Rev. Fitotec. Mex. 18: 77-84, 1995

PRECISION Y SESGO DE ESTIMADORES DE DIFERENCIAS DE MEDIAS DE SINTETICOS COMPLEMENTARIOS MEDIANOS

PRECISION AND BIAS ESTIMATORS OF MEANS DIFFERENCES OF MEDIAN COMPLEMENTARY SYNTHETICS

Jaime Sahagún Castellanos[1]

RESUMEN

El método de estimación del valor de materiales genéticos como progenitores de un sintético, consistente en valorar su desempeño en sintéticos complementarios medianos (SCM) mediante la evaluación de una familia proveniente de cada material, permite la inclusión de números (N) muy elevados de materiales candidatos a formar sintéticos. Con el fin de contribuir a la caracterización del método de sintéticos complementarios medianos (MSCM), en este estudio se deriva un estimador insesgado de la diferencia de las medias de dos SCM y se compara la precisión de esta estimación con la obtenida cuando se aplica el método de Wright y el de Márquez-Sánchez. Se encontró que, con base en la varianza del estimador, el método más preciso es el de Márquez-Sánchez y el menos preciso es MSCM. Sin embargo, cuando N es muy grande (N > 16) desde un punto de vista práctico sólo el método MSCM puede ser razonablemente aplicado, debido a su menor número de tratamientos. Por otra parte, cuando se asigna un número igual de parcelas experimentales a los tres métodos, el estimador de Wright es el de la menor varianza.

PALABRAS CLAVE ADICIONALES

Comparación de medias, predicción, sintéticos complementarios, sintéticos complementarios medianos.

SUMMARY

The method of evaluation of the parental value of genetic materials consisting in the assessment of their performance in median complementary synthetics (MCS), which is accomplished by testing a family from each material, allows the evaluation of a high number (N) of candidates to build up synthetic varieties. In order to contribute to the characterization of the method of median complementary synthetics (MMCS), an unbiased estimator of the difference between two MSC means is developed; its precision is compared against the precision obtained with estimators from the methods proposed by Wright (MW) and Márquez-Sánchez (MMS). Based on the variance of the estimators, it was found that MMS was the most precise method and the MW was intermediate. However, when N is larger than 16, only the MSCM can be reasonably applied because of its lower number of entries. When the comparison of the precision was made under the same number of experimental units for each method, the variance for the MW was the lowest.

ADDITIONAL INDEX WORDS

Comparison of means, prediction, complementary synthetics, median complementary synthetics.

INTRODUCCION

Considerando que la probabilidad de desarrollar una variedad sintética superior aumenta en la medida en que puedan ser evaluados más progenitores potenciales, Sahagún (1994a) propuso el método de evaluación de N progenitores por su comportamiento en sintéticos complementarios (complementarios en el sentido de que uno está formado por n progenitores y el otro por los N-n progenitores restantes). El método consiste en cruzar cada progenitor potencial, independientemente de su grado de heterocigosis, con la población constituida por los N progenitores en consideración. Posterior-

[1] Departamento de Fitotecnia. Universidad Autónoma Chapingo. C.P. 56230. Chapingo, Méx. México.

Fecha de recepción: 13 de enero de 1995

mente, las N cruzas así formadas se evaluan en experimentos de campo. Siendo N par, la diferencia entre la media de las cruzas de N/2 progenitores cualquiera y la de los N/2 restantes es un estimador de la misma diferencia entre las medias de dos sintéticos complementarios que, por estar constituido por la mitad de los progenitores se denominan medianos (Sahagún, 1994a). Este estimador puede ser la base para definir, mediante una prueba de hipótesis, cual es de los dos el mejor.

Por lo que respecta a ventajas, este método permite la evaluación de una cantidad considerable de progenitores potenciales; así, mientras que con este método el experimento de evaluación sólo incluye N materiales, con el método de Wright se requiere de la evaluación de N(N+1)/2. El método de los sintéticos complementarios, sin embargo, es poco conocido y se requiere de estudios que permitan conocer mejor sus cualidades.

