



**Universidad Nacional Autónoma de México
Facultad de Ciencias**

Examen 2 A

Realizado por

Cícero Álvarez Alicia Guadalupe 318010807

Hernández Alva Luis Ángel 315251674

Isunza Alvarez Marcos Guillermo 419002921

Regalado Urbina Brandon Imanol 317312878

Profesores

Gonzalo Pérez de la Cruz

Noe Eusebio Amador González

César Humberto Valle Márquez

Asignatura

Seminario de Estadística I:

Aprendizaje Estadístico Automatizado

29 de noviembre de 2023

Estimación del parámetro tau usando el método Monte Carlo y bootstrap no paramétrico.

a. Método Monte Carlo

Para estimar el parámetro que nos solicitan se generan diez mil muestras, cada muestra tiene 25 observaciones, de la variable aleatoria $\hat{\tau} \sim \text{Poi}(\theta = 0)$.

De este modo, al estimar $E(\hat{\tau})$ y $V(\hat{\tau})$ se obtienen los siguientes resultados:

$$\begin{aligned}\mathbb{E}[\hat{\tau}] &\approx \frac{\sum_{i=1}^{10000} \hat{\tau}_i}{10000} & \& \\ &= 0.2230542, & & \\ \mathbb{V}[\hat{\tau}] &= \mathbb{E}[\hat{\tau}^2] - \mathbb{E}[\hat{\tau}]^2 & & \\ &\approx 0.0030194.\end{aligned}$$

El histograma de las muestras generadas se pueden ver en la siguiente figura:

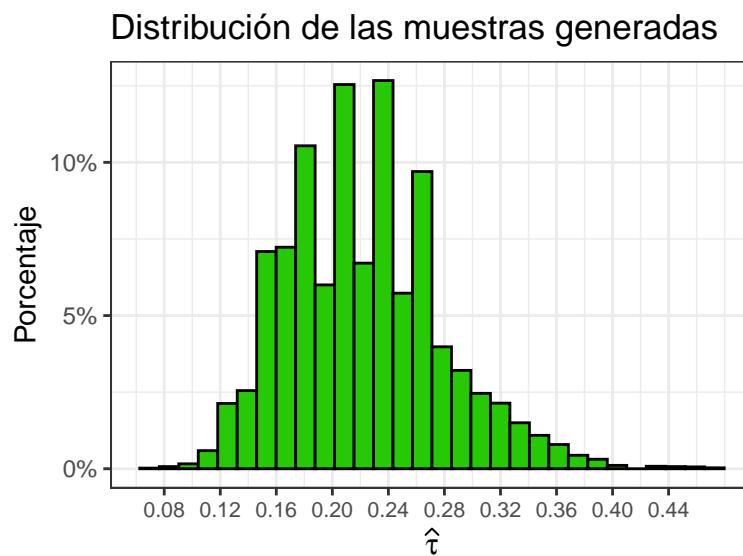


Figura 1: Histograma para las muestras generadas por Monte Carlo

Para los observar los detalles del código utilizado para la estimación del parámetro se pueden consultar en el archivo RMarkdown en los chunks *EstamacionDetau* y *PrimerHistograma* en las líneas 47 y 83 respectivamente.

b. Bootstrap no paramétrico

En contraste, para aplicar *bootstrap* sólo se toma una muestra de 25 números aleatorios de una distribución Poisson(θ), con $\theta = 1.5$.

De aquí se tiene que

$$\mathbb{E}[\hat{\tau}] \approx 0.2396035 \tag{1}$$

$$\mathbb{V}[\hat{\tau}] \approx 0.00317630 \tag{2}$$

Los detalles se encuentran en el chunk *Bootstrap* la línea 107 del archivo RMarkdwon.

