Algunos conceptos básicos de estadística paramétrica para genética cuantitativa

Promedio aritmético de una variable X en una muestra aleatoria de una población.

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + X_3 + \cdots + X_n);\ \bar{X} = \frac{1}{n}\sum X_1$$

Por ejemplo, la media de los valores 4, 5, 6, 7, 8

$$\bar{X} = \frac{1}{n}(4 + 5 + 6 + 7 + 8) = 6$$

¿cómo se desvía cada valor (X_i) de su media?

$$\sum (X_1 - \bar{X}) = 0$$

$$\sum \quad [(4 - 6) + 5 - 6) + (6 - 6) + (7 - 6) + (8 - 6) = 0$$

Media geométrica (G): La media geométrica de *n* valores, es igual a la raíz *n* del producto de tales valores:

G= $[4\times5\times6\times7\times8]^{1/n}$ = $[6720]^{1/5}$=5.827

Varianza y *Desviación estándar*. Medimos la variación de los individuos, en X, respecto de su promedio, relativos a su media. Sin embargo, no usamos las deviaciones sino el cuadrado de éstas (**Sum of Squares** =*SS*)

$$\sum (X_1 - \bar{X})^2 = SS_x$$

$$SS_x = \sum [(4 - 6)^2 + (5 - 6)^2 + (6 - 6)^2 + (7 - 6)^2 + (8 - 6)^2] = 10$$

$$\sum (X_i - \bar{X})^2 = SS_x = \sum (X_i)^2 - \frac{(\sum X_i)^2}{n}$$

Y el promedio de las desviaciones cuadradas se denomina la *Varianza*:

$V_x = \sigma_x^2 = \frac{1}{n}\sum(X_i - \bar{X})^2 = \frac{1}{n}SS_x$, ***poblacional*** o

$s_x^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$ ***muestral***, dado que la variación de los individuos alrededor de una muestra es menor que su variación alrededor de la media poblacional.

La raíz cuadrada de la varianza es la *desviación estándar* y ésta se representa como *s* o σ, dependiendo si se trata de una muestra o de una población, respectivamente.

En ocasiones los valores de los organismos se agrupan en *clases,* por ejemplo:

Media de la clase	80	90	100	110	120	$\bar{X}$ = 100
No. Inds.	3	12	18	12	3	N=48

Para calcular la SS_x incluimos entonces la frecuencia (*f*) de las observaciones de cada clase:

$$SS_x = \sum(fX_i)^2 - \frac{(\sum fX_i)^2}{n}$$

Es decir:

$SS_x = (3 \bullet 80)^2 + (12 \bullet 90)^2 + (18 \bullet 100)^2 + (12 \bullet 110)^2 + (3 \bullet 120)^2 - (4800^2/48) = 4800$

La varianza es por tanto, $s_x^2 = \frac{1}{n-1}SS_x$

$$s_x^2 = \frac{4880}{47} = 102.13$$

$$s = \sqrt{s_x^2}$$

Y la desviación estándar,

$$s = \sqrt{102.13} \quad = 10.11$$

La distribución normal. La variación en los caracteres métricos de los individuales de las poblaciones naturales o derivados de cruzamientos se distribuye cercano a una función matemática normal. En genética cuantitativa, la distribución obedece a los valores genéticos y ambientales que pueden desviarse por encima o por debajo del promedio de la población. Por ejemplo, el número de semillas por fruto en la especie *Datura inoxia* se distribuye de forma continua, con un promedio de 120 y una desviación estándar de 61.1641 (n = 140 frutos).

