0	10	8	8	4	7	0	7
3	0	9	0	7	7	1	7	4	5	6	0	5	6	8	9	6	7	0	2	4
4	8	8	6	9	12	10	9	1	5	7	7	4	4	3	4	5	9	3	2	0
ΣCT	16	33	17	27	38	14	21	8	10	34	20	18	21	26	22	19	21	16	7	14
ΣCL	93				81				82				88				58			

*: ΣCT = Sumatorio para cada trabajador o subgrupo, ΣCL = Sumatorio para cada laboratorio o grupo.

Juegos de hipótesis:

1. Caso de los promedios (m) de los cinco laboratorios (suma de CL´s):
   Ho: $m_{LabA} = m_{Lab\ B} = m_{Lab\ C} = m_{LabD} = m_{Lab\ E}$
   Ha: Al menos una media es diferente de los demás medias
2. Caso de de los 20 trabajadores de todos los laboratorios (suma de CT´s):
   Ho: $m_{T\ I} = m\ m_{TII} = m_{TIII} = m_{\ TIV} = \ldots = m_{\ TIV}$ del ultimo laboratorio
   Ha: Al menos una media es diferente de los demás medias
3. Caso de los Trabajadores (T) de cada laboratorio (Lab):

   3A. Los Trabajadores del laboratorio A:
   Ho: $m_{\ TI\ Lab\ A} = m_{\ TII\ Lab\ A} = m_{\ TIII\ Lab\ A} = m_{\ TIV\ Lab\ A}$
   Ha: Al menos una media es diferente

   3B. Los Trabajadores del laboratorio B:
   Ho: $m_{\ TI\ Lab\ B} = m_{\ TII\ Lab\ B} = m_{\ TIII\ Lab\ B} = m_{\ TIV\ Lab\ B}$
   Ha: Al menos una media es diferente

   3C. Los Trabajadores del laboratorio C:
   Ho: $m_{\ TI\ Lab\ C} = m_{\ TII\ Lab\ C} = m_{\ TIII\ Lab\ C} = m_{\ TIV\ Lab\ C}$

Ha: Al menos una media es diferente

3D. Los Trabajadores del laboratorio D:

Ho: $m_{TI\ Lab\ D} = m_{TII\ Lab\ D} = m_{TIII\ Lab\ D} = m_{TIV\ Lab\ D}$

Ha: Al menos una media es diferente

3E. Los Trabajadores del laboratorio E:

Ho: $m_{TI\ Lab\ E} = m_{TII\ Lab\ E} = m_{TIII\ Lab\ E} = m_{TIV\ Lab\ E}$

Ha: Al menos una media es diferente

El templete o el modelo de la tabal de ANOVA para el caso a ANOVA anidada se demuestra en la Tabla 2.

Tabla 2. Modelo de la tabla de anova para anova anidada*.

FV	Gl	SC	V	$F_{Calculada}$
Laboratorio (grupos)	k-1	$\Sigma(Tcg)^2/r_{CG} - (\Sigma x_i)^2/N_T$	SC_{Lab}/gl_{Lab}	V_{Lab}/V_{Error}
GP´s combinados	$N_{\text{Total de grupos de Trabajadores}} - N_{Labs}$	ΣSC_{Labs}	$SC_{GP's}/gl_{GP's}$	$V_{\Sigma J}/V_{Error}$
Lab A	k-1	$\Sigma(T_{LabA})^2/r_{LabA} - (\Sigma x_{LabA})^2/r_{LabA}$	SC_{GPA}/gl_{LabA}	V_{GPA}/V_{Error}
Lab B	k-1	$\Sigma(T_{LabB})^2/r_{LabB} - (\Sigma x_{LabB})^2/r_{LabB}$	SC_{GPb}/gl_{LabB}	V_{GPB}/V_{Error}
Lab C	k-1	$\Sigma(T_{LabC})^2/r_{LabC} - (\Sigma x_{LabC})^2/r_{LabC}$	SC_{GPC}/gl_{LabC}	V_{GPC}/V_{Error}
Lab D	k-1	$\Sigma(T_{LabD})^2/r_{LabD} - (\Sigma x_{LabD})^2/r_{LabD}$	SC_{GPD}/gl_{LabD}	V_{GPD}/V_{Error}
Lab E	k-1	$\Sigma(T_{LabE})^2/r_{LabE} - (\Sigma x_{LabE})^2/r_{LabE}$	SC_{GPE}/gl_{LabE}	V_{GPE}/V_{Error}
Error	SC_T-SC_{Lab}-SC_{GP}	$SC_T - (SC_{Lab} + SC_{GP's})$	--	--
Total	$N_{Total} - 1$	$\Sigma x_i^2 - (\Sigma x_i)^2/N_T$	--	--

