



La esperanza del cuadrado medio

Expectation of Square Means



Luis F Restrepo B¹*, Estad, Esp estad bioma.

¹ Grupo de Investigación Grica, Facultad de Ciencias Agrarias, Universidad de Antioquia, AA 1226, Medellín, Colombia.
lusitano@agronica.udea.edu.co

(Recibido: 31 mayo, 2006; aceptado: 26 abril, 2007)

Resumen

La esperanza del cuadrado medio es fundamental en el análisis de la varianza, para establecer la prueba F adecuada en el contraste de factores dispuestos en el diagrama de estructura bajo un modelo de clasificación experimental elegido por el investigador. Desconocer el cuadrado medio es estar a ciegas en la interpretación de resultados, ya que el desconocimiento de dicho componente puede dar lugar a conclusiones no adecuadas, lo cual es grave en la toma de decisiones generadas a partir del análisis estadístico. Para construir el cuadrado medio se requiere conocer el tipo de efecto, si es aleatorio o fijo, o un componente mixto.

Palabras clave: *esperanza del cuadrado medio, factor, F calculado*

Summary

Expectation of the square mean is fundamental in the analysis of variance in order to establish the F test adapted in the contrast of factors arranged in the diagram of structure under a model of experimental classification chosen by the researcher. The lack of knowledge of the mean square is being blind at the time of interpretation of results, because the misunderstanding of such a component can give inappropriate conclusions which in turn get bias in statistic analysis-base decisions. To construct the mean square it is necessary to previously known if the effect is random, fixed or mixed.

Key words: *factor, F calculated, main, square.*

Introducción.

Los cuadrados medios esperados son calculados de acuerdo con el tipo de factor replicación e interacción; sirven para ver la variación entre grupos

y para establecer la variación dentro de los grupos considerados en un proceso investigativo mediante la utilización del diseño experimental. Existen tres tipos de factores: fijos, aleatorios y mixtos (1, 12, 19).

* Autor para el envío de la correspondencia y la solicitud de separatas. Facultad de Ciencias Agrarias, Universidad de Antioquia, AA 1226, Medellín, Colombia. E-mail: lusitano@agronica.udea.edu.co

Los factores fijos en un experimento tienen en cuenta todos los tratamientos de interés para el investigador, es decir actúan con la población de tratamientos (1, 2, 3). Se considera factor a cada una de las variables que intervienen en el diseño con sus respectivos niveles o dosificaciones, los cuales son cada uno de los valores o la escala que toma el factor.

Los factores aleatorios son aquellos en los que sólo se toma una parte de niveles y por lo tanto se trabaja con un subconjunto del conjunto total de tratamientos (11, 13).

Los factores mixtos se representan cuando en un experimento están presentes tanto factores fijos como aleatorios, anotando que la interacción entre un factor fijo y uno aleatorio es aleatoria.

Análisis de la varianza

Cuando las mediciones resultantes de la aplicación de una serie de tratamientos son de índole continua, discreta y en algunos casos particulares de naturaleza cualitativa y si se cumplen ciertas suposiciones, se puede emplear una metodología conocida como análisis de la varianza.

El análisis de la varianza es utilizado para dar respuesta a dos clases de planteamientos:

Clase I. Los parámetros correspondientes son varianzas y sus magnitudes absolutas y relativas son de interés; el objetivo es la detección y estimación de componentes de varianza, los cuales son atribuibles a la variación aleatoria de las características de los individuos de un tipo particular con respecto a los valores medios de estas características en la población. Los supuestos sobre los que se basa el análisis son:

1. γ_{ij} son variables aleatorias las cuales se distribuyen alrededor de la media $\mu..$ (promedio general).
2. Las variables aleatorias γ_{ij} son sumas de componentes donde estos componentes son aleatorizados (9).
3. La covarianza entre componentes son nulas (4). La covarianza indica la direccionalidad de la relación entre variables; se puede dar una relación directa

o positiva o puede ser de naturaleza inversa o negativa y resulta cuando una variable aumenta y la otra disminuye.

