

Revisión Teórica

Indicar el valor de verdad de las siguientes premisas

- (a) Los outliers sólo pueden ser detectados a través de métodos univariados o bivariados.
- (b) En el ACP siempre la suma de los autovalores de la matriz de varianzas y covarianzas es igual a la suma de las varianzas
- (c) La prueba de independencia testea la existencia de asociación entre variables categóricas. Rechaza la hipótesis de independencia cuando el estadístico toma un valor “grande” (p-valor menor que 0.05)
- (d) El ACP se basa en el análisis de la matriz de varianzas y covarianzas.
- (e) Uno de los criterios para determinar con cuántas componentes quedarse en ACP consiste en tomar un número proporcional a la cantidad de variables en el problema
- (f) Hacer ACP con la matriz de correlaciones es lo mismo que hacerlo con la matriz de covarianzas utilizando variables estandarizadas
- (g) El vector de medias de una matriz de datos de $n \times p$ tiene n componentes
- (h) El AFC requiere que las variables sean independientes para que tenga sentido aplicarse
- (i) En AFC, la suma de los autovalores es igual al valor de Chi cuadrado
- (j) El análisis de correspondencias se aplica solamente cuando los perfiles fila y columna son idénticos o paralelos.

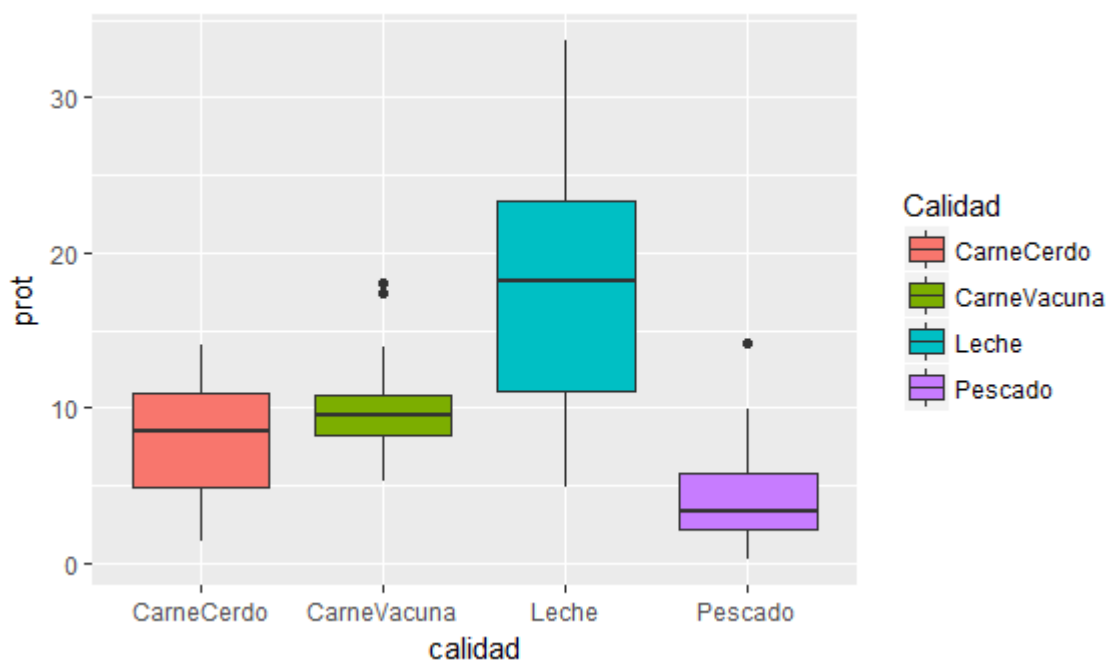
1 Responder las siguientes preguntas

- (a) Cuándo se utiliza la distancia Chi cuadrado? Qué propiedad particular la hace útil en ese contexto?
- (b) Establezca dos similitudes y dos diferencias entre los test de homogeneidad y los de Independencia?
- (c) Con qué objetivo se calculan los residuos de las celdas de un test de independencia?
- (d) Qué cuantifica la inercia en un análisis de correspondencias?

