

9. ANÁLISIS DE CORRELACIONES CANÓNICAS (ACC)

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

GENERALIZACIÓN DEL ANÁLISIS DE REGRESIÓN MÚLTIPLE

- En algunos conjuntos de datos multivariados, las variables se dividen naturalmente en dos grupos.
- **Objetivo del ACC:** Investigar la relación existente entre dos grupos de variables.
- **Ejemplo:** (Tabla 1.3 Distribución de una mariposa) El conjunto de datos consta de 16 colonias de mariposas, de las que se tienen datos para 4 variables ambientales y 6 frecuencias genéticas. **¿Existe alguna relación entre estos dos conjuntos de variables?**

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

GENERALIZACIÓN DEL ANÁLISIS DE REGRESIÓN MÚLTIPLE

- **Ejemplo (Hotelling, 1936):** Resultados de una prueba de velocidad de lectura (X_1), capacidad de lectura (X_2), velocidad de cálculo aritmético (Y_1) y capacidad de cálculo aritmético (Y_2), practicada a 140 escolares de 7o. grado.
- ¿La habilidad de lectura (medida por X_1 y X_2) está o no relacionada con la habilidad en aritmética (medida por Y_1 y Y_2)?

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

GENERALIZACIÓN DEL ANÁLISIS DE REGRESIÓN MÚLTIPLE

- Para responder a esta interrogante, el ACC busca una combinación lineal de X_1 y X_2 :

$$U = a_1X_1 + a_2X_2$$

y una combinación lineal de Y_1 y Y_2 :

$$V = b_1Y_1 + b_2Y_2$$

seleccionados de modo que la correlación entre U y V sea la mayor posible.

NOTA: Es el mismo principio que en el ACP, excepto que maximiza la correlación en vez de la varianza.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

GENERALIZACIÓN DEL ANÁLISIS DE REGRESIÓN MÚLTIPLE

- Con X_1 , X_2 , Y_1 e Y_2 estandarizadas, Hotelling encontró la mejor elección para U y V :

$$U = -2,78X_1 + 2,27X_2$$

$$V = -2,44Y_1 + 1,00Y_2$$

en que las dos variables tienen correlación 0,62.

U mide las diferencias entre capacidad y velocidad de lectura, y **V** mide la diferencia entre capacidad y velocidad aritmética. Vale decir que los alumnos con grandes diferencias entre X_1 y X_2 , también tienden a grandes diferencias entre Y_1 e Y_2 . Es en este aspecto que lectura y aritmética muestran la mayor correlación.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

GENERALIZACIÓN DEL ANÁLISIS DE REGRESIÓN MÚLTIPLE

- En el **ANÁLISIS DE REGRESIÓN MÚLTIPLE** se analiza la relación existente entre una variable **Y** y dos o más variables **X_1, X_2, \dots, X_p** .
- El **ANÁLISIS DE CORRELACIONES CANÓNICAS** es una generalización de la regresión múltiple, en que **varias variables Y** se relacionan simultáneamente **con varias variables X**.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

GENERALIZACIÓN DEL ANÁLISIS DE REGRESIÓN MÚLTIPLE

Correlaciones canónicas

- Si se tienen **p** variables X_1, X_2, \dots, X_p y **q** variables Y_1, Y_2, \dots, Y_q , existen **r = Mín (p, q)** combinaciones lineales:

$$U_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p$$

$$U_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p$$

⋮

⋮

⋮

$$U_r = a_{r1}X_1 + a_{r2}X_2 + \dots + a_{rp}X_p$$

$$V_1 = b_{11}Y_1 + b_{12}Y_2 + \dots + b_{1q}Y_q$$

$$V_2 = b_{21}Y_1 + b_{22}Y_2 + \dots + b_{2q}Y_q$$

⋮

⋮

$$V_r = b_{r1}Y_1 + b_{r2}Y_2 + \dots + b_{rq}Y_q$$

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

GENERALIZACIÓN DEL ANÁLISIS DE REGRESIÓN MÚLTIPLE

Estas combinaciones lineales se seleccionan de modo que

- **la correlación entre U_1 y V_1 sea la máxima posible;**
- **la correlación entre U_2 y V_2 sea la segunda máxima correlación posible, sujeto a que estas dos variables no estén correlacionadas con U_1 y V_1 ;**
- **la correlación entre U_3 y V_3 sea la tercera máxima correlación posible, sujeto a que estas dos variables no estén correlacionadas con U_1, V_1, U_2 y V_2 ;**
- y así sucesivamente.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

