

**RELACION ENTRE 17 CARACTERES MORFOLOGICOS,
FENOLOGICOS Y QUIMICOS Y EL CONTENIDO
DE NITROGENO EN LA SEMILLA DE UNA SERIE
DE LINEAS DE SOYA (Glycine max (L.) Merr.)**

ADALBERTO BENAVIDES MENDOZA

T E S I S

**PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE
MAESTRO EN CIENCIAS
EN FITOMEJORAMIENTO**



**Universidad Autónoma Agraria
Antonio Narro**

PROGRAMA DE GRADUADOS

Buenavista, Saltillo, Coah.

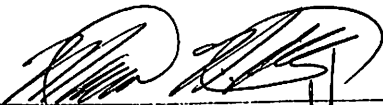
ABRIL, 1991

Tesis elaborada bajo la supervisión del comité particular de
asesoría y aprobada como requisito parcial, para optar
al grado de

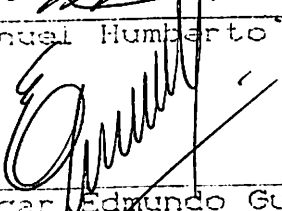
MAESTRO EN CIENCIAS
EN FITOMEJORAMIENTO

COMITE PARTICULAR


Asesor principal:


Bíol. M. C. Manuel Humberto Reyes Valdés

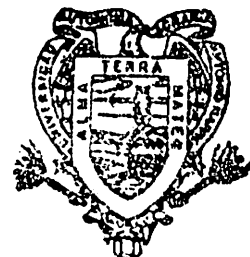
Asesor:


Ing. M. C. Edgar Edmundo Guzmán Medrano

Asesor:


Dra. Diana Jasso de Rodríguez


Dr. José Manuel Fernández Brondo
Subdirector de Asuntos de Postgrado



BIBLIOTECA
EGIDIO G. REBONATO
BANCO DE TESIS
U.A.A.A.N.

Buenavista, Saltillo, Coahuila. Abril 1991

AGRADECIMIENTOS

Al M.C. Humberto Reyes Valdés y al Dr. Eleuterio López Pérez por su inestimable apoyo a lo largo de mis estudios.

A mis padres por toda la ayuda y comprensión que me otorgaron.

A todas las personas que, por falta de espacio, no puedo incluir pero que ayudaron en una forma u otra en la realización de este trabajo.

DEDICATORIA

A mi esposa Laura

A mis hijos Adalberto y Andrés

A mis padres y a mis hermanos

COMPENDIO

Relacion Entre 17 Caracteres Morfológicos, Fenológicos y Químicos y el Contenido de Nitrógeno en la Semilla de una Serie de Líneas de Soya (*Glycine max* (L.) Merr.).

POR

ADALBERTO BENAVIDES MENDOZA

MAESTRIA

FITOMEJORAMIENTO

UNIVERSIDAD AUTONOMA AGRARIA ANTONIO NARRO
BUENAVISTA, SALTILLO, COAHUILA. ABRIL 1991.

Biol. M.C. Manuel Humberto Reyes Valdés - Asesor -

Palabras clave: Soya, nitrógeno, proteína, calidad de la semilla, rendimiento.

El presente trabajo se llevó a cabo con la finalidad de investigar la variación en el contenido de proteína en las semillas (expresado como porcentaje de nitrógeno) en una serie de líneas y cultivares de soya así como detectar y cuantificar las relaciones existentes entre el carácter mencionado y un conjunto de variables fenológicas, morfológicas y químicas. Esto último se realizó con el propósito de marcar las variables más útiles como predictoras del contenido de proteína en las semillas. Lo realizado para el anterior carácter se llevó a cabo con la variable peso seco de las semillas, con la finalidad de que sirviese como punto de comparación y validación.

No se encontraron diferencias significativas entre la mayoría de los cultivares y líneas para los caracteres contenido porcentual de nitrógeno en las semillas y peso seco de las mismas. Este hecho indica, al parecer, que el espectro de cultivares y líneas utilizadas no fue lo suficientemente amplio.

Se encontró una relación positiva entre los caracteres peso seco de las semillas y contenido porcentual de nitrógeno en las mismas. Este resultado se encuentra acorde con la hipótesis que indica que los dos caracteres son genéticamente independientes pero ligados fisiológicamente durante el desarrollo. Esta hipótesis excluye también la presencia de una contradicción energética entre los dos caracteres mencionados.

ABSTRACT

Relationship Between 17 Morphological, Phenological and Chemical Traits and Nitrogen Content in Seed of Some Soybean (*Glycine max* (L.) Merr.) Lines.

By

ADALBERTO BENAVIDES MENDOZA

MASTER'S DEGREE

PLANT BREEDING

UNIVERSIDAD AUTONOMA AGRARIA ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA. APRIL 1991

Biol. M.C. Manuel Humberto Reyes Valdes -Advisor-

Key Words: Soybean, nitrogen, protein, seed quality, yield.

The objectives of this work were to investigate the variation in protein content (in nitrogen percentage basis) in some soybean lines and cultivars as well as to detect and measure the relationships between the above trait and a set of phenological, morphological and chemical characters. It was done with the aim of marking the most useful predictor variables of protein content in seeds.

As a mean of comparison, the above procedure was done with seeds dry weight too.

There were not found significant differences among lines and cultivars for seed nitrogen percentage and dry weight. It seems, then, that the lines and cultivars spectrum was not wide enough.

A positive relationship was found between dry weight and nitrogen percentage. This result agrees with a hypothesis which indicates that both traits are independent from the genetic point of view, but physiologically linked during the development. Such an hypothesis exclude an energetic contradiction between both traits too.

The results of the determination of nitrogen percentage indicated the existence of a set of important variables. These were: days to R6-R7, seeds dry weight, nitrogen percentage in vegetative structures, pods dry weight, days to R1, and seed number per plant. However, the statistical tests indicated that these "important" variables are little useful for prediction or indirect selection.

It was concluded that the indirect selection approach for nitrogen percentage, by using correlated traits, would not work well, at least for the traits and varieties considered in this research.

INDICE DE CONTENIDO

	Página
INDICE DE CUADROS	xiii
INDICE DE FIGURAS	xv
INTRODUCCION	1
REVISION DE LITERATURA	4
MATERIALES Y METODOS	14
MATERIAL BIOLÓGICO	14
PROCEDIMIENTO EXPERIMENTAL	15
ANÁLISIS ESTADÍSTICO	20
DESCRIPCIÓN DE LAS TÉCNICAS ESTADÍSTICAS.....	23
RESULTADOS	29
DISCUSION	73
CONCLUSIONES	83
RESUMEN	85
LITERATURA CITADA	88
APÉNDICES	91

INDICE DE CUADROS

Cuadro No.		Página
4.1	Valores promedio del contenido porcentual de nitrógeno por maceta para cada línea o cultivar	36
4.2	Valores promedio de peso seco de semillas (en gramos) por maceta para cada línea o cultivar	36
4.3	Resultados de la prueba de Mann-Whitney para el carácter contenido porcentual de nitrógeno en las semillas	38
4.4	Resultados de la prueba de Mann-Whitney para el carácter peso seco de las semillas.....	39
4.5	Listado de los valores marcados como significativos de la correlación lineal simple y la correlación parcial entre la variable contenido porcentual de nitrógeno en las semillas y las restantes características	41
4.6	Listado de los valores marcados como significativos de la correlación lineal simple y la correlación parcial entre la variable peso seco de las semillas y las restantes características	42
4.7	Valores característicos de la matriz de correlaciones	44
4.8	Valores de las dos primeras componentes principales considerando las 17 variables	46
4.9	Coeficientes de los vectores característicos utilizados en la	

	construcción de la primera y segunda componentes principales	49
4.10	Valores estimados de los coeficientes de regresión parcial con el porcentaje de nitrógeno en las semillas como variable de respuesta	51
4.11	Análisis de varianza para la regresión de la variable porcentaje de nitrógeno en las semillas sobre las restantes variables y partición de la suma de cuadrados de regresión para realizar pruebas secuenciales de F	54
4.12	VARIABLES incluidas en el modelo restringido obtenido después de llevar a cabo selección de variables empleando el algoritmo <i>backward</i> en la regresión de la variable porcentaje de nitrógeno en las semillas sobre las restantes variables	55
4.13	Valores de los coeficientes de regresión parcial estandarizados para valores del coeficiente de sesgo, w , de cero a uno. La regresión es de la variable porcentaje de nitrógeno en las semillas sobre las restantes variables	56
4.14	Valores de los coeficientes de regresión parcial estandarizados obtenidos con un ajuste para cuadrados mínimos. La regresión es de la variable porcentaje de nitrógeno en las semillas sobre las restantes variables	59
4.15	Resumen de las variables seleccionadas como importantes para cada una de las técnicas utilizadas. La variable de respuesta fue el porcentaje de nitrógeno en las semillas	60
4.16	Valores estimados de los coeficientes de regresión parcial con el peso seco de las semillas como variable de respuesta	62
4.17	Análisis de varianza para la regresión de la variable peso seco de las semillas sobre las restantes variables y partición de la suma de cuadrados de regresión para realizar pruebas secuenciales de F	65

4.18	Variables incluidas en el modelo restringido obtenido después de llevar a cabo selección de variables empleando el algoritmo <i>backward</i> . La regresión es de la variable peso seco de las semillas sobre las restantes variables	66
4.19	Valores de los coeficientes de regresión parcial estandarizados para valores del coeficiente de sesgo, w , de cero a uno. La regresión es de la variable peso seco de las semillas sobre las restantes variables	68
4.20	Valores de los coeficientes de regresión parcial estandarizados obtenidos con un ajuste para cuadrados mínimos. La regresión es de la variable peso seco de las semillas sobre las restantes variables	70
4.21	Resumen de las variables seleccionadas como importantes para cada una de las técnicas utilizadas. La variable de respuesta fue el peso seco de las semillas	71

INDICE DE FIGURAS

Figura No.		Página
1	Gráfica de los valores de porcentaje de nitrógeno en las semillas contra el peso seco de las mismas.....	30
2.A	Serie cronológica de los datos de porcentaje de nitrógeno en las semillas.....	32
2.B	Serie cronológica de los datos de peso seco de las semillas.....	33
3.	Histograma en tres dimensiones que muestra la distribución conjunta de las variables porcentaje de nitrógeno en las semillas y peso seco de las semillas.....	34
4.	Ordenación gráfica de los individuos utilizando la primera y segunda componentes principales.....	47
5.	<i>Ridge trace</i> para la regresión múltiple de la variable porcentaje de nitrógeno en las semillas sobre las restantes variables.....	58
6.	<i>Ridge trace</i> para la regresión múltiple de	

la variable peso seco de las semillas
sobre las restantes variables.....

69

INTRODUCCION

Desde los inicios de la agricultura las leguminosas han ocupado un lugar importante en la alimentación humana. En cada cultura agrícola del pasado se encontraron, invariablemente, una o más especies de leguminosas en combinación con alguna o algunas especies de gramíneas. En la actualidad el cultivo de leguminosas no ha perdido su importancia, además de que su uso se ha diversificado notablemente. Una especie, *Glycine max* (L.) Merr., la soya, sobresale en el mercado agrícola mundial en su tradicional uso como fuente de proteína y en el relativamente moderno uso como fuente de aceite.

La semilla de soya es una buena fuente de proteína, útil tanto para la alimentación humana como en la preparación de alimentos concentrados para animales. Entre las leguminosas la soya es la especie que, en promedio, presenta proteína de mayor calidad y contenidos de proteína más altos en la semilla. Existe, sin embargo, el problema de que el carácter contenido de proteína es complejo y su expresión fenotípica se ve muy influenciada tanto por el ambiente externo como por el ambiente interno de la planta. El mejorador de plantas, cuyo interés se enfoca en elevar la producción de proteínas por parte de las leguminosas, se

enfrenta a una serie de dilemas, conocidos desde hace bastantes años pero no resueltos aún, que dificultan su labor. Esto es resultado, al parecer, de la presencia de un vacío en conocimientos básicos acerca de los mecanismos de regulación genético moleculares y fisiológicos relacionados con la incorporación, transporte y reparto de nitrógeno, materia prima de las proteínas. Este vacío es, sin embargo, difícil de llenar. La complejidad de estos mecanismos de regulación es apabullante y seguramente llevará bastantes años el llegar a conocerlos a fondo y manipularlos.

Los estudios realizados hasta el momento, con el empleo de las herramientas clásicas del mejorador de plantas, se topan una y otra vez con la presencia de una relación inversa entre el rendimiento total y el contenido porcentual de proteína producidos por un cultivar o cultivares dados. Aunque existe controversia acerca de si esta relación inversa tiene una base genética y/o fisiológica o es simplemente un artefacto resultante de ciertas condiciones del ambiente durante el desarrollo de la planta, la tendencia moderna es considerar como el correcto este último enfoque. Se sigue entonces que existe, al parecer, la posibilidad de obtener cultivares de soya, o de otras especies de cultivo, cuya semilla contenga cantidades mayores de proteína sin que ocurra un abatimiento en el rendimiento total.

Objetivos

Uno de los objetivos del presente trabajo fue investigar si existen diferencias significativas para el contenido de proteína en las semillas (expresado como contenido de nitrógeno en las mismas) en una serie de genotipos de soya. El segundo objetivo fue determinar la presencia o ausencia de la relación inversa entre el contenido de proteína en las semillas y el rendimiento de las mismas. El tercer objetivo fue detectar y cuantificar las relaciones existentes entre un conjunto de variables morfológicas y fenológicas y el carácter contenido de proteína en la semilla, esto con el propósito de sugerir posibles formas de selección indirecta.

REVISIÓN DE LITERATURA

Se sabe que existe gran cantidad de factores que condicionan la expresión final de un carácter complejo. En el caso del rendimiento de un cultivo esta expresión final es una cierta cantidad de material acumulado y es por ello dependiente de todos los procesos de absorción, traslado, asimilación y almacenamiento de materia así como de aquellos que implican captación y utilización de energía. Es un hecho conocido el que la eficiencia de producción de materia seca de las especies de cultivo actualmente utilizadas es muy inferior a la teóricamente posible. Según Loomis y Williams (1963) esta baja eficiencia es resultado principalmente de tres limitantes: cantidad de área foliar, disposición espacial de la misma y aporte de CO_2 a la planta.

Si bien la eficiencia en la producción, por parte de un cultivo, se ha enfocado tradicionalmente hacia la obtención de mayores rendimientos brutos, en los últimos años se ha reconocido la necesidad de considerar también la calidad del producto. Es claro que la definición de calidad varía con el uso que se le dará al producto final pero, en general, haciendo referencia a las leguminosas de grano y pensando en términos de alimentación humana y de animales no rumiantes, la calidad es función del contenido de proteína,

de la digestibilidad y de la composición de aminoácidos (Byers *et al.*, 1977).

La soya (*Glycine max* (L.) Merr.) es una leguminosa cuyo cultivo se lleva a cabo para la obtención de aceite y proteína principalmente. Tanto el rendimiento como la calidad de la semilla dependen fuertemente de la eficiencia general de la planta. La eficiencia de la producción de materia seca por parte de una planta se relaciona con múltiples factores que al final convergen hacia dos procesos principales: fotosíntesis y asimilación de nitrógeno. Ambos procesos se relacionan de manera recíproca y compleja; la tasa y eficiencia fotosintética dependen del área foliar, del número y disposición de las ramificaciones, de la tasa de aparición y duración foliar, del contenido de proteína foliar, etc. La asimilación de nitrógeno, en forma de NO_3^- , NH_4^+ ó N_2 , depende básicamente de la disponibilidad de energía para los procesos de absorción, reducción, translocación y asimilación (Shibles *et al.*, 1975). Jackson *et al.* (1986) indicaron que la interdependencia entre fotosíntesis y absorción de nitrógeno es muy compleja y va más allá de la simple disponibilidad de energía, involucrando al parecer un factor, o factores, que modulan la utilización de la energía en el proceso de absorción y transporte del nitrógeno.

Un punto importante en la mayoría de los programas de mejoramiento de la soya lo constituye el incremento del rendimiento y del contenido de proteína en la semilla (Hartwig, 1969; Shibles *et al.*, 1975). Sin embargo, la

presencia de interrelaciones muy complejas entre los procesos fotosintético y de asimilación de nitrógeno han causado que hasta el momento esta labor no haya tenido un avance espectacular.

A través de una buena cantidad de años se han llevado a cabo muchos estudios sobre mejoramiento en la soya. Los enfoques básicos han sido incremento del rendimiento, incremento de la cantidad de proteína en la semilla o una mezcla de ambos. Los resultados de muchos estudios indican la presencia de una relación inversa entre el rendimiento y el contenido de proteína en la semilla (Kwon y Torrie, 1964; Hartwig, 1969).

✓ En un artículo clásico, Sinclair y DeWit (1975) publicaron los resultados de un estudio acerca de la relación entre el rendimiento de grano y el contenido de proteína en la semilla. La conclusión presentada fue que no es posible el incremento simultáneo en estos dos caracteres por ser incompatibles desde un punto de vista energético. De acuerdo a los mencionados autores la incompatibilidad tendría como base la competencia por cetoácidos y por la energía derivada de los fotosintatos entre los sistemas de síntesis de carbohidratos y de proteínas; un impedimento adicional sería que un incremento en la concentración de proteína en la semilla requeriría mayor cantidad de N absorbido del suelo, proceso cuyo costo energético es relativamente alto.