- (e) **Cómo deben tratarse los registros de una base que se consideren outliers? Si se desea detectarlos en forma multivariada cómo conviene hacerlo?**
- (f) **Para que se realizan transformaciones por fila? Y por columna?**
- (g) **Qué características tienen las matrices de covarianzas? Y las de correlación?**
- (h) **Como deben interpretarse los biplot de ACP?**
- (i) **Como deben interpretarse los biplot simétricos de AC?**

Revisión Práctica

1. La matriz de datos contiene el consumo per cápita de diferentes tipos de proteínas(vacuna, cerdo, frutas y vegetales, pescado, huevo, leche, cereal, embutido) en un conjunto de países. Albania, Austria, Bélgica, Bulgaria, Checoslovaquia, Dinamarca, Alemania E, Finlandia, Francia, Grecia, Hungría, Irlanda, Italia, P.Bajos, Noruega, Polonia, Portugal, Rumania, España, Suecia, Suiza, Inglaterra, Rusia, AlemaniaO.



- (a) Comente acerca de la forma de distribución de los diferentes consumos de proteínas de estos países.

(b) Interprete el vector de medias:

Table 1: Vector de Medias

CarneVacuna	10.054167
CarneCerdo	8.016667
Huevos	3.008333
Leche	17.429167
Pescado	4.4375
Cereal	31.2625
Embutidos	4.3292
FrutosSecos	2.962
FrutasyVegetales	4.175

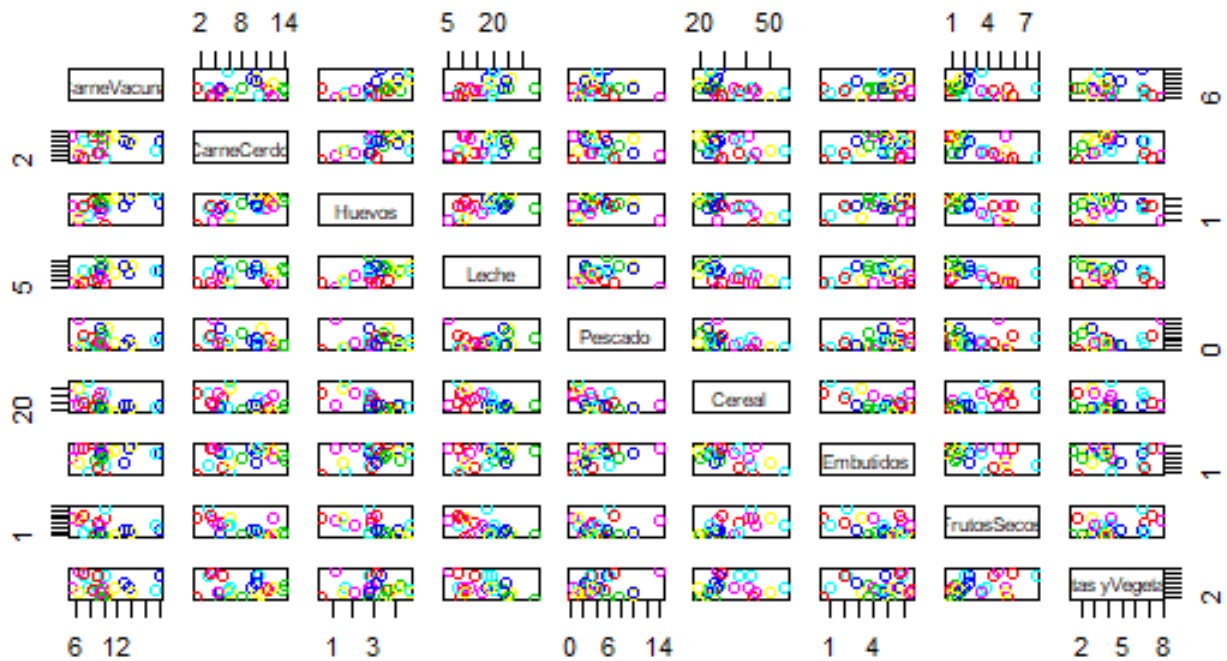
- (c) Puede a partir de esta información indicar si sería conveniente aplicar una técnica de reducción de la dimensión?.
- (d) Calcule la traza de la matriz de covarianzas y la de la matriz de correlación y explique su significado.