A continuación se presenta, en un histograma, la visualización de las muestras obtenidas usando el método Bootstrap no paramétrico:

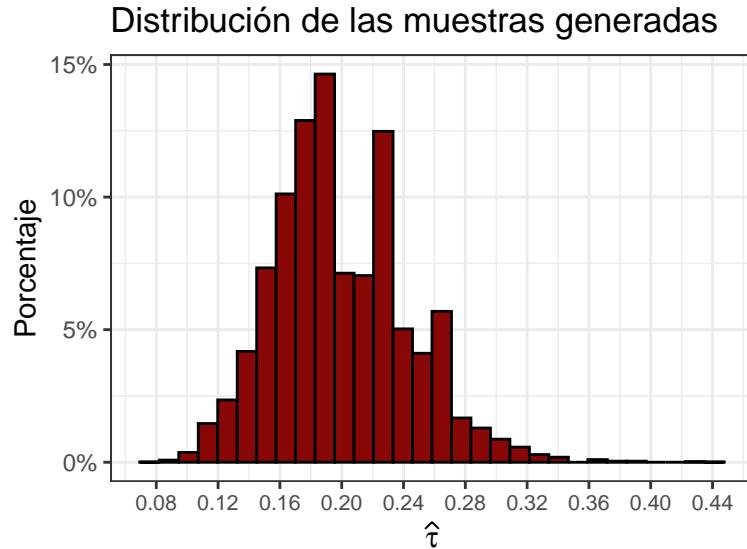


Figura 2: Histograma para las muestras generadas por *bootstrapping*

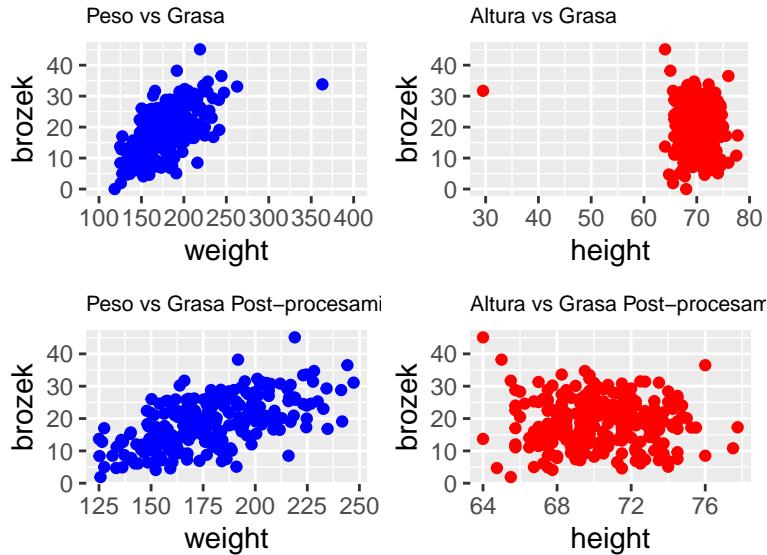
En este caso, los métodos difirieron en aproximadamente 0.016 para la esperanza del estimador y 0.00015 para su varianza, además hay una similaridad visual hallada. Como se esperaba, ambos se asemejan a una distribución Poisson con parámetros ligeramente modificados.

La cercanía que se alcanzó es sorprendente, especialmente al considerar que el método bootstrap solo requiere datos en su forma más básica para ofrecer una estimación aceptable del parámetro y su distribución. A pesar de ello, la escasa cantidad de datos de entrada (25) contribuye a que los resultados varíen para cada aplicación. Por lo tanto, a medida que se reduce el número de observaciones, se vuelve crucial la calidad de la muestra proporcionada. Por otro lado, en repeticiones consecutivas, Monte Carlo demostró ser más coherente en sus resultados.

Selección de variables clínicas para modelar el promedio del porcentaje de grasa corporal en hombres

Es de interés responder a la siguiente pregunta: ¿Qué variables clínicas modelan de manera óptima el promedio del porcentaje de grasa corporal en los hombres?