αχγπασρθϖω

Análisis de la varianza

El análisis de la varianza se emplea con éxito en ensayos comparativos sobre el rendimiento (*yield*) de variedades de cultivos. Sin embargo, se sabe que existe una amplia variación en el rendimiento entre parcelas, a pesar de que éstas sean del mismo tamaño y se siembren con la misma variedad. Existen distintos factores que explican esta variación, entre los que se cuentan la uniformidad del suelo, el tamaño de partículas del suelo, la humedad, etc. que a pesar de una apariencia uniforme no existe homogeneidad por lo que esto es muy relevante al aplicar técnicas de cultivo con el objetivo de comparar variedades (Figura arroz). Ciertos factores afectan el desarrollo de las plantas, por ejemplo, el sitio (crecimiento), la hora de siembra (germinación), etc., por lo que el comparar distintas variedades en el mismo o en distintos campos o parcelas puede ser un problema complejo. (por supuesto que el análisis de la varianza se aplica hoy a infinidad de comparaciones, no sólo las de carácter biológico o genético, para lo cual fue diseñado; sin embargo, analizaremos un ejemplo agronómico constado con la genética cuantitativa.)

Para reducir el efecto de estos problemas, se debe sembrar de cada *variedad* (genotipo) un número suficiente de parcelas que se denominan *replicaciones*. Con esto se logra que, en el promedio de las parcelas, los efectos favorables y desfavorables tienden a compensarse (=0). Por lo tanto, si tenemos un número suficiente de replicaciones se puede conocer el rendimiento de las variedades con mayor seguridad. No obstante, para generalizar el valor de una variedad a una región, se debería analizar durante varios años y en distintos lugares su rendimiento. Por ejemplo, si analiza 10 variedades x 10 localidades x 5 replicaciones x 5 años, arroja un total de 2500 parcelas, un número imposible de trabajar, o 250 parcelas por variedad, por año. Quizá con 25 parcelas por variedad podría obtenerse un estimado razonable; aun así, puede ser difícil de lograr.

En la práctica, muy relevante la disposición de las parcelas para hacer las comparaciones de *yield* (y). Suponga que queremos comparar **5** variedades (i.e., genotipos o líneas) de una especie de planta y ponemos **cuatro** r*eplicaciones* de cada variedad. Creamos entonces un arreglo de las variedades en el terreno. Empleamos en el campo lo que se denomina *bloques* (*blocks*), cada bloque contiene una parcela para cada una de las variedades bajo análisis (Figura anexa, parcelas sembradas con distintas líneas de arroz).

En este arreglo se debe hacer un esfuerzo para que la variación *dentro de cada bloque* sea lo más pequeña posible, aunque puede haber variación *entre bloques* porque además se pretende que las conclusiones sean válidas para una amplitud de condiciones (ver foto). Las condiciones físicas del terreno ofrecen variación espacial aún en parcelas aparentemente homogéneas. Si considere experimentación en ecosistemas naturales (imagen) la variación entre unidades de muestreo (parcelas o bloques), debe ser enrome y la replicación necesaria mayor.

Dentro de los bloques, las parcelas deberán tener una disposición *aleatoria* (de hecho, el diseño se denomina técnicamente *randomized blocks*), lo que supone una *distribución aleatoria* de los *efectos ambientales*; por ejemplo, la variación en la humedad, fertilidad, o pendiente del suelo. Sin embargo, estos factores no varían necesariamente de un punto a otro, sino en gradientes, generados por pendientes ligeras, orientación o cercanía a un arroyo, y otros factores. Esta variación es considerada en la variación *entre bloques*. Suponiendo la existencia de un gradiente de humedad (ver foto) a partir del arroyo, no sería adecuado trazar los bloques de forma perpendicular al cauce del arroyo.

Disposición en terreno de cuatro bloques, cada uno con cinco parcelas donde se sembraron cinco genotipos (A – E)