*: K = # de trabajadores por cada laboratorio, $r_{LabX} = 16$

Solución:

SC Totales

$SC_{Totales} = \Sigma x_i^2 - (\Sigma x_i)^2/N_T$
$\Sigma x_i = 402$
$\Sigma x_i^2 = 2{,}990$
$FC = (\Sigma x_i)^2/ N_T = (402)^2 /80 = 161{,}604/80 = 2{,}020.05$
$SC_{Totales} = 2{,}990 - 2{,}020.05 = 969.95$

SC de los 5 laboratorios
$SC_{Labs} = \Sigma(Tcg)^2/r_{CG} - (\Sigma x_i)^2/N_T$
$\Sigma(Tcg)^2/r_{CG} = [\{(93)^2 + (81)^2 + (82)^2 + (88)^2 + (58)^2\} / 16] = 2{,}065.125$
$FC_{Labs} = (\Sigma x_i)^2/ N_T = (402)^2 /80 = 161{,}604/80 = 2{,}020.05$
$SC_{Labs} = 2{,}065.125 - 2{,}020.05$
$SC_{Labs} = 45.075$

SC de GP del Labo A
$SC_{Lab\ A} = \Sigma(T_{LabA})^2/r_{LabA} - (\Sigma x_{LabA})^2/r_{LabA}$
$\Sigma(T_{Lab\ A})^2/r_{LabA} = [\{(16)^2 + (33)^2 + (17)^2 + (27)^2\} / 4] = 590.75$
$FC_{Lab\ A} = (\Sigma x_{Lab.A})^2/r_{Lab.A} = (93)^2/16 = 540.5625$
$SC_{Lab\ A} = 590.75-540.5625$
$SC_{Lab\ A} = 50.1875$

SC de GP del Lab B
$SC_{Lab\ B} = \Sigma(T_{LabB})^2/r_{LabB} - (\Sigma x_{LabB})^2/r_{LabB}$
$\Sigma(T_{LabB})^2/r_{LabB} = [\{(38)^2 + (14)^2 + (21)^2 + (8)^2\} / 4] = 536.25$
$FC_{LabB} = (\Sigma x_{Lab.B})^2/r_{Lab.B} = (81)^2/16 = 410.0625$
$SC_{Lab\ B} = 536.725 - 410.0625$
$SC_{Lab\ B} = 126.1875$

SC de GP del Lab C
$SC_{Lab\ C} = \Sigma(T_{LabC})^2/r_{LaC} - (\Sigma x_{LaC})^2/r_{LabC}$
$\Sigma(T_{LabC})^2/r_{Dist\ C} = [\{(34)^2 + (20)^2 + (18)^2 + (21)^2\} / 4] = 495.00$
$FC_{LabC} = (\Sigma x_{LabC})^2/r_{LabC} = (82)^2/16 = 420.25$
$SC_{LabC} = 495.00 - 420.25$
$SC_{Lab\ C} = 74.75$