Como parte de un análisis exploratorio de datos preliminar se presentan las siguientes gráficas:



Dado que en el conjunto de datos original se identificaron valores inusuales para las variables de *weight* y *height*, los cuales se alejaban del resto de valores, se realizó primeramente un preprocesamiento de los datos. Esto consistió en la eliminación de aquellas observaciones que mostraban un peso superior a 250 lbs, una altura inferior a 60 pulgadas o un valor de cero para la variable “brozek”. Las primeras dos gráficas representan la relación entre las variables indicadas antes de eliminar las observaciones atípicas y las segundas gráficas representan la relación entre las variables después de haber eliminado las obsevaciones atípicas. [1]

En un incio, al comienzo del análisis, se considera el modelo de regresión lineal simple para datos continuos con distribución Gaussiana y liga identidad.

$$E(\hat{brozek}) = \beta_0 + \beta_1 * x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p$$

En el modelo anterior solo se consideran los componentes principales. Aquí $E(\hat{brozek})$ representa el promedio del porcentaje de grasa corporal en hombres y x_p representan a las variables que ayudan a modelar de mejor manera el promedio del porcentaje de grasa corporal. De acuerdo a lo solicitado por los investigadores, se procedió a ajustar distintos modelos lineales generalizados del promedio de porcentaje de grasa corporal en hombres después de obtener las variables pertinentes mediante distintos métodos de selección. Los resultados se resumen en las siguientes dos tablas:

No. de modelo	Método de selección de variables	BIC
1	Mejor subconjunto	1412.142
2	Forward	1446.202
3	Backward	1455.516
4	Lasso simple	1413.107
5	Forward con interacciones	1416.311
6	Backward con interacciones	1405.596
7	Lasso con interacciones	1411.985
8	Backward, distribución Gamma, liga logarítmica	1490.06
9	Mejor subconjunto, variables al cuadrado	1423.089

Cuadro 1: Resultados de los métodos de selección

No. de modelo	Forma matemática del modelo
1	$E(\hat{brozek}) = \beta_0 + \beta_1 \cdot \text{height} + \beta_2 \cdot \text{abdom} + \beta_3 \cdot \text{wrist}$
2	$E(\hat{brozek}) = \beta_0 + \beta_1 \cdot \text{weight} + \beta_2 \cdot \text{abdom} + \beta_3 \cdot \text{wrist}$
3	$E(\hat{brozek}) = \beta_0 + \beta_1 \cdot \text{weight} + \beta_2 \cdot \text{abdom} + \beta_3 \cdot \text{wrist}$
4	$E(\hat{brozek}) = \beta_0 + \beta_1 \cdot \text{age} + \beta_2 \cdot \text{height} + \beta_3 \cdot \text{abdom} + \beta_4 \cdot \text{wrist}$
5	$E(\hat{brozek}) = \beta_0 + \beta_1 \cdot \text{hip} + \beta_2 \cdot \text{height:hip} + \beta_3 \cdot \text{neck:abdom} + \beta_4 \cdot \text{neck:hip}$
6	$E(\hat{brozek}) = \beta_0 + \beta_1 \cdot \text{abdom} + \beta_2 \cdot \text{height:wrist} + \beta_3 \cdot \text{chest:hip}$
7	$E(\hat{brozek}) = \beta_0 + \beta_1 \cdot \text{abdom} + \beta_2 \cdot \text{age:abdom} + \beta_3 \cdot \text{age:thigh} + \beta_4 \cdot \text{height:wrist}$
8	$E(\hat{brozek}) = \beta_0 + \beta_1 \cdot \text{hip} + \beta_2 \cdot \text{height:hip} + \beta_3 \cdot \text{neck:abdom} + \beta_4 \cdot \text{neck:hip}$
9	$E(\hat{brozek}) = \beta_0 + \beta_1 \cdot \text{height} + \beta_2 \cdot \text{height}^2 + \beta_3 \cdot \text{abdom} + \beta_4 \cdot \text{abdom}^2 + \beta_5 \cdot \text{wrist} + \beta_6 \cdot \text{wrist}^2$

Cuadro 2: Tabla con número de modelo y forma matemática.

En el **Cuadro 1** se muestran los modelos obtenidos mediante diferentes métodos de selección de variables. En el **Cuadro 2** se muestran las variables que fueron seleccionadas para cada modelo. Primero nos enfocaremos en los primeros siete modelos de la tabla, es decir los modelos sin modificaciones y después discutiremos los dos modelos con modificaciones.