Table 2: Matriz de Covarianzas

	CVac	CCerdo	Huevos	Leche	Pesc	Cereal	Embut	Fr.Secos	Fr y veg
CVac	9,53	1,16	1,71	9,76	-0,17	-12,28	0,42	-1,64	-0,64
CCerdo	1,16	12,75	2,25	6,17	-3,27	-13,25	1,66	-4,15	-0,51
Huevos	1,71	2,25	1,07	3,84	-0,03	-6,68	0,7	-1	-0,16
Leche	9,76	6,17	3,84	46,05	2,03	-36,87	2,07	-7,58	-5,32
Pesc	-0,17	-3,27	-0,03	2,03	10,55	-15,16	1,96	-0,55	1,42
Cereal	-12,28	-13,25	-6,68	-36,87	-15,16	92,32	-7,92	11,03	1,81
Embut	0,42	1,66	0,7	2,07	1,96	-7,92	2,5	-1,34	0,19
Fr.Secos	-1,64	-4,15	-1	-7,58	-0,55	11,03	-1,34	3,5	1,39
fr y veg	-0,64	-0,51	-0,16	-5,32	1,42	1,81	0,19	1,39	3,09

(e) calcule el elemento (1,3) de la matriz de correlación.

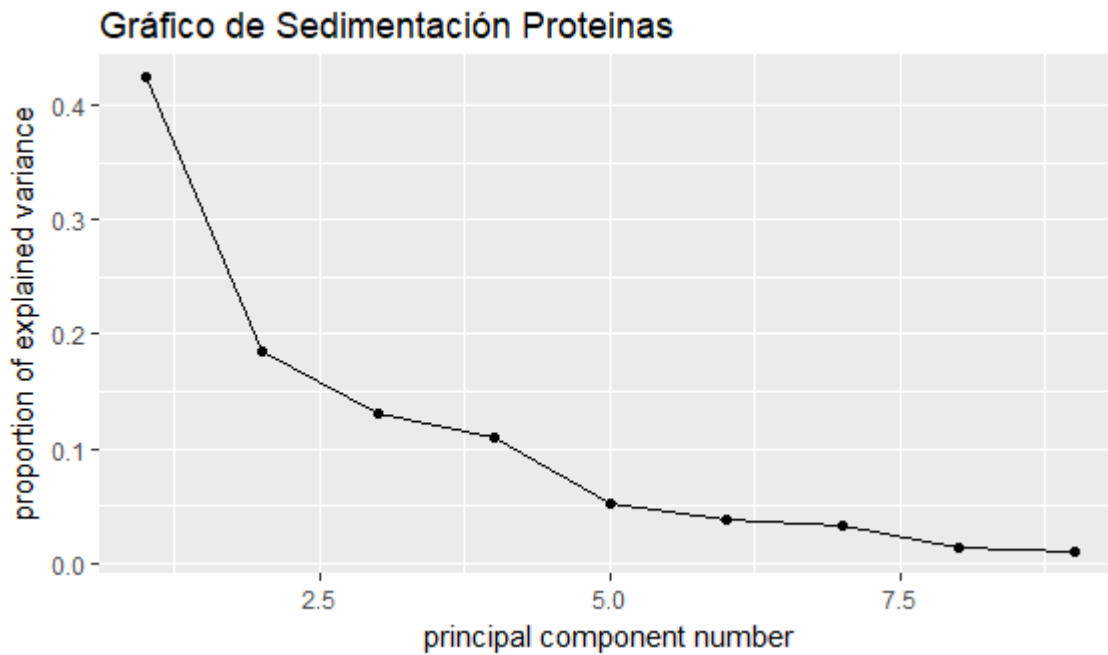
(f) Qué permite apreciar el siguiente gráfico?