1	2	3	4
D=30	E=32	C=32	A=32
C=34	A=25	B=46	D=47
B=45	D=41	E=37	C=45
A= 29	C=39	A=27	E=45
E=41	B=39	D=42	B=52

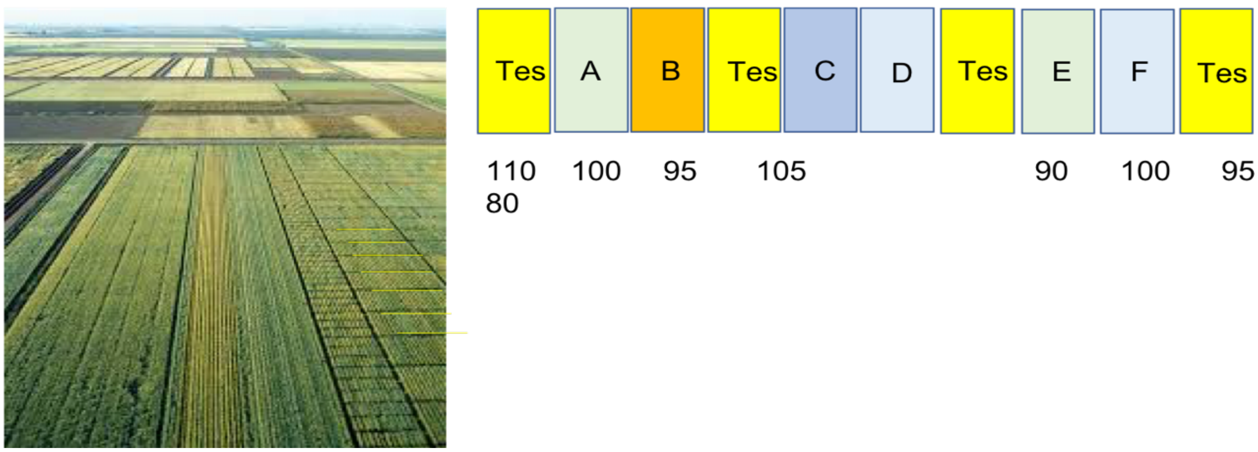
$\bar{X}_{Bloques}$: 35.8 35.2 36.8 44.2

$\bar{X}_{genotipos}$: A= 28.25 B=45.50 C=37.50 D=40.00 E=38.75

$\bar{\bar{X}} = 38.00$

También se puede considerar la variación ambiental cuando se comparan genotipos respecto a genotipos estándar o *testigos*. Por ejemplo, si se desea incrementar el rendimiento respecto a un genotipo superior. En la figura, el rendimiento de los genotipos bajo análisis es similar a juzgar por sus promedios (A, B, E, F); no obstante, el rendimiento de los testigos se reduce hacia las parcelas de la derecha, por lo que la calidad ambiental se reduce. Por tanto, es muy probable que los genotipos E y F sean relativamente mejores que los testigos. (Puede verificarlo con las proporciones respecto al promedio de testigos vecinos.)

Comparación del rendimiento de genotipos respecto a al promedio de los genotipos testigos



Análisis de la variación

La variación en torno a la media la expresamos usando el cuadrado de las variaciones individuales. Esto de entrada da más peso a las variaciones más grandes. Si se usa la sumatoria de las desviaciones al cuadrado, en lugar de la sumatoria de las desviaciones, entonces el promedio, la sumatoria y el producto siempre tienen el valor de cero. La razón más importante, sin embargo, es que la suma de las desviaciones al cuadrado y la varianza son *aditivas*. Si hay dos causas que afectan de forma independiente a un carácter, cada una produce una varianza de 100, y el carácter tendrá una varianza de 100 + 100 =200. La desviación estándar del carácter en lugar de 10 +10, tendrá un valor de raíz de 200 = 14.14. Ya que los efectos de factores causales, no correlacionados, teóricamente se suman, entonces es posible analizar los datos mediante el análisis de la varianza. Veamos el siguiente ejemplo.

Modelo de dos factores

Podemos ver el procedimiento de análisis de la variación cuando dos factores afectan un carácter. Tomemos los datos de las parcelas del primer cuadro ilustrado arriba, y calculemos los promedios por *bloque*. Esto implica que no consideremos en primera instancia a las variedades. Si recordemos los valores de la tabla son los siguientes:

		Bloques		
Variedad	**1**	**2**	**3**	**4**
A	29	25	27	32
B	45	39	46	52
C	34	39	32	45
D	30	41	42	47
E	41	32	37	45

Los datos están ordenados ahora por variedad, en las filas, y por bloques, en las columnas. Los valores individuales en cada bloque corresponden a una parcela. Las unidades corresponden al rendimiento o *fitness*, de cada variedad.

Podemos entonces calcular tanto la suma de valores de Bloques = **Tb**, y su promedio =**Tb/a**, donde **a** es el número de parcelas por bloque (en este caso 5).