De acuerdo a los modelos que se observan en la tabla, la variable que más veces aparece es *abdom*, la cual aparece un total de 6 veces, sin contar aquellas veces en las que aparece interactuando con otra variable, seguida de la variable *wrist*, la cual aparece 4 veces, sin contar interacciones. Además los coeficientes asociados a la variable *abdom* son positivos en los primeros cuatro modelos por lo que podemos afirmar que la variable más significativa asociada a un incremento del valor promedio de la grasa corporal en los hombres es la medida del abdomen [2]. En cambio, en los primeros cuatro modelos de la tabla que incluyen a la variable *wrist* los coeficientes asociados a la variable *wrist* son negativos en los modelos en que esta variable aparece y por lo tanto es posible inferir que a medida que la medición de la muñeca aumenta entonces habrá una disminución del valor promedio de la grasa corporal en hombres [3]. En conclusión las variables clínicas más significativas para modelar el promedio de la grasa corporal en los hombres son la medida del abdomen y de la muñeca. Las variables que no aparecen entre las seleccionadas y que podemos observar en la forma matemática de los modelos que se encuentran en el Cuadro 2 son: adipos, knee, ankle, biceps, forearm. Por lo tanto ninguna de las variables anteriormente mencionadas tiene algún efecto sobre el valor promedio del porcentaje de grasa en hombres.

Además, según el criterio BIC, se observa que al considerar modelos que toman en cuenta las interacciones entre las variables, el BIC de los modelos con interacciones disminuye. Esto indica que, al incorporar interacciones, el modelo del promedio de porcentaje de grasa corporal en hombres se ajusta mejor a los datos.

Se realizaron modificaciones a los modelos hasta entonces obtenidos. Una modificación consistió en cambiar la distribución de la variable respuesta y considerar una función liga logarítmica, la otra modificación consistió en incluir las variables al cuadrado. Se tomaron los mejores modelos de cada modificación de acuerdo al criterio BIC. [4] Para el modelo ajustado cuyas variables se obtuvieron mediante el método Backward y fueron modificados mediante la distribución Gamma con liga logarítmica, no se observa una mejora en comparación con los demás modelos, ya que su puntuación de acuerdo al criterio BIC aumentó. Siguiendo la misma lógica, en el caso del modelo cuyas variables se obtuvieron mediante el método de Mejor subconjunto y se modificó al incluir las variables al cuadrado, solo representa una mejora frente a los modelos con variables seleccionadas mediante Forward y Backward.

Por otra parte, de los modelos observados en el **Cuadro 1** se puede ver que aquel que tiene un menor valor BIC es el modelo cuya selección de variables se realizó con el método stepwise conocido como *Backward* incluyendo interacciones. Los coeficientes de este modelo pueden ser interpretados de la siguiente manera: Un aumento del 100 % de la variable *hip* se asocia a un aumento del 114 % en el promedio de la grasa corporal de los hombres dejando al resto de variables fijas, a su vez las variables asociadas a la altura y a la medida del cuello (*height* y *neck*) interactúan con la variable *hip*, que representa la medida de la cadera, el valor del coeficiente asociado es negativo, en particular, un aumento de una unidad para la variable asociada a la interacción *hip:height* se asocia a una disminución del 0.4 % del promedio de la grasa corporal en hombres y un aumento en una unidad a la variable asociada a la interacción *hip:neck* está asociada a una disminución del 2 % para el promedio de la grasa corporal en hombres, por otro lado la interacción *neck:abdom* está asociada a un aumento del 2 % en el promedio de la grasa corporal para hombres. [5].

En este caso se puede observar que los decrementos causados por las interacciones *hip:height* y *hip:neck* así los aumentos asociados a *neck:abdom* son de una magnitud muy pequeña como para que las variables sean consideradas estadísticamente relevantes. Sin embargo, dentro del modelo elegido, el aumento causado por la variable *hip* supera el 100 % y por tanto se considera a la medida de la cadera como una variable significativa. En consecuencia, esto indica que la medida de la cadera está asociada positivamente a un aumento del promedio de porcentaje de grasa corporal en los hombres.