GENERALIZACIÓN DEL ANÁLISIS DE REGRESIÓN MÚLTIPLE

- Cada par de variables canónicas $(U_1, V_1), (U_2, V_2), \dots, (U_r, V_r)$ representan una "dimensión" independiente en la relación entre los dos conjuntos de variables (X_1, X_2, \dots, X_p) y (Y_1, Y_2, \dots, Y_q) .
- El primer par (U_1, V_1) tiene la mayor correlación posible y es, por lo tanto, el más importante. El segundo par (U_2, V_2) tiene la segunda mayor correlación posible y, por lo tanto, es el segundo par más importante. etc. etc.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

PROCEDIMIENTO PARA UN ANÁLISIS DE CORRELACIONES CANÓNICAS

- Supongamos que la matriz de correlación $(p+q) \times (p+q)$ entre las variables $X_1, X_2, \dots, X_p, Y_1, Y_2, \dots, Y_q$ tiene la siguiente forma cuando se calcula a partir de una muestra:

$$\begin{array}{c}
 X_1 \quad X_2 \dots X_p \quad Y_1 \quad Y_2 \dots Y_q \\
 \begin{array}{c}
 X_1 \\
 X_2 \\
 \cdot \\
 \cdot \\
 \cdot \\
 X_p \\
 Y_1 \\
 Y_2 \\
 \cdot \\
 \cdot \\
 \cdot \\
 Y_q
 \end{array}
 \left[\begin{array}{cc}
 & \\
 \begin{array}{c}
 p \times p \text{ matrix} \\
 \mathbf{A}
 \end{array} & \begin{array}{c}
 p \times q \text{ matrix} \\
 \mathbf{C}
 \end{array} \\
 \hline
 \begin{array}{c}
 q \times p \text{ matrix} \\
 \mathbf{C}'
 \end{array} & \begin{array}{c}
 q \times q \text{ matrix} \\
 \mathbf{B}
 \end{array}
 \end{array}
 \right.
 \end{array}$$

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

PROCEDIMIENTO PARA UN ANÁLISIS DE CORRELACIONES CANÓNICAS

- A partir de la matriz anterior puede calcularse la **matriz** $\mathbf{B}^{-1}\mathbf{C}'\mathbf{A}^{-1}\mathbf{C}$, de dimensión $\mathbf{q} \times \mathbf{q}$.
- Y seguidamente los valores propios de esta matriz:
$$(\mathbf{B}^{-1}\mathbf{C}'\mathbf{A}^{-1}\mathbf{C} - \lambda\mathbf{I})\mathbf{b} = \mathbf{0}$$
- Resulta que los valores propios $\lambda_1 > \lambda_2 > \dots > \lambda_r$ son entonces los cuadrados de las correlaciones entre las variables canónicas y
- los elementos de los correspondientes vectores propios, $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_r$ son los coeficientes de las variables Y para las variables canónicas.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

PROCEDIMIENTO PARA UN ANÁLISIS DE CORRELACIONES CANÓNICAS

- Los coeficientes de la i-ésima variable canónica U_i , para las variables X, están dados por los elementos del vector:
$$\mathbf{a}_i = \mathbf{A}^{-1}\mathbf{C}\mathbf{b}_i$$
- Los cálculos anteriores presuponen que las variables X e Y están estandarizadas (media cero y desviación estándar unitaria)

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

PROCEDIMIENTO PARA UN ANÁLISIS DE CORRELACIONES CANÓNICAS

- El i -ésimo par de variables canónicas resulta:

$$U_i = \mathbf{a}'_i \mathbf{X}$$

$$V_i = \mathbf{b}'_i \mathbf{Y}$$

con $\mathbf{a}'_i = (a_{i1}, a_{i2}, \dots, a_{ip})$

$$\mathbf{b}'_i = (b_{i1}, b_{i2}, \dots, b_{iq})$$

$$\mathbf{X}' = (x_1, x_2, \dots, x_p)$$

$$\mathbf{Y}' = (y_1, y_2, \dots, y_q)$$

donde \mathbf{X} e \mathbf{Y} son vectores de datos estandarizados.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