De los sintéticos complementarios, los denominados medianos fueron sugeridos como los de mayor interés porque cuando N es un número par se hace posible comparar, con una misma base, a las medias de los sintéticos complementarios en consideración (Sahagún, 1994a). Sin embargo, en el trabajo anterior (Sahagún, 1994a) no se determinaron las particularidades de las ventajas estadísticas que en términos de sesgo se tienen con este tipo de sintéticos complementarios. Con respecto a la diferencia entre las medias de estos sintéticos, Sahagún (1994a) encontró que su precisión es más reducida que la que se obtiene cuando la estimación se hace con el método de Wright (Wright, 1922) y que la precisión de éste es, a su vez, menor que la que se obtiene cuando se hace con el método de Márquez-Sánchez (1992). Sin embargo, debe tomarse en consideración el hecho de que con los sintéticos complementarios la demanda de unidades experimentales es bastante menor. Por otra parte, aunque el método de los sintéticos complementarios medianos (MSCM) produce una estimación de la diferencia de sus medias, el fitomejorador también requiere determinar si esa diferencia es, o no, estadísticamente significativa, para así tomar decisiones con bases probabilísticas.

De cara a la problemática anterior, este estudio fue emprendido con el fin de: (1) Estudiar las propiedades del método de sintéticos complementarios (MSC) como estimador de la diferencia de las medias de rendimiento de cada pareja de este tipo de sintéticos; (2) Comparar la precisión, en igualdad de unidades experimentales, de la estimación de esta diferencia que produce el MSC con la que se obtiene cuando la estimación se hace con la fórmula de Wright (1922) y con la fórmula de Márquez-Sánchez (1992), y (3) Describir una metodología para comparar estadísticamente las medias de dos sintéticos complementarios medianos cuando la diferencia de sus medias de rendimiento se estima con este enfoque metodológico.

MARCO TEORICO Y RESULTADOS

La fórmula de Wright establece que el rendimiento predicho de una variedad sintética (Y_{2W}) se puede obtener según la expresión

$$Y_{2W} = Y_{1B} - (Y_{1B} - Y_0)/n \qquad (1)$$

en donde Y_{1B} es el promedio experimental de los cruzamientos entre n progenitores y Y_0 es el promedio experimental de éstos. Por su parte, la fórmula que Márquez-Sánchez (1992) derivó para predecir el rendimiento de una variedad sintética (Y_{2L}) es

$$Y_{2L} = Y_{1B} - [(Y_{1B}-Y_0)/n]-[(Y_0-S_1)/mn] \qquad (2)$$

en donde los términos nuevos son m, el número de plantas por cada uno de los n progenitores, y S_1, el promedio de los valores experimentales de las n familias obtenidas cuando cada progenitor se autofecunda.

Representando el valor genotípico del p ésimo individuo (p = 1,2,...,m) del progenitor i (i = 1,2,...,n) como Y_{pi1pi2}, de acuerdo con el arreglo genotípico del sintético, que se expresa como (Sahagún, 1994b)

$$\frac{1}{(2mn)^2} \sum_p \sum_q \sum_i \sum_j \sum_k \sum_l A_{pik} A_{qjl},$$

la media esperada de rendimiento del sintético se puede expresar en la forma

$$Y_2 = \frac{1}{(2mn)^2} \sum_p^m \sum_q^m \sum_i^n \sum_j^n \sum_k^2 \sum_l^2 Y_{pikqjl}. \quad (3)$$

Para enfatizar que el número de progenitores potenciales es grande y mayor que n, Sahagún (1994a) representó con N a este número y, por analogía con (3), generó la expresión

$$Y_2 (2Nm)^2 = \sum_p^m \sum_q^m \sum_I^N \sum_J^N \sum_k^2 \sum_l^2 Y_{pIkqJl}$$

$$= \sum_p^m \sum_q^m \sum_i^n \sum_j^n \sum_k^2 \sum_l^2 Y_{pikqjl} \; +$$

$$\sum_p^m \sum_q^m \sum_i^n \sum_{j'} \sum_k^2 \sum_l^2 Y_{pikqj'l} \cdot \; +$$

$$\sum_p^m \sum_q^m \sum_{i'} \sum_j^n \sum_k^2 \sum_l^2 Y_{pi'kqjl} \; +$$

$$\sum_p^m \sum_q^m \sum_{i'} \sum_{j'} \sum_k^2 \sum_l^2 Y_{pi'kqj'l}. \quad (4)$$