Referencias a los chunks

- [1] El preprocesamiento de los datos se puede consultar en los chunks de código *Preprocesamiento y RemocionCasosExtraños* que se encuentran en las líneas 56 y 98 del archivo RMarkdown.
- [2]. Los valores estimados del coeficiente asociado a la variable *abdom* para los 6 modelos en los que aparece la variable se pueden consultar en los summarys de los modelos ajustados dentro de los chunks de código: *MejorSubconjunto*, *Forward*, *Backward*, *Lasso*, *backwardInteracciones* y *LassoConInteracciones* en las líneas 188, 221, 250, 325, 354 y 384 respectivamente.
- [3]. Los valores estimados del coeficiente asociado a la variable *wrist* para los 4 modelos en los que aparece la variable *wrist* por si sola se pueden consultar en los summarys de los modelos ajustados dentro de los chunks de código: *MejorSubconjunto*, *Forward*, *Backward*, *Lasso* en las líneas 188, 221, 250, y 325 respectivamente.
- [4]. Los chuncks de las modificaciones son *Modificacionfuncionliga* y *ModificacionCuadratica* y se pueden consultar en las líneas 438 y 513 del archivo RMarkdown..
- [5]. Los valores estimados de los coeficientes asociados a las variables *hip*, *hip:height*, *hip:neck* y *neck:abdom* se pueden consultar en el summary del modelo ajustado dentro del chunk de código *backwardInteracciones* en la línea de código 354 del archivo RMarkdown..

Reducción de dimensionalidad para test de personalidad.

Se realizó un preprocesamiento donde se filtraron las variables de nuestro interés y se exploraron dimensiones interesantes con los datos de manera continua y ordinal, se trabajó datos con y sin escalar. Adicional se renombraron las variables para fines prácticos y de interpretación.

Con ayuda de la librería factoextra se obtuvieron los *Componentes Principales*, procedemos a usar la función fviz_eig como un índice para el número a considerar de estas y en @ref(fig:Grafica13) se muestran los resultados, para los datos con o sin escalar se sugiere entre 4 o 5 componentes pues después de estos ya no hay mucho cambio en la varianza que aportan. Además se acumula en los tres casos un aproximado de 60 % de varianza total cuando consideramos 4 componentes. Más detalles en Chunk factoCP.