PROCEDIMIENTO PARA UN ANÁLISIS DE CORRELACIONES CANÓNICAS

- Las **varianzas de U_i y V_i** dependen de la escala adoptada para el **valor propio b_i** . Sin embargo, a partir de los datos puede calcularse la desviación estándar de U_i y dividir los valores de a_{ij} por ésta. Esto produce una **variable canónica U_i escalada, con varianza unitaria**. De modo similar, si los valores de b_{ij} se dividen por la desviación estándar de V_i , esto produce una **variable canónica V_i escalada, con varianza unitaria**.
- Esta forma de estandarización de las variables canónicas no es esencial, puesto que el escalado no afecta la correlación entre U_i y V_i . Sin embargo, puede ser útil cuando se examinan los valores numéricos que asumen las variables canónicas para las observaciones, utilizando los datos disponibles.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

PRUEBA DE SIGNIFICANCIA

Si existen r valores propios, luego existen r pares de variables canónicas. Sin embargo, algunos de esos pares pueden reflejar correlaciones demasiado pequeñas para ser estadísticamente significativas.

Prueba de Bartlett (1947)

- **Objetivo:** Una prueba aproximada para analizar la relación entre las variables \mathbf{X} como un todo y las variables \mathbf{Y} como un todo, para el caso en que los datos provienen de una muestra aleatoria con distribución normal multivariada.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

PRUEBA DE SIGNIFICANCIA

- **Procedimiento:**

Cálculo del estadígrafo: $X^2 = -\left[n - \frac{1}{2}(p + q + 3)\right] \sum_{i=1}^r \ln(1 - \lambda_i)$

donde n es el número de casos para los cuales se disponen datos.

El estadígrafo se compara con el porcentaje de puntos de la distribución χ^2 con pq g.d.l. Un valor significativamente grande, proporciona evidencia de que por lo menos una de las r correlaciones canónicas es significativa. Un resultado no significativo indica que incluso la correlación canónica más grande puede explicarse solo por variaciones de muestreo.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

PRUEBA DE SIGNIFICANCIA

Algunos autores sugieren que esta prueba se puede extender para probar la importancia de cada una de las correlaciones canónicas.

Las sugerencias implican:

1) Compare la i -ésima contribución:

$$-\left[n - \frac{1}{2}(p + q + 3)\right] \sum_{i=1}^r \ln(1 - \lambda_i)$$

con los puntos porcentuales de la distribución χ^2 con $p+q-2i+1$ g.d.l.

2) Compare la suma desde la $(i+1)$ -ésima hasta la r -ésima contribución con los puntos porcentuales de la distribución χ^2 con $(p-1)(q-1)$ g.d.l.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

PRUEBA DE SIGNIFICANCIA

Aquí, se supone que el **primer enfoque** está probando la i -ésima correlación canónica directamente, mientras que el **segundo** está probando la significación de las $(i + 1)$ ésimas hasta r -ésimas correlaciones canónicas como un todo.

La razón por la cual **estas pruebas no son confiables** es esencialmente la misma que ya se discutió para la prueba que se utiliza con el AFD. Esto es, que la i -ésima mayor correlación canónica de la muestra puede, de hecho, haber surgido de una correlación canónica de una población que no es la i -ésima más grande. Por lo tanto, la asociación entre las r contribuciones y las correlaciones canónicas de la r población es difusa (Harris 1985, p. 211).

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

INTERPRETACIÓN DE LAS VARIABLES CANÓNICAS

- Luego de decidir cuales variables canónicas son significativas, a continuación debe interpretarse su significado: **¿Qué miden?**

$$U_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$$

$$V_i = b_{i1}Y_1 + b_{i2}Y_2 + \dots + b_{iq}Y_q$$

- A primera vista puede parecer que esta es una pregunta trivial. En particular, si U_i puede describirse en términos de X variables, con coeficientes a_{ij} grandes, y V_i en términos de Y variables, con coeficientes b_{ij} grandes. "Grandes" significa positivo grande o negativo grande.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

INTERPRETACIÓN DE LAS VARIABLES CANÓNICAS

- Lamentablemente, **las correlaciones entre las variables X e Y pueden perturbar este proceso de interpretación.** Por ejemplo, puede suceder que $a_{i1} > 0$ y que la correlación simple entre U_i y X_1 sea negativa. Esta aparente contradicción puede ocurrir debido a que X_1 esté fuertemente correlacionada con una o más de las otras variables X y que, parte del efecto de X_1 esté siendo computado por los coeficientes de las otras variables X . De hecho, si una de las variables X es casi una combinación lineal de las otras variables X , luego existirá una variedad infinita de combinaciones lineales de las variables X , algunos de ellas con valores de los coeficientes a_{ij} muy diferentes, que producen virtualmente el mismo valor de U_i . La misma consideración es válida para las combinaciones lineales de las variables Y .