4. Los desvíos $(\mu_i - \mu..)$ $(\mu_j - \mu..)$ y ε_{ij} se distribuyen en forma normal. Los ε_{ij} son de naturaleza aleatorios y constituyen la cuantificación estadística de todos los factores no controlados que inciden sobre la variable respuesta.

Con los supuestos anteriores se satisfacen las inferencias sobre los componentes de varianza.

Para establecer las esperanzas de las medias de cuadrado se simbolizan, F como factor fijo y R factor aleatorio; además deberá aparecer el subíndice que caracteriza a cada factor con el respectivo número de niveles.

La hipótesis asociada a la clase I es (15):

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \dots = \sigma_k^2$$

$$H_1 : \text{Al menos un } \sigma_i^2 \text{ difiere, } i = 1, 2, \dots, k$$

H0: indica que existe homogeneidad de varianzas.

H1: indica que al menos una varianza difiere.

Clase II. Consiste en la determinación y estimación de ciertas relaciones entre las medias de subconjuntos del universo considerado. Se hace inferencia o estimación sobre la diferencia de los efectos medios de tratamientos, se exploran comparaciones entre dos o mas efectos medios. Los parámetros son medias, para elaborar inferencias es necesario tener presente los supuestos:

1. La variable dependiente es aleatoria, la cual se distribuye alrededor de los verdaderos valores medios μ_{ij} los cuales son constantes (10).
2. Los parámetros de μ_{ij} se hallan relacionados en forma lineal (5).
3. Las variables aleatorias γ_{ij} son homogéneas y mutuamente incorrelacionadas.
4. Los $\gamma_{ij} \sim NC(\mu_{ij}, \sigma^2)$. La hipótesis para efecto fijo se basa sobre efectos promedios y se establece así:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

$$H_1 : \text{al menos un } \mu_i \text{ difiere con } i = 1, 2, \dots, k$$

H0: indica que el efecto promedio de los tratamientos es igual.

H1: indica que al menos un tratamiento difiere en el efecto promedio.

También se puede simbolizar así:

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \dots = \tau_k$$

$$H_1 : \text{al menos un } \tau_i \text{ difiere, } i = 1, 2, \dots, k$$

Procedimiento para construir el cuadrado medio esperado.

1. Determine el diagrama de estructura.
2. Establezca el modelo lineal correspondiente en forma aditiva.
3. Definidos los factores en cuanto al tipo, tenga en cuenta que la interacción de un factor fijo y uno aleatorio resulta aleatoria.
4. Escriba 1 (uno) si uno de los subíndices del componente fila es igual al subíndice de la columna y si el factor es aleatorio (15, 17).
5. Escriba 0 (cero) si uno de los subíndices del componente fila coincide con el subíndice de la columna y si el factor es fijo (16).
6. Si los subíndices del componente fila no coinciden con los subíndices de la columna coloque el número de niveles correspondientes al encabezado de la columna (1).
7. Para obtener el valor esperado de la media de cuadrados de cualquier componente del modelo, primero se cubren todas las columnas encabezadas por los subíndices activos de este componente (14). A continuación en cada renglón que contenga al menos los mismos subíndices que el componente considerado, se calcula el producto de los números visibles y se multiplica por el factor fijo o aleatorio apropiado. La suma de estas cantidades corresponde al valor esperado de la media de cuadrados del componente del modelo considerado.
8. El número de fuentes de variación depende de cuántos son los factores y cómo están dispuestos, si cruzados o anidados, o si son fijos o aleatorios (17).
9. Existen varias reglas prácticas para determinar el cuadrado medio esperado, para ello se dispone de las reglas de Schultz, Bennett, Franklin, Cornfield y Tukey (6, 7, 8).