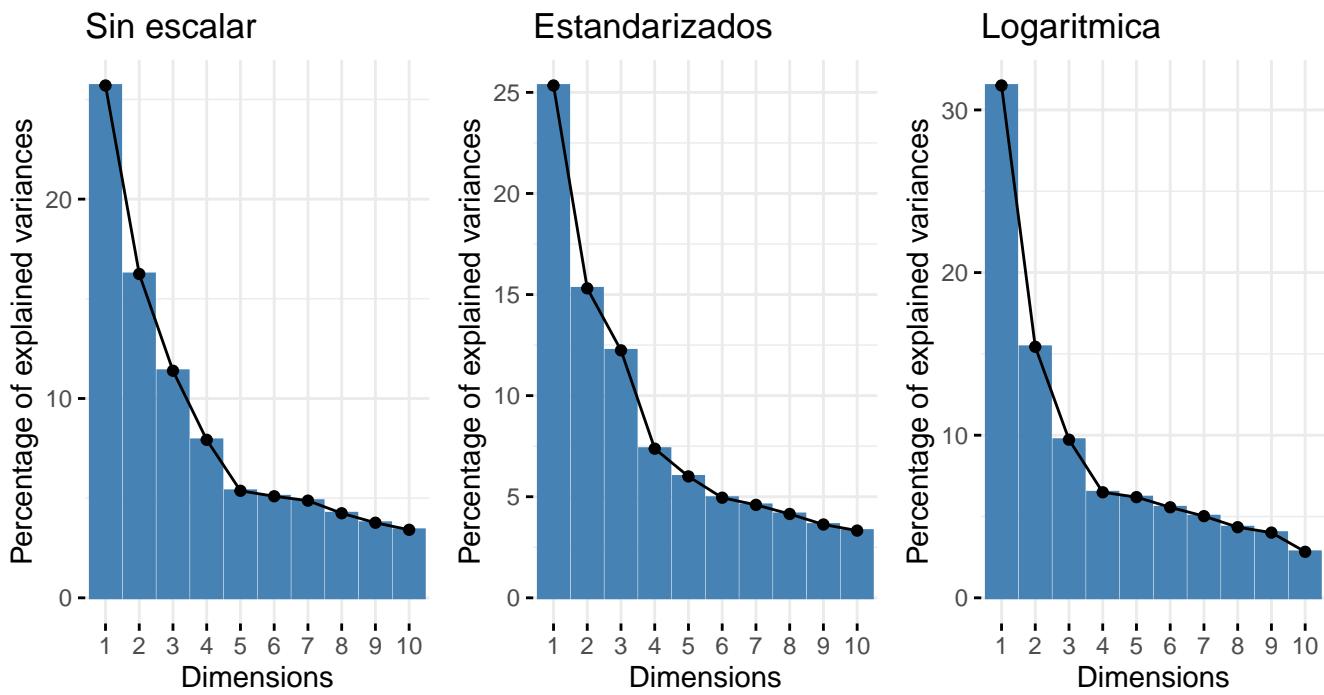


Figura 1: Índices para número de componentes principales

Tras revisar las correlaciones con las variables originales podemos darnos una idea mejor de las dimensiones que encontramos, son muy similares salvo algunas variables pero en esencia podemos considerar 4 tipos de personalidades que se derivan de lo siguiente:

1.– Para los datos sin escalar, las variables Victimista, Peleonero, Frio, Rudo y Entusiasta son las que más pesan en el componente 1, las variables Parlanchín, Reservado, Asertivo y Tímido son las de mayor peso para el componente 2, mientras que para el componente 3 las de mayor peso son Exhaustivo, Perseverante y Eficiente, para el componente 4 las de mayor peso son Descuidado e Indulgente.

2.– Para los datos estandarizados, las variables Victimista, Exhaustivo, Peleonero, Confiable, Entusiasta, Frío, Eficiente y Rudo son las de mayor peso en el componente 1, las variables Parlanchín, Reservado, Asertivo y Tímido son las de mayor peso, para el componente 3 son Exhaustivo, Frio, Perseverante y Eficiente, por el último el componente 4 tiene de nuevo a Descuidado e Indulgente como las de mayor peso.

3.– Para los datos en escala logarítmica, Victimista, Peleonero, Rudo y Frío son las de mayor peso en el componente 1, Parlanchín, Reservado, Tímido y Asertivo en el componente 2, Descuidado, Perseverante y Frío son las de mayor peso en el componente 3 y por último, el componente 4 tiene correlaciones muy bajas (menores a 0.5) y las de mayor peso son Asertivo, Perseverante y Entusiasta.

En el Chunk Cor se encuentran las correlaciones de todas las variables con cada componente, anteriormente únicamente se tomó en cuenta correlaciones mayores a 0.5 en valor absoluto, para mayor interpretabilidad tenemos la Gráfica @ref(fig:Grafica23), en esta sólo se presentan los datos originales y los de escala logarítmica, son las proyecciones de las variables de mayor peso en los primeros 2 componentes pues estos rescatan la mayor varianza, podemos guiarnos por el sentido y magnitud de las fechas para intuir la influencia de cada variables en cada componente, por ejemplo entre más Peleonero te consideres, mayor valor tendrá el primero componente. Algo a notar es que las direcciones en ambas graficas son muy similares pero el primer componente en escala logarítmica rescata más varianza.