Sinclair y DeWit (1975) marcan a la soya como una especie única en cuanto a sus características de composición de la semilla así como en cuanto a sus requerimientos energéticos y de nitrógeno. Esta singularidad llevó a los autores a lanzar la hipótesis de que la soya es una especie "autodestructiva", es decir, que requiere movilizar grandes cantidades de nitrógeno del tejido foliar hacia las partes reproductivas durante el período de llenado de las semillas para cubrir los fuertes requerimientos de nitrógeno en estas últimas. El posible impacto de esta característica fue probado utilizando un modelo dinámico y los resultados fueron publicados en 1976 (Sinclair y DeWit, 1976). El resultado más relevante fue el que la fuente principal de variación en el rendimiento es la duración del desarrollo de las semillas y que para conseguir alargar este período no basta con incrementar la tasa y eficiencia fotosintética, sino asociar este incremento con un aumento en la asimilación de nitrógeno. Otro resultado interesante derivado del uso del modelo dinámico fue que al aumentar la tasa fotosintética, sin asociar este aumento con una mayor tasa de incorporación de N, se obtenía una disminución en el rendimiento, a causa de que el nitrógeno necesario era movilizado a partir del tejido foliar y esta movilización causaba una rápida senescencia. Recientemente, Salado-Navarro *et al.* (1985) obtuvieron resultados que apoyan la hipótesis de Sinclair y DeWit (1975, 1976).

Resultados análogos fueron reportados por Westermann *et al.* (1985) en un estudio realizado con varios cultivares de frijol (*Phaseolus vulgaris* L.). Los datos de estos autores indican que tanto la actividad fotosintética como la actividad de fijación de nitrógeno durante el llenado de las semillas (periodos R6 a R9) tienen una influencia significativa sobre el rendimiento y la concentración final de nitrógeno en las semillas.

Al contrario de Sinclair y DeWit (1975, 1976), Hanson *et al.* (1961) indicaron que existe la posibilidad de conseguir incrementos tanto en contenido de proteína como en rendimiento de grano. Según estos autores, no existe una contradicción energética inherente entre ambos caracteres y la relación inversa resulta más bien de restricciones ambientales y fisiológicas. Aunque existe controversia a este respecto, la tendencia a través de los años es aceptar como correcto lo marcado por Hanson *et al.* (1961).

En varios estudios, tanto en leguminosas como en cereales, se ha encontrado que la relación inversa entre el rendimiento y el contenido de proteína se debe sobre todo a factores ambientales y fisiológicos, más que a factores energéticos o genéticos *per se*. Adams (1967) y Bhatia y Rabson (1976) indicaron que las correlaciones negativas entre diferentes fracciones o componentes del rendimiento surgen bajo varios tipos de estrés ambiental y que se derivan de la acción de sistemas genéticamente independientes pero que interactúan uno con otro a través

del desarrollo de la planta. Johnson et al. (1969) mencionaron que la relación inversa entre rendimiento y concentración de proteína en el grano no es constante y que en general depende de la fertilidad del suelo, de la disponibilidad de agua y otros factores ambientales. En el caso de los cereales, Kibite y Evans (1984) y Cox et al. (1985), en estudios realizados con varias líneas y cultivares de trigo (*Triticum aestivum* L.), encontraron que las correlaciones genéticas entre rendimiento de grano y contenido de proteína del mismo, aunque siempre negativas, generalmente fueron de valor bajo pudiendo ser significativas o no. Las conclusiones obtenidas en esos estudios fueron que existe la posibilidad de seleccionar líneas con alto rendimiento y alto contenido de proteína en las semillas (Cox et al., 1985) y que la relación inversa entre rendimiento de grano y concentración de proteína en las semillas, aunque fenotípicamente real, no fue causada al parecer por factores genéticos (Kibite y Evans, 1984).

En un artículo reciente Sinclair y Horie (1989) anotaron las diferencias encontradas en cuanto a un carácter llamado Uso Eficiente de la Radiación (RUE) con relación al contenido de nitrógeno por unidad de área foliar. Las diferencias encontradas separaron consistentemente a las especies C_3 de las C_4 , siendo estas últimas las más eficientes en el uso de la radiación solar en relación a un contenido dado de nitrógeno foliar. De las especies estudiadas el maíz (*Zea mays* L.) y la soya fueron la más

eficiente y la menos eficiente, respectivamente, en cuanto a la utilización de energía solar.

Muchos estudios llevados a cabo en los últimos años remarcan las relaciones existentes entre los procesos de fotosíntesis y asimilación de nitrógeno, así como los indicativos fisiológicos o morfológicos de dichas relaciones. Paralelamente, los criterios para seleccionar cultivares de una especie dada han ampliado su cobertura hasta incluir características fisiológicas y morfológicas (Evans, 1975).

Blad y Baker (1972) en un estudio con varios cultivares de soya, encontraron que las plantas con hojas pequeñas y orientación tendiente a ser vertical permitían una penetración más eficaz de la luz hacia las partes bajo el dosel lo cual se traducía en una saturación lumínica más homogénea. Estos autores concluyeron que la selección para obtener cultivares con mayor tasa fotosintética podía realizarse con base en el carácter tamaño y orientación de las hojas. Egli *et al.* (1989), en un estudio llevado a cabo en la soya, encontraron que la capacidad de rendimiento de un genotipo dado se relaciona fuertemente con el número de células contenidas en los cotiledones. A su vez, este carácter es influido de manera importante por el ambiente fisiológico presente durante el desarrollo de la semilla, específicamente en la fase cuando ocurren las divisiones celulares. En un estudio con varios cultivares de trigo, Lawlor *et al.* (1989) encontraron que en las plantas

sometidas a deficiencias de nitrógeno los principales efectos fueron la reducción del tamaño de las hojas y la reducción del número y volumen de las células de las mismas. Se encontró también que el contenido de clorofila y proteína foliar disminuyó frente al estrés de nitrógeno.

Algunos factores internos muy complejos, como la tasa de reparto de fotosintatos, se han señalado también como puntos determinantes del rendimiento de proteína y de materia seca en muchas especies de cultivo (Evans, 1975). A este respecto, Gent y Kiyomoto (1939) estudiaron varios cultivares de trigo con diferentes índices de cosecha; se encontró que los altos índices de cosecha se asociaban básicamente con dos características: un reparto mayor de fotosintatos hacia las semillas y una mayor retención de la proteína foliar durante el período de desarrollo de la semilla, con lo cual se retardaba la senescencia. Hanway y Weber (1971) encontraron, en ocho cultivares de soya, diferencias fuertes en la tasa diaria de acumulación de materia seca durante el desarrollo vegetativo. Algo interesante fue que la tasa diaria de acumulación de materia seca en las semillas fue muy similar en todos los cultivares; las diferencias en rendimiento entre los ocho cultivares fueron debidas principalmente a la duración del período de acumulación de materia seca en las semillas.

Como se mencionó, los criterios y métodos propuestos para selección de cultivares han ampliado bastante el espectro de caracteres considerados. Holbrook et al. (1989)

aplicaron exitosamente dos ciclos de selección para incremento de rendimiento y proteína total, manteniendo constante la concentración de proteína en las semillas, utilizando índices de selección con restricción. El uso de caracteres fisiológicos para selección indirecta ha dado resultados menos consistentes. Por ejemplo, Sherrard et al. (1986) y Buttery y Buzzell (1988) reportaron que el uso de caracteres fisiológicos, como el contenido de nitrógeno en las partes foliares, como criterio de selección para alto rendimiento, es poco recomendable a causa de las fuertes influencias ambientales que se observan sobre esos caracteres.

Una alternativa, aplicada con más éxito, es la de realizar selección sobre caracteres morfológicos relacionados de alguna manera con la obtención de rendimientos altos; algo como el estudio ya mencionado de Blad y Baker (1972). Ledent y Moss (1979) realizaron un estudio de este tipo con varios cultivares de trigo utilizando 37 variables morfológicas y el rendimiento de grano. Los resultados del análisis indicaron que cuatro de las 37 variables se relacionaron fuertemente con el rendimiento. En un estudio con soya, Williams et al. (1979) investigaron la relación entre el rendimiento de grano y cinco variables, cuatro de ellas variables fenológicas. Los resultados obtenidos por estos investigadores indicaron que la fecha de siembra, la duración del período de desarrollo de la semilla y el peso de 200 semillas fueron, entre los

considerados, los más fuertemente relacionados con el rendimiento de grano. Carbonell Guevara y Bartual Pastor (1983) llevaron a cabo un estudio sobre las relaciones entre 25 variables, fenológicas y morfológicas, y el rendimiento y calidad de la semilla en 125 líneas de soya sembradas en dos años consecutivos. Los resultados de los autores indicaron que los caracteres relacionados fuertemente con el rendimiento de grano fueron la longitud del período de desarrollo de la semilla, el número de nudos reproductivos, la altura de la planta y la altura en la que aparece el primer fruto en la madurez. Por otro lado, los caracteres relacionados con el contenido de proteína en las semillas fueron: longitud del período vegetativo, tamaño y calidad de la semilla, diámetro foliar en el momento de la formación de la semilla y encamado temprano.

MATERIALES Y METODOS

Material Biológico

El material biológico utilizado en el presente trabajo fue un grupo de cultivares y líneas avanzadas proporcionadas por el Ing. Mario Cruz Fernández, Responsable del Programa de Soya en el Campo Experimental Ebano adscrito al CIFAP Región Pánuco. Los materiales utilizados fueron los siguientes:

1. Tapachula 86.
2. Santa Rosa.
3. Júpiter.
4. UFV-1.
5. H80-25-35.
6. H80-25-39.
7. H82-16-71.
8. H86-51-50.

Sin embargo, debido a una serie de problemas, seguramente de adaptación, los datos del cultivar Tapachula 86 no fue posible obtenerlos. El suelo utilizado fue uno con gran cantidad de mantillo y fue colectado en una zona forestal cercana a Jaguey de Ferniza, Municipio de Saltillo, Coahuila. El análisis del suelo, llevado a cabo en el

Laboratorio de Calidad de Aguas del Departamento de Riego y Drenaje de la Universidad Autónoma Agraria Antonio Narro, indicó como características relevantes un contenido alto de materia orgánica y nitrógeno, un pH de seis punto siete y un contenido bajo de hierro.

Procedimiento Experimental

El trabajo experimental se llevó a cabo en uno de los invernaderos de la Universidad Autónoma Agraria Antonio Narro durante el año de 1989. El suelo utilizado se cribó para eliminar hojas, ramas y demás componentes mayores del suelo y con el material remanente se llenaron macetas de material plástico. La fecha de siembra fue el 22 de abril y la fecha de cosecha fue variable de acuerdo con la línea o cultivar. La cosecha consistió en el total de material reproductivo más el vegetativo, excepto la raíz, y se llevó a cabo un poco antes del período R7, según lo definen Fehr *et al.* (1971). Como fertilizante se aplicó solución nutritiva de Hogland; la solución se aplicó un total de cuatro veces, 25 mililitros por maceta en cada ocasión, al presentarse indicios de clorosis en las plantas.

El diseño experimental utilizado fue el completamente aleatorio con cuatro repeticiones. Cada cultivar o línea se consideró un tratamiento y cada maceta, con aclareo a cuatro plantas, se consideró una repetición.

Para cada una de las plantas se obtuvieron datos acerca de 18 variables de tipo morfológico, fenológico, de peso seco y de contenido de nitrógeno. Las variables consideradas fueron las siguientes:

1. X_1 : días a R1. El número de días transcurridos desde la fecha de siembra hasta el período R1 según Fehr *et al.* (1971).
2. X_2 : días a R6-R7. El número de días transcurridos desde la fecha de siembra hasta un poco antes del período R7 según Fehr *et al.* (1971).
3. X_3 : altura de planta. La distancia lineal desde el nivel superior de la tierra de la maceta hasta la parte más alta de la planta sin considerar si el tallo es recto o no.
4. X_4 : número de nudos. El número de nudos presentes en la fecha de la cosecha.
5. X_5 : número de nudos reproductivos. El número de nudos con una o más vainas en la fecha de la cosecha.
6. X_6 : número de vainas. El número total de vainas presentes hasta la fecha de la cosecha.
7. X_7 : número de trifoliolos. El número total de trifoliolos presentes en la fecha de la cosecha. No se consideró material foliar caído.
8. X_8 : número de ramificaciones. El número total de ramificaciones presentes hasta la fecha de la

cosecha. Se consideró una ramificación cualquier derivación del tallo principal con dos o más trifoliolos.

9. X_p : número de semillas. El número de semillas cosechadas por planta.

10. X_{10} : longitud foliar acumulada. La distancia lineal desde la base hasta el ápice de la lámina foliar siguiendo la nervadura central. La suma de este valor para todos los foliolos en una planta constituyó la longitud foliar acumulada para esa planta.

11. X_{11} : ancho foliar acumulado. La distancia lineal desde un borde a otro de la lámina foliar en la parte más ancha de la misma. La suma de este valor para todos los foliolos de una planta constituyó el ancho foliar acumulado para esa planta.

12. X_{12} : porcentaje de germinación de las semillas progenitoras. El porcentaje promedio de germinación obtenido en un ensayo de germinación en laboratorio con una muestra de las semillas progenitoras.

13. X_{13} : peso seco de las vainas. El peso seco de las vainas separadas de las semillas. El material vegetal se secó en una estufa a una temperatura de 70 grados centígrados hasta alcanzar peso constante.

14. X_{14} : peso seco de las semillas. El peso seco de las semillas cosechadas. El tratamiento de secado fue el mismo indicado para las vainas.
15. X_{15} : peso seco del tallo, ramificaciones y partes foliares. El peso seco de todo el material vegetativo excepto la raíz. El tratamiento de secado fue el mismo indicado para las vainas.
16. X_{16} : peso promedio de las semillas progenitoras. El peso promedio de una semilla del lote de semillas progenitoras obtenido a partir de una muestra aleatoria.
17. X_{17} : porcentaje de nitrógeno en las semillas. El contenido porcentual de nitrógeno en una muestra representativa del lote de las semillas cosechadas. Este dato se obtuvo, cuando la cantidad de semilla fue suficiente, a partir del porcentaje de proteína (considerando el factor 6.25) con un aparato analizador de calidad de granos marca NEOTEC GOA-31EL o bien, cuando se dispuso de poca semilla, realizando un análisis Kjeldahl.
18. X_{18} : porcentaje de nitrógeno en el tallo, ramificaciones y partes foliares. El contenido porcentual promedio de nitrógeno en el tejido vegetativo excepto la raíz. Este dato se obtuvo de la forma indicada para la característica anterior.

Los análisis químicos necesarios para la obtención de los datos de porcentaje de nitrógeno se realizaron en el Laboratorio de Análisis Químicos del Departamento de Fitomejoramiento de la Universidad Autónoma Agraria Antonio Narro.

Análisis Estadístico

Con acuerdo a los objetivos del trabajo el análisis estadístico se dividió en dos partes principales; en cada una de esas partes los datos se consideraron de diferente manera. Para investigar si se presentaron diferencias significativas en el contenido porcentual de nitrógeno en las semillas se consideró el promedio de esta característica para cada maceta. En este caso también se incluyó, como información potencialmente útil, el carácter peso seco de las semillas cosechadas. En cambio, para investigar las relaciones entre los dos caracteres anteriores entre si y con las restantes 13 variables se utilizaron los datos de plantas individuales.

En conjunto, el análisis de los datos consistió de lo siguiente:

1. Análisis gráfico de cada carácter considerado. Es útil como primer acercamiento y permite detectar fácilmente desviaciones de una distribución como la normal. El método utilizado fue graficar la frecuencia acumulada para el carácter considerado en una gráfica de probabilidad normal. La desviación de una línea recta indica desviación de una distribución como la normal.

2. Análisis gráfico bivariado. Consiste en graficar los dos caracteres considerados como más importantes, el contenido porcentual de nitrógeno en la semilla y el peso seco de las semillas, contra los restantes caracteres. El análisis de la dispersión en cada gráfica permite saber si es necesaria o no la inclusión de términos de segundo orden o mayor, o bien si es necesario el uso de coeficientes no lineales en un modelo de regresión. El análisis gráfico bivariado permite también el análisis de los caracteres en una serie cronológica.

3. Análisis gráfico tridimensional. Se llevó a cabo de dos maneras. La primera, para aquellos casos en que es importante conocer la correlación lineal simple entre dos caracteres, se utilizó un histograma en tres dimensiones el cual permite detectar si la distribución conjunta de las variables es como una normal bivariada. El segundo caso fue el anotar los valores de las variables peso seco de las semillas y porcentaje de nitrógeno en las semillas contra los valores de cada una de las restantes variables en un espacio con tres ejes coordenados. La gráfica resultante permite el conocer la distribución conjunta de las dos variables consideradas como más importantes (el porcentaje de nitrógeno en las semillas y el

peso seco de las mismas) en conjunto con cada una de las otras variables.

4. Análisis estadístico específico según el objetivo a considerar. Para el primer objetivo del trabajo el análisis estadístico consistió en una prueba de Kruskal-Wallis. Para el segundo y tercer objetivos el análisis estadístico fue múltiple y consistía originalmente en la obtención de una matriz de distancias de Mahalanobis (D^2) y prueba sobre las distancias con la T^2 de Hotelling. Esta parte del análisis, sin embargo, se eliminó ya que una prueba sobre los datos para averiguar si su distribución conjunta era como una normal multivariada reveló una desviación muy fuerte. En lugar de la D^2 de Mahalanobis se utilizó una gráfica con las dos primeras componentes principales para visualizar la dispersión de los individuos. Otra parte del análisis para el segundo y tercer objetivos fue el uso de regresión lineal múltiple en sus modalidades Gauss-Markov y *Ridge* con selección de variables "importantes". Se consideró también la obtención de una matriz de correlaciones y la realización de una regresión múltiple sobre la misma.

El análisis de los datos se llevó a cabo en el Centro de Cómputo del Departamento de Fitomejoramiento de la Universidad Autónoma Agraria Antonio Narro utilizando el

paquete estadístico de Hintze (1985) y el de STSC, Inc. and Statistical Graphics Corporation (1987).

Descripción de las Técnicas Estadísticas

Prueba de Kruskal-Wallis.