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

INTERPRETACIÓN DE LAS VARIANTES CANÓNICAS

- Los problemas de interpretación que surgen con las variables X o Y altamente correlacionadas le son familiares a los usuarios del análisis de regresión múltiple. Exactamente los mismos problemas surgen con la estimación de los coeficientes de regresión.
- Realmente, un comentario razonable parece ser que si las variables X o Y están fuertemente correlacionadas puede no haber manera alguna de dilucidar sus contribuciones a las variables canónicas. No obstante, no hay dudas respecto a que cualquier persona continuará intentando efectuar interpretaciones bajo estas circunstancias.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

INTERPRETACIÓN DE LAS VARIANTES CANÓNICAS

Algunos autores han sugerido que es mejor describir las variables canónicas analizando sus correlaciones con las variables X e Y , en lugar de los coeficientes a_{ij} y b_{ij} . Por ejemplo, si U_i está positivamente fuertemente correlacionado con X_1 , luego puede considerarse que U_i refleja en gran medida a X_1 . Del mismo modo, si V_i está negativamente fuertemente correlacionada con Y_1 , puede considerarse que V_i refleja en gran medida lo contrario que Y_1 .

Esta propuesta tiene el mérito de considerar todas las variables con las que las variables canónicas parecen estar correlacionadas.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

APLICACIONES DEL ACC

Ej. 3: Distribución de una mariposa

- Estudio de **16 colonias** de la mariposa *Euphydryas editha* en California y Oregon.
- **4 variables ambientales** (altitud, precipitación anual, temperatura máxima y mínima anual)
- **6 variables genéticas** (porcentajes de 6 genes de genesfosfoglucoasa-isomerasa [Pgi])
- Cualquier relación significativa entre las variables ambientales y genéticas tiene interés puesto que indica la adaptación de E. editha al medio ambiente local.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

APLICACIONES DEL ACC

- Computacionalmente, en este ACC:
 - X** = variables ambientales
 - Y** = frecuencias genéticas
- Además, puesto que la suma de las variables **Y** representa el 100%, no tiene sentido incluirlas a todas en el análisis, puesto que permitiría obtener diferentes combinaciones lineales de estas variables, para tener la misma correlación con una combinación de las variables X. En efecto,

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

APLICACIONES DEL ACC

asumamos que el primer par (U_1, V_1) sea:

$$V_1 = b_{11}Y_1 + b_{12}Y_2 + \dots + b_{16}Y_6$$

Luego V_1 se puede reescribir sustituyendo Y_1 por 100 menos la suma de las otras variables, con lo que:

$$V_1 = 100b_{11} + (b_{12} - b_{11})Y_2 + \dots + (b_{16} - b_{11})Y_6$$

Esto significa que la correlación entre U_1 y V_1 es la misma que entre

$$(b_{12} - b_{11})Y_2 + \dots + (b_{16} - b_{11})Y_6 \text{ y } U_1$$

Puesto que la constante $100b_{11}$ en la segunda combinación lineal no afecta la correlación.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

APLICACIONES DEL ACC

Este problema se supera eliminando una de las frecuencias génicas del análisis. En este caso, se omitió la frecuencia del gen 1.30.

