



TRABAJO PRÁCTICO

Parte I

Ejercicio 1: El archivo nombrado DAMASCOS contiene datos referido a una caracterización morfológica extraída de la publicación: Badenes et al (1993). Caracterización pomológica de variedades y clones de albaricoquero. *Investigación Agraria, Producción y Protección Vegetales*, 8: 55-65.

El material vegetal sobre el que se trabajó pertenece a una colección de damascos (*Prunus armeniaca* L.) autóctonos de la Comunidad Valenciana (España), seleccionados por la Unidad de Cultivos Leñosos del Departamento de Producción Vegetal de la Escuela Técnica Superior de Ingenieros Agrónomos (Universidad Politécnica de Valencia), en la zona de cultivo del damasco de las comarcas Rivera Alta y Vall d'Albaida.

La muestra estuvo conformada por 18 variedades distintas. Con el fin de caracterizar la colección de variedades se tomaron muestras de flores, hojas y frutos registrándose los principales caracteres cuantitativos de estos órganos utilizados por la UPOV (Unión para la protección de las obtenciones vegetales) para la patente de variedades, referidos a la especie de damasco.

Las flores se recogieron en plena floración, eligiéndose aquellas totalmente desarrolladas y abiertas. Los caracteres medidos fueron: tamaño de la flor (TAMFLOR), longitud de pétalo (LONGPET) y ancho de pétalo (ANCHOPET).

Las hojas fueron muestreadas en el mes de mayo, se eligieron hojas adultas procedentes de ramos mixtos, aproximadamente de la parte central de éstos, y se seleccionaron aquellas en buen estado fisiológico y sanitario. Los caracteres medidos fueron: superficie de hoja (SUPHOJA), relación longitud y ancho de hoja (LONANCHO), relación peciolo-limbo (PECLIMBO).

Por otra parte, los frutos se recolectaron maduros, tomando como referencia la desaparición total del color verde, se muestrearon al azar de todas las caras del árbol. Los caracteres registrados fueron: peso de fruto (PESOF), longitud de fruto (LONGF), ancho de fruto (ANCHOF), espesor de fruto (ESPESORF), peso del endocarpio (PESOEND), longitud del endocarpio (LONGEND) y ancho del endocarpio (ANCHOEND).

El tamaño de la muestra para cada órgano y variedad fue de 25 repeticiones.

Todos los registros de peso están expresados en gramos, los de longitud en milímetros y superficie en centímetros cuadrados.

El archivo contiene las medias de cada uno de los caracteres mencionados para todas las variedades bajo estudio.

- ¿Cuáles son los valores de n y p ? ¿Cuanto vale y que indica el valor x_{32} ? ¿Y el vector \mathbf{x}_6 ?
- ¿Cómo clasificaría las variables sobre las que se está trabajando?
- Encuentre el vector de medias y matriz de varianzas-covariancias asociados a la tabla de datos?
- ¿Podría decir cuál y cuáles variables son las más dispersas?
- ¿Estandarice las variables por media y desvío? Ahora puede responder al inciso (d)?
- ¿Halle la matriz de correlación. ¿Que variables son las más relacionadas?
- ¿Pueden dividirse las variables en subgrupos, de modo que las variables dentro de un mismo subgrupo tengan elevadas correlaciones entre sí y que las que se encuentren en subgrupos diferentes tengan bajas correlaciones? Si es así, ¿cuáles variables quedan en cada uno de los subgrupos?
- Encuentre la matriz que mide el grado de similitud entre las variedades en función de la distancia euclídea calculada sobre los datos originales.
- ¿Podría decir cuáles son los tres pares de variedades que presentan mayor semejanza?

- j) Repita lo realizado en el inciso (h) pero sobre las variables estandarizadas por media y desvío estándar. Son las mismas las tres variedades más parecidas ? Comente al respecto.
- k) Mida el grado de concordancia entre ambas matrices de distancia.
- l) Realice un Análisis de Componentes Principales utilizando de la matriz de correlaciones.
- m) Analice los porcentaje de variabilidad explicada por los primeros ejes principales.
- n) Establezca intuitivamente grupos de variedades similares según su cercanía en el plano principal.
- o) Encuentre e interprete gradientes de las variables originales en el plano principal en función de sus cargas sobre las dos primeras componentes.
- p) Caracterice los grupos determinados en el inciso (n) según los gradientes descriptos en (o).
- q) Superponga en la representación del plano principal un MST. Comente al respecto, haría algún reagrupamiento ?
- r) Con el software R realice el ACP recurriendo a operaciones con matrices (decomposición espectral)
- s) Verifique que con el enfoque Biplot (DVS) llega a los mismos resultados
- t) Obtenga 4 dendogramas ultramétricos según diferentes criterios de encadenamiento (SIMPLE, COMPUESTO, UPGMA y WARD).
- u) Asocie a cada uno de los árboles obtenidos en el inciso anterior la matriz cofenética correspondiente. Que miden los elementos de estas matrices ?
- v) Cuantifique la concordancia entre la matriz de distancia que dio origen a los dendogramas y las 4 matrices cofenéticas. A que conclusión llega ?
- w) Existe algún punto de corte sobre el índice de jerarquización del dendograma UPGMA que origine los mismos agrupamiento de variedades obtenidos en Análisis de Componentes Principales ?
- x) Mida el grado de concordancia entre los resultados obtenidos por Componentes y por Cluster UPGMA
- y) Halle el dendograma aditivo Neighbor Joining, representa mejor la configuración de variedades que el Cluster UPGMA ?