Esta prueba es descrita por Hollander y Wolfe (1973). La prueba es apropiada para un diseño completamente aleatorio y se utilizó en lugar de un análisis de varianza de una vía por presentar los datos problemas fuertes de heterogeneidad de varianzas y desviación de una distribución como una normal. La prueba se aplicó utilizando datos de promedios de parcelas. El procedimiento de la prueba indica asignar rangos a todas las observaciones en orden ascendente, obtener la suma de los rangos para cada tratamiento y calcular un estadístico de prueba para la hipótesis nula de que los tratamientos tienen la misma localización. El estadístico de prueba es:

$$H = \frac{12}{n(n+1)} \sum_j \frac{R_j^2}{n_j} - 3(n+1)$$

en donde n_j es el número de observaciones en el j -ésimo tratamiento, R_j es la suma de los rangos para el j -ésimo tratamiento y n es la suma de los n_j , es decir, el total de observaciones. Cuando los n_j son pequeños el estadístico H se distribuye como una χ^2 con $l-1$ grados de libertad, en donde l es el número de tratamientos.

Obtención de los Estimadores de la Matriz de Covarianzas y de la Matriz de Correlaciones

Si X es una matriz de orden $(n \times k)$, en donde n es el número de observaciones y k es el número de variables consideradas, y la matriz contiene datos corregidos para la media, entonces el estimador de la matriz de covarianzas es:

$$S = \left(\frac{1}{n-1} \right) X'X$$

y el estimador de la matriz de correlaciones es:

$$R = \left(\frac{1}{n-1} \right) X'X$$

cuando X es una matriz de datos estandarizados.

Obtención de las Correlaciones Parciales

Sea C la inversa de la matriz de correlaciones. El resultado de la operación siguiente, utilizando los elementos de la matriz C :

$$\frac{-C_{jl}}{(C_{jj}C_{ll})^{1/2}}$$

es la correlación parcial entre la j -ésima y la l -ésima variable.

Prueba de Normalidad Multivariada

La prueba es descrita por Senn (1989) y se realiza utilizando el escalar resultante del siguiente producto:

$$Z'S^{-1}Z$$

en donde Z es un vector de medias estandarizadas de orden $(k \times 1)$, k es el número de variables consideradas y S es el estimador de la matriz de covarianzas de orden $(k \times k)$. El escalar resultado del producto tiene, cuando la hipótesis nula de que los datos se distribuyen como una normal multivariada es cierta, una distribución χ^2 con k grados de libertad.

Análisis de Componentes Principales

Esta técnica es descrita por Kendall y Stuart (1976).

A los escalares que satisfacen la ecuación característica:

$$|ZI - R| = 0$$

se les denomina valores característicos de R . En esta igualdad I es una matriz identidad de orden k , Z es una matriz diagonal que contiene a los escalares λ_j , $j=1, \dots, k$, que son los valores característicos de R , R es una matriz de correlaciones de orden $(k \times k)$ y 0 es un vector nulo de

orden $(k \times 1)$. El desarrollo del determinante anterior permite obtener el polinomio característico de R cuyas raíces son, como se mencionó, los valores característicos de R . Por otro lado, los vectores que satisfacen la siguiente igualdad:

$$(ZI - R)L = 0$$

se les denomina vectores característicos de R . En este caso L es una matriz de coeficientes de orden $(k \times k)$.

La primera y segunda componentes principales se obtienen de la siguiente forma:

$\xi_1 = Xl_1$, es la primera componente principal,

$\xi_2 = Xl_2$, es la segunda componente principal,

en donde l_1 y l_2 son, respectivamente, el primero y segundo vectores columna de L y X es una matriz de datos estandarizados de orden $(n \times k)$.

Regresión Lineal Múltiple

El modelo lineal de primer orden es:

$$Y = X\beta + \varepsilon$$

en donde Y es un vector de valores de la variable de respuesta, o variable dependiente, de orden $(n \times 1)$, X es la matriz de datos de las variables independientes de orden $(n$

$x(k+1)$, β es un vector de orden $((k+1) \times 1)$ que contiene los valores de los coeficientes de regresión y ε es un vector de desviaciones aleatorias de orden $(n \times 1)$.

El vector de estimaciones de los coeficientes de regresión, b , obtenido a partir de un ajuste para cuadrados mínimos, se define como sigue:

$$b = (X'X)^{-1}X'Y$$

en donde $X'X$ es una matriz que se supone de rango completo.

Existe la posibilidad de reducir la dimensionalidad del hiperplano de ajuste de la regresión disminuyendo el orden de la matriz X de $(n \times (k+1))$ a $(p \times (k+1))$, $p < k$. Para ello se aplicó, además de otros métodos, un procedimiento automático conocido como *selección backward*. El algoritmo tiene como base la aplicación del principio de la suma de cuadrados extra así como la utilización de pruebas parciales de F . El tema es tratado con mayor detalle por Draper y Smith (1981).

La obtención de las matrices de covarianzas de los residuales y de covarianzas de los coeficientes de regresión parcial, necesarias para las pruebas de t sobre estos últimos, son tratadas en detalle por Draper y Smith (1981).

Por otro lado, el vector de estimadores sesgados, los coeficientes *ridge*, se obtiene de la siguiente manera:

$$b^R = (X'X + wI)^{-1}X'Y \quad , \quad w \geq 0$$

en donde la matriz X y el vector Y se encuentran en forma estandarizada. I es una matriz identidad de orden k y w es un coeficiente de sesgo. La regresión ridge es la forma recomendada para aquellos casos en que la matriz $X'X$ no es de rango completo. En el escrito de Hoerl y Kennard (1970) se revisa este tema con mayores detalles.

Regresión Sobre la Matriz de Correlaciones

El llevar a cabo una regresión sobre la matriz de correlaciones permite seleccionar variables importantes con acuerdo al valor absoluto de los coeficientes de regresión. También permite aminorar problemas no muy fuertes de multicolinealidad.

La estimación de los coeficientes de regresión parcial estandarizados se consigue con:

$$b^* = R^{-1}D$$

en donde R es una matriz de correlaciones entre las variables independientes, D es un vector de correlaciones, de orden $(k \times 1)$, entre la variable dependiente y las variables independientes y b^* es un vector de coeficientes de orden $(k \times 1)$. Puede verse fácilmente que $b^* = b^R$ para el caso de $w = 0$.

RESULTADOS

Las gráficas de probabilidad normal indicaron problemas de desviación de una distribución como la normal en todas las variables excepto número de nudos, número de nudos reproductivos y número de ramificaciones. El problema fue especialmente fuerte en las variables peso seco de las semillas, longitud y ancho foliar acumulados, peso seco de las vainas y peso seco del tallo, ramificaciones y partes foliares.

En el análisis gráfico bidimensional se incluyeron las gráficas de las variables peso seco de las semillas y porcentaje de nitrógeno en las semillas contra las restantes variables. En ninguna de las gráficas se encontró indicación de la presencia de relaciones no lineales. Es interesante sobre todo la gráfica en donde aparece la variable porcentaje de nitrógeno en las semillas contra el peso seco de las semillas (Figura 1) en donde se aprecia la presencia de tres grupos de plantas: uno en donde se tienen valores bajos para ambas variables, otro en el cual se tienen valores altos de contenido porcentual de nitrógeno pero valores bajos del peso seco de las semillas y otro en el cual ambas variables toman valores altos.

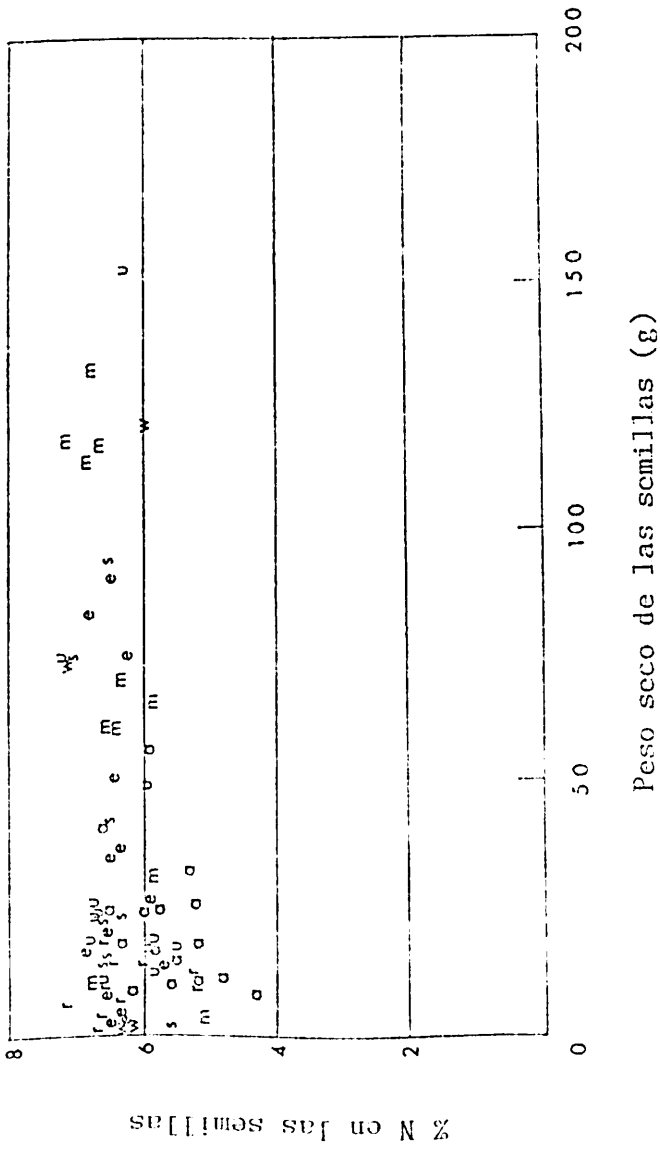
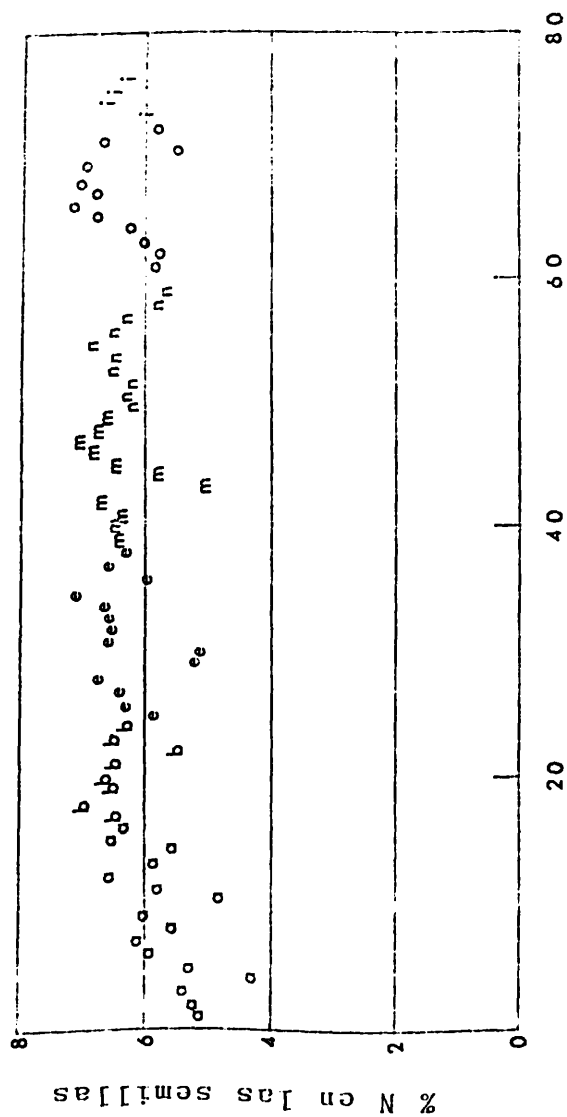


Figura 1. Gráfica de los valores de porcentaje de nitrógeno en las semillas contra el peso seco de las mismas. (a) H80-2535, (s) H80-2639, (r) H82-1671, (m) H86-5150, (c) Santa Rosa, (u) UFV-1, (w) Júpiter.

El análisis gráfico bidimensional constó también del acomodo de los datos de porcentaje de nitrógeno en las semillas y de peso seco de las mismas en forma de una serie cronológica utilizando como índice temporal los datos de días a R1 (Figuras 2.A y 2.B). En la figura 2.A, correspondiente a la variable porcentaje de nitrógeno en las semillas, puede observarse que se presentó bastante dispersión de los valores en relación a su media. A pesar de este hecho es clara una tendencia ligera al incremento del porcentaje de nitrógeno en las semillas conforme se incrementa el período desde la fecha de siembra hasta R1. En la Figura 2.B, que corresponde a la variable peso seco de las semillas, se presentó el mismo caso de una dispersión muy fuerte de los datos con respecto a su media: en esta oportunidad, sin embargo, no se apreció una tendencia temporal obvia.

En el análisis gráfico tridimensional en su primera variante se realizó un histograma en tres dimensiones para las variables porcentaje de nitrógeno en las semillas y peso seco de las semillas (Figura 3). puede apreciarse en esta gráfica que existe una fuerte desviación de la dispersión esperada para una distribución normal bivariada. En la otra variante del análisis gráfico tridimensional se encontró que la dispersión de los datos en un espacio generado con tres ejes coordenados, con las variables porcentaje de nitrógeno en las semillas y peso seco de las semillas en dos de los ejes contra las restantes variables colocadas de manera



Secuencia de Individuos Considerando Días a R1

Figura 2A. Serie cronológica de los datos de porcentaje de N en las semillas. El índice temporal es la variable días a R1. (a) H80-2535, (b) H80-2639, (e) Santa Rosa, (m) H86-5150, (n) H82-1671, (o) UFV-1, (i) Júpiter.

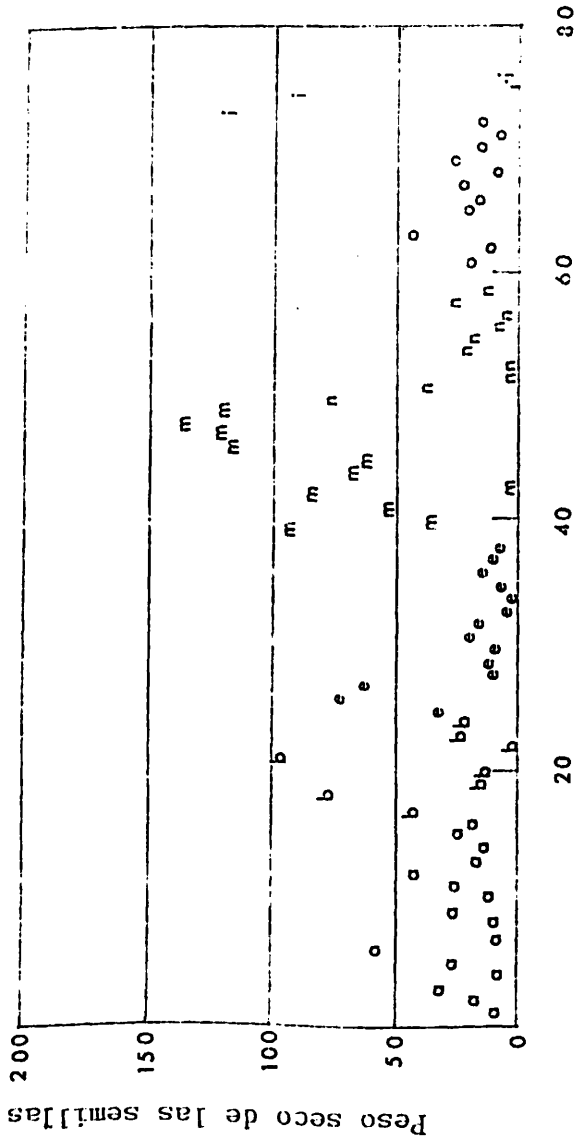


Figura 2B. Serie cronológica de los datos de peso seco de las semillas. El índice temporal es la variable días a RI. (a) H80-2535, (b) H80-2639, (c) Santa Rosa, (m) H86-5150, (n) H82-1671, (o) UFV-1, (i) Júpiter.

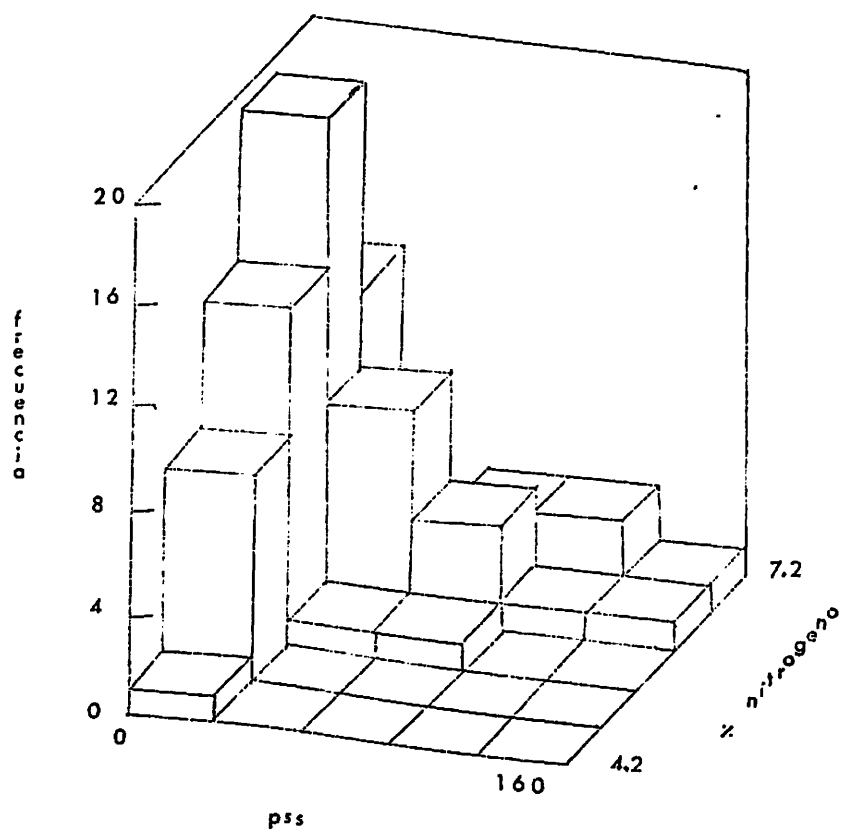


Figura 3. Histograma en tres dimensiones que muestra la distribución conjunta de las variables porcentaje de nitrógeno en las semillas y peso seco de las semillas.

consecutiva en el tercer eje, tuvo por lo general dependencia de los valores de la variable peso seco de las semillas, no ocurriendo así con la variable porcentaje de nitrógeno en las semillas. En buena parte de los casos se observó que la dispersión daba lugar a un cierto número de grupos de individuos, de dos a seis, con diferente comportamiento respecto a las variables consideradas.