en donde $_{i', j' = n+1, n+2, \ldots, N}$. De los cuatro términos de (4), los dos primeros representan el total (S_{iJ}) de los valores genotípicos de las cruzas de cada uno de los progenitores $_{1, 2, \ldots, n}$ con los N progenitores potenciales. La suma de los dos restantes términos de (4) se refieren al total ($S_{i' \cdot J}$) de las cruzas de los progenitores $_{n+1, n+2, \ldots, N}$ con los N progenitores potenciales. Así, en ausencia de efectos maternos,

$$S_{iJ} - S_{i' \cdot J} = \sum_p^m \sum_q^m \sum_i^n \sum_j^n \sum_k^2 \sum_l^2 Y_{pikqjl} \; -$$

$$\sum_p^m \sum_q^m \sum_{i'} \sum_{j'} \sum_k^2 \sum_l^2 Y_{pi'kqj'l}. \quad (5)$$

Propiedades de la estimación de la diferencia de dos sintéticos complementarios medianos

Para determinar exactamente al estimador insesgado de la diferencia entre dos sintéticos complementarios medianos (SCM), a continuación se deriva la esperanza matemática (E) de tal diferencia, correspondiente a la diferencia de totales en (5). En esta derivación los subindices i, i', I, j, j', J se utilizan para identificar al sintético o a la población de que se trate. Se considera al valor experimental como una variable aleatoria ($\hat{S}$) formada por una componente de origen genético (S) y una de origen ambiental (E). Esta última, como es usual, se considera como una variable aleatoria normal e independientemente distribuida con media cero y varianza σ_ε^2. Por ejemplo, con $\hat{S}_{iJ}$ y $\hat{S}_{ij}$ se hará referencia a dos valores experimentales cuyos valores esperados son, respectivamente: $E(\hat{S}_{iJ}) = E(S_{iJ} + E_{ij}) = S_{iJ}$ y $E(\hat{S}_{ij}) = E(S_{ij}) + E_{ij}) = S_{ij}$. S_{iJ} fue definido en (4) y S_{ij} es el total de los valores genotípicos del sintético formado por las n líneas i (i = 1, 2, ..., n). En cada caso, la derivación de un valor esperado se hace de manera que aparezca explícitamente, en su caso, el sintético ahí contenido. Así,

$$E[\hat{S}_{ij}] = \frac{(2nm)^2}{(2m)^2Nn}(\hat{S}_{ij}) + [1 - \frac{(2nm)^2}{(2m)^2Nn}E(\hat{S}_{ij'})$$

$$= \frac{n}{N}Y_{2,n} + (1 - \frac{n}{N})Y_{2,n(N-n)} \qquad (6)$$

en donde $Y_{2,n}$ es el rendimiento esperado del sintético que formarían los n progenitores y $Y_{2,n(N-n)}$ es el rendimiento esperado de las cruzas de cada uno de estos n progenitores con los progenitores restantes. Con respecto a la media del grupo de familias formadas cruzando a cada uno de los restantes N-n progenitores con la población del total de progenitores, su valor promedio es

$$E[\hat{S}_{i'j}] = [1 - \frac{(2m(N-m))^2}{(2m)^2N(N-n)}]E(\hat{S}_{i'j}) + \frac{[2m(N-n)]^2}{(2m)^2N(n-n)}E(\hat{S}_{i'j'})$$

$$= (1 - \frac{N-n}{N})Y_{2,(N-n)n} + \frac{N-n}{N}Y_{2,N-n} \qquad (7)$$

en donde, $Y_{2,N-n}$ es el rendimiento esperado del sintético complementario del sintético cuyo rendimiento esperado es $Y_{2,n}$ y $Y_{2,(N-n)n}$ es el rendimiento esperado de las cruzas de estos N-n progenitores con los n progenitores restantes. Así, con base en (6) y (7), y en ausencia de efectos maternos, la diferencia de medias es

$$E[\hat{S}_{ij}] - E[\hat{S}_{i'j}] = \frac{n}{N}Y_{2,n} - [1 - \frac{N-n}{N}]Y_{2,N-n} + [\frac{N-2n}{N}]Y_{2,N-n}. \qquad (8)$$