SC de GP del Lab D
$SC_{Lab\ D} = \Sigma(T_{LabD})^2/r_{LabD} - (\Sigma x_{LabD})^2/r_{LabD}$
$\Sigma(T_{LabD})^2/r_{LabD} = [\{(21)^2 + (26)^2 + (22)^2 + (19)^2\} / 4] = 490.50$
$FC_{LabD} = (\Sigma x_{LabD})^2/r_{LabD} = (88)^2/16 = 484.00$
$SC_{Lab\ D} = 490.50 - 484.00$
$SC_{Lab\ D} = 6.50$

SC de GP del Lab E
$SC_{LabE} = \Sigma(T_{LabE})^2/r_{LabE} - (\Sigma x_{LabE})^2/r_{LabE}$
$\Sigma(T_{LabE})^2/r_{LabE} = [\{(21)^2 + (16)^2 + (7)^2 + (14)^2\} / 4] = 235.50$

$FC_{LabE} = (\Sigma x_{LabE})^2/r_{LabE} = (58)^2/16 = 210.25$
$SC_{LabE} = 235.50 - 210.20$
$SC_{LabE} = 25.25$

SC del Error
$SC_E = SC_{Total} - (\Sigma SC_{Labs}) = 969.95 - (45.075 + 50.1875 + 126.1875 + 74.75 + 6.5 + 25.25) = 642$

SC de GP'$s_{Combinados}$ = Suma de SC de los todos (5) laboratorios
SC de GP'$s_{Combinados}$ = 50.1875 + 126.1875 + 74.75 + 6.5 + 25.25 = 282.875

En la Tabla 3 se demuestra los datos del ANOVA para los laboratorios.

Tabla 3. Tabla anova para datos de los laboratorios.

FV	gl	SC	V	$F_{Calculada}$	$F_{Tabulada}$
Laboratorio	4	45.075	11.269	1.053	2.53
GP´s combinados	15	282.875	18.858	1.763	1.84
GP del Lab A	3	50.1875	16.729	1.563	2.76
GP del Lab B	3	126.188	42.063	3.931*	2.76
GP del Lab C	3	74.750	24.917	2.329	2.76
GP del Lab D	3	6.500	2.167	0.202	2.76
GP del Lab E	3	25.250	8.417	0.787	2.76
Error	60	642.000	10.700	--	--
Total	79	969.950	--	--	--

Conclusiones:

Los resultados del análisis de ANOVA anidada indican los siguientes hallazgos.

1. El desempeño (promedio de análisis) es igual entre los 5 laboratorios.
2. El promedio de desempeño de todos los trabajadores (combinados) es igual, es decir, no hay diferencia entre los promedios de desempeño de suma de los 20 GP´s.
3. Finalmente, para cada laboratorio individual, con la excepción del laboratorio B, los promedios de las análisis de los demás laboratorios (A, C, D y E) son iguales. En otras

palabras, solamente existe diferencia significativa entre el desempeño de los trabajadores del laboratorio B.

Por tanto, se puede utilizar la técnica de ANOVA anidada para las situaciones de la presencia de los subgrupos dentro de los grupos, con el objetivo del determinar la diferencia estadística entre los promedios tanto de sub-grupos como de los grupos. Esta técnica tiene la ventaja de ahorrar el esfuerzo, ya que en lugar de realizar varios ANOVA's, el investigador se puede llegar a conclusiones deseadas realizando solamente un análisis de varianza denominada ANOVA anidada.

Referencias

Badii, M.H., J. Castillo, J. Landeros & K. Cortez. 2009. Papel de la estadística en la investigación científica. Pp. 1-43. In: M.H. Badii & J.Castillo (eds). Desarrollo Sustentable: Métodos, Aplicaciones y Perspectivas. UANL. Monterrey.

Casella, G. & R.L. Berger. 2002. Statistical Inference, 2nd. Edition. Cengage Learning, Autralia.

Pagano, M. & K. Gauvreau. 2000. Principles of Biostatistics. 2nd. Edition. Cengage Learning. USA.

Steel, R:G:D: & J.H. Torrie. 1960. Principles and Procedures of Statistics. McGraw-Hill. N.Y.

Zar, J.H. 1996. Biostatistical Analysis. 3rd. Edition. Prentice Hall. New jersey.