10. En los experimentos con factores fijos las conclusiones son sólo aplicables a los niveles o tratamientos incluidos (19).
11. En los experimentos con factores aleatorios las conclusiones son aplicables a todos los niveles del factor.
12. El cuadrado medio esperado tiene como componente el cuadrado medio del error experimental.
13. Los subíndices de cada término del modelo deben dividirse en tres clases: 1) activos, son aquellos que se hallan en el término y no están entre paréntesis, 2) pasivos, son aquellos subíndices que se encuentran entre paréntesis, y 3) ausentes, son los subíndices que están presentes en el modelo pero que no ocurren en ese término en particular.

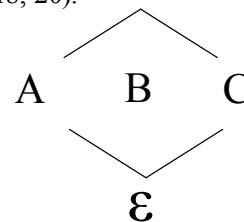
Por ejemplo:

τ_{ij} : acá i, j representan subíndices activos y k está ausente.

$\epsilon_{k(ij)}$ son subíndices pasivos y k está activo.

14. Si todos los factores son fijos, cada factor en la construcción de cuadrado medio tiene en cuenta el cuadrado medio del error experimental y la variabilidad asociada al factor con sus coeficientes correspondientes a la fila.

Para construir el cuadrado medio esperado se parte de un arreglo factorial 2^3 en un diseño completamente aleatorizado. Los factores se representan por A, B, C y son fijos, el diagrama de estructura es (18, 20).



Que corresponde al modelo lineal:

$$y_{ijk} = \mu + A_i + B_j + C_s + Ab_{ij} + AC_{is} + BC_{js} + ABC_{ijs} + \epsilon_{k(ijs)}$$

Para proceder a la construcción se elabora una tabla de la manera siguiente: en la tabla 1 se puede observar que cada factor tiene asociado un subíndice

y están presentes todas las fuentes de variación del modelo lineal aditivo, salvo la del total. Se pone en el encabezado de la tabla de doble entrada cuatro filas; en la primera, se representa el subíndice de los distintos factores; en la segunda, el tipo de factor donde F representa que es un factor fijo y A que es aleatorio; en la tercera, se coloca una simbología

internacional relacionada con cada factor; y en la cuarta, el número de niveles de los factores y las replicaciones asociadas al error experimental. Se ubica cero si coincide el subíndice del factor con el encabezado de la columna si el factor es fijo y uno si es aleatorio. Luego se procede a construir la esperanza del cuadrado medio así (véase Tabla 2).

Tabla 1. Subíndices de los factores A, B y C fijos.

	l	j	s	k
	F	F	F	A
Fuente de variación	a	b	c	r
	α	β	τ	
A_i	0	b	c	r
B_j	a	0	C	r
C_s	a	b	0	r
AB_{ij}	0	0	C	r
AC_{is}	0	b	0	r
BC_{js}	a	0	0	r
ABC_{ijs}	0	0	0	r
$\epsilon_{k(ijs)}$	1	1	1	1

Tabla 2. Construcción del cuadrado medio.

	l	j	s	k
	F	F	F	A
Fuente de variación	a	b	c	r
A_i				$\sigma^2 + bcr \sum \alpha^2 / (a-1)$
B_j				$\sigma^2 + acr \sum \beta^2 / (b-1)$
C_s				$\sigma^2 + abr \sum \tau^2 / (c-1)$
AB_{ij}				$\sigma^2 + cr \sum (\alpha\beta)^2 / (a-1)(b-1)$
AC_{is}				$\sigma^2 + br \sum (\alpha\tau)^2 / (a-1)(c-1)$
BC_{js}				$\sigma^2 + ar \sum (\beta\tau)^2 / (b-1)(c-1)$
ABC_{ijs}				$\sigma^2 + r \frac{\sum (\alpha\beta\tau)^2}{(a-1)(b-1)(c-1)}$
$\epsilon_{k(ijs)}$				σ^2

Como se puede apreciar, cada factor tiene el cuadrado medio del error experimental adicionándole los niveles de la fuente de variación, ubicados en las celdas de las distintas columnas relacionadas con el componente y esto se multiplica por la sumatoria cuadrada de los factores inmersos en la fuente de variación, denotados con la convección internacional, relacionados por los grados de libertad asociados con la fuente de variación. Ahora si todos los factores son aleatorios el cuadrado medio para el ejemplo citado será (véase Tabla 3).