En cuanto a las pruebas de diferencias entre cultivares, además de incluir a la variable porcentaje de nitrógeno en las semillas, correspondiente al primer objetivo del trabajo, se consideró también la variable peso seco de las semillas por ser la misma una fuente de información potencialmente útil. En el Cuadro 4.1 se anotan los valores promedio por maceta para el porcentaje de nitrógeno en las semillas para cada línea o cultivar. En el Cuadro 4.2 se anota esto mismo pero para el carácter peso seco de las semillas. En la parte inferior de los cuadros mencionados aparece el valor del estadístico H de Kruskal-Wallis y la probabilidad asociada con dicho valor. Puede apreciarse, a partir de las probabilidades de los valores de H , que para ambas variables no se presentaron diferencias entre los cultivares o líneas que pudieran marcarse como significativas. Sin embargo, la prueba de Kruskal-Wallis detecta únicamente diferencias globales entre los grupos considerados por lo que, para detectar posibles diferencias individuales, se aplicó la prueba de Mann-Whitney (Hollander y Wolfe, 1973) la cual es parecida

Cuadro 4.1. Valores promedio del contenido porcentual de nitrógeno por maceta para cada línea o cultivar.

Tratamiento						
H80-2535	H80-2639	H82-1671	H86-5150	Sta. Rosa	UFV-1	Júpiter
4.9925	6.8125	5.6900	5.4500	6.2200	5.9680	6.3375
5.7350	6.7300	6.7275	6.4900	6.5175	6.8100	
5.8000	6.6250	5.9300	6.8425	6.0825	6.6150	
6.0650	6.2050	6.4938	6.3350	6.5100	6.2150	
Promedio						
5.6481	6.5931	6.2227	6.2919	6.3325	6.4013	6.3375
$H = 8.4592$ $P(X^2: 8.4592) = 0.2057$						

Cuadro 4.2. Valores promedio de peso seco de semillas (en gramos) por maceta para cada línea o cultivar.

Tratamiento						
H80-2535	H80-2639	H82-1671	H86-5150	Sta. Rosa	UFV-1	Júpiter
16.9000	22.7515	13.4000	35.2024	56.3000	58.1121	53.2337
25.4250	12.3235	6.7827	61.6281	11.0022	20.0274	
25.8250	37.5788	14.5581	121.7919	13.0259	22.1165	
18.7500	36.6250	29.3131	43.6280	35.8032	18.3335	
Promedio						
21.7250	27.3197	16.0142	65.5626	36.5331	29.6474	53.2337
$H = 8.4592$ $P(X^2: 8.6262) = 0.1957$						

a la de Kruskal-Wallis pero aplicada a dos tratamientos o muestras. En el Cuadro 4.3 se anotan los resultados para el carácter porcentaje de nitrógeno en las semillas y en el Cuadro 4.4 los resultados para el carácter peso seco de las semillas. En general, en ambos casos, se presentaron pocas diferencias entre cultivares o líneas, siendo la mayor parte de ellos estadísticamente homogéneos para las dos variables consideradas.

En consideración al segundo y tercer objetivos se llevó a cabo una prueba de distribución normal multivariada. El resultado mostró una desviación muy fuerte de lo que se esperaría bajo esta distribución. El valor de χ^2 fue de 2.6222×10^6 con 18 grados de libertad, valor al que le corresponde una probabilidad extremadamente baja. Tal como se mencionó, esta desviación tan fuerte obligó a un cambio en los planes originales de análisis, cambio que resultó en llevar a cabo una ordenación de los individuos utilizando componentes principales en lugar de realizar la ordenación con la matriz de distancias de Mahalanobis, método estadísticamente más potente.

Para investigar las relaciones entre las variables se hizo uso de la siguiente serie de análisis: de correlación lineal simple, de correlación parcial, de regresión lineal múltiple con selección de variables, de regresión *ridge* con selección subjetiva de variables y de regresión lineal múltiple sobre la matriz de correlaciones. Los resultados de cada uno de ellos se conjuntaron

Cuadro 4.3. Resultados de la prueba de Mann-Whitney para el carácter contenido porcentual de nitrógeno en las semillas. En la tabla se anotan los pares de líneas o cultivares comparados, el valor del estadístico T de Mann-Whitney y el valor Z asociados a este valor de T. Las probabilidades marcadas con un asterisco son las consideradas significativas.

Cultivares comparados	T	Z	P($> Z $)
H80-2535 vs. H80-2639	0	-2.3094	0.0209*
H80-2535 vs. H82-1671	4	-1.1547	0.2482
H80-2535 vs. H86-5150	3	-1.4434	0.1489
H80-2535 vs. Santa Rosa	0	-2.3094	0.0209*
H80-2535 vs. UFV-1	1	-2.0207	0.0433*
H80-2535 vs. Júpiter	0	-1.4142	0.1573
H80-2639 vs. H82-1671	13	1.4434	0.1489
H80-2639 vs. H86-5150	10	0.5774	0.5637
H80-2639 vs. Santa Rosa	13	1.4434	0.1489
H80-2639 vs. UFV-1	11	0.8560	0.3935
H80-2639 vs. Júpiter	3	0.7071	0.4795
H82-1671 vs. H86-5150	8	0	1.0000
H82-1671 vs. Santa Rosa	6	-0.5774	0.5637
H82-1671 vs. UFV-1	6	-0.5774	0.5637
H82-1671 vs. Júpiter	2	0	1.0000
H86-5150 vs. Santa Rosa	8	0	1.0000
H86-5150 vs. UFV-1	8	0	1.0000
H86-5150 vs. Júpiter	2	0	1.0000
Santa Rosa vs. UFV-1	7	-0.2887	0.7728
Santa Rosa vs. Júpiter	2	0	1.0000
UFV vs. Júpiter	2	0	1.0000

Cuadro 4.4. Resultados de la prueba de Mann-Whitney para el carácter peso seco de las semillas. En la tabla se anotan los pares de líneas o cultivares comparados, el valor del estadístico T de Mann-Whitney y el valor Z asociado a este valor de T. Las probabilidades marcadas con un asterisco son las consideradas significativas.

Cultivares comparados	T	Z	P(> Z)
H80-2535 vs. H80-2539	6	-0.5774	0.5637
H80-2535 vs. H82-1671	12	1.1547	0.2482
H80-2535 vs. H86-5150	0	-2.3094	0.0209 *
H80-2535 vs. Santa Rosa	8	3.4641	6×10^{-4} *
H80-2535 vs. UFV-1	7	-0.2887	0.7728
H80-2535 vs. Júpiter	0	-1.4142	0.1573
H80-2639 vs. H82-1671	12	1.1547	0.2482
H80-2639 vs. H86-5150	2	-1.7321	0.0833 *
H80-2639 vs. Santa Rosa	7	-0.2887	0.7728
H80-2639 vs. UFV-1	9	0.2887	0.7728
H80-2639 vs. Júpiter	0	-1.4142	0.1573
H82-1671 vs. H86-5150	0	-2.3094	0.0209 *
H82-1671 vs. Santa Rosa	6	-0.5774	0.5637
H82-1671 vs. UFV-1	3	-1.4434	0.1489
H82-1671 vs. Júpiter	0	-1.4142	0.1573
H86-5150 vs. Santa Rosa	11	0.8660	0.3885
H86-5150 vs. UFV-1	14	1.7321	0.0833 *
H86-5150 vs. Júpiter	2	0	1.0000
Santa Rosa vs. UFV-1	7	-0.2887	0.7728
Santa Rosa vs. Júpiter	2	0	1.0000
UFV vs. Júpiter	1	-0.7070	0.4795

para comparación y se señalaron cuales de las variables investigadas aparecieron como más importantes en la determinación del contenido porcentual de nitrógeno en las semillas. De nuevo se consideró la variable peso seco de las semillas como una variable de respuesta por considerarse una fuente de información potencialmente útil.

La matriz de correlaciones lineales simples se anexa en el Apéndice del escrito en el Cuadro 1. Lo mismo se llevó a cabo con la matriz de correlaciones parciales, la cual aparece en el Cuadro 2 del Apéndice. Lo que se anota a continuación, en los Cuadros 4.5 y 4.6 son las correlaciones lineales simples y las correlaciones parciales que aparecieron como significativas.

Para el caso de la variable porcentaje de nitrógeno en las semillas, el número de variables cuyo coeficiente de correlación lineal simple con la misma fue marcado como significativo fue de nueve, mientras que en el caso de la correlación parcial se tuvieron únicamente cuatro variables en esta situación. Es de notarse el caso de la variable X_{10} , el porcentaje de nitrógeno en las partes vegetativas, la cual mostró un coeficiente de correlación lineal simple marcado como no significativo y que, sin embargo, apareció como la variable con mayor coeficiente de correlación parcial. Para la variable peso seco de las semillas se presentó una situación análoga: el número de variables anotadas con un coeficiente de correlación lineal simple significativo fue de 12, mientras que el número de variables

Cuadro 4.5. Listado de los valores marcados como significativos de la correlación lineal simple y la correlación parcial entre la variable contenido porcentual de nitrógeno en las semillas y las restantes características.

Variable	Valor de r^{**}	Valor de r parcial**
1. Dias a R1	0.5725*	-0.0933ns
2. Dias a R6-R7	0.4317*	0.2116*
3. No. de vainas	0.2320*	-0.0736ns
4. No. de semillas	0.2045*	0.1714ns
12. Porcentaje de germinación de la semilla progenitora	-0.2402*	0.1124ns
13. Peso seco de las vainas	0.1918*	-0.2696*
14. Peso seco de las semillas	0.2999*	0.3685*
15. Peso seco del tallo, ramificaciones y partes foliares	0.1999*	-0.0546ns
16. Peso promedio de semilla progenitora	-0.2996*	0.0422ns
18. Porcentaje de nitrógeno en el tallo, ramificaciones y partes foliares	0.0363ns	0.4633*

**Los valores marcados con un asterisco son aquellos para los que la $POD \leq 0.1$.

00155

BANCO DE TESIS

Cuadro 4.8. Listado de los valores marcados como significativos de la correlación lineal simple y la correlación parcial entre la variable peso seco de la semilla y las restantes características.

Variabile	Valor de r**	Valór de r parcial**
1. Días a R1	0.1242	-0.2530*
3. Altura de planta	0.2254*	0.2803*
4. No. de nudos	0.4070*	-0.4429*
5. No. de nudos reproductivos	0.4513*	0.1898
6. No. de vainas	0.9294*	0.0516
7. No. de trifolios	0.7840*	-0.1802
8. No. de ramificaciones	0.5776*	0.1557
9. No. de semillas	0.9504*	0.1894
10. Longitud foliar acumulada	0.8835*	0.1058
11. Ancho foliar acumulado	0.8646*	-0.2744*
13. Peso seco de las vainas	0.9700*	0.5045*
15. Peso seco del tallo, ramificaciones y partes foliares	0.9184*	0.2203*
17. Porcentaje de nitrógeno en la semilla	0.2999*	0.3685*
18. Porcentaje de nitrógeno en el tallo, ramificaciones y partes foliares	-0.0831	-0.2204**

** Los valores marcados con un asterisco son aquellas para los que la $P < 0.1$.

con un coeficiente de correlación parcial significativo fue de ocho, dos de ellas, las variables días a R1 y contenido porcentual de nitrógeno en las partes vegetativas, con un coeficiente de correlación lineal simple marcado como no significativo. Es de notar el valor del coeficiente de correlación lineal simple y del coeficiente de correlación parcial entre la variable peso seco de las semillas y contenido de nitrógeno de las mismas; en ambos casos el valor fue significativo y positivo, aunque de valor bajo.

Son notables también en los Cuadros 4.5 y 4.6 los fuertes cambios en el valor, e incluso en el signo, entre los coeficientes de correlación lineal simple y los coeficientes de correlación parcial para la mayor parte de las variables anotadas. Lo que indica este hecho es que existen una buena cantidad de efectos indirectos entre las variables consideradas como variables de estímulo y por ello el cambio en los valores de correlación con las variables de respuesta, contenido porcentual de nitrógeno en las semillas y peso seco de las mismas, al aplicar la correlación parcial.

La ordenación con componentes principales se llevó a cabo con los valores y vectores característicos obtenidos de la matriz de correlaciones. Los valores característicos se anotan en el Cuadro 4.7; la suma de los recíprocos de los valores característicos indica la presencia de multicolinealidad fuerte, hecho marcado anteriormente por los cambios en el valor y signo de los coeficientes de

Cuadro 4.7. Valores característicos de la matriz de correlaciones.

Valor característico	% de la variación total	Recíproco del valor característico
$\lambda_1 = 8.614018$	47.66	0.1161
$\lambda_2 = 3.880391$	21.56	0.2577
$\lambda_3 = 1.344246$	7.47	0.7439
$\lambda_4 = 0.972821$	5.40	1.0279
$\lambda_5 = 0.949791$	5.28	1.0529
$\lambda_6 = 0.826394$	3.48	1.5964
$\lambda_7 = 0.482125$	2.68	2.0742
$\lambda_8 = 0.356658$	1.98	2.8022
$\lambda_9 = 0.311847$	1.73	3.2067
$\lambda_{10} = 0.191795$	1.07	5.2139
$\lambda_{11} = 9.5849 \times 10^{-2}$	0.53	10.4331
$\lambda_{12} = 8.5020 \times 10^{-2}$	0.47	11.7619
$\lambda_{13} = 3.6998 \times 10^{-2}$	0.21	27.0285
$\lambda_{14} = 2.1817 \times 10^{-2}$	0.12	45.8356
$\lambda_{15} = 1.3121 \times 10^{-2}$	0.07	76.2137
$\lambda_{16} = 5.3086 \times 10^{-3}$	0.05	120.3572
$\lambda_{17} = 5.1760 \times 10^{-3}$	0.03	193.1994
$\lambda_{18} = 3.4072 \times 10^{-3}$	0.02	293.4961
$\sum_{j=1}^k \lambda_j = 18$	100	$\sum_{j=1}^k \frac{1}{\lambda_j} = 796.4176$

correlación parcial en relación a los coeficientes de correlación lineal simple. Los valores en la parte central del Cuadro 4.7 son las relaciones entre el valor de cada raíz característica y la suma de todos ellos e indican la cantidad porcentual de la varianza total que es "asimilada" por cada una de las funciones lineales llamadas componentes principales. Para el caso de las dos primeras componentes principales, que son las utilizadas en la ordenación de los individuos, el porcentaje de varianza acumulado entre ambas es de 69.4 por ciento, cantidad bastante aceptable.

Los valores de las dos primeras componentes principales se anotan en el Cuadro 4.8. Estos valores son útiles para realizar una discriminación gráfica de los individuos utilizando las componentes principales como valores índice. En la Figura 4 puede apreciarse la distribución de los individuos así como las agrupaciones subjetivas realizadas.

En la gráfica de las dos componentes principales son patentes varios hechos:

1. La primera componente principal es poco útil como discriminante entre las diferentes líneas y cultivares. En cambio, la segunda componente principal separa el conjunto de individuos básicamente en tres grupos: uno de ellos, en la parte de abajo y a la derecha es la línea H20-25-35; el segundo grupo es un conglomerado de líneas y cultivares (en donde sin embargo pueden

Cuadro 4.8. Valores de las dos primeras componentes principales considerando las 17 variables.

No.*	Primera componente principal	Segunda componente principal	No.*	Primera componente principal	Segunda componente principal
1	1.74	-3.13	38	-6.07	-0.09
2	0.67	-3.19	39	-7.01	0.28
3	-1.81	-3.40	40	-6.74	0.55
4	1.54	-3.29	41	-5.91	0.10
5	-0.83	-4.25	42	0.10	-0.55
6	-2.33	-3.35	43	-2.86	-0.36
7	2.03	-2.34	44	-2.94	-0.21
8	2.80	-2.61	45	-0.73	-0.16
9	0.41	-2.98	46	-4.25	0.49
10	2.01	-3.24	47	-1.01	0.32
11	1.00	-2.83	48	2.20	1.04
12	-0.67	-2.87	49	3.30	1.58
13	1.04	-3.09	50	1.49	0.67
14	0.89	-3.31	51	1.15	0.92
15	0.79	-2.62	52	2.65	0.96
16	1.45	-2.40	53	2.73	0.84
17	-0.47	-0.16	54	1.50	1.45
18	-3.33	-0.65	55	2.64	1.33
19	1.49	-0.15	56	-3.77	1.29
20	2.50	0.54	57	-0.82	0.01
21	-5.56	-0.63	58	-1.39	1.06
22	3.78	0.58	59	-4.08	0.53
23	0.40	-0.27	60	0.71	1.97
24	0.52	-0.36	61	1.84	2.17
25	0.28	-0.64	62	-0.66	1.98
26	2.23	0.04	63	-7.36	2.80
27	0.76	-0.03	64	1.21	1.27
28	2.83	0.68	65	1.83	1.33
29	3.61	0.75	66	1.47	1.63
30	2.47	-0.57	67	2.29	1.82
31	2.30	-0.11	68	1.58	1.73
32	-3.83	-0.65	69	2.25	1.52
33	3.18	0.56	70	2.69	1.96
34	2.35	-0.36	71	2.22	1.56
35	1.66	-0.42	72	-6.87	3.45
36	-3.91	-0.80	73	-2.81	4.53
37	-1.11	-0.38	74	3.75	5.05
			75	2.78	4.36

* Los números de 1 a 16 corresponden a la línea H80-2535; 17 a 24 corresponden a H80-2630; 25 a 34 corresponden a H82-1671; 35 a 45 corresponden a H80-5150; 46 a 59 corresponden al cultivar Santa Rosa; 60 a 71 corresponden a UFV-1 y 72 a 75 al cultivar Júpiter.