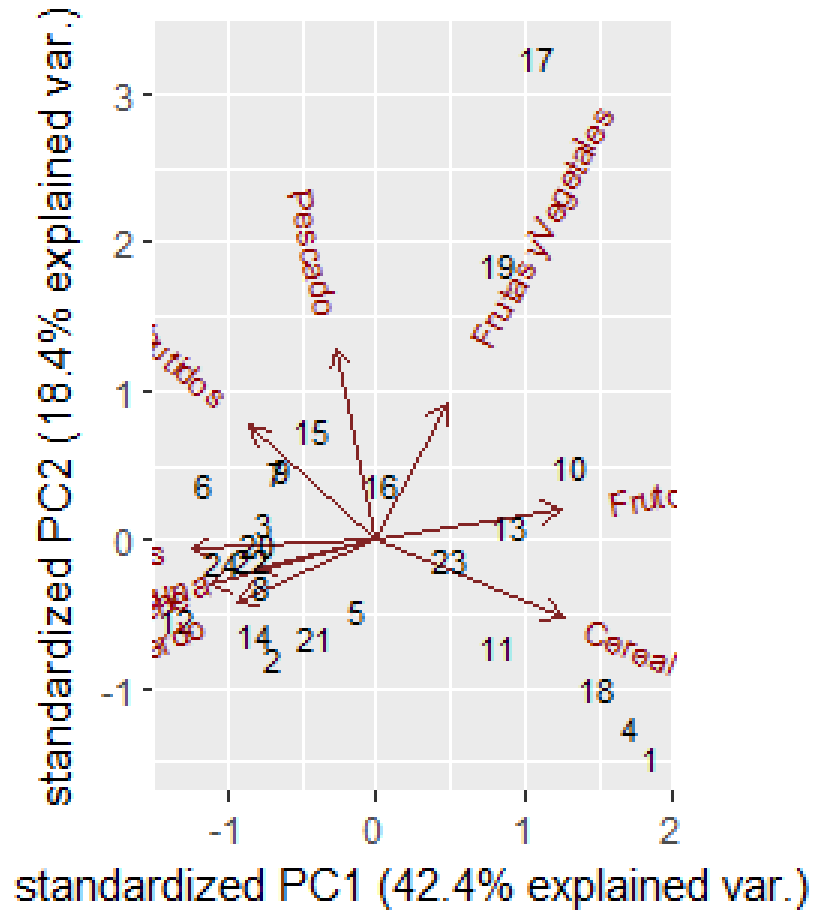
(g) A partir de las siguientes salidas, responda:

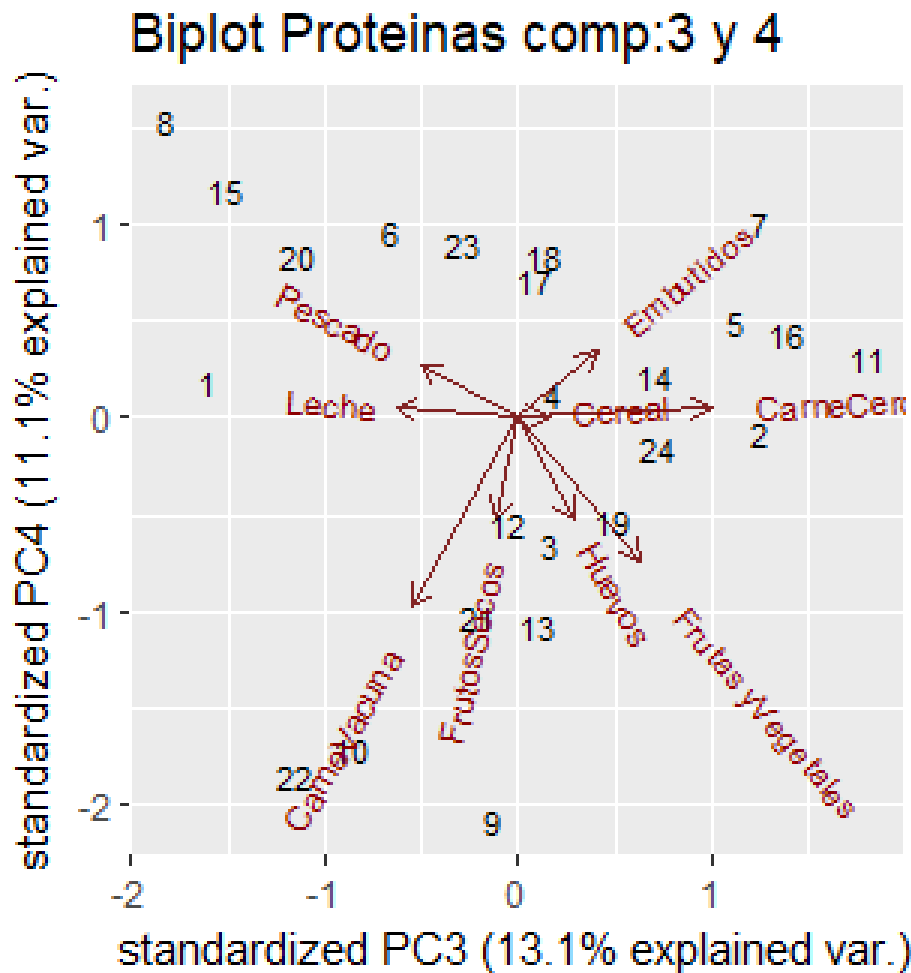
- Con cuántas componentes se quedaría? En qué criterio basa su decisión?
- Qué explica la primer componente principal?
- Qué países tiene consumos de proteínas similares?
- Qué consumos de proteínas se asocian positivamente? Y negativamente?
- Qué países consumen más pescado? Cuáles menos? Encuentra para esto alguna justificación.

