Covarianza entre promedios

R.J.C. Cantet

Departamento de Producción Animal, Fac. Agr. UBA

1. **Introducción**

En la predicción del valor de cría de un individuo intervienen, generalmente, varias fuentes de información, como por ejemplo, su dato personal, o los promedios de sus hijos, o el promedio de sus nietos, o el de sus medio hermanos paternos (MHP), etc. Como las distintas fuentes se hallan emparentadas generarán covarianzas genéticas, que se traducirán en covarianzas fenotípicas. El propósito de esta nota es revisar la metodología de derivación de las covarianzas entre dichos promedios para modelos que posean o no efectos maternos. Un problema que se produce para obtener las covarianzas entre promedios es la notación de subíndices dentro del modelo. Adoptaremos como enfoque general utilizar un subíndice más de lo necesario que indexará el individuo al cual pertenece el promedio y no los individuos cuyos datos constituyen el promedio. Así, con efectos directos unicamente necesitaremos 2 subíndices: j = individuo al que pertenece el dato, e i = individuo cuyo promedio me interesa calcular. Con efectos maternos necesitaremos 3 subíndices: k = individuo al que pertenece el dato, i = individuo cuyo promedio me interesa calcular y j = madre de i, posiblemente relacionada con k.

2. **Covarianzas sin efectos maternos**

Considre el modelo:

$$P_{ij} = \mu_{ij} + a_{ij} + e_{ij}$$

$$a_{ij} \sim (0,\sigma^2_A) \quad e_{ij} \sim (0,\sigma^2_e) \quad cov(a_{ij},e_{ij}) = 0 \quad P_{ij} \sim (0,\sigma^2)$$

No haremos en este punto ninguna suposición sobre si j es hijo, nieto o sobrino, etc, de i salvo que la relación aditiva entre i y j es A_{ij}. Sean $P_{i.}$ y $P_{i'.}$ iguales a:

$$P_{i.} = \frac{\sum_{j=1}^{m} P_{ij}}{m} \qquad P_{i'.} = \frac{\sum_{j'=1}^{n} P_{i'j'}}{n}$$

Estos promedios pueden ser de cualquier tipo de individuos. Solo basta especificar que la relación entre i e i' es $A_{ii'}$. Como ejemplo, suponga que i es el padre de j; i' es un hijo de i y j' es un hijo de i'. Consecuentemente $P_{i.}$ es el promedio de los hijos de i y $P_{i'.}$ el promedio de un grupo de nietos de i: aquellos con padre i'. Nuestro objetivo es calcular cov($P_{i.}$, $P_{i'.}$). Procederemos así:

$$cov(P_{i.},P_{i'.}) = cov\left[\frac{\sum_{j=1}^{m} P_{ij}}{m}, \frac{\sum_{j'=1}^{n} P_{i'j'}}{n}\right]$$

$$= \left(\frac{1}{mn}\right) cov\left[\sum_{j=1}^{m} P_{ij}, \sum_{j'=1}^{n} P_{i'j'}\right]$$

$$= \left(\frac{1}{mn}\right) cov\left[\sum_{j=1}^{m} (a_{ij} + e_{ij}), \sum_{j'=1}^{n} (a_{i'j'} + e_{i'j'})\right]$$

En este punto es conveniente notar, que los residuales de distintos individuos son independientes, e independientes del valor de cría, con lo cual la expresión anterior es:

$$= \left(\frac{1}{mn}\right) \text{cov}\left[\sum_{j=1}^{m} a_{ij}, \sum_{j'=1}^{n} a_{i'j'}\right] \qquad [1]$$

Ahora bien, los a_{ij} son los valores de cría de los individuos j = 1, 2..m, todos relacionados del mismo modo (igual A_{ij}) con i. Otro tanto ocurre con los j'. Es entonces que cualquier covarianza entre los valores de cría de los distintos animales en $P_{i.}$ y $P_{i'.}$ son todas iguales comportándose como "constantes" al sumar sobre j o j': las variables aleatorias son distintas pero las covarianzas son iguales. Además:

$$\text{cov}\,(a_{ij}, a_{i'j'}) = a_{ij,i'j'}\,\sigma_A^2 \qquad [2]$$

Por lo tanto usando [2] en [1] obtenemos finalmente:

$$\left(\frac{1}{mn}\right) \text{cov}(\,m\,a_{ij}, n\,a_{i'j'}\,) = \left(\frac{mn}{mn}\right) \text{cov}(a_{ij}, a_{i'j'}) = A_{ij,i'j'}\,\sigma_A^2 \qquad [3]$$