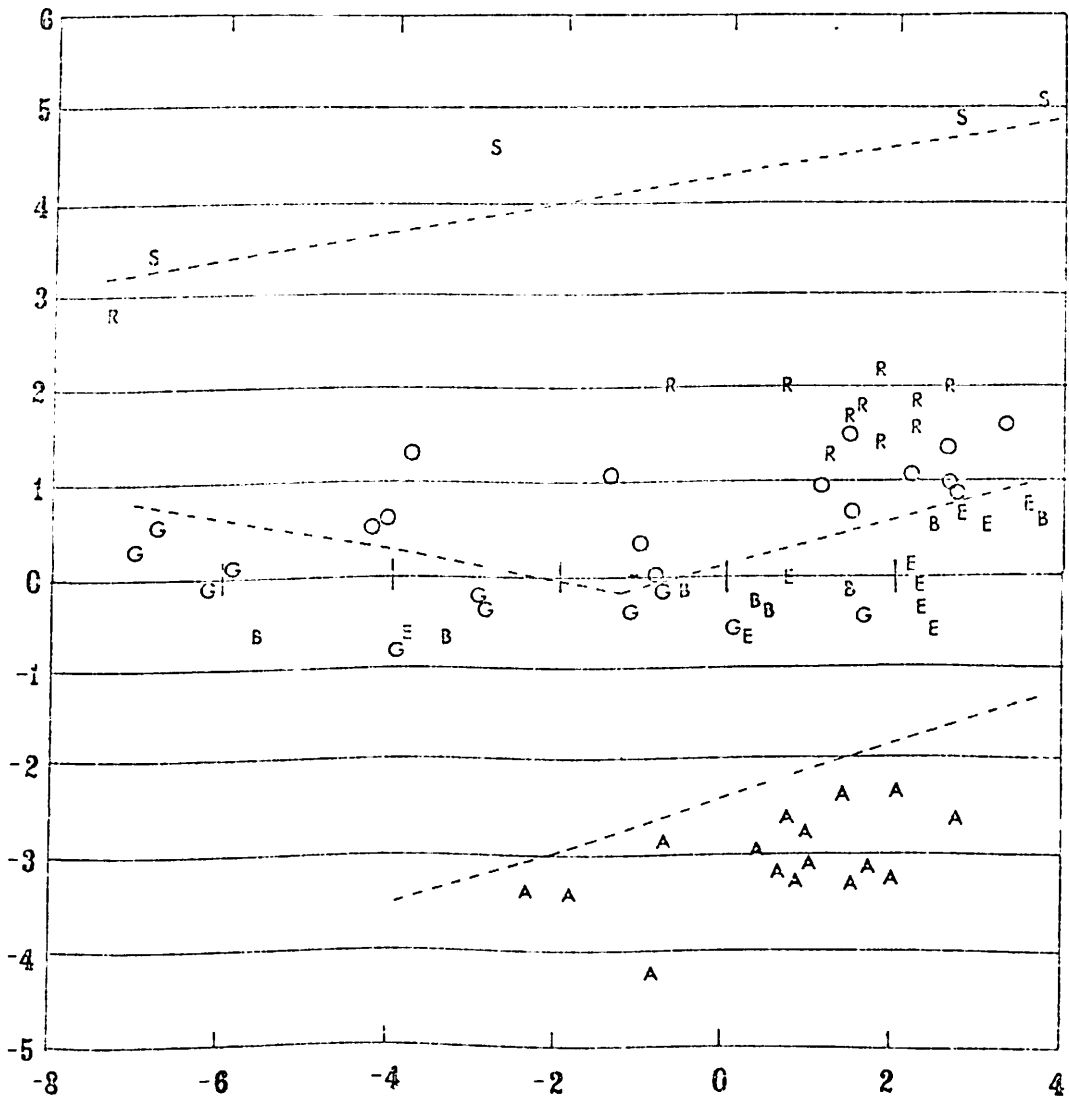


Figura 4. Ordenación gráfica de los individuos utilizando la primera y segunda componentes principales. Los valores anotados sobre el eje de las abscisas corresponden a los de la primera componente principal; los anotados sobre el eje de las ordenadas corresponden a los de la segunda componente principal. Las líneas discontinuas dentro de la tabla demarcan agrupaciones de individuos.

separarse las líneas de los cultivares) que abarca a H80-25-39, H82-16-71, H86-51-50, Santa Rosa y UFV-1; el tercer grupo, por último, lo constituye el cultivar Júpiter.

2. Si se observa la distribución de los puntos en la gráfica, puede observarse que, a excepción de la línea H80-25-35 y del cultivar Santa Rosa, los individuos de cada línea o cultivar presentaron una distribución muy heterogénea a lo largo de la primera componente principal. Sobre la segunda componente principal, en cambio, la distribución de cada línea o cultivar fue homogénea.

3. Si se comparan los valores de porcentaje de nitrógeno en las semillas con la distribución de las líneas y cultivares, puede apreciarse que existe cierta tendencia a que se incremente el porcentaje de nitrógeno conforme se incrementa el valor de la segunda componente principal.

La agrupación de los individuos con base en los valores de sus dos primeras componentes principales puede interpretarse de manera subjetiva analizando los coeficientes de los vectores característicos considerados, los cuales se anotan en el Cuadro 4.9. Con base en los valores de estos coeficientes, la primera componente principal puede considerarse una componente de cantidad de material acumulado, mientras que la segunda componente

Cuadro 4.9. Coeficientes de los vectores característicos utilizados en la construcción de la primera y segunda componentes principales. Los valores subrayados son los que influyen fuertemente en el valor de la componente principal.

variable	Primer vector característico	Segundo vector característico
X ₁ : Dias a R1.	-5.47x10 ⁻³	<u>0.4501</u>
X ₂ : Dias a R6-R7	-2.65x10 ⁻²	<u>0.4323</u>
X ₃ : Altura de planta	-9.48x10 ⁻²	-0.1654
X ₄ : Número de nudos	-0.2057	-0.2015
X ₅ : Número de nudos reproductivos	-0.1905	-0.2813
X ₆ : Número de vainas	<u>0.3235</u>	8.32x10 ⁻²
X ₇ : Número de raicillos	<u>-0.3115</u>	-5.68x10 ⁻²
X ₈ : Número de ramificaciones	-0.2415	-2.95x10 ⁻²
X ₉ : Número de semillas	<u>-0.3219</u>	1.53x10 ⁻²
X ₁₀ : Longitud foliar acumulado	<u>-0.3255</u>	8.20x10 ⁻³
X ₁₁ : Ancho foliar acumulado	<u>-0.3255</u>	1.53x10 ⁻²
X ₁₂ : % germinación de la semilla progenitora	6.21x10 ⁻²	<u>-0.3183</u>
X ₁₃ : Peso seco de las vainas	<u>-0.3296</u>	4.74x10 ⁻²
X ₁₄ : Peso seco de las semillas	<u>-0.3153</u>	8.28x10 ⁻²
X ₁₅ : Peso seco vegetativo	<u>-0.3250</u>	0.0662
X ₁₆ : Peso promedio de la semilla progenitora	-6.81x10 ⁻³	<u>-0.4155</u>
X ₁₇ : % N en las semillas	-6.93x10 ⁻²	0.2304
X ₁₈ : % N en las partes vegetativas	-2.03x10 ⁻²	-0.2429

principal puede considerarse una componente de viabilidad de la semilla progenitora y de extensión temporal del ciclo biológico de los individuos. Es interesante la tendencia observada en este caso ya que de acuerdo con la distribución sobre la segunda componente principal, se presentó cierta tendencia al incremento del porcentaje de nitrógeno en las semillas conforme disminuyó la calidad para la siembra de la semilla progenitora, expresada como porcentaje de germinación y peso promedio de la semilla. Las correlaciones lineales simples de estas dos variables con la variable porcentaje de nitrógeno fueron ambas significativas y negativas; las correlaciones parciales, por otro lado, fueron marcadas como no significativas por lo que, al parecer, los efectos de estas variables sobre la variable de respuesta fueron de tipo indirecto.

En cuanto al análisis de regresión lineal múltiple, considerando a la variable porcentaje de nitrógeno en las semillas como la variable de respuesta, la siguiente serie de cuadros ofrece un resumen de los resultados. En el Cuadro 4.10 se anotan los valores estimados de los coeficientes de regresión parcial para el modelo completo con 17 variables. En las columnas a la derecha de las estimaciones de los coeficientes de regresión parcial aparecen los valores de t y el intervalo de confianza de 0.95 para los coeficientes de regresión. Las probabilidades asociadas con los valores de t que aparecen marcadas con un asterisco indican cuales coeficientes de regresión se consideran diferentes a cero.

Cuadro 4.10. Valores estimados de los coeficientes de regresión parcial con el porcentaje de nitrógeno en las semillas como variable de respuesta. A la derecha del valor estimado aparecen el valor de t y su probabilidad asociada así como el intervalo de confianza para los coeficientes de regresión. Los valores marcados con un asterisco son los considerados significativos.

Variable	Estimación	t	P(> t)	Intervalo de confianza de 0.95	
				Límite inferior	Límite superior
X ₁ :Días a R1	-0.0198	-0.7371	0.4524	-0.0758	0.0359
X ₂ :Días a R5-R7	0.0409	1.5842	0.1077*	-0.0092	0.0910
X ₃ :Altura de planta	0.0050	0.9266	0.3550	-0.0058	0.0157
X ₄ :Número de nudos	0.0050	0.1873	0.8521	-0.0484	0.0584
X ₅ :Número de nudos reproductivos	-0.0138	-0.5162	0.6077	-0.0673	0.0397
X ₆ :Número de vainas	-0.0021	-0.5955	0.5539	-0.0091	0.0050
X ₇ :Número de trifolios	0.0037	0.6413	0.5236	-0.0079	0.0154
X ₈ :Número de ramificaciones	0.0073	0.3228	0.7480	-0.0380	0.0525
X ₉ :Número de semillas	0.0030	1.3136	0.1942	-0.0016	0.0076
X ₁₀ :Longitud foliar acumulada	-0.0004	-1.0365	0.3034	-0.0012	0.0004
X ₁₁ :Ancho foliar acumulado	0.0001	0.0769	0.9390	-0.0020	0.0021
X ₁₂ :Porcentaje de germinación de la					

Cuadro 4.10.....continuación

Variable	Estimación	t	P(ot)	Intervalo de confianza de 0.95	
				Límite inferior	Límite superior
semilla P02 genitora	0.0097	0.8548	0.3995	-0.0125	0.0319
X13:Peso seco de las vainas	-0.0030	-0.1139	0.9099*	-0.1041	-0.0025
X14:Peso seco de las se- millas	0.0259	3.4505	0.0001*	0.0089	0.0449
X15:Peso seco de las par- tes vegeta- tivas	-0.0043	-0.4131	0.6811	-0.0254	0.0167
X16:Peso prome- dio de la semilla progenito- ra	2.6835	0.3188	0.7511	-14.1735	19.5454
X18:Porcentaje de nitróge- no en las partes ve- getativas	0.6117	3.9472	0.0002*	0.3013	0.9221

Estimación de $\beta_0 = -1.3112$ Valor de R^2 ajustado = 0.4108

Si a partir de estos datos se quisiera construir un modelo restringido, este involucraría a las variables marcadas que son: días a R6-R7, peso seco de las vainas, peso seco de las semillas y porcentaje de nitrógeno en las partes vegetativas. El valor ajustado del coeficiente de determinación múltiple para el modelo completo fue 0.4108.

En el Cuadro 4.11 aparece el análisis de varianza para la regresión y el resultado de la partición de la suma de cuadrados de regresión para llevar a cabo pruebas secuenciales de F. Las variables seleccionadas para incluirse en un modelo restringido fueron, en este caso, días a R1, días a R6-R7, altura de planta, número de trifoliolos, longitud foliar acumulada, porcentaje de germinación de la semilla progenitora, peso seco de las semillas y porcentaje de nitrógeno en las partes vegetativas.

Los resultados de la aplicación del algoritmo *backward* para selección de variables se presentan en el Cuadro 4.12. En este caso el modelo restringido incluye a las variables días a R6-R7, número de semillas, peso seco de las vainas, peso seco de las semillas y porcentaje de nitrógeno en las partes vegetativas. El valor ajustado del coeficiente de determinación múltiple para el modelo restringido fue de 0.4669, lo que constituyó una mejora sobre el valor obtenido para el modelo completo.

En el Cuadro 4.13 aparecen los valores de los coeficientes de regresión parcial estandarizados para

Cuadro 4.11. Análisis de varianza para la regresión de la variable porcentaje de nitrógeno en las semillas sobre las restantes variables y partición de la suma de cuadrados de regresión para realizar pruebas secuenciales de F.

Análisis de varianza para la regresión:					
Fuente	g. l.	S.C.	CM	F	P
Regresión	17	13.4077	0.7887	4.0344	$P < 10^{-4}$
Residual	57	11.1411	0.1955		
Total ajustado	74	24.5488			

Partición de la suma de cuadrados de regresión utilizando sumas de cuadrados condicionales:

Fuente	g. l.	S.C.	F	P
X ₁	1	3.4068	17.43	0.0001*
X ₂	1	1.6759	8.57	0.0049*
X ₃	1	0.6752	3.46	0.0680*
X ₄	1	0.0230	0.12	0.7364
X ₅	1	0.0489	0.25	0.6244
X ₆	1	0.4395	2.25	0.1392
X ₇	1	0.5603	2.87	0.0959*
X ₈	1	0.1570	0.80	0.3824
X ₉	1	0.0075	0.04	0.8461
X ₁₀	1	0.5474	2.80	0.0997*
X ₁₁	1	0.0997	0.51	0.4856
X ₁₂	1	0.7795	3.99	0.0506*
X ₁₃	1	0.3025	1.55	0.2186
X ₁₄	1	1.2116	6.20	0.0157*
X ₁₅	1	0.0380	0.20	0.4892
X ₁₆	1	0.3294	1.68	0.2001
X ₁₇	1	3.0453	15.53	0.0002*

Cuadro 4.12. Variables incluidas en el modelo restringido obtenido después de llevar a cabo selección de variables empleando el algoritmo backward en la regresión de la variable porcentaje de nitrógeno en las semillas sobre las restantes variables.

Variable	Estimación	t	P(>t)	Intervalo de confianza de 0.95	
				Límite inferior	Límite superior
X2: días = R6-R7	0.0180	5.3162	$P < 10^{-4}$	0.0113	0.0248
X9: número de semillas	0.0016	1.6129	0.113	-0.0004	0.0036
X13: peso seco de las vainas	-0.0793	-4.5739	$P < 10^{-4}$	-0.1139	-0.0447
X14: peso seco de las semillas	0.0309	5.1663	$P < 10^{-4}$	0.0189	0.0439
X18: porcentaje de nitrógeno en las partes vegetativas	0.5280	4.3583	$P < 10^{-4}$	0.2863	0.7693

Estimación de $\beta_0 = 2.5382$
 Valor de R^2 ajustado = 0.4669

Tabla 4.13. Valores de los coeficientes de regresión parcial estandarizados para valores del coeficiente de escape, w , de cero a uno. La regresión es de la variable porcentaje de nitrógeno en los sedillos sobre las restantes variables.

w	Coeficientes								
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
0.0	0.1846	0.1181	0.1113	0.0813	-0.1114	-0.1812	0.8000	0.0585	1.1797
0.1	0.1813	0.1089	0.1010	0.0813	-0.0705	0.1810	-0.1182	0.1810	0.1710
0.2	0.1813	0.1078	0.1008	0.0813	-0.0700	0.1810	-0.0189	0.1089	0.1113
0.3	0.1811	0.1081	0.1089	-0.0877	-0.0178	0.0108	-0.0808	0.1084	1.0108
0.4	0.1808	0.1088	0.0912	-0.0881	-0.0177	0.0780	-0.0878	0.0808	0.0718
0.5	0.1288	0.1008	0.0818	-0.1000	-0.0281	0.1000	-0.0878	0.0718	0.1018
0.6	0.1288	0.1747	0.0718	-0.0481	-0.0188	0.0802	-0.0488	0.0808	0.0881
0.7	0.1188	0.1818	0.0887	-0.0488	-0.0288	0.0847	-0.0488	0.1888	0.0800
0.8	0.1118	0.1888	0.0881	-0.0401	-0.0281	0.0811	-0.0882	0.0802	0.0481
0.9	0.1072	0.1847	0.0848	-0.0878	-0.0281	0.0482	-0.0840	0.0808	0.0481
1.0	0.1088	0.1878	0.0802	-0.0888	-0.0278	0.0488	-0.0804	0.0471	0.0808

w	Coeficientes								
	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}
0.0	-0.7378	0.0882	0.1707	-1.6881	1.6882	-0.8202	0.0703	0.5084	
0.1	-0.1881	-0.1881	0.0440	-0.1198	0.4878	-0.1818	-0.0820	0.3888	
0.2	-0.0888	-0.1810	0.0178	-0.0418	0.3108	-0.0707	-0.0871	0.3080	
0.3	-0.0718	-0.0888	0.0080	-0.0182	0.2888	-0.0424	-0.0821	0.2818	
0.4	-0.0881	-0.0781	-0.0081	-0.0380	0.1888	-0.0282	-0.0781	0.2188	
0.5	-0.0888	-0.1087	-0.0187	-0.0381	0.1887	0.0188	-0.0740	0.1882	
0.6	-0.0882	-0.0481	-0.0187	0.0142	0.1881	-0.0080	-0.0721	0.1880	
0.7	-0.0180	-0.0880	-0.0080	0.0181	0.1818	-0.0088	-0.0888	0.1481	
0.8	-0.0288	-0.0818	-0.0088	0.0187	0.1108	-0.0002	-0.0878	0.1808	
0.9	-0.0188	-0.0171	-0.0081	0.0181	0.1087	0.0028	-0.0880	0.1181	
1.0	-0.0188	-0.0181	-0.0081	0.0112	0.1118	0.0042	-0.0842	0.1078	

diferentes valores del coeficiente de sesgo w . El comportamiento del sistema de coeficientes *ridge* puede apreciarse en la Figura 5, en la cual aparece la *ridge trace* para la regresión múltiple de la variable porcentaje de nitrógeno en las semillas sobre las restantes variables. El sistema de coeficientes *ridge* presenta un comportamiento estable con un valor del coeficiente de sesgo de aproximadamente 0.7. Las variables que en este punto aparecen como las indicadas para formar un modelo restringido son: días a R1, días a R6-R7, peso seco de las semillas y porcentaje de nitrógeno en las partes vegetativas.