Para el caso particular en que N = 2n; es decir, para el caso de sintéticos complementarios medianos (SCM):

$$2[E(\hat{S}_{ij}) - E(\hat{S}_{i'j})] = Y_{2n} - Y_{2,N-n}. \qquad (9)$$

Representando como T_i y $T_{i'}$ al rendimiento correspondiente a la cruza del progenitor i (i = 1, 2, ..., n) y a la del i ' = (i ' = n+1, n+2, ..., N) con los N progenitores, respectivamente, entonces

$$E\{2[\sum_i T_i/n - \sum_{i'} T_{i'}/n]\} = Y_{2,n} - Y_{2,N-n}.$$

En este contexto, el concepto de estimador insesgado (Mood *et al.*, 1974), es muy útil. Ellos establecen que un estimador $\hat{h}(\theta_1, \theta_2, ..., \theta_u)$ de una función paramétrica $h(\theta_1, \theta_2, ..., \theta_u)$ es un estimador insesgado si $E[\hat{h}(\theta_1, \theta_2, ..., \theta_u)] = h(\theta_1, \theta_2, ..., \theta_u)$. Así, el doble de la diferencia entre las medias de las n progenies tipo i y de las N-n tipo i', cuando N = 2n es un estimador insesgado de la diferencia de las medias de dos SCM. Aqui cada T_i y $T_{i'}$ constituye una unidad de evaluación; es decir, el experimento para evaluar los materiales correspondientes al método de los sintéticos complementarios medianos (MSCM) tendría N = 2n tratamientos. Así, la varianza del estimador de la diferencia de medias de dos SCM estimada por MSCM [V(SCM)], para una varianza σ^2 del error experimental y una repetición, es

$$V[SCM] = V[2(\sum_i T_i - \sum_{i'} T_{i'})/n]$$

$$= 4[(n+n)\sigma^2]/n^2$$

$$= \frac{4^2\sigma^2}{N}. \qquad (10)$$

Esta (10) es la varianza calculada con base en una parcela. Si se consideran medias de r repeticiones se tendría que dividir entre r.

$$V[SCM] = \frac{4^2\sigma^2}{Nr}.$$

Comparación de dos Medias de SCM

Para comparar las medias de dos sintéticos complementarios medianos (con medias

V_1 y V_2) por medio de una prueba de t de Student de dos colas, si la evaluación de las N familias se hace en un experimento con r repeticiones, el cuadrado medio del error experimental se representa por CME, y la t calculada (tc) se obtiene como sigue

$$tc = \frac{2\,[\sum_i T_i - \sum_{i'} T_{i'}]/n}{[\frac{16CME}{rN}]^{0.5}},$$

y las medias de los sintéticos serán diferentes con un nivel α de significancia si (Tc) $\geq$ Tt en donde el valor Tt es el valor tabulado de la distribución t de Student con los grados de libertad que tiene el error y que delimita hacia su derecha una cola cuya área es $\alpha/2$.

Comparación de la Precisión

Para comparar la precisión del MSCM en la estimación de una diferencia de medias entre SCM con respecto a las fórmulas (1) y (2), se emplea a la varianza de la diferencia entre las medias experimentales de cada uno de los tres métodos. Para el caso de la fórmula (1), si Y_{2w} y Y'_{2w} son las medias de dos SCM de tamaño N/2 entonces en virtud de que Y_{2w} y Y'_{2w} no tienen tratamiento alguno en común pero tienen la misma estructura, entonces

$$\begin{aligned} V[Y_{2w} - Y'_{2w}] &= 2V[Y_{2w}] \\ &= 2V[\frac{(N/2)-1}{N/2} Y_1 + \frac{1}{N/2} Y_0] \\ &= 2\,[\frac{(4)\,[(N/2)-1]}{N^2} \cdot \frac{(2)\,(2)}{N} \sigma^2 + \frac{4}{N^2}\,\frac{2}{N}\,\sigma^2] \\ &= \frac{16(N-1)}{N^3}\,\sigma^2 . \end{aligned} \qquad (11)$$