De forma horizontal, en las filas, podemos también obtener la suma por cada genotipo **Ta** y su media **Ta/b**, donde **b** es el número de bloques (en este caso 4). Finalmente obtenemos la Gran Media o Media total $\bar{X} = T/n$, donde n = a*b; n = 4*5=20. Por tanto, la media total =38.

			Bloques			Suma Genotipo	Media Genotipo
	Variedad	**1**	**2**	**3**	**4**	**Ta**	**Media X= Ta/b**
	A	29	25	27	32	113	28.25
	B	45	39	46	52	182	45.5
	C	34	39	32	45	150	37.5
	D	30	41	42	47	160	40
	E	41	32	37	45	155	38.75
Suma Bloque	**Suma Tb**	179	176	184	221	**760**	
Media Bloque	**X =Tb/a**	35.8	35.2	36.8	44.2		**38**

$$V_x = \sigma_x^2 = \frac{1}{n}\sum(X_i - \bar{X})^2 = \frac{1}{n}SS_x$$

Podemos estimar entonces la variabilidad de las parcelas, obteniendo la *desviación estándar* y el *coeficiente de variación*. Estimamos primero cómo se desvía cada valor de parcela respecto de la media total (=38):

Desviaciones de cada parcela de la media Total

Variedad	**1**	**2**	**3**	**4**	**Ta**	**X= Ta/b**

A	-9	-13	-11	-6	-39	-9.75
B	7	1	8	14	30	7.5
C	-4	1	-6	7	-2	-0.5
D	-8	3	4	9	8	2
E	3	-6	-1	7	3	0.75
Tb-aX	-11	-14	-6	31	0	0
Xb-X	-2.2	-2.8	-1.2	6.2	0	0

Por ejemplo, la parcela (A) en el bloque 1 se desvía -9 unidades respecto de la gran media (29-38=-9), y así sucesivamente. En la penúltima columna está la suma total de las desviaciones de cada variedad, mientras que en la penúltima fila esta la suma total de las desviaciones de cada bloque. La última columna y fila dan el promedio de desviación de cada variedad y de cada bloque, respectivamente, de la media total. Nótese que la suma de las desviaciones promedio es cero en ambos casos. Ahora calculamos la suma de cuadrados total de las desviaciones:

$SS_x = \sum(X_i - \bar{X})^2 = -9^2 + -13^2 + -11^2 + \ldots + -1^2 + 7^2 = 1084$

Esta suma dividida entre los *grados de libertad* (*n*-1) nos da la varianza:

$SS_x = \frac{1}{n-1}\sum(X_i - \bar{X})^2 = 57.0526$, su raíz cuadrada, la desviación estándar:

$s = \sqrt{57.0526} = 7.5331$. Por último, el *coeficiente de variación* (%), es la razón de la desviación estándar sobre la media lo que indica un valor de C. V. (%) = 19.9. Este valor, de entrada, señala una heterogeneidad en las parcelas experimentales.

Componentes de las desviaciones

Consideremos ahora que sembramos 5 genotipos o variedades en 20 parcelas, con una variedad por parcela y cada variedad representada en cuatro bloques, como se ilustra en el primer cuadro de arriba. Por lo tanto, cada parcela se desvía de la media total por la misma cantidad que ya indicamos en la tabla inmediata superior. No obstante, las desviaciones individuales (de parcelas), se reparten en tres efectos: el debido a las variedades, a los bloques y a las parcelas.

Aquí es necesario indicar que suponemos que el efecto de las variedades son genéticas (genotípicas), que entre bloques se deben a variación en el suelo y otras variables ambientales y el efecto de las parcelas a diferencias de suelo dentro de los bloques (por ejemplo, a variación en el número de plantas, dados por germinación); a la heterogeneidad a error de muestreo por heterogeneidad fenotípica entre las variedades, y a interacciones bloque x variedad.