Los valores de los coeficientes de regresión parcial estandarizados, para el valor cero del coeficiente de sesgo, son los obtenidos en la regresión sobre la matriz de correlaciones con ajuste para cuadrados mínimos y aparecen en el Cuadro 4.14. Con acuerdo a estos valores las variables más importantes en la determinación del porcentaje de nitrógeno en las semillas son, en orden de importancia: peso seco de las semillas, peso seco de las vainas, número de semillas, días a R6-R7, longitud foliar acumulada, porcentaje de nitrógeno en las partes vegetativas, número de vainas y días a R1.

En el Cuadro 4.15 se anota un resumen de los resultados de los procedimientos de selección de variables. En la parte de abajo de este cuadro se anota el número de veces que cada una de las variables aparece dentro de un

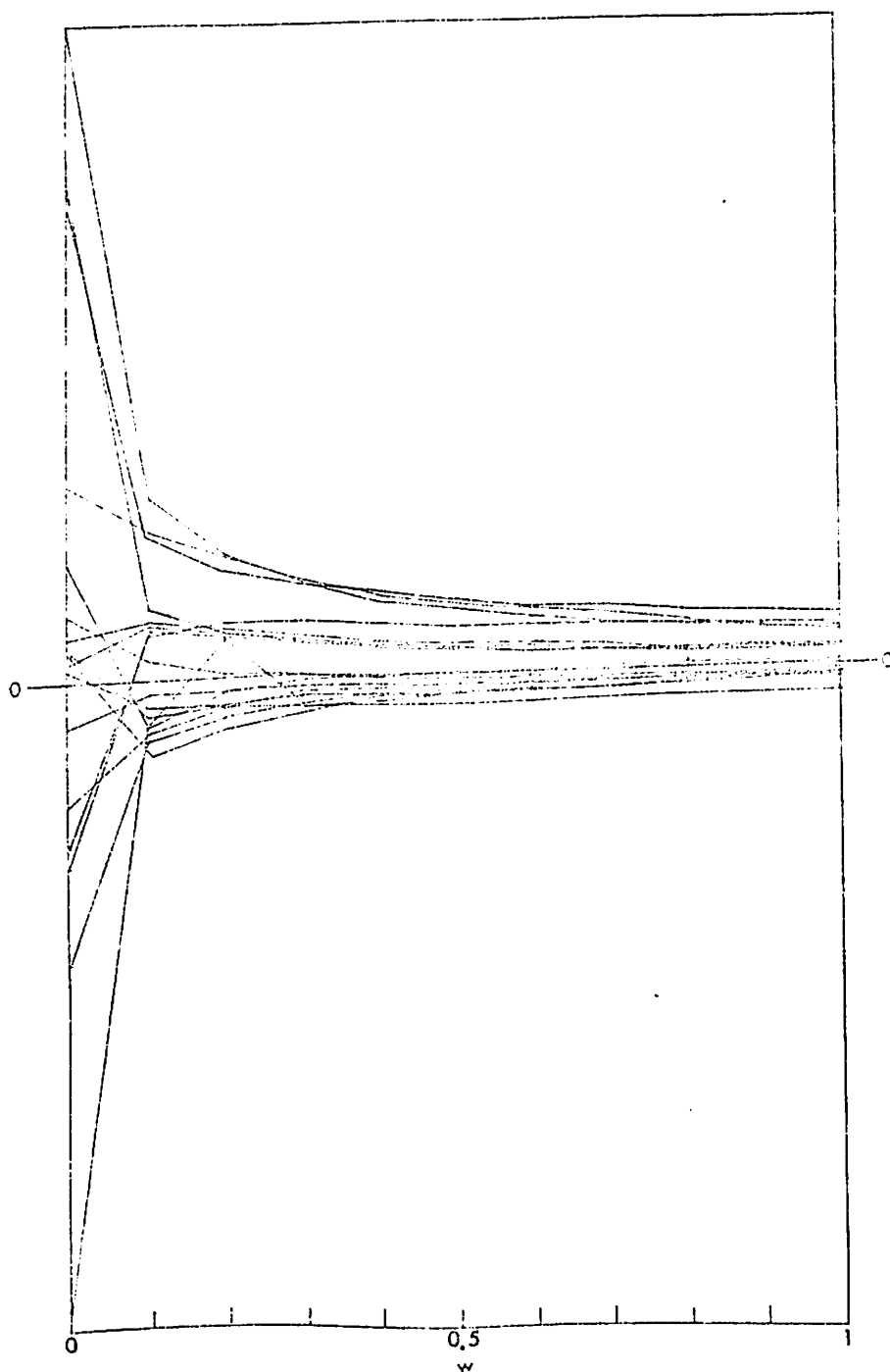


Figura 5. Ridge trace para la regresión múltiple de la variable porcentaje de nitrógeno en las semillas sobre las restantes variables. Los valores sobre el eje de las abscisas corresponden al coeficiente de sesgo, w ; los valores sobre el eje de las ordenadas son los de los coeficientes ridge.

Cuadro 4.14. Valores de los coeficientes de regresión parcial estandarizados obtenidos con un ajuste para cuadrados mínimos. La regresión es de la variable porcentaje de nitrógeno en las semillas sobre las restantes variables.

Variable	Valor del coeficiente**
X1: días a R1	-0.4545*
X2: días a R5-R7	1.2128*
X3: altura de planta	0.1125
X4: número de nudos	0.0318
X5: número de nudos reproductivos	-0.1154
X6: número de vainas	-0.4612*
X7: número de trifoliolos	0.3000
X8: número de ramificaciones	0.0535
X9: número de semillas	1.2590*
X10: longitud foliar acumulada	-0.7378*
X11: ancho foliar acumulado	0.0652
X12: porcentaje de germinación de la semilla progenitora	0.1707
X13: peso seco de las vainas	-1.6331*
X14: peso seco de las semillas	1.6852*
X15: peso seco de las partes vegetativas	-0.3202
X16: peso promedio de la semilla progenitora	0.0703
X18: porcentaje de nitrógeno en las partes vegetativas	0.5054*

** Los valores marcados son los seleccionados para formar un modelo restringido.

Cuadro 4.15. Resumen de las variables seleccionadas como importantes para cada una de las técnicas utilizadas. La variable de respuesta fue el porcentaje de nitrógeno en las semillas.

Técnica	Variables seleccionadas*
Correlación lineal simple	X ₁ , X ₂ , X ₆ , X ₉ , X ₁₂ , X ₁₃ , X ₁₄ , X ₁₅ , X ₁₆
Correlación parcial	X ₂ , X ₁₃ , X ₁₄ , X ₁₈
Valor de t para coeficientes de regresión parcial	X ₂ , X ₁₃ , X ₁₄ , X ₁₈
Pruebas secuenciales de F	X ₁ , X ₂ , X ₃ , X ₇ , X ₁₀ , X ₁₂ , X ₁₄ , X ₁₈
Selección backward	X ₂ , X ₉ , X ₁₃ , X ₁₄ , X ₁₈
Coefficientes <u>ridge</u> para w=0.7	X ₁ , X ₂ , X ₁₄ , X ₁₈
Coefficientes de regresión parcial estandarizados	X ₁ , X ₂ , X ₆ , X ₉ , X ₁₀ , X ₁₃ , X ₁₄ , X ₁₈

* Incluye métodos de selección objetivos y subjetivos

Variable	Número de veces que aparece en un modelo restringido	Variable	Número de veces que aparece en un modelo restringido
X ₁	4	X ₁₂	2
X ₂	7	X ₁₃	6
X ₃	1	X ₁₄	7
X ₆	2	X ₁₅	1
X ₇	1	X ₁₆	1
X ₉	3	X ₁₈	6
X ₁₀	2		

modelo restringido para cada una de las técnicas de selección de variables. En el caso de las variables días a RS-R7 y peso seco de las semillas existió acuerdo para su inclusión para todas las técnicas; para las variables porcentaje de nitrógeno en las partes vegetativas y peso seco de las vainas existió acuerdo para su inclusión en seis de las siete técnicas; las variables días a R1 y número de semillas fueron marcadas para incluirse en un modelo restringido por cuatro y por tres de las siete técnicas utilizadas, respectivamente.

En consideración al análisis de regresión múltiple, considerando a la variable peso seco de las semillas como la variable de respuesta, los resultados se presentan en los Cuadros 4.16 y siguientes. En el Cuadro 4.16 se anotan los valores estimados de los coeficientes de regresión parcial para el modelo completo con 17 variables. En las columnas a la derecha de las estimaciones de los coeficientes de regresión parcial aparecen los valores de t y el intervalo de confianza de 0.95 para los coeficientes de regresión. Las probabilidades asociadas con los valores de t que aparecen marcadas con un asterisco indican cuales coeficientes de regresión se consideran diferentes a cero. Si a partir de estos datos se quisiera construir un modelo restringido, este involucraría a las variables marcadas que son: días a R1, altura de planta, número de nudos, ancho foliar acumulado, peso seco de las vainas, peso seco de las partes vegetativas, porcentaje de nitrógeno en las semillas y

Cuadro 4.16. Valores estimados de los coeficientes de regresión parcial con el peso seco de las semillas como variable de respuesta. A la derecha del valor estimado aparecen el valor de t y su probabilidad asociada así como el intervalo de confianza para los coeficientes de regresión.

Variables	Estimación	t	P(> t)	Intervalo de confianza de 0.95	
				Límite inferior	Límite superior
X ₁ : días a R1.	-0.7359	-1.9737	0.0533*	-1.4828	0.0109
X ₂ : días a R5-R7	0.4103	1.1856	0.2407	-0.2832	1.1043
X ₃ : altura de planta	0.1620	2.2928	0.0256*	0.0205	0.3034
X ₄ : número de nudos	-1.2210	-3.7293	0.0004*	-1.8768	-0.5633
X ₅ : número de nudos reproductivos	0.3874	1.0659	0.2909	-0.3406	1.1155
X ₆ : número de vainas	0.0188	0.3898	0.6981	-0.0777	0.1153
X ₇ : número de trifolias	-0.0785	-0.9913	0.3257	-0.2372	0.0801
X ₈ : número de ramificaciones	0.3641	1.1898	0.2390	-0.2488	0.9770
X ₉ : número de semillas	0.0409	1.2973	0.1997	-0.0233	0.1041
X ₁₀ : longitud foliar acumulada	0.0042	0.8034	0.4251	-0.0062	0.0146
X ₁₁ : ancho foliar acumulado	-0.0290	-2.1547	0.0354*	-0.0559	-0.0020

...

Cuadro 4.16.....continuación

Variable	Estimación	t	P(>t)	Intervalo de confianza de 0.95	
				Límite inferior	Límite superior
X12: porcentaje de germinación de la semilla progenitora	0.0652	0.4616	0.6462	-0.2300	0.3677
X13: peso seco de las vainas	1.3731	4.4117	$PC10^{-4}$ *	0.7497	1.9965
X15: peso seco de las partes vegetativas	0.2396	1.7051	0.0936*	-0.0419	0.5211
X16: peso promedio de la semilla progenitora	-133.4003	-1.1691	0.2472	-361.9480	95.1477
X17: porcentaje de nitrógeno en las semillas	5.0495	2.9926	0.0041	1.6699	8.4291
X18: porcentaje de nitrógeno en las partes vegetativas	-3.9878	-1.7051	0.0934*	-8.6694	0.6938

Estimación de $\beta_0 = -20.6130$
 Valor de R^2 ajustado = 0.9719

porcentaje de nitrógeno en las partes vegetativas. El valor ajustado del coeficiente de determinación múltiple para el modelo completo fue 0.9719.

En el Cuadro 4.17 aparece el análisis de varianza para la regresión y el resultado de la partición de la suma de cuadrados de regresión para llevar a cabo pruebas secuenciales de F. Las variables seleccionadas para incluirse en un modelo restringido fueron, en este caso, días a R1, días a R6-R7, altura de planta, número de nudos, número de nudos reproductivos, número de vainas, número de trifoliolos, número de ramificaciones, número de semillas, peso seco de las vainas, porcentaje de nitrógeno en las semillas y porcentaje de nitrógeno en las partes vegetativas.

Los resultados de la aplicación del algoritmo *backward* de selección de variables se presentan en el Cuadro 4.18. En este caso el modelo restringido incluye a las variables días a R1, días a R6-R7, altura de planta, número de nudos, número de nudos reproductivos, número de semillas, ancho foliar acumulado, peso seco de las vainas, peso seco de las partes vegetativas, porcentaje de nitrógeno en las semillas y porcentaje de nitrógeno en las partes vegetativas. El valor ajustado del coeficiente de determinación múltiple para este modelo restringido fue de 0.9730, valor ligeramente mayor al valor obtenido con el modelo completo.

Cuadro 4.17. Análisis de varianza para la regresión de la variable peso seco de las semillas sobre las restantes variables y partición de la suma de cuadrados de regresión para realizar pruebas secuenciales de F.

Análisis de varianza para la regresión:					
Fuente	g.l.	SC	CM	F	P
Regresión	17	84443.80	5555.52	151.862	$P < 10^{-4}$
Residual	57	2092.11	36.7036		
<hr/>					
Total ajustado	74	86535.91			
Partición de la suma de cuadrados de regresión utilizando sumas de cuadrados condicionales:					
Fuente	g.l.	SC	F	P	
X ₁	1	1488.8038	40.8600	$P < 10^{-4}$ *	
X ₂	1	6095.1568	166.0600	$P < 10^{-4}$ *	
X ₃	1	9927.6778	270.4600	$P < 10^{-4}$ *	
X ₄	1	12054.0799	328.4200	$P < 10^{-4}$ *	
X ₅	1	10281.1729	280.1100	$P < 10^{-4}$ *	
X ₆	1	47970.6678	1306.9700	$P < 10^{-4}$ *	
X ₇	1	1574.5949	42.9000	$P < 10^{-4}$ *	
X ₈	1	139.0345	3.7900	0.0586*	
X ₉	1	2464.9599	67.1600	$P < 10^{-4}$ *	
X ₁₀	1	0.0666	0.0018	0.9666	
X ₁₁	1	10.3575	0.2800	0.6030	
X ₁₂	1	50.7037	1.3900	0.2447	
X ₁₃	1	1885.5592	51.3700	$P < 10^{-4}$ *	
X ₁₅	1	93.3004	2.5400	0.1164	
X ₁₆	1	70.1374	1.9100	0.1723	
X ₁₇	1	230.7083	6.2900	0.0150*	
X ₁₈	1	106.8335	2.9100	0.0934*	

Cuadro 4.18. Variables incluidas en el modelo restringido obtenido después de llevar a cabo selección de variables empleando el algoritmo backward. La regresión es de la variable peso seco de las semillas sobre las restantes variables.

Variable	Estimación	t	Pr(≥t)	Intervalo de confianza de 0.95	
				Límite inferior	Límite superior
X1: días a R1	-0.7867	-3.2867	0.0009	-1.2878	-0.3300
X2: días a R2-R1	0.5681	1.2218	0.0019	0.2124	0.9110
X3: altura de planta	0.1418	2.1918	0.0311	0.0128	0.2712
X4: número de nudos	-1.2421	-4.0831	0.0002	-1.8577	-0.6265
X5: número de nudos reproductivos	0.5802	2.0424	0.0488	0.0120	1.1085
X6: número de semillas	0.0532	3.0881	0.0050	0.0188	0.0878
X11: ancho foliar acumulado	-0.0258	-3.6307	0.0003	-0.0393	-0.0123
X13: peso seco de las vainas	1.3923	5.4881	$P < 10^{-4}$	0.8852	1.8993
X15: peso seco de las partes vegetativas	0.2214	2.5650	0.0127	0.0489	0.3938
X17: porcentaje de nitrógeno en las semillas	8.1653	3.2580	0.0021	1.9480	8.3826
X19: porcentaje de nitrógeno en las partes vegetativas.	-4.4687	-2.0837	0.0432	-8.7970	-0.1405

Estimación de $\beta_0 = -55.0673$
 Valor de R^2 ajustado = 0.9530

En el Cuadro 4.19 aparecen los valores de los coeficientes de regresión parcial estandarizados para diferentes valores del coeficiente de sesgo w . El comportamiento del sistema de coeficientes *ridge* puede apreciarse en la Figura 6, en la cual aparece la *ridge trace* para la regresión múltiple de la variable peso seco de las semillas sobre las restantes variables. El sistema de coeficientes *ridge* presenta un comportamiento estable con un valor del coeficiente de sesgo de aproximadamente 0.6. Las variables que en este punto aparecen como las indicadas para formar un modelo restringido son: número de nudos reproductivos, número de vainas, número de semillas, longitud foliar acumulada, ancho foliar acumulado, peso seco de las vainas, peso seco de las partes vegetativas y porcentaje de nitrógeno en las semillas.

Los valores de los coeficientes de regresión parcial estandarizados, para el valor cero del coeficiente de sesgo, son los obtenidos en la regresión sobre la matriz de correlaciones con ajuste para cuadrados mínimos y aparecen en el Cuadro 4.20. De acuerdo con estos valores las variables más importantes en la determinación del peso seco de las semillas son, en orden de importancia: peso seco de las vainas, ancho foliar acumulado, peso seco de las partes vegetativas, número de semillas, días a R1 y días a R6-R7.