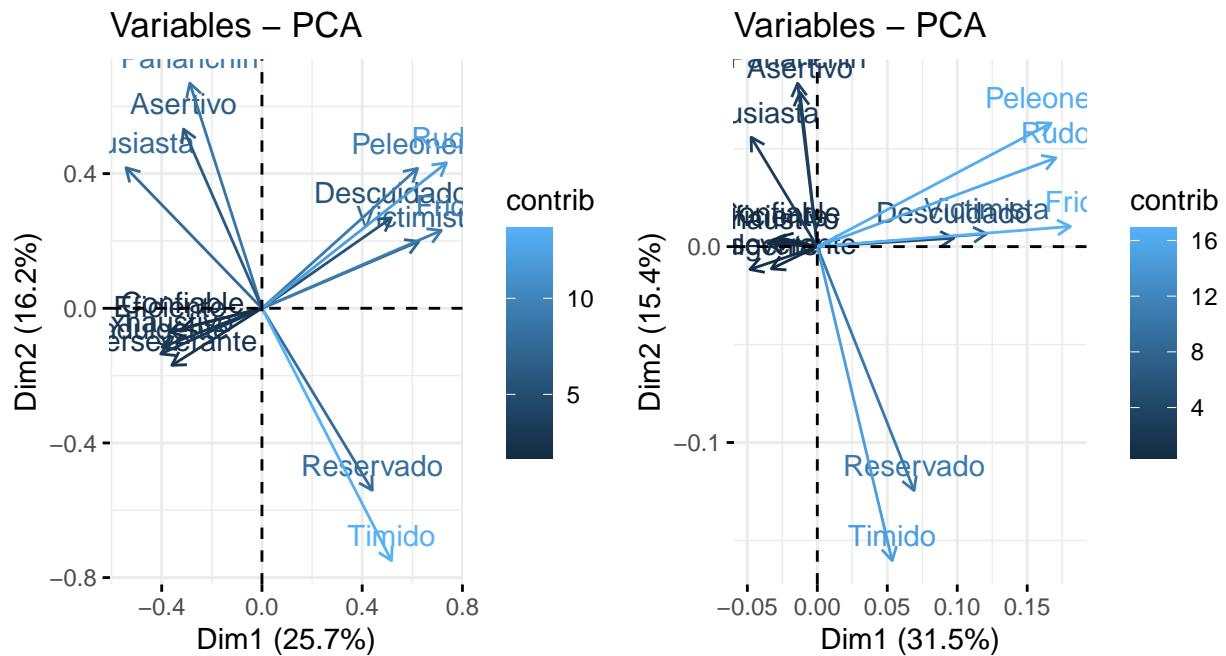
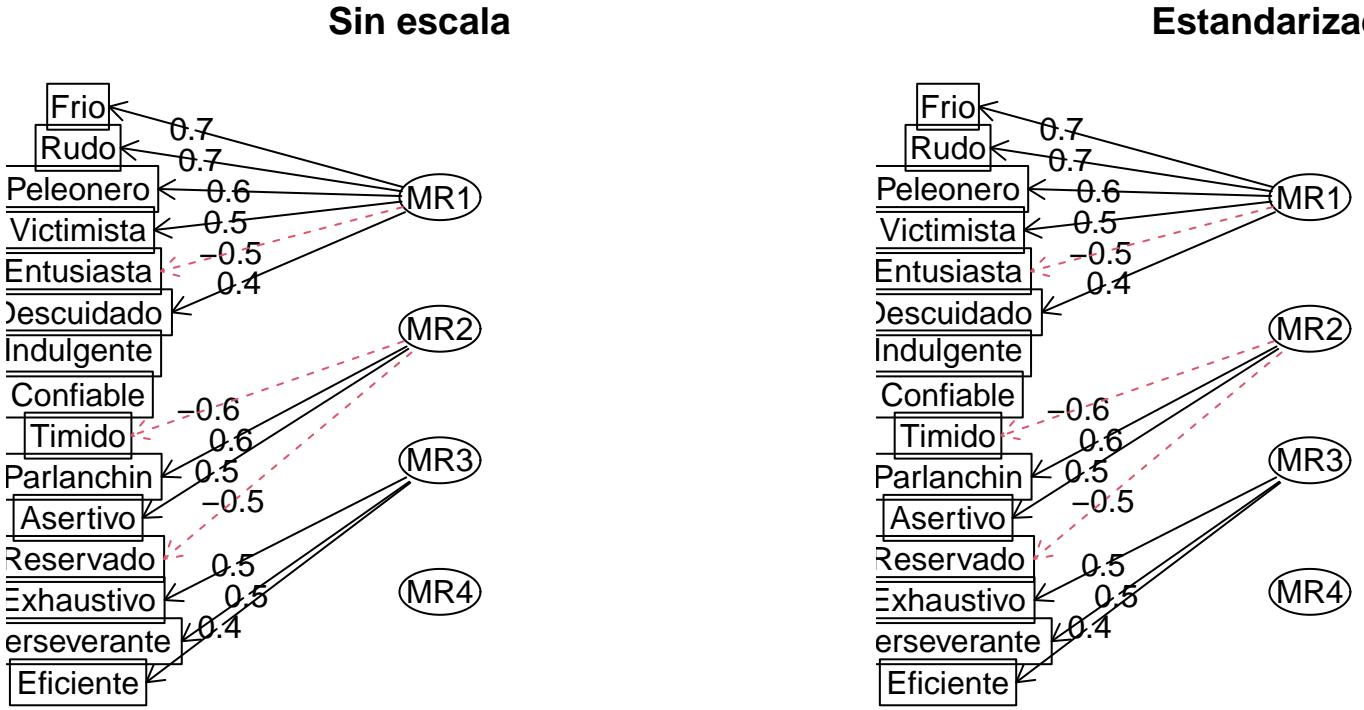