Es importante entonces dividir las desviaciones de las parcelas (en esas tres causas hipotéticas):

$(X - \bar{X}) = (\bar{X}_a - \bar{X}) + (\bar{X}_b - \bar{X}) + (X - \hat{X})$

Donde X es el valor de una parcela individual, $\bar{X}$ es la media total, $\bar{X}_a$es la media de una variedad cualquiera, $\bar{X}_b$ es la media del bloque y $\hat{X}$ es el valor esperado de una parcela cualquiera si los efectos de la variedad y el bloque *fueran estrictamente aditivos*. En general los subíndices a y b indican los factores del experimento.

Si esto es así, podemos entonces calcular el valor esperado (o la desviación esperada) para cada parcela individual.

Tomemos como ejemplo la parcela 1 del bloque 1. Esta tiene una desviación de la media de -9. La media de las desviaciones de la Variedad A con respecto a la media total es de -9.75 y la media de las desviaciones del bloque 1 con respecto a la media total es de -2.2. Por tano, la ecuación es:

$(X - \hat{X}) = (X - \bar{X}) - [(\bar{X}_a - \bar{X}) + (\bar{X}_b - \bar{X})]$

(-9) = (-9.75) + (-2.2) + $(X - \hat{X})$,

$(X - \hat{X})$ = +2.95

El valor esperado para la primera parcela es: (29-38) = (28.25-38) + (35.8-38) + (29-$\hat{X}$),

Entonces: $\hat{X}$ = $(\bar{X}) - [\bar{X}_a + \bar{X}_b]$ = 26.05

En resumen, usando los valores correspondientes, se pueden calcular las desviaciones esperadas para cada parcela:

Desviaciones de cada parcela de los valores esperados

Variedad	1	2	3	4	Ta
A	2.95	-0.45	-0.05	-2.45	0
B	1.7	-3.7	1.7	0.3	0
C	-1.3	4.3	-4.3	1.3	0
D	-7.8	3.8	3.2	0.8	0
E	4.45	-3.95	-0.55	0.05	0
	0	0	0	0	0

La suma de la última fila y columna son cero.

Fraccionamiento de la SS_{Total} en sus componentes

Tomemos la SS_{total}, cuyo valor es de 1084, y obtengamos la importancia relativa de los tres efectos en la variabilidad de el rendimiento de las parcelas. Si los efectos son independientes, podemos escribir la siguiente ecuación:

	Efectos de los componentes en la *SS*		
SS_{Total}	Variedades	Bloques	Parcelas
$\sum^{ab}(X-\bar{X})^2 =$	$\sum^{ab}(\bar{X}_a-\bar{X})^2 +$	$\sum^{ab}(\bar{X}_b-\bar{X})^2 +$	$\sum^{ab}\left(\bar{X}_b-\hat{X}\right)^2$

De forma análoga a la suma de las desviaciones. La SS_{Total} se puede descomponer en tres fracciones (variedades, bloques y parcelas). La sumatoria, *ab*, es total e incluye el mismo número de desviaciones cuadradas (i.e., 20). Por ejemplo, en el caso de las variedades (*a* =5) las desviaciones promedio están basadas en cuatro parcelas (*b* = 4)) y la desviación media de los 4 bloques en 5 parcelas (*a*=5). Si la ecuación es válida para AB parcelas, entonces es equivalente a

$$\sum^{ab}(X-\bar{X})^2 = \quad b\sum^{a}(\bar{X}_a-\bar{X})^2 + \quad a\sum^{b}(\bar{X}_b-\bar{X})^2 + \quad \sum^{ab}\left(\bar{X}_b-\hat{X}\right)^2$$

O bien como:

$$\sum(X-\bar{X})^2 = \quad b\sum(\bar{X}_a-\bar{X})^2 + \quad a\sum(\bar{X}_b-\bar{X})^2 + \quad \sum\left(\bar{X}_b-\hat{X}\right)^2$$

Por facilidad, se expresa generalmente como:

$$SS_x = SS_{ab} = SS_a + SS_b + SS_{a\cdot b}$$

Mientras que SS_{ab}, indica una combinación de factores *ab* (i.e., parcelas), pero $SS_{a\cdot b}$ indica una interacción estadística de factores y se calcula también por resta. Es una cantidad compleja que generalmente suponemos que resulta de ciertas combinaciones de *a* y *b*, pero puede depender de otros factores no contemplados en el esquema experimental de los dos efectos definidos de antemano.