Variables del ACC

X Variables:

X_1 =altitud ; X_2 =precipitación anual ; X_3 = temperatura máxima anual ; X_4 =temperatura mínima anual

Y Variables:

Y_1 =combinación de genes de baja frecuencia de movilidad 0.40 y 0.60 ; Y_2 =movilidad 0.80 ; Y_3 =movilidad 1.0 ; Y_4 =movilidad 1.6

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

APLICACIONES DEL ACC

- ☐ Valores estandarizados de X_1 a X_4 e Y_1 a Y_4
- ☐ Matriz de correlaciones para las variables X_1 a X_4 e Y_1 a Y_4 , para datos de *Euphydryas editha*, dividida en las submatrices A, B, C y C':

Variables	X1	X2	X3	X4	Y1	Y2	Y3	Y4
X1	1	0,567	-0,828	-0,936	-0,201	-0,573	0,727	-0,458
X2	0,567	1	-0,479	-0,705	-0,468	-0,550	0,699	-0,138
X3	A -0,828	-0,479	1	0,719	C 0,224	0,536	-0,717	0,438
X4	-0,936	-0,705	0,719	1	0,246	0,593	-0,759	0,412
Y1	-0,201	-0,468	0,224	0,246	1	0,638	-0,561	-0,584
Y2	C' -0,573	-0,550	0,536	0,593	B 0,638	1	-0,823	-0,127
Y3	0,727	0,699	-0,717	-0,759	-0,561	-0,823	1	-0,264
Y4	-0,458	-0,138	0,438	0,412	-0,584	-0,127	-0,264	1

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

APLICACIONES DEL ACC

- ☐ Valores propios de $(B^{-1}C'A^{-1}C-\lambda I)b=0$:

Valores propios:				
	F1	F2	F3	F4
Valor propio	0,743	0,202	0,149	0,008
Variabilidad (%)	67,459	18,336	13,497	0,709
% acumulado	67,459	85,795	99,291	100,000

- ☐ Los valores propios $\lambda_1 > \lambda_2 > \dots > \lambda_r$ son los cuadrados de las correlaciones canónicas entre las variables canónicas:

Correlaciones canónicas:				
	F1	F2	F3	F4
	0,862	0,449	0,386	0,088

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

APLICACIONES DEL ACC

□ Variables canónicas:

Coeficientes canónicos estandarizados (Y1):				
	F1	F2	F3	F4
X1	-0,124	2,394	-2,979	-1,374
X2	-0,293	-0,692	-1,350	-0,242
X3	0,468	0,468	-0,580	-1,702
X4	0,260	1,362	-3,551	0,083

Coeficientes canónicos estandarizados (Y2):				
	F1	F2	F3	F4
Y1	0,548	-1,727	3,503	-0,659
Y2	0,422	-2,241	1,320	1,412
Y3	-0,089	-3,808	3,790	0,511
Y4	0,825	-2,818	2,778	-0,631

Hay 4 CC porque este es el mínimo de la cantidad de variables X y el número de variables Y (que resultan ser 4)

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

APLICACIONES DEL ACC

□ Prueba de Bartlett

Aunque los CC son bastante grandes, no son estadísticamente significativos, debido al pequeño tamaño de la muestra (16 observaciones). Se encuentra: $X^2=18,34$ con 16 g.d.l.; la probabilidad de un valor tan grande de la distribución X^2 es de aproximadamente 0,30

- Dejando a un lado la falta de significancia, puede tener interés interpretar el primer par de variables canónicas:

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

APLICACIONES DEL ACC

Interpretación de las variantes canónicas

$$U_1 = -0.12X_1 - 0.29X_2 + 0.47X_3 + 0.26X_4$$

$$V_1 = 0.55Y_1 + 0.42Y_2 - 0.09Y_3 + 0.82Y_4$$

(U₁, V₁) La ecuación de U₁ muestra un contraste entre, por una parte, X₃ (temperatura máxima) y X₄ (temperatura mínima) y, por otra parte, X₂ (precipitación). Para V₁ se tiene un coeficiente moderado a grande para Y₁ (movilidad 0.40 y 0.60), Y₂ (movilidad 0.80) y Y₄ (movilidad 1.16) y un coeficiente pequeño y negativo para Y₃ (movilidad 1.00). Parecería que el gen de movilidad 0.40, 0.60, 0.80 y 1.16 tienden a ser frecuentes en colonias con altas temperaturas y baja precipitación.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

APLICACIONES DEL ACC

Correlaciones entre U₁ y las variables X:

Correlaciones entre las variables de entrada y las variables canónicas (Y1):

	F1	F2	F3	F4	
X1	-0,921	0,339	0,059	-0,180	Altitud
X2	-0,771	-0,517	-0,261	-0,265	Precipitación
X3	0,898	-0,203	-0,021	-0,389	t máx
X4	0,919	-0,055	-0,228	0,316	t mín