Si para estimar la diferencia entre las medias de dos SCM se utiliza la fórmula (2) y siendo Y_{2L} y Y'_{2L} las medias experimentales de esos SCM, siguiendo la estrategia anterior se obtiene

$$\begin{aligned} V\,[Y_{2L} - Y'_{2L}] &= 2V\,[Y_{2L}] \\ &= 2V\,[\frac{(N/2)-1}{N/2} Y_1 + \frac{2}{mN} S_1 + \frac{2(m-1)}{mN} Y_0] \\ &= 2\,[\frac{(4)(2)[(N/2)-1]}{N^2\,.\,N} + \frac{4}{(mN)^2} \cdot \frac{2}{N} + \frac{(4(m-1)^2}{(mN)^2} \cdot \frac{2}{N}]\sigma^2 \\ &= [\frac{16(N-1)}{N^3} - \frac{32(m-1)}{m^2N^3}]\sigma^2 . \end{aligned} \qquad (12)$$

DISCUSION

De acuerdo con los resultados (8) y (9), se deduce que el MSCM produce un estimador insesgado de la diferencia de dos SC únicamente cuando N es un número par y n=N/2. En el caso más general, con $2[\sum_i T_i/n - \sum_{i'} T_{i'}/(N-n)]$ se tendrá un estimador cuya varianza es

$$\begin{aligned} V\,[2(\sum_i T_i/n - \sum_{i'} T_{i'}/(N-n))] &= 4[\frac{1}{n} + \frac{1}{N-n}]\sigma^2 \\ &= \frac{4N}{n(N-n)}\,\sigma^2 . \end{aligned} \qquad (13)$$

Esta varianza (13) alcanza su mínimo cuando n = N/2. Por supuesto, aqui N debe ser un número par.

Así, aunque las expresiones (4) y (5) sugieren que se puede estimar la diferencia de las medias de dos SC cualesquiera, la estimación es insesgada y de mínima varianza si los dos sintéticos son complementarios medianos. En cambio, con la diferencia de medias experimentales $\Sigma T_i/n$ - $\Sigma T_{i'}/n$ propuesta previamente por Sahagún (1994a) se estima la semidiferencia de las medias de dos SCM. Por ello, la varianza que se reporta en ese estudio es diferente que la que aquí se muestra en (10).

Con respecto a la estimación de la diferencia de dos SCM con las fórmulas (1) y (2), la precisión que se obtiene con el método MSCM, de acuerdo con los resultados (10), (11) y (12), es la más baja. Este es el precio que se paga por la capacidad de evaluar en forma sencilla a un número muy elevado de progenitores potenciales con base en su comportamiento en sintéticos complementarios medianos. En el Cuadro 1 se muestra una comparación de la capacidad del método MSCM para evaluar progenitores potenciales con relación a la de los métodos de Wright (MW) y de Márquez-Sanchez (MMS). Por ejemplo, para evaluar 14 progenitores en sintéticos complementarios medianos, el MSCM requiere de un experimento con 14 tratamientos mientras que el MMS requiere de 119; a cambio, el MMS tiene una varianza que es apenas 6.6% de la varianza obtenida con MSCM.

Desde un punto de vista práctico MMS y MW no son aconsejables para trabajar con 16 o más progenitores; en cambio, con un número menor de progenitores pueden utilizarse para estimar la media de cualquier sintético en forma individual y, consecuentemente, de cualquier combinación lineal de los sintéticos posibles. Con MSCM, por su parte, sólo se pueden estimar diferencias entre sintéticos complementarios medianos, cualquiera que sea la magnitud de N.

Cuadro 1. Número de tratamientos a evaluar y coeficientes de la varianza de tres métodos para estimar la diferencia entre las medias de dos sintéticos complementarios medianos para diferentes números de progenitores potenciales (N).