Para obtener las *SS* tomemos primero el caso de las variedades. En la Tabla se pueden ver los valores, última columna, de cada variedad:

$SS_a = b\sum^a(\bar{X}_a-\bar{X})^2 = 4(-9.75)^2 + 4(7.5)^2 + 4(-0.5)^2 + 4(2)^2 + 4(0.75)^2 = 624.5$

$SS_b = a\sum^b(\bar{X}_b-\bar{X})^2 = 5(-2.2)^2 + 5(-2.8)^2 + 5(-1.2)^2 + 5(6.2)^2 = 262.8$

$SS_{a.b} = b\sum^a(\bar{X}_a-\bar{X})^2 = (2.95)^2 + (-0.45)^2 + (-0.05)^2 + \cdots + (-0.55)^2 + (0.05)^2 = 196.7$

Ya que

$SS_{ab} = SS_a + SS_b + SS_{a \cdot b}$, también se puede obtener por diferencia conociendo la *SS* total. Así, si SS_{Total} = 1084, entonces

$SS_{a \cdot b}$ = 1084 – [624.5 – 262.5] = 196.7

Es muy importante notar que estas son las contribuciones de los efectos a la suma de cuadrados total, las cuales podemos ponerlas en forma tabular de la siguiente manera:

Componentes de variación	Suma de Cuadradas		*d.f.*
	%	Absoluta	
Variedades, SS_a	624.5	57.6	
Bloques, SS_b	262.8	24.2	
Parcelas, $SS_{a \cdot b}$	196.7	18.1	
Total, SS_{ab}	1084	100	

Grados de libertad

En estadística los grados de libertad es el número de miembros de una serie de datos que se pueden llenar con valores arbitrarios, sin violar las condiciones preestablecidas. Por ejemplo, los grados de libertad para el caso de un locus con 2 alelos es 1. Ya que la suma de ambos alelos es 1, una vez establecido el valor de primer alelo el otro no puede ser un valor arbitrario. Seis personas apuestan a adivinar el número de puntos al lanzar un dado; el que atine se lleva la apuesta. Todos eligen su número previo al lanzamiento, no obstante, solo existen 5 grados de libertad porque el sexto queda definido por *default* (claro que todos tienen la misma probabilidad de ganar (1/6), si el dado tiene un balance perfecto, lo que no ocurre, por lo general.

Volviendo a nuestro problema, si suponemos que el promedio de las desviaciones en rendimiento de todas las 20 parcelas es cero, entonces sólo 19 pueden variar libremente pero no el vigésimo valor, que se establece automáticamente para que la suma sea cero. Entonces, los grados de libertad de la SS_{Total} es 19.

Por esta razón, los grados de libertad de las desviaciones de las variedades y de los bloques, con respecto a su media, son *a*-1 y *b*-1, respectivamente. Podemos volver ahora a nuestra tabla y asentar los grados de libertad (*degrees of freedom* o *d.f.*). Como se aprecia de la Tabla, del total de observaciones, tenemos 19 grados de libertad para la suma de cuadrados totales, 4 para variedades y 3 para los bloques: 19-4-3 = 12. Éste es el número de grados de libertad de la suma de cuadrados de parcelas (del error o residual, Parcelas, $SS_{a \cdot b}$). Los grados de libertad para las parcelas se establece como el producto de (*a*-1)(*b*-1) = 12 (columnas menos 1 por hileras menos 1).