En el Cuadro 4.21 se anotan los resultados de los procedimientos de selección de variables. En la parte de abajo de este cuadro se anota el número de veces que cada

Cuadro 4.19. Valores de los coeficientes de regresión parcial estandarizados para valores del coeficiente de sesgo, w , de cero a uno. La regresión es de la variable peso seco de las semillas sobre las restantes variables.

w	Coeficientes								
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
0.0	-0.2975	0.1951	0.0538	-0.1241	0.0517	0.0690	-0.1008	0.0426	0.2718
0.1	-0.0815	-0.0042	0.0342	-0.1004	0.0973	0.1873	-0.0516	0.0004	0.2418
0.2	-0.0329	-0.0023	0.0299	-0.0791	0.0941	0.1688	-0.0188	0.0011	0.2126
0.3	-0.0264	0.0017	0.0287	-0.0614	0.0878	0.1570	0.0055	0.0059	0.1941
0.4	-0.0180	0.0053	0.0234	-0.0483	0.0819	0.1439	0.0208	0.0107	0.1809
0.5	-0.0120	0.0083	0.0282	-0.0389	0.0770	0.1428	0.0313	0.0151	0.1710
0.6	-0.0074	0.0107	0.0280	-0.0310	0.0719	0.1379	0.0358	0.0188	0.1630
0.7	-0.0039	0.0126	0.0277	-0.0249	0.0695	0.1338	0.0443	0.0220	0.1565
0.8	-0.0011	0.0141	0.0275	-0.0191	0.0665	0.1301	0.0486	0.0247	0.1509
0.9	0.0011	0.0154	0.0272	-0.0145	0.0640	0.1271	0.0518	0.0270	0.1461
1.0	0.0030	0.0164	0.0269	-0.0106	0.0619	0.1243	0.0544	0.0290	0.1418

w	Coeficientes							
	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₅	X ₁₆	X ₁₇	X ₁₈
0.0	0.1251	-0.3238	0.0201	0.5678	0.2821	-0.0555	0.0805	-0.0515
0.1	0.0697	-0.0256	-0.0373	0.3431	0.1620	-0.0486	0.1038	-0.0560
0.2	0.0233	0.0147	-0.0344	0.2751	0.1509	-0.0454	0.1008	-0.0551
0.3	0.0899	0.0467	-0.0320	0.2405	0.1432	-0.0434	0.0952	-0.0799
0.4	0.0934	0.0589	0.0304	0.2181	0.1375	-0.0418	0.0897	-0.0742
0.5	0.0953	0.0664	-0.0293	0.2021	0.1330	-0.0405	0.0848	-0.0689
0.6	0.0961	0.0713	-0.0285	0.1837	0.1293	-0.0391	0.0804	-0.0641
0.7	0.0966	0.0747	-0.0280	0.1601	0.1261	-0.0381	0.0756	-0.0601
0.8	0.0967	0.0770	-0.0278	0.1720	0.1231	-0.0373	0.0711	-0.0564
0.9	0.0964	0.0787	0.0273	0.1651	0.1207	-0.0360	0.0702	-0.0531
1.0	0.0960	0.0798	-0.0270	0.1591	0.1184	-0.0351	0.0675	-0.0503

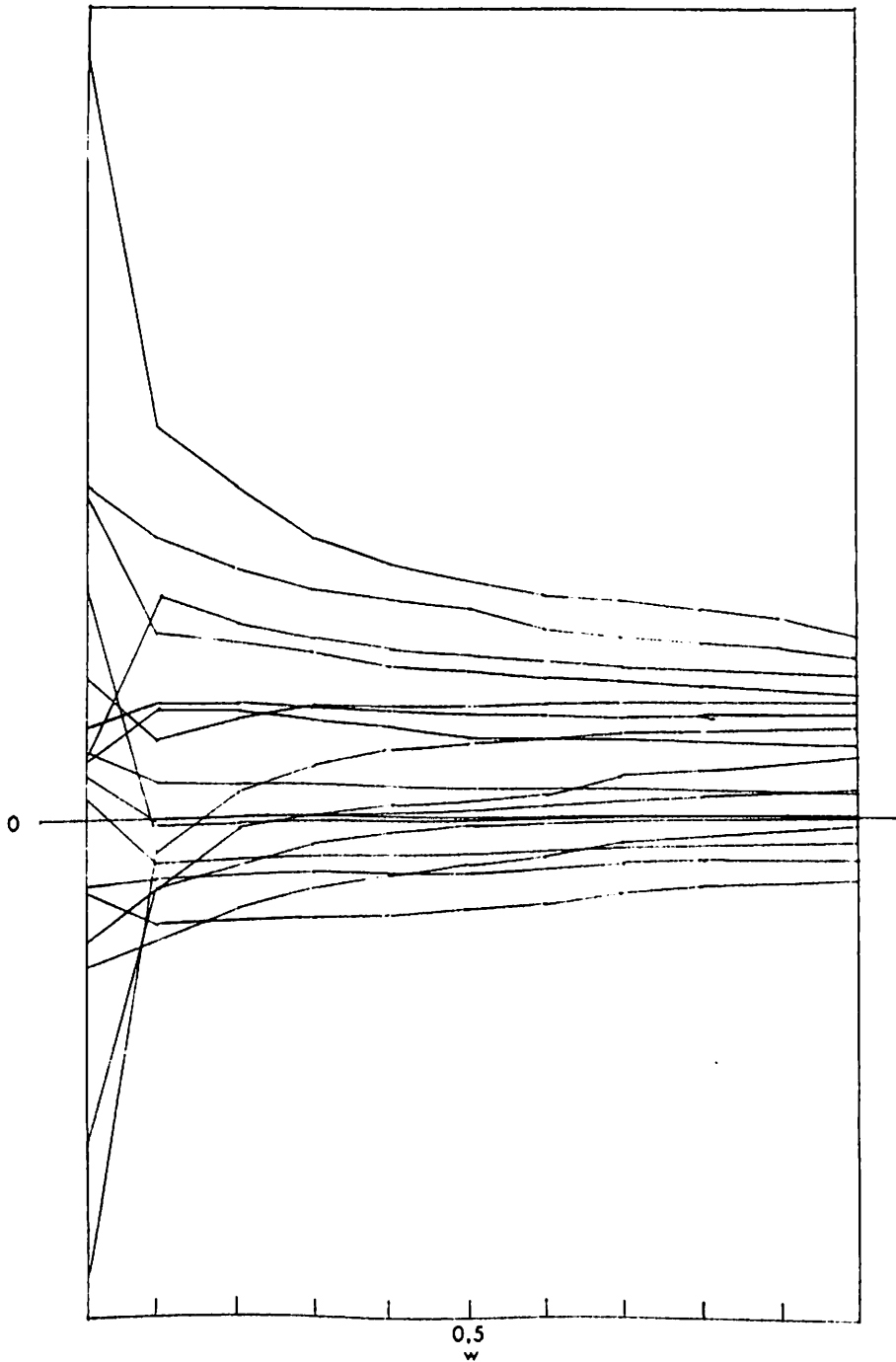


Figura 6. Ridge trace para la regresión múltiple de la variable peso seco de las semillas sobre las restantes variables. Los valores sobre el eje de las abscisas corresponden al coeficiente de sesgo, w , los valores sobre el eje de las ordenadas son los de los coeficientes ridge.

Cuadro 4.20. Valores de los coeficientes de regresión parcial estandarizados obtenidos con un ajuste para cuadrados mínimos. La regresión es de la variable peso seco de las semillas sobre las restantes variables.

Variable	Valor del coeficiente**
X ₁ : días a R1	-0.2575*
X ₂ : días a R6-R7	0.1952*
X ₃ : altura de planta	0.0568
X ₄ : número de nudos	-0.1241
X ₅ : número de nudos reproductivos	0.0517
X ₆ : número de vainas	0.0690
X ₇ : número de trifoliolos	-0.1008
X ₈ : número de ramificaciones	0.0426
X ₉ : número de semillas	0.2718*
X ₁₀ : longitud foliar acumulada	0.1252
X ₁₁ : ancho foliar acumulado	-0.3838*
X ₁₂ : porcentaje de germinación de la semilla progenitora	0.0202
X ₁₃ : peso seco de las vainas	0.6678*
X ₁₅ : peso seco de las partes vegetativas	0.2821*
X ₁₆ : peso promedio de la semilla progenitora	-0.0558
X ₁₇ : porcentaje de nitrógeno en las semillas	0.0805
X ₁₈ : porcentaje de nitrógeno en las partes vegetativas	-0.0525

** Los valores marcados son los seleccionados para formar un modelo restringido.

Cuadro 4.21. Resumen de las variables seleccionadas como importantes para cada una de las técnicas utilizadas. La variable de respuesta fue el peso seco de las semillas.

Técnica	Variables seleccionadas
Correlación lineal simple	X3, X4, X5, X6, X7, X8, X9, X10, X11, X13, X15, X17
Correlación parcial	X1, X3, X4, X11, X13, X15, X17, X18
Valor de t para coeficientes de regresión parcial	X1, X3, X4, X11, X13, X15, X17, X18
Pruebas secuenciales de F	X1, X2, X3, X4, X5, X6, X7, X8, X9, X13, X17, X18
Selección backward	X1, X2, X3, X4, X5, X9, X11, X13, X15, X17, X18
Coeficientes ridge para $w=0.6$	X5, X6, X9, X10, X11, X13, X15, X17
Coeficientes de regresión parcial estandarizados	X1, X2, X9, X11, X13, X15

Variable	Número de veces que aparece en un modelo restringido	Variable	Numero de veces que aparece en un modelo restringido
X1	5	X9	5
X2	3	X10	2
X3	5	X11	6
X4	5	X13	7
X5	4	X15	6
X6	3	X17	6
X7	2	X18	4
X8	2		

una de las variables aparece dentro de un modelo restringido para cada una de las técnicas de selección de variables. Para la variable peso seco de las vainas existió acuerdo entre todas las técnicas para incluirla en un modelo restringido; para las variables días a R1, altura de planta, número de nudos, número de nudos reproductivos, número de semillas, ancho foliar acumulado, peso seco de las partes vegetativas, porcentaje de nitrógeno en las semillas y porcentaje de nitrógeno en las partes vegetativas existió acuerdo para su inclusión en un modelo restringido en al menos cuatro de las técnicas.

DISCUSION

De acuerdo con los resultados de la prueba de Kruskal-Wallis no se presentaron diferencias significativas entre las diferentes líneas y cultivares para los caracteres porcentaje de nitrógeno en las semillas y peso seco de las semillas. La prueba de Mann-Whitney para comparar las líneas o cultivares de par en par marcó muy pocas diferencias significativas. Este resultado no es sorprendente si se considera el bajo número de líneas y cultivares utilizados.

En cuanto a la relación entre los caracteres porcentaje de nitrógeno en las semillas y el peso seco de las mismas, los resultados de este trabajo apoyan la hipótesis de que la relación inversa reportada por investigadores como Sinclair y DeWit (1975, 1976), Salado-Navarro *et al.* (1985), Kwon y Torrie (1964) y Hartwig (1969) es resultado no de una contradicción energética básica, sino de restricciones ambientales y fisiológicas tal como ha sido marcado por Hanson *et al.* (1961), Adams (1967), Bhatia y Rabson (1976) y Johnson *et al.* (1969). En el presente estudio los coeficientes de correlación lineal simple y de correlación parcial entre los caracteres arriba mencionados fueron de signo positivo y significativos, si bien ambos fueron de valor bajo. Un comportamiento análogo

se observó en los resultados de los análisis de regresión, en donde al considerar cada una de estas variables como variable de respuesta, la otra apareció como variable importante y con un coeficiente de regresión de signo positivo.

Los resultados arriba mencionados indican que es posible la selección de cultivares de alto rendimiento y con un contenido alto de proteína en las semillas. Sin embargo, tal parece que la selección debe realizarse en ambientes sin restricciones o bien empleando técnicas especiales de selección como la indicada en el reporte de Holbrook et al. (1989).

Respecto a la relación existente entre el porcentaje de nitrógeno en las semillas, el peso seco de las semillas y los restantes caracteres, un punto importante se refiere a lo adecuado de la selección de los caracteres estudiados. Si se observa el valor del coeficiente de determinación múltiple en el Cuadro 4.16, en donde se anotan los resultados del análisis de regresión lineal múltiple de la variable peso seco de las semillas sobre las restantes 17 variables, puede notarse que su valor, 0.9717, es indicativo de que la función lineal generada con las variables bajo uso es muy adecuada para explicar o predecir los valores observados del peso seco de las semillas. Lo mismo es indicado por el criterio, que debe usarse conjuntamente o en sustitución al del coeficiente de determinación múltiple, desarrollado por Wetz (1964), el cual indica que, para que

una ecuación de regresión se considere un predictor adecuado (en el sentido de que el rango de valores de la respuesta predichos por la ecuación sea sustancial comparado con el error estándar de la respuesta), el valor F observado (CM Regresión/CM Residual) debe de exceder no meramente el valor de un punto F seleccionado, sino al menos cuatro veces el punto seleccionado. El valor de F observado (Cuadro 4.17) con 17 y 57 grados de libertad fue de 151.362, valor mucho mayor al valor de F tabular que fue de 1.84.

En el Cuadro 4.18, por otro lado, aparecen las variables incluidas en un modelo restringido con 11 variables. En este caso el valor del coeficiente de determinación múltiple aumentó ligeramente de 0.9717 a 0.9730. Como puede observarse, las variables incluidas se refieren básicamente a área foliar, peso seco y número de estructuras (cantidad de materia), contenido de nitrógeno en las semillas y en las partes vegetativas y variables fenológicas.

Lo anotado antes puede explicarse a partir de la consideración de lo que es una estructura. Una estructura dada puede definirse como la forma transitoria que adquiere un proceso. En este caso existen una serie de eventos fisiológicos (que a su vez dependen de una serie de eventos moleculares) que son visualizados o representados como estructuras morfológicas o cambios temporales en las mismas (caracteres fenológicos). Lo que ocurrió para el caso de la variable peso seco de las semillas es que los caracteres

morfológicos y fenológicos elegidos representaron adecuadamente, indirecta o directamente, la determinación fisiológica del proceso de acumulación de material en las semillas. Lo anotado en el párrafo anterior también es indicativo de que la selección de líneas con una combinación adecuada de duración de ciertas fases del desarrollo y cantidad de material vegetativo y reproductivo pueden dar lugar a la obtención de cultivares de alto rendimiento. Estos resultados están acordes con lo marcado por Blad y Baker (1972), Ledent y Moss (1979), Williams *et al.* (1979), Hanway y Weber (1971) y Carbonell Guevara y Bartual Pastor (1983).

En consideración a la regresión lineal múltiple de la variable porcentaje de nitrógeno en las semillas sobre las restantes 17 variables, los resultados fueron muy diferentes a lo observado en el caso de la variable peso seco de las semillas. El valor observado de F para el modelo completo (Cuadro 4.11) fue de 4.0344, valor muy bajo como para considerar a la ecuación de regresión un predictor adecuado según el criterio de Wetz (1964). Por otro lado, si se examinan los Cuadros 4.10 y 4.12, en donde aparecen los resultados del ajuste para el modelo completo y para el modelo restringido, respectivamente, puede verse que los valores del coeficiente de determinación múltiple son bastante bajos, 0.4108 y 0.4669, respectivamente, además de que, para el modelo restringido, se observaron muy pocas variables. La muy baja correlación entre las observaciones

de la variable de respuesta y las funciones lineales de las variables seleccionadas indica que, al parecer, la representación de los procesos fisiológicos que determinan el contenido de nitrógeno en la semilla, por parte de los caracteres morfológicos y bioquímicos considerados en este estudio no fue adecuada. Es necesario entonces, tal parece, el estudiar otros caracteres morfológicos, incluso al nivel microscópico, o remitirse directamente a los caracteres fisiológicos.

A no ser que el resultado obtenido en cuanto a la determinación del contenido porcentual de nitrógeno en las semillas sea un artefacto resultante del pequeño número de líneas y cultivares utilizados, es dudoso que sea posible llevar a cabo selección indirecta sobre el carácter contenido de nitrógeno en las semillas utilizando caracteres morfológicos fácilmente medibles como los utilizados en este trabajo. Con toda seguridad que pueden encontrarse uno o más caracteres que funcionen como indicadores indirectos del contenido porcentual de nitrógeno en las semillas, pero el problema es que debe evitarse que estos caracteres se coloquen en la categoría de caracteres microscópicos o macroscópicos difíciles de cuantificar o apreciar.

Es interesante notar que las variables seleccionadas como importantes en la determinación del porcentaje de nitrógeno se encuentren todas en el conjunto seleccionado para el caso de la variable peso seco de las semillas. Al parecer este hecho es indicativo de que existen puntos

comunes de determinación, seguramente en cuanto a la cantidad de energía y materia disponibles, para ambos conjuntos de procesos. En el caso del porcentaje de nitrógeno en las semillas, de acuerdo con los resultados de Gent y Kiyomoto (1989) y a lo mencionado por Evans (1975), existen seguramente factores muy complejos no considerados y tal vez separados en algún grado de los procesos que determinan la acumulación de materia no protéica en la semilla.

Existe cierto grado de acuerdo en cuanto a las variables incluidas en el modelo restringido de regresión, considerando el porcentaje de nitrógeno en las semillas como la variable de respuesta, y las variables marcadas como importantes en el estudio realizado por Carbonell Guevara y Bartual Pastor (1983). Dos de las variables incluidas en el modelo restringido, el peso seco de las vainas y el porcentaje de nitrógeno en las partes vegetativas, no aparecen incluidas en los resultados del mencionado estudio.

Los Cuadros 4.15 y 4.21 constituyen el resumen de los resultados obtenidos con la aplicación de diferentes técnicas de selección de variables. En general, las conclusiones obtenidas a partir de estos cuadros son las mismas mencionadas hasta el momento, excepto para el caso de las dos variables que expresan la calidad de la semilla progenitora (el peso promedio de la misma y su porcentaje de germinación) y que fueron marcadas como importantes por algunos de los métodos de selección. El punto interesante es

que en ambos casos, tanto para el peso promedio de la semilla progenitora como para el porcentaje de germinación de la misma, se presentó una relación inversa entre estos caracteres y el porcentaje de nitrógeno en las semillas. Este hecho es también patente en la gráfica de ordenación con las dos primeras componentes principales y, más claramente, en los valores para los coeficientes de estas variables en el segundo vector característico de la matriz de correlaciones. Tal vez valga la pena la posterior investigación de la relación entre estas variables y el contenido de nitrógeno en las semillas, ya que pudiera ser indicativo de algún mecanismo de regulación importante.