Tabla 3. Valores asignados a los factores A, B y C aleatorios.

Fuente de variación	i	j	s	k
	A	A	A	A
	α	β	τ	
	a	b	c	r
A_i	1	b	c	r
B_j	a	1	c	r
C_s	a	b	1	r
AB_{ij}	1	1	c	r
AC_{is}	1	b	1	r
BC_{js}	a	1	1	r
ABC_{ijs}	1	1	1	r
$\epsilon_{k(ijs)}$	1	1	1	1

Se coloca 1 por ser aleatorio el factor, el cuadrado medio se observa en la tabla 4.

Para un diseño experimental en el que los factores son aleatorios, la esperanza del cuadrado medio de una fuente de variación se halla adicionándole a la varianza del error experimental, el componente de variabilidad del factor con sus respectivos niveles y replicas pertenecientes a la fila donde esta el factor y se le suma la variabilidad de las fuentes de variación donde esta el factor; así el cuadrado medio es acumulativo ya que contempla las interacciones donde está el factor presente.

Por ejemplo, el factor A tiene presente en la elaboración del cuadrado medio la variabilidad del error experimental, más las variabilidades de las interacciones ABC, AB y AC con sus respectivos coeficientes.

Tabla 4. Esperanza del cuadrado medio con factores aleatorios.

Fuente de variación	i	j	s	k
	A	A	A	A
	α	β	τ	
	a	b	c	r
A_i	$\sigma^2 + r\sigma^2_{\alpha\beta\tau} + br\sigma^2_{\alpha\tau} + cr\sigma^2_{\alpha\beta} + bcr\sigma^2_{\alpha}$			
B_j	$\sigma^2 + r\sigma^2_{\alpha\beta\tau} + ar\sigma^2_{\beta\tau} + cr\sigma^2_{\alpha\beta} + acr\sigma^2_{\beta}$			
C_s	$\sigma^2 + r\sigma^2_{\alpha\beta\tau} + ar\sigma^2_{\beta\tau} + br\sigma^2_{\alpha\tau} + abr\sigma^2_{\tau}$			
AB_{ij}	$\sigma^2 + r\sigma^2_{\alpha\beta\tau} + \chi\rho\sigma^2_{\alpha\beta}$			
AC_{is}	$\sigma^2 + r\sigma^2_{\alpha\beta\tau} + \beta r\sigma^2_{\alpha\tau}$			
BC_{js}	$\sigma^2 + r\sigma^2_{\alpha\beta\tau} + ar\sigma^2_{\beta\tau}$			
ABC_{ijs}	$\sigma^2 + r\sigma^2_{\alpha\beta\tau}$			
$\epsilon_{k(ijs)}$	σ^2			

Ahora suponga que A es un factor fijo y B, C aleatorios (véase Tabla 5).

Tabla 5. Valores asociados con los factores mixtos.

Fuente de variación	i	j	S	K
	F	A	A	A
	α	β	τ	
	a	b	C	r
A_i	0	b	c	r
B_j	a	1	c	r
C_s	a	b	1	r
AB_{ij}	0	1	c	r
AC_{is}	0	b	1	r
BC_{js}	a	1	1	r
ABC_{ijs}	0	1	1	r
$\epsilon_{k(ijs)}$	1	1	1	1

El cuadrado medio es (véase Tabla 6).

La fuente de variación BC no acumula todo el componente de la interacción ABC ya que BC es aleatorio, pero debido a que tanto B como C son factores aleatorios, la interacción ABC a pesar de que es aleatoria posee un factor que es fijo A y por este motivo no acumula.