Componentes de variación	Suma de Cuadradas	*d.f.*

	Absoluta	%	
Variedades, SS_a	624.5	57.6	4
Bloques, SS_b	262.8	24.2	3
Parcelas, $SS_{a \cdot b}$	196.7	18.1	12
Total, SS_{ab}	1084	100	19

Cuadrado medio

Si dividimos la *SS* entre el número de grados de libertad obtenemos un valor denominado el cuadrado medio, *CM* (propiamente, es el promedio de las desviaciones al cuadrado) que equivalente a la varianza. En nuestro caso, obtenemos los cuadrados medios para variedades y bloques y parcelas. El cuadrado medio de las parcelas se considera una medida del "error experimental" y lo denominamos CM_{Error}, $CM_{a \cdot b}$ σ^2_{error} o s^2_{error}, y representa la varianza *residua*l o aquella no explicada por los factores incluidos en nuestro modelo. Este lo empleamos para evaluar a los otros componentes, dividiéndolos entre el CM_{Error}, que denominamos *F*-ratio. Volvamos a nuestra tabla para hacer esos cálculos.

Componentes de variación	*SS*, Suma de Cuadradas		*d.f.*	*CM*, Cuadrado Medio	*F-ratio*	$F_{0.05}$ (ν_1, ν_2)
	Absoluta	%				
Variedades, SS_a	624.5	57.6	4	156.12	9.52	$F^4_{12} = 3.259$
Bloques, SS_b	262.8	24.2	3	87.60	5.34	$F^3_{12} = 3.490$
Parcelas, $SS_{a \cdot b}$	196.7	18.1	12	16.39		
Total, SS_{ab}	1084	100	19			

Apreciamos que las razones son mayores que 1. No obstante, la significancia de *F* depende de los grados de libertad del numerador, el *efecto*, y el denominador, el *residual*. Por ejemplo, para el caso de las variedades, comparamos el valor de *F* con aquel de la distribución de probabilidades con F^4_{12}, para el caso de las variedades (Tabla estadística anexa), y con F^3_{12} para el caso de los bloques. Los valores respectivos se anotan en la tabla para la una distribución de *F* con $\alpha = 0.05$; es decir esperamos que las razones de varianza estimadas estén en la región de la distribución de *F* que contiene el 5% de las probabilidades. Observamos que la razón de *F* para variedades excede ampliamente el valor de F podemos afirmar que las variedades difieren significativamente en su rendimiento (*yield*); aún a un nivel $\alpha = 0.05$ la razón de *F* es mayor que el de tablas. En el caso de la razón de *F* de bloques, encontramos que a un nivel de 5%, existen diferencias entre bloque (no así al 1%).

Se define como $F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$ Escriba aquí la ecuación.
Donde

$$s_1^2 = \frac{\sum(X_1 - \bar{X}_1)^2}{n_1 - 1}\text{ , y } s_2^2 = \frac{\sum(X_2 - \bar{X}_2)^2}{n_2 - 1}$$

y grados de libertar $v_1 = n_1$-1 y $v_2 = n_2$-1.

Cuando $\sigma_2^2 = \sigma_2^2,\ F = \frac{s_1^2}{s_2^2}$.

La distribución de F está sesgada a la derecha, pero su sesgo disminuye conforme se incrementan los grados de libertad. El valor de F es siempre positivo 0. F toma valores entre 0 e ∞. El valor crítico de F se obtiene de las tablas.

La varianza explicada *Y*

La proporción de la variación en la respuesta (y) alrededor de la media que puede atribuirse a los efectos incluidos en el experimento, más que al azar, la estimamos mediante la R^2, y es la razón entre la *SS* atribuibles a los efectos (modelo; i.e., variedades y bloques) entre la SS_{Total}. En nuestro ejemplo $R^2 = \frac{SS_a + SS_b}{SS_{Total}} = \frac{624.5+262.8}{1084} = 0.818$. Con el objeto de hacer comparables los resultados con otros modelos que incluyen un número distinto de parámetros (efectos), se calcula una ajustada, la cual usa los cuadrados medios del error y total, por tanto considera los grados de liberta, en lugar de la suma de cuadrados. Se calcula como $R^2_{Ajustada}$ = 1-$\frac{CM_{Error}}{CM_{Total}} = \frac{16.39}{57.05} = 0.7127$

Valores críticos superiores de la distribución de F con ν_1 *d. f.* en el numerador y ν_2 en el denominador, a un nivel de significancia del 5%

$$F_{0.05}\,(\nu_1,\ \nu_2)$$

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.882	240.543	241.882
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.220
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165