Esto sugiere que U₁ se interpreta mejor como una medida de altas temperaturas y baja altitud y precipitación.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

APLICACIONES DEL ACC

□ Correlaciones entre V_1 y las variables Y:

Correlaciones entre las variables de entrada y las variables canónicas (Y2):

	F1	F2	F3	F4	
Y1	0,384	0,627	0,595	0,324	Movilidad 0.40/0.60
Y2	0,740	0,151	0,081	0,651	Movilidad 0.80
Y3	-0,961	-0,251	0,005	-0,116	Movilidad 1.00
Y4	0,475	-0,520	-0,436	-0,560	Movilidad 1.16

- V_1 parece asociado con una falta de genes de movilidad 1.0

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

APLICACIONES DEL ACC

Las interpretaciones de U_1 y V_1 no son las mismas cuando se hacen basándose en los coeficientes de las funciones canónicas, que si se basan en correlaciones.

Para U_1 , la diferencia no es grande y solo concierne la altitud, pero para V_1 , la importancia de los genes de movilidad 1.0 es muy diferente. En general, las interpretaciones basadas en correlaciones parecen mejores puesto que se corresponden con lo que se observa en los datos.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

APLICACIONES DEL ACC

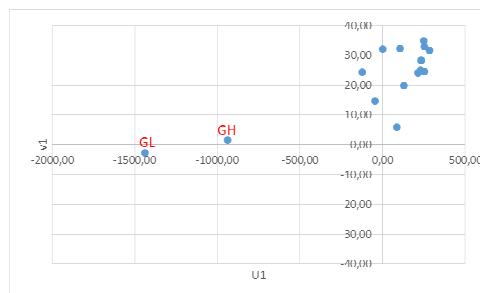
Por ejemplo, la colonia GL tiene la altitud más alta, la precipitación alta, las temperaturas más bajas y la frecuencia más alta de genes de movilidad 1.0. Esto se compara con la colonia UO con baja altitud, baja precipitación, alta temperatura y la menor frecuencia de movilidad de genes 1.0.

Sin embargo, como se mencionó anteriormente, **existen problemas reales con la interpretación de variables canónicas cuando las variables con las que están construidas tienen altas correlaciones.**

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

APLICACIONES DEL ACC

Un gráfico de los valores de las colonias para las variables canónicas U1 y V1 muestra que las colonias GL y GH son diferentes en comparación con las otras colonias. La frecuencia de genes de movilidad 1.0 es inusualmente alta.



AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

APLICACIONES DEL ACC

Software

La opción de ACC no está tan ampliamente disponible en paquetes estadísticos como las opciones para análisis multivariado consideradas hasta ahora.

XLSTAT no incluye ACC en el paquete básico, es necesario adquirir el paquete Premium.

Bibliografía

Giffins, R. (1985) *Canonical Analysis: a Review with applications in Ecology*. Springer-Verlag, Berlin.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

ACC - RESUMEN

- ✓ El ACC puede considerarse como una generalización de la regresión múltiple, que permite relacionar varias variables dependientes (Y) con varias variables explicativas (X). Alternativamente, puede verse como un método para tratar las relaciones entre dos grupos de variables. Involucra la búsqueda de combinaciones lineales de las variables X ($U_1, U_2, \text{etc.}$) que tienen las máximas correlaciones posibles con combinaciones lineales de las variables Y ($V_1, V_2, \text{etc.}$).
- ✓ El proceso de cálculo de las variables U y V , denominadas variables canónicas, es un problema de valores propios.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.

ACC - RESUMEN

- ✓ Se presentaron tests de significancia estadística, para determinar si las correlaciones entre las variables U y V son mayores de lo esperado. Sin embargo, existen cuestionamientos sobre la validez de estos tests, particularmente para muestras pequeñas.
- ✓ Las variables U y V se pueden interpretar en términos de los coeficientes que tienen para las variables X e Y, respectivamente, pero esto puede ocasionar problemas. Por lo tanto, las correlaciones entre las variables U y X y las correlaciones entre las variables V e Y, se utilizan a menudo en lugar del proceso de interpretación.

AMARN 2018 - IMFIA.FI.UDELAR -
Ing. Luis Silveira, Ph.D.