Tabla 6. Esperanza del cuadrado medio para A factor fijo, B y C aleatorios.

	i	j	S	k
Fuente de variación	F	A	A	A
	α	β	τ	
	a	b	C	r
A_i	$\sigma^2 + r\sigma^2_{\alpha\beta\tau} + rb\sigma^2_{\alpha\tau} + rc\sigma^2_{\alpha\beta} + rbc \frac{\sum \sigma^2}{a-1}$			
B_j	$\sigma^2 + r\sigma^2_{\beta\tau} + rac\sigma^2_{\beta}$			
C_s	$\sigma^2 + r\sigma^2_{\beta\tau} + rab\sigma^2_{\tau}$			
AB_{ij}	$\sigma^2 + r\sigma^2_{\alpha\beta\tau} + rc\sigma^2_{\alpha\beta}$			
AC_{is}	$\sigma^2 + r\sigma^2_{\alpha\beta\tau} + rb\sigma^2_{\alpha\tau}$			
BC_{js}	$\sigma^2 + r\sigma^2_{\beta\tau}$			
ABC_{ijs}	$\sigma^2 + r\sigma^2_{\alpha\beta\tau}$			
$\epsilon_{k(ijs)}$	σ^2			

Si A y B son fijos y C un factor aleatorio queda (véase Tabla 7).

Tabla 7. Valores asociados con los subíndices donde A y B son fijos, C aleatorio.

	i	j	s	K
Fuente de variación	F	F	A	A
	α	β	τ	
	a	b	c	r
A_i	0	b	c	r
B_j	a	0	c	r
C_s	a	b	1	r
AB_{ij}	0	0	c	r
AC_{is}	0	b	1	r
BC_{js}	a	0	1	r
ABC_{ijs}	0	0	1	r
$\epsilon_{k(ijs)}$	1	1	1	1

El cuadrado medio es (véase Tabla 8).

En el diseño estadístico mixto el factor aleatorio sólo acumula componentes aleatorios puros; es decir, donde todos los factores de la interacción sean aleatorios y tengan relación con el factor.

El factor fijo acumula las interacciones donde está el factor pero no existen factores fijos en la interacción.

El cuadrado medio para un factorial anidado con A y B cruzados y C anidado dentro de B, además A fijo, B y C aleatorios.

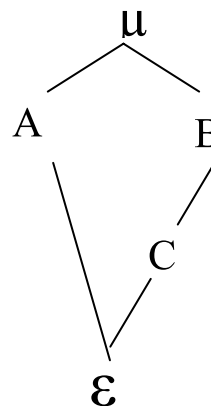


Tabla 8. Esperanza del cuadrado medio con A, B factores fijos y C factor aleatorio.

	i	j	s	k
Fuente de variación	F	F	A	A
	α	β	τ	
	a	b	c	r
A_i	$\sigma^2 + r\beta\sigma^2_{\alpha\tau} + rbc \frac{\sum \sigma^2}{a-1}$			
B_j	$\sigma^2 + r\alpha\sigma^2_{\beta\tau} + rac \sum \beta^2 / (b-1)$			
C_s	$\sigma^2 + rab\sigma^2_{\tau}$			
AB_{ij}	$\sigma^2 + r\sigma^2_{\alpha\beta\tau} + rc \frac{\sum (\alpha\beta)^2}{(a-1)(b-1)}$			
AC_{is}	$\sigma^2 + r\beta\sigma^2_{\alpha\tau}$			
BC_{js}	$\sigma^2 + r\alpha\sigma^2_{\beta\tau}$			
ABC_{ijs}	$\sigma^2 + r\sigma^2_{\alpha\beta\tau}$			
$\epsilon_{k(ijs)}$	σ^2			

Tabla 9. Esperanza del cuadrado medio con componentes anidados.

	i	j	s	k
Fuente de variación	F	F	A	A
	α	β	τ	
	a	b	c	r
A	$\sigma^2 + r\sigma^2_{ac(b)} + cr\sigma^2_{ab} + bcr \frac{\sum \sigma^2}{(a-1)}$			
B	$\sigma^2 + r\sigma^2_{ac(b)} + ar\sigma^2_{c(b)} + cr\sigma^2_{ab} + acr\sigma^2_b$			
AB	$\sigma^2 + r\sigma^2_{ac(b)} + cr\sigma^2_{ab}$			
C/B	$\sigma^2 + r\sigma^2_{ac(b)} + ar\sigma^2_{c(b)}$			
AC/B	$\sigma^2 + r\sigma^2_{ac(b)}$			
$\epsilon_{k(ijs)}$	σ^2			

Si A y B son fijos y C aleatorios se tiene (véase Tabla 10).