El análisis crítico de los resultados debe contemplar la adecuación de los datos a las suposiciones sobre las que se asientan los modelos matemáticos base de las técnicas estadísticas utilizadas. Si la distribución de los datos viola fuertemente los supuestos entonces lo que se detecta no son valores "significativos" de diferencias o de estimaciones de parámetros, sino la violación de los supuestos. El principal problema presente en la serie de datos colectados para este estudio fue una desviación muy fuerte de una distribución como la normal, presencia de multicolinealidad y, para el caso de las medias de tratamientos, fuerte heterogeneidad de varianzas.

En el caso de la comparación de las medias de tratamientos el problema se resolvió fácilmente con la aplicación de dos pruebas de distribución libre. En el caso

de los análisis de correlación se comparó el valor de un cierto número de coeficientes de correlación de Pearson con los valores de los correspondientes coeficientes de correlación de Spearman. Como no se encontraron diferencias en cuanto a cuales coeficientes de correlación eran marcados como significativos y cuales no, tal vez a causa de que la prueba de t es robusta para tamaños de muestra mayores a 30 (Box, 1953), se conservó la matriz de correlaciones de Pearson por su utilidad en pruebas posteriores y para propósitos de comparación en cuanto a su utilidad como método para seleccionar variables.

En el caso de la relación entre las variables de respuesta y el conjunto de variables predictoras, el problema de hecho no se resolvió. No se encontró en este caso un método no paramétrico que realizara lo que lleva a cabo un análisis de regresión múltiple. Lo que si se llevó a cabo fue investigar la magnitud del problema. Para ello el método básico fue el análisis gráfico de los residuales en busca de desviaciones de una distribución como la normal por parte de los mismos, autocorrelación y necesidad de términos funcionales extra. No se detectó en ningún caso la presencia de autocorrelación ni la carencia de términos funcionales extra. Las gráficas de probabilidad normal para los residuales de las regresiones llevadas a cabo no mostraron problemas excesivamente fuertes, tanto para el caso de los modelos completos como para los modelos restringidos. En general, si los residuales no indican problemas fuertes se

supone que los datos son adecuados para las pruebas estadísticas utilizadas en los análisis de regresión (Draper y Smith, 1981).

Un tipo de problema cuya presencia no es obvia a partir del análisis de los residuales es el de la multicolinealidad. en este caso, tanto para la regresión del porcentaje de nitrógeno en las semillas como para la regresión del peso seco de las semillas, la matriz $X'X$ se presentó mal condicionada. Este hecho constituyó la justificación para el uso de la regresión *ridge*. En general, el uso de los coeficientes *ridge* llevó a disminuir el número de variables incluidas en los modelos restringidos, tal como puede apreciarse en los Cuadros 4.15 y 4.21, sobre todo para el caso de la variable porcentaje de nitrógeno en las semillas.

En general, se presentó cierto acuerdo entre los diferentes métodos de selección de variables en cuanto a cuales variables incluir en los modelos restringidos. El hecho de utilizar varias técnicas de selección de variables se deriva de que no existe ninguna regla obvia que indique, *a priori*, cual método de selección de variables debe utilizarse con preferencia sobre otros; el enfoque utilizado en este trabajo es usar varios métodos de selección de variables y comparar los resultados obtenidos con todos ellos. El proceso es una especie de validación cruzada y se espera por ello que las variables seleccionadas como importantes en la determinación de las variables de

respuesta lo sean realmente.

El caso de la regresión *ridge* es especial; si bien las estimaciones de los coeficientes de regresión *ridge* no son estimaciones insesgadas se espera, sin embargo, que la distancia entre el estimador *ridge* y el valor poblacional del coeficiente de regresión sea menor que la obtenida con el ajuste Gauss-Markov. En este sentido, los estimadores *ridge* pueden considerarse como los más adecuados para el tipo de matriz de datos de que se dispone. A pesar de esta ventaja, un problema fuerte con los estimadores *ridge* es la complejidad de su distribución; por esto es que no se posee alguna prueba estadística asociada con los mismos y ello origina que la selección de variables sea totalmente subjetiva.

Por último, es recomendable al parecer el aplicar el método de estimación *Jackknife* para la eliminación del sesgo en los estimadores resultante de la presencia de multicolinealidad (Hollander y Wolfe, 1973). Sin embargo, la estimación *Jackknife* es un método óptimo con muestras muy grandes y por ello no es recomendable en un estudio como el presente. Por otro lado, el esfuerzo de cálculo requerido es grande incluso cuando se dispone de un computador de alta velocidad y no es recomendable cuando se dispone de muchas variables.

CONCLUSIONES

No se presentaron diferencias significativas entre la mayoría de los cultivares y líneas para los caracteres contenido porcentual de nitrógeno y peso seco de las semillas.

No se encontró una relación inversa entre los caracteres peso seco de las semillas y contenido porcentual de nitrógeno en las mismas. Para el caso de la determinación del contenido porcentual de nitrógeno solo unas pocas variables, días a R6-R7 (x_2), peso seco de las semillas (x_{14}), porcentaje de nitrógeno en las partes vegetativas (x_{10}), peso seco de las vainas (x_{19}), días a R1 (x_1) y número de semillas (x_9), en ese orden, fueron marcadas como importantes, si bien los valores de los coeficientes de correlación lineal simple, de correlación parcial, de determinación múltiple y de F indicaron que el uso de estas variables "importantes" es poco recomendable para propósitos de predicción o de selección indirecta sobre el carácter contenido porcentual de nitrógeno en las semillas.

Lo contrario ocurrió para el carácter peso seco de las semillas, los valores de los coeficientes de correlación lineal simple, de correlación parcial, de determinación múltiple y de F indicaron que el uso de las variables

marcadas como importantes, es decir, el peso seco de las vainas (x_{13}), ancho foliar acumulado (x_{14}), peso seco de las partes foliares (x_{15}), porcentaje de nitrógeno en las semillas (x_{17}), días a R1 (x_1), altura de planta (x_9), número de nudos (x_4) y número de semillas (x_9) es muy recomendable para propósitos de predicción o de selección indirecta sobre el carácter peso seco de las semillas.

La conclusión principal que se sigue de los resultados es que el enfoque de selección indirecta sobre el carácter contenido porcentual de nitrógeno en las semillas, utilizando caracteres morfológicos y fenológicos como indicadores, al parecer es poco viable. Esta conclusión, claro está, es aplicable únicamente al espectro de caracteres considerados en este estudio.

RESUMEN

El presente estudio se desarrolló con los objetivos de investigar la variación en el carácter contenido porcentual de nitrógeno en las semillas en un grupo de genotipos de soya e investigar las relaciones entre este carácter y una serie de variables fenológicas, morfológicas y químicas. Lo mismo se llevó a cabo para el carácter peso seco de las semillas; el análisis de los datos se hizo así ya que, siendo el rendimiento de semilla de soya una característica bastante estudiada, cualquier resultado anómalo derivado del análisis indicaría un problema con el muestreo, con la forma en que se desarrolló el experimento o con el análisis de los datos.

El material biológico utilizado estuvo constituido por 7 cultivares y líneas avanzadas de soya las cuales se enumeran a continuación: Santa Rosa, Júpiter, UFV-1, H80-25-35, H80-26-39, H82-16-71, H86-51-50. El cultivar Tapachula-86 estuvo incluido inicialmente pero no fue posible obtener los datos del mismo ya que no mostró adaptación al ambiente en donde se desarrolló el experimento.

Las variables consideradas en el trabajo fueron: porcentaje de nitrógeno en las semillas, peso seco de las semillas, días a R1, días a R6-R7, altura de planta, número de nudos, número de nudos reproductivos, número de vainas, número de trifoliolos, número de ramificaciones, número de semillas, longitud foliar acumulada, ancho foliar acumulado, porcentaje de germinación de las semillas progenitoras, peso seco de las vainas, peso promedio de las semillas progenitoras y peso seco y porcentaje de nitrógeno para las partes vegetativas exceptuando la raíz.

El experimento se desarrolló en un invernadero utilizando un diseño completamente aleatorio con cuatro repeticiones por cultivar o línea y cuatro individuos por repetición. Los análisis consistieron en análisis gráficos preliminares y análisis estadísticos. Las diferencias entre cultivares respecto a las dos características de interés fueron analizadas con las pruebas de Kruskal-Wallis y de Mann-Whitney. No se encontraron diferencias significativas para la mayoría de los cultivares para porcentaje de nitrógeno en las semillas y peso seco de las mismas.

Para investigar las relaciones entre las variables se realizó una validación cruzada de los resultados utilizando diferentes técnicas de selección de variables. Las técnicas usadas fueron: correlación lineal simple, correlación lineal parcial, regresión lineal simple con selección de variables, regresión sobre la matriz de

correlaciones y regresión ridge.

No se detectó una relación inversa entre la variable peso seco de las semillas y contenido de nitrógeno de las mismas. La función lineal construida con las variables marcadas como importantes en la determinación del porcentaje de nitrógeno en las semillas (días a R6-R7, peso seco de las semillas, porcentaje de nitrógeno en las partes vegetativas, peso seco de las vainas, días a R1 y número de semillas) se mostró poco útil para propósitos de predicción. Se sigue de este resultado que la selección indirecta para el carácter contenido porcentual de nitrógeno en las semillas, utilizando los caracteres marcados como importantes, es poco recomendable.

LITERATURA CITADA

- Adams, M.W. 1967. Basis of Yield Component Compensation in Crop Plants with Special Reference to the Field Bean, *Phaseolus vulgaris*. Crop Sci. 7(5):505-510. United States of America.
- Bhatia, C.R. and R. Rabson. 1976. Bioenergetic Considerations in Cereal Breeding for Protein Improvement. Science. 194(4272):1418-1421. United States of America.
- Blad, B.L. and D.G. Baker. 1972. Orientation and Distribution of Leaves within Soybean Canopies. Agron. J. 64(1):26-29. United States of America.
- Box, G.E.P. 1953. Non-Normality and Tests of Variances. Biometrika. 40(3):318-335. U.K.
- Buttery, B.R. and R.I. Buzzell. 1988. Soybean Leaf Nitrogen in Relation to Photosynthetic Rate and Yield. Can. J. Plant Sci. 68(3):793-795. Canada.
- Byers, M., M.A. Kirkman and B.J. Mifflin. 1977. Factors Affecting the Quality and Yield of Seed Protein. In: Norton, G. (Ed.) Plant Proteins. Butterworth & Co. (Publishers) Ltd. London. U.K. p. 227-243.
- Carbonell Guevara, E.A. y R. Bartual Pastor. 1983. Valoración Agronómica y Clasificación de una Colección de Líneas de Soja Sembradas en el Bajo Guadalquivir. Publicación del Instituto Nacional de Investigaciones Agrarias. Madrid. España. 56 p.
- Cox, M.C., C.O. Qualset and D.W. Rains. 1985. Genetic Variation for Nitrogen Assimilation and Translocation in Wheat. II. Nitrogen Assimilation in Relation to Grain Yield and Protein. Crop Sci. 25(3):435-440. United States of America.
- Draper, N.R. and H. Smith. 1981. Applied Regression Analysis. John Wiley & Sons, Inc. New York. United States of America. 709 p.
- Egli, D.B., E.L. Ramseur, Y. Zhen-Wen and C.H. Sullivan. 1989. Source-Sink Alterations Affect the Number of

Cells in Soybean Cotyledons. Crop Sci. 29(3):732-735. United States of America.

- Evans, L.T. 1975. The Physiological Basis of Crop Yield. In: Evans, L.T. (Ed.) Crop Physiology: Some Case Histories. Cambridge University Press, London. U.K. p. 327-355.
- Fehr, W.R., C.E. Caviness, D.T. Burmood and J.S. Pennington. 1971. Stage of Development Descriptions for Soybeans, *Glycine max* (L.) Merrill. Crop Sci. 11(6):929-931. United States of America.
- Gent, M.P.N. and R.K. Kiyomoto. 1989. Assimilation and Distribution of Photosynthate in Winter Wheat Cultivars Differing in Harvest Index. Crop Sci. 29(1):120-125. United States of America.
- Hanson, W.D., R.C. Leffel and R.H. Howell. 1961. Genetic Analysis of Energy Production in the Soybean. Crop Sci. 1(2):121-126. United States of America.
- Hanway, J.J. and C.R. Weber. 1971. Dry Matter Accumulation in Eight Soybean (*Glycine max* (L.) Merrill) Varieties. Agron. J. 63(2): 227-230. United States of America.
- Hartwig, E.E. 1969. Breeding Soybeans for High Protein Content and Quality. In: FAO/IAEA Division of Atomic Energy in Food and Agriculture (Ed.). New Approaches to Breeding for Improved Plant Protein. International Atomic Energy Agency (IAEA). Vienna. Austria. p. 67-70.
- Hintze, J.L. 1985. Number Cruncher Statistical System (NCSS). Version 4.2. 7/85. Dr. Jerry L. Hintze. Kaysville, Utah 84037. United States of America.
- Hoerl, A.E. and R.W. Kennard. 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics 12(1):55-67. United States of America.
- Holbrook, C.C., J.W. Burton and T.E. Carter. 1989. Evaluation of Recurrent Restricted Index Selection for Increasing Yield While Holding Seed Protein Constant in Soybean. Crop Sci. 29(2):324-329. United States of America.
- Hollander, M. and D.A. Wolfe. 1973. Nonparametric Statistical Methods. John Wiley & Sons, Inc. New York. U.S.A. 503 p.

- Jackson, W.A., W.L. Pan, R.H. Moll and E.J. Kamprath. 1986. Uptake, Translocation, and Reduction of Nitrate. In: Neyra, C.A. (Ed.) Biochemical Basis of Plant Breeding. CRC Press, Inc. Boca Raton, Florida. United States of America. p.73-108.
- Jonhson, V.A., P.J. Mattern, D.A. Whited and J.W. Schmidt. 1969. Breeding for High Protein Content and Quality in Wheat. In: FAO/IAEA Division of Atomic Energy in Food and Agriculture (Ed.). New Approaches to Breeding for Improved Plant Protein. International Atomic Energy Agency (IAEA). Vienna. Austria. p. 20-40.
- Kendall, M. and A. Stuart. 1976. The Advanced Theory of Statistics. Vol. 3. Charles Griffin & Company Limited. London. 585 p.
- Kibite, S. and L.E. Evans. 1984. Causes of Negative Correlations Between Grain Yield and Grain Protein Concentration in Common Wheat. *Euphytica* 33(3):801-810. The Netherlands.
- Kwon, S.H. and J.H. Torrie. 1964. Heritability of and Interrelationships Among Traits of Two Soybean Populations. *Crop Sci.* 4(2): 196-198. United States of America.
- Lawlor, D.W., M. Kontturi and A.T. Young. 1989. Photosynthesis by Flag Leaves of Wheat in Relation to Protein, Ribulose Bisphosphate Carboxylase Activity and Nitrogen Supply. *J. Exp. Bot.* 40(210):43-52. United States of America.
- Ledent, J.F. and D.N. Moss. 1979. Relation of Morphological Characters and Shoot Yield in Wheat. *Crop Sci.* 19(4): 445-451. United States of America.
- Loomis, R.S. and W.A. Williams. 1963. Maximum Crop Productivity: an Estimate. *Crop Sci.* 3(1):67-72. United States of America.
- Salado-Navarro, L.R., K. Hinson and T.R. Sinclair. 1985. Nitrogen Partitioning and Dry Matter Allocation in Soybeans with Different Seed Protein Concentration. *Crop Sci.* 25(3):451-455. United States of America.
- Seen, S. 1989. Combining Outcome Measures: Statistical Power is Irrelevant. *Biometrics* 45(5):1027-1028. United States of America.

- Sherrard, J.H., R.J. Lambert, F.E. Below, R.T. Dunand, M.J. Messmer, M.R. Willman, C.S. Winkels and R.H. Hageman. 1986. Use of Physiological Traits, Especially Those of Nitrogen Metabolism for Selection in Maize. In: Neyra, C.A. (Ed.) Biochemical Basis of Plant Breeding. CRC Press, Inc. Boca Raton, Florida. United States of America. p. 109-130.
- Shibles, R., I.C. Anderson and A.H. Gibson. 1975. Soybean. In: Evans, L.T. (Ed.) Crop Physiology: Some Case Histories. Cambridge University Press, London. U.K. p. 151-189.
- Sinclair, T.R. and C.T. DeWit. 1975. Photosynthate and Nitrogen Requirements for Seed Production by Various Crops. Science 189(4202):565-567. United States of America.
- _____. 1976. Analysis of the Carbon and Nitrogen limitations to Soybean Yield. Agron. J. 68(2):319-324. United States of America.
- _____ and T. Horie. 1989. Leaf Nitrogen, Photosynthesis, and Crop Radiation Use Efficiency: A Review. Crop Sci. 29(1):90-98. United States of America.
- STSC, Inc. and Statistical Graphics Corporation. 1987. STATGRAPHICS. Statistical Graphics System. Version 2.6. United States of America.
- Westermann, D.T., L.K. Porter and W.A. O'Deen. 1985. Nitrogen Partitioning and Mobilization Patterns in Bean Plants. Crop Sci. 25(1):225-229. United States of America.
- Wetz, J.M. 1964. Criteria for Judging Adequacy of Estimation by an Approximating Response Function. Ph.D. Thesis. College of Engineering. University of Wisconsin. Madison, Wisconsin. U.S.A. 118 p.
- Williams, W.A., C.O. Qualset and S. Geng. 1979. Ridge Regression for Extracting Soybean Yield Factors. Crop Sci. 19(6):869-873. United States of America.