Ejercicio 2: Sobre 41 variedades de pepino (*Cucumis sativus*) correspondientes a cuatro tipos de frutos diferentes (holandés, francés, pepinillo y α - β) se registraron además ocho caracteres agronómicos relacionados con variables cualitativas: expresión del sexo (SEXO), color de ornamenta (CORNAMEN), tipo de ornamenta (TORNAMEN), forma del extremo de pedúnculo (PEDUNC), presencia o ausencia de verrugas (VERRUGAS) y moteado (MOTEADO) en fruto, y resistencia o susceptibilidad a Cladospora (CLADOSP) y Virus del Mosaico (CMV). Estos datos que se presentan en el archivo CUALITATIVAS fueron suministrados por la Subdirección General de Semillas y Plantas de Vivero del Ministerio de Agricultura, Pesca y Alimentación y el Laboratorio de Genética del Instituto Valenciano de Investigaciones Agrarias (España).

Las codificaciones de las variables son las siguientes:

VAR: variedad (con una letra inicial se indica su pertenencia al tipo de fruto)

TIPO: tipo de fruto: 1 holandés, 2 francés, 3 pepinillo, 4 alfa-beta

SEXO: expresión del sexo: 1 flores masculinas principalmente; 2 50% flores masculinas; 3 flores femeninas principalmente

CORNAMEN: color de ornamenta: 1 negra, 2 blanca

TORNAMEN: tipo de ornamentación: 1 pelos, 2 espinas

PEDUNC: extremo pedúnculo: 1 con cuello, 2 agudo, 3 obtuso

VERRUGAS: verrugas en fruto: 0 ausente, 1 presente

MOTEADO: moteado en fruto: 0 ausente, 1 presente

CLADOSP: resistencia a cladospora: 0 ausente, 1 presente, (-1) no se tiene información

CMV: resistencia virus mosaico del coliflor (CMV): idem

- a) Cuantifique (en forma manual) la similitud entre las variedades correspondientes a la primera y segunda fila en función del porcentaje de caracteres comunes respecto al número de caracteres totales. Idem las variedades asociadas a las filas 12 y 13 (incluir la variable TIPO).
- b) Halle una matriz de similitud entre variedades en función del coeficiente SM generalizado.
- c) Aplique Análisis de Coordenadas principales para representar en un espacio bidimensional la semejanza entre las variedades.
- d) Conforme grupos de variedades según su homogeneidad en la caracterización agronómica cualitativa.
- e) Encuentre el dendrograma ultramétrico con ligamiento UPGMA correspondiente
- f) Mida a través de su matriz cofilética la concordancia con la matriz de distancias que le dio origen
- g) Cuantifique concordancia entre plano principal de ACoordP y Cluster

Ejercicio 3: En el archivo CUANTITATIVAS se presentan los datos correspondientes a la caracterización agronómica cuantitativa de las 41 variedades de pepino tratadas en el ejercicio anterior. Los caracteres cuantitativos analizados fueron: número de flores femeninas por nudo (FLORES), número de espinas en el ovario (ESPINAS), número de aristas (ARISTAS) y estrías (ESTRIAS) en fruto, longitud de fruto (FRUTO), intensidad de cuello (CUELLO) e intensidad de color de cuello (CCUELLO).

- a) Aplique sobre estos datos un Análisis de Componentes Principales a partir de matriz de correlaciones.
- b) Realice la representación de las variedades en el plano principal, encuentre grupos y caracterícelos
- c) Evalúe visualmente y a través de la correlación entre matrices la concordancia general entre esta configuración y la hallada en función de los caracteres cualitativos

Ejercicio 4: Con los resultados del Análisis de Coordenadas Principales y Análisis de Componentes Principales obtenidos en los dos ejercicios anteriores

- a) Halle la configuración de consenso recurriendo a Análisis de Procrustes Generalizados
- b) Podría decir para cual o cuales variedades existe mayor concordancia entre la caracterización cuantitativa y cualitativa ?
- c) Idem para las variedades con mayor discrepancia entre ambas configuraciones.
- d) Grafique sólo los puntos de consenso y diga si pudo conformar grupos asociados a los tipos de frutos. Compare con los resultados de los ejercicios 2 y 3.