Tabla 10. Esperanza del cuadrado medio, con factores mixtos y presencia de anidamiento.

Fuente de variación	i	j	s	k
	F	F	A	A
	α	β	τ	
	a	b	c	r
A	$\sigma^2 + r\sigma^2_{ac(b)} + bcr \frac{\sum \sigma^2}{(a-1)}$			
B	$\sigma^2 + r\sigma^2_{ac(b)} + ar\sigma^2_{c(b)} + acr \sum \beta^2 / (b-1)$			
AB	$\sigma^2 + r\sigma^2_{ac(b)} + cr \sum (\alpha\beta)^2 / (a-1)(b-1)$			
C/B	$\sigma^2 + r\sigma^2_{\alpha c(\beta)} + \alpha r\sigma^2_{c(\beta)}$			
AC/B	$\sigma^2 + r\sigma^2_{\alpha c(\beta)}$			
$\epsilon_{k(ijs)}$	σ^2			

Conclusión

Al efectuar un análisis de la varianza se debe tener presente el tipo de factor o factores involucrados en el diseño de clasificación

experimental, a fin de poder generar la esperanza del cuadrado medio adecuada y así llegar a conclusiones coherentes en el análisis de la información.

Referencias

- Calzada BJ. Métodos estadísticos para la investigación. 1. ed. Universidad de la Molina. Lima. 1970. 643p.
- Cochran WG, Cox GM. Experimental designs. 2. ed. New York, John Wiley, 1957. 611p.
- Cochran WG, Cox GM. Diseños experimentales. Trillas. México D. F. 1981.611p.
- Cordeiro GM. Modelos lineares generalizados. Unicamp, Campinas.1986. 286p.
- Dobson F. An introduction to linear models, Chadpman-Hall. Second edition New York.1991. 221p.
- Federer WT. Experimental design. New York, the MacMillan Company, 1953.554p.
- Gaylor DW, Hopper FN. Estimating the degrees of freedom for linear combination of mean squares by satherthwaite's formula, Technometrics. 1969; 4:691-706.
- Hinkelmann K, Kempthorne O. Design and analysis of experiments. Vol 1, John Wiley, New York.1994.512p.
- Kempthorne O, Folks L. Probability statistics and data analysis. Ames, Iowa State University Press. Iowa. 1961. 555p.
- Lemma AF. Hipoteses estadísticas con muestras desequilibradas, Fac Sci. Agrom. Gembloux, Bruselas.1991.64p.
- Little TM, Hills FJ. Métodos estadísticos aplicados en agricultura, trillas México D. F. 1976. 270p.
- Martinez GA. Diseños experimentales, Colegio de Postgrado de Chapingo. México D. F. 1983. 1058p.
- McCullag P, Nelder J. Generalized linear models. Prentice Hall, London,1989. 511p.
- Montgomery DC. Design and analysis of experiments. New York, John Wiley, 1981. 418p.

15. Ostle B. Estadística aplicada. Limusa-Wiley S.A. México. D. F.1973. 629 p.
16. Senedecor GW. Métodos Estadísticos aplicados a la investigación agrícola y biológica. Continental S.A. México D. F. 1970. 503p.
17. Sokal R, Rohlf J. Introducción a la bioestadística. Reverte S.A. New Cork.1980. 363p.
18. Throckmorton TN Structures of clasification data, unpublished Ph D, dissertation. State University. Dept Statistical. Iowa.1961; 32:927-950.
19. Torrie JH. Bioestadística. Principios y procedimientos. M.cGraw-Hill. México D. F. 1981.622p.
20. Warren H, Taylor Jr, H.G Hilton. Astructure diagram symbolization for analysis of variance.1981;35:85-93.