El resultado indica que las covarianzas entre promedios **no dependen del número de datos en dichos promedios: solo de las relaciones aditivas entre los individuos constituyentes en el promedio ($A_{ij,i'j'}$).** Por ejemplo, si j <- i <- B -> C -> i'-> j', entonces $A_{ij,i'j'}$ = 1/32, y cov($P_{i.}$, $P_{i'.}$) = (1/32) σ_A^2. Es importante notar que, en el caso de que i -> i' (i sea el padre o la madre de i') el razonamiento anterior es correcto siempre y cuando el dato de i'($P_{ii'}$) no esté incluido dentro de $P_{i.}$. En caso contrario:

$$\left(\frac{1}{mn}\right) \left[\ \text{cov}(a_{ii'}\sum_{j'=1}^{n} a_{i'j'}) + \text{cov}\,(\sum_{j=1}^{m-1} a_{ij}, \sum_{j'=1}^{n} a_{i'j'})\ \right] =$$

$$\left(\frac{1}{mn}\right) [\ n\ \text{cov}(a_{ii'}, a_{i'j'}) + (m-1)n\ \text{cov}\,(\,a_{ij}, a_{i'j'})\ ] =$$

$$\frac{[\, n\, A_{ii',i'j'} \; + \; (m-1) n \; A_{ij,i'j'}]\sigma^2_A}{m\, n} \qquad [4]$$

Note que si el dato del individuo j' no está en el promedio $P_{i.}$, entonces todas las relaciones son iguales a $A_{ij,i'j'}$; en particular $A_{ii',i'j'}$ = $A_{ij,i'j'}$ y [3] se hace igual a [4].

3. **Covarianzas con efectos maternos**

Por simplicidad solo consideraremos el caso en el cual ningún individuo está involucrado en ambos promedios: en uno como dato y en el otro como madre j o como individuo k, cuyo promedio nos interesa. El modelo con efectos maternos para el individuo k con madre j, asociado al individuo i es:

$$P_{ijk} = \mu_{ijk} + a_{oik} + a_{mjk} + e_{mjk} + e_{oijk}$$

donde P_{ijk} es el fenotipo del individuo i con madre j; a_o, a_m, e_o y e_m son los valores de cria y desvíos ambientales, directos y maternos, respectivamente. Para simplificar la notación emplearemos aqulla que venimos utilizando en el curso, o sea: ik = X; ij = W; i'k' = Y e i'j' = Z. Entonces la covarianza entre dos promedios en las que aparecen efectos maternos es:

$$\operatorname{cov}(P_{i.}, P_{i'.}) = \operatorname{cov}\left[\frac{\sum_{k=1}^{m} P_{ijk}}{m}, \frac{\sum_{k'=1}^{n} P_{i'j'k'}}{n} \right]$$

$$= \left(\frac{1}{mn} \right) \operatorname{cov}\left[\sum_{k=1}^{m} (a_{oX} + a_{mW} + e_{mW}), \sum_{k'=1}^{n} (a_{oY} + a_{mZ} + e_{mZ}) \right]$$

$$= \left(\frac{mn}{mn}\right) [cov(a_{oX}, a_{oY}) + (cov(a_{oX},a_{mZ}) + cov(a_{mW},a_{oY}))+$$

$$+ cov(a_{mW},a_{mZ}) + \delta_{WZ}\sigma^2_{Em}]$$

Por lo que, y finalmente, llegamos a:

$$cov(P_{i.},P_{i'.}) = A_{XY}\ \sigma^2_{Ao} + [A_{XZ} + A_{YW}]\sigma_{AoAm} + A_{WZ}\sigma^2_{Am} + \delta_{WZ}\sigma^2_{Em} \qquad [5]$$

Note que esta expresión es general como para comprender los casos de transplante embrionario. Por ejemplo, sea P1 es el promedio de n1 hijos machos del toro α, cada uno con distinta madre; y sea P2 el promedio de n2 individuos criados en el vientre de las hijas hembras de α, las cuales actuan como receptoras (además cada madre tiene un solo ternero). El animal Y es hijo biológico de S y D, los cuales no están emparentados con α. Entonces:

W -> X <- α -> Z -r-> Y.

La expresión -r-> indica que el aporte es como "receptora" implicando que, al pasar por un -r->, la relación es 0. Por ejemplo, $A_{XY} = 0$. Además $A_{XZ} = \frac{1}{4}$; $A_{YW} = 0 = A_{WZ}$ y $\delta_{WZ} = 0$, porque son dos madres distintas. Finalmente usando dichas relaciones en [5], tenemos que cov(P1, P2) = $\frac{1}{4}\ \sigma_{AoAm}$.

Otro ejemplo: suponga que P1 es el promedio de n1 hijos del toro α, cada uno con distinta madre; y P2 el promedio de n2 nietos maternos (de las hijas hembras) de α, ninguna de las cuales posee datos en P1 (además cada madre tiene un solo ternero). Entonces:

W -> X <- α -> Z -> Y.

Entonces, $A_{XY} = 1/8$; $A_{XZ} = \frac{1}{4}$; $A_{YW} = 0 = A_{WZ}$ y $\delta_{WZ} = 0$, con lo que cov(P1, P2) = (1/8) σ^2_{Ao} + (1/4) σ_{AoAm}.