Tratamientos[1]				Coeficientes de la varianza[2]		
N	MSCM	MW	MMS	MSCM	MW	MMS
4	4	10	14	4.000	0.750	0.705
6	6	21	27	2.667	0.379	0.357
8	8	36	44	2.000	0.219	0.213
10	10	55	65	1.600	0.144	0.138
12	12	66	78	1.333	0.102	0.100
14	14	105	119	1.142	0.076	0.075
16	16	136	152	1.000	0.059	0.058
18	18	171	189	0.889	0.047	0.047
20	20	210	230	0.800	0.038	0.038
30	30	465	495	0.533	0.017	0.017
40	40	820	860	0.400	0.010	0.010
60	60	1830	1890	0.267	0.002	0.002
100	100	5050	5150	0.160	0.002	0.002

[1] MSCM = método de sintéticos complementarios medianos
MW = método de Writht (1922)
MMS = método de Márquez-Sánchez (1992)

[2] Para MMS se utilizó m = 10 y para los tres métodos se hizo el cálculo con base en una repetición.

Resultan apropiadas dos consideraciones adicionales en relación a las varianzas de los estimadores. La varianza σ^2 a que se hizo referencia fue la varianza del error del experimento de evaluación y fue siempre considerada constante, independientemente del número de materiales (tratamientos) en evaluación. Cabría esperar que la varianza aumente conforme se eleve el número de tratamientos, ya que esto implica un incremento en el bloque completo (o incompleto) a utilizarse. Esto sería una ventaja (difícil de cuantificar) del MSCM porque utilizaría bloques más pequeños.

Haciendo la comparación de los tres métodos en condiciones de igualdad en unidades experimentales, mediante el manejo de la cantidad de repeticiones, se encontró que la fórmula de Wright (1922) sería la de mayor precisión y la de precisión intermedia sería la de Márquez-Sánchez (1992) como se aprecia en el Cuadro 2. En estos calculos se consideró como base el número de parcelas que para una repetición requiere el MMS. Por ejemplo, para el caso N = 4, MMS requiere 14 unidades experimentales en tanto que MSCM requiere sólo 4; es decir, para tener un mismo número de unidades experimentales el MSCM debería usar 3.5 repeticiones por cada repetición de MMS, con lo cual su varianza resultaría ser 1.143.

De la información del Cuadro 2 se puede concluir que si N es de una magnitud tal que la aplicación de MW y MMS resulta factible y es práctica, para la comparación de SCM el orden de preferencia es MMW, MMS y MSCM. Evidentemente, si N es tal que la disponibilidad de recursos no permite el uso de MW y MMS, MSCM es una alternativa que puede ser viable.

Cuadro 2. Coeficientes de la varianza de tres estimadores de la diferencia entre las medias de dos sintéticos complementarios medianos para diferentes tamaños de sintéticos (N) en igualdad de número de parcelas.

N	Métodos[1]		
	MSCM	MW	MMS
4	1.143	0.536	0.705
6	0.593	0.288	0.357
8	0.364	0.179	0.213
10	0.246	0.122	0.138
12	0.205	0.086	0.100
14	0.135	0.067	0.075
16	0.105	0.053	0.058
18	0.093	0.043	0.047
20	0.069	0.035	0.038
30	0.032	0.016	0.017
40	0.019	0.0095	0.0099
60	0.008	0.0019	0.002
100	0.003	0.00157	0.002

[1] MSCM = método de sintéticos complementarios medianos
MW = método de Wright (1922)
MMS = método de Márquez-Sánchez (1992)

BIBLIOGRAFIA

Márquez-Sánchez, F. 1992. On the yield prediction of composite varieties. Maydica 37:271-274.

Mood, A.M., F.A. Graybill and D.C. Boes. 1974. Introduction to the theory of Statistics. Third edition. McGraw Hill 564 pp.

Sahagún C., J. 1994a. Incorporación de recursos genéticos al desarrollo de nuevos cultivares: Aspectos de Genética Cuantitativa. 107-113. En: Cuevas, S.J.A., E. Estrada y E. Cedillo (eds.). Memorias del I Simposium Internacional sobre Etnobotánica en Mesoamérica. Chapingo, Méx. marzo, 1994.

__________**. 1994b.** Sobre el cálculo de coeficientes de endogamia de variedades sintéticas. Agrociencia serie Fitociencia 5: 67-78.

Wright, S. 1922. The effects of inbreeding and crossbreeding in guinea pigs. USDA Bull. 1121.