Ejercicio 5: Posteriormente sobre estas 41 variedades de pepino se realizó una caracterización molecular por la técnica de RAPDs (Random Amplified Polymorphic DNA Sequence), utilizando 11 cebadores que originaron 33 polimorfismos. Para cada variedad se analizaron de cuatro a seis plantas registrándose la frecuencia relativa de presencia de banda para cada marcador.

El archivo RAPDS contiene la información donde se indican las 33 bandas con B01, B02, B03, , B033 ; y en las filas las variedades.

- a) Halle la distancia genética de Prevosti entre variedades.
- b) Podría aplicar el coeficiente de similitud SM ? Porque ?
- c) Realice un Análisis de Coordenadas Principales para encontrar la configuración de las variedades de pepino en función de esta caracterización molecular. Encuentra asociaciones en función del tipo de pepino?
- d) Encuentre el dendrograma ultramétrico con ligamiento UPGMA
- e) Mida a través de su matriz cofilética la concordancia con la matriz de distancias que le dio origen
- f) Halle el dendrograma aditivo neighbor Joining
- g) Mida su concordancia con matriz de distancia original
- h) Relacione ambos dendogramas y saque conclusiones

NOTA: Archivos CUALITATIVAS, CUANTITATIVAS y RAPDS extraídos parcialmente de las publicaciones:

Bernet et al (2003). Applicability of molecular markers in the context of protection of new varieties of Cucumis sativus L. Plant Breeding, 122:146-152.

Bramardi et al (2005). Simultaneous agronomic and molecular of genotypes via the generalised analysis procrustes analysis: an application to varieties of cucumber. Crop Science. 45:1603-1609.

Ejercicio 6:

- Mida la concordancia entre las configuraciones en el plano principal obtenidas en los ejercicios 2, 3 y 5
- Mida la concordancia entre la caracterización agronómica (cualitativa+cuantitativa) y molecular (planos principal)
- Con APG halle el consenso entre las configuraciones (plano principal) obtenidas en base a datos cualitativos, cuantitativos y moleculares.
- Identifique si hay algún tipo de pepino para el cual hay más discrepancia entre estas tres caracterizaciones
- Finalice el análisis con un cluster UPGMA obtenido a partir de la configuración de consenso

Ejercicio 7: En los archivos MAIZ1 y MAIZ2 se presentan los datos correspondientes a la caracterización de 31 poblaciones nativas de maíz conservadas en el Banco de Germoplasma de INTA Pergamino, en función de 10 variables cuantitativas, en dos ambientes: Pergamino (Ambiente 1) y Ferré (Ambiente 2).

Los descriptores o variables observadas fueron:

LMZ	Longitud de mazorca (cm)	GRXM	Granos por metro
DMZ	Diámetro de mazorca (mm)	PESO	Peso de 1000 granos (gr)
AGR	Ancho de grano (mm)	REND	Rendimiento (Kg/ha)
LGR	Largo de grano (mm)	ALTPL	Altura de planta (cm)
PROL	Prolificidad (espigas/planta)	ALTMZ	Altura de inserción de la mazorca (cm)

- Cuantifique la concordancia de la configuración de poblaciones nativas de maíz en ambos ambientes en el espacio original mediante el coeficiente de correlación de Pearson entre matrices de distancias euclídeas estandarizadas entre individuos, y utilizando el coeficiente R_v .
- Realice un ACP para cada ambiente, compare semejanzas y diferencias entre ambas caracterizaciones tanto para individuos como para variables.
- Cuantifique la relación de las dos configuraciones en el plano principal originado por los ACP
- Encuentra indicios de interacción tanto genotipo-ambiente como variable-ambiente ?
- Como se quiere encontrar una caracterización 'media' o 'promedio' para las 31 poblaciones en función de la información dada en ambos ambientes proceda a realizar una ACP sobre el promedio de las variables para ambos ambientes
- Ahora concatene ambos archivos por filas y columnas y realice un ACP para ambas situaciones, encuentra respuesta para las interacciones planteadas en el inciso b ?
- Aplicar APG para tener otra visualización de la interacción genotipo-ambiente (retenga todas las dimensiones de ambas configuraciones).
- Visualizar del gráfico correspondiente las 3 poblaciones con mayor efecto ambiente y las 3 con menor efecto
- Comparar lo encontrado en h) con el ANOVA correspondiente. Realice comentarios.

Datos extraídos de: MARTICORENA, M.; BRAMARDI, S.; DEFACIO, R. (2010). Characterization of maize populations in different environmental conditions by means of Three-Mode Principal Components Analysis. *Revista Ciencia e Investigación Agraria*. 37(3): 93-105.