	PC1	PC2	PC3	PC4	PC5	PC6
CarneVacuna	-0.2791	-0.1006	-0.3357	-0.6533	0.3127	-0.4648
CarneCerdo	-0.3179	-0.2167	0.6186	0.0325	-0.2849	-0.1588
Huevos	-0.4276	-0.0336	0.1793	-0.3474	0.0157	0.4100
Leche	-0.3847	-0.1593	-0.3871	0.0317	-0.2364	0.5350
Pescado	-0.0919	0.6686	-0.3082	0.1774	-0.2648	-0.1575
Cereal	0.4336	-0.2703	0.0975	0.0081	0.2626	0.0539
Embutidos	-0.2915	0.3983	0.2500	0.2282	0.7433	0.1800
FrutosSecos	0.4276	0.1069	-0.0687	-0.3501	0.0764	0.4935
Frutasy Veg	0.1606	0.4757	0.3885	-0.4943	-0.2598	0.0280



Biplot Proteinas comp:1 y 2





(h) Se encuestó a 200 mujeres acerca de sus preferencias a la hora de comprar ropa. Se registraron las respuestas a las siguientes variables:

- Edad: se categorizó en adolescentes < jóvenes < adultas < mayores.
- Ahorro: se categorizaron las respuestas en ahorrativas y derrochonas.
- Pago: con dos niveles efectivo y tarjeta.
- Compañía: considerando dos categorías solas y acompañadas.

La matriz de Burt correspondiente es:

Table 3: Matriz de Burt

	adol	adult	jov	may	efect	tarj	ahorr	derroch	solas	amigas	Total
adolesc	33	0	0	0	5	28	2	31	29	4	132
adultas	0	27	0	0	23	4	15	12	12	15	108
jovenes	0	0	82	0	39	43	22	60	64	18	328
mayores	0	0	0	58	44	14	41	17	26	32	232
efect	5	23	39	44	111	0	62	49	60	51	444
tarj	28	4	43	14	0	89	18	71	71	18	356
ahorrat	2	15	22	41	62	18	80	0	29	51	320
derroch	31	12	60	17	49	71	0	120	102	18	480
solas	29	12	64	26	60	71	29	102	131	0	524
amigas	4	15	18	32	51	18	51	18	0	69	276
Total	132	108	328	232	444	356	320	480	524	276	3200

- Explique por qué la matriz de Burt es simétrica.
- Señale el significado de los bloques señalados en color.
- Explique los valores resaltados.
- Que celdas contribuyen importantemente al estadístico de Chi cuadrado?
- Considera que la edad esta asociada de alguna forma con el medio de pago preferido? Y con la compañía elegida para salir de compras? En este contexto, tiene sentido el análisis de correspondencias?
- Estas relaciones se ven reflejadas de alguna forma en el biplot simétrico?
- Cuáles son los scores de adultas? Y de con amigas?

Table 4: Autovalores

Eigenvalues									
	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8	Dim.9
Variance	0.479	0.405	0.311	0.25	0.25	0.25	0.25	0.25	0.25
% of variance	13.688	11.574	8.889	7.143	7.143	7.143	7.143	7.143	7.143
Cumulative % variance	13.688	25.262	34.151	41.294	48.437	55.58	62.723	69.865	77.008

Table 5: Coordenadas

	Dim.1	Dim.2
pago	0.217	0.629
ahorro	0.385	0.029
compania	0.472	0.15
frec	0.843	0.813

MCA factor map