Figura 2: Proyección en componentes

Para continuar con el análisis consideramos el enfoque de *Análisis Factorial Exploratorio*, para ello nos apoyamos de la librería psych y la función fa. De nuevo consideramos datos sin escalar, estandarizados y con escala logarítmica, optamos por seguir la recomendación de parallel y considerar 4 factores (Chunk AFE).



De las gráficas anteriores podemos notar, 3 componentes parecen ser suficiente para resumir la información, a diferencia de componentes principales hemos reducido un poco más la dimensionalidad, además los resultados son muy similares a CP pues las variables de mayor peso se repiten casi todas. Aquí tenemos que el MR1 está asociado con una personalidad “brusca” donde ser frío o rudo la aumentan pero considerarse entusiasta la disminuye, el MR2 lo relacionamos con habilidades sociales, entre más callado menor es este valor y por último el MR3 lo asociamos con el desempeño y dedicación que le ponemos a las cosas, es una relación monótona creciente.

Para decidirnos por un modelo se probaron varias rotaciones como varimax y simplimax, también se consideraron a las variables como ordinales y de nuevo con ayuda de fa se obtuvieron las variables latentes mientras que con principal las componentes principales, más detalles en Chunk RotacionesCP, RotacionesAFE y Ordinales. Optamos por un modelo de Componente principales pues estos recuperan más varianza y dentro de estos el que usa la rotación “cluster” y maneja las variables como ordinales es el mejor rankeado pues recupera un 66 % de varianza total, además nos restringimos a considerar sólo 3 componentes pues el cuarto sólo está relacionado con una variable.

Components Analysis

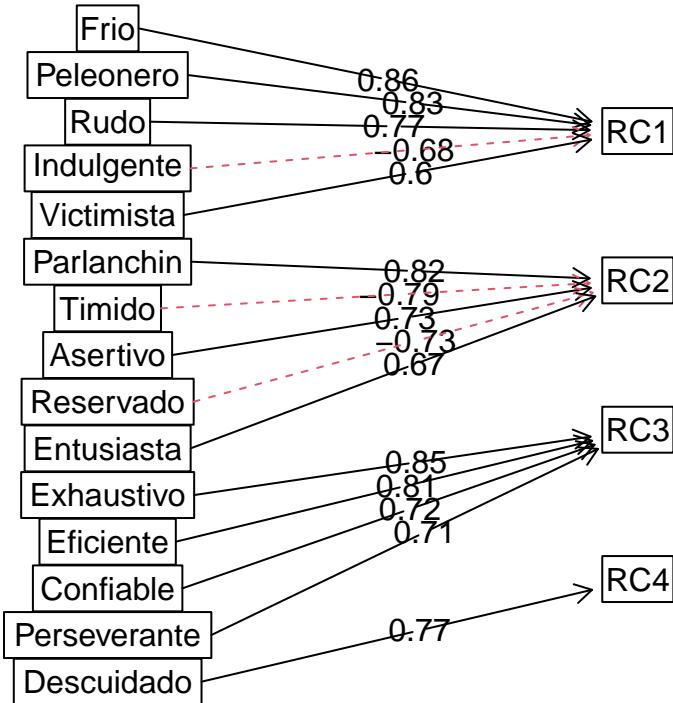


Figura 3: Componentes principales modelo seleccionado

Ya con nuestro modelo seleccionado pasamos a la interpretación, según las Gráfica @ref(fig:Grafica43):

Entre más Victimista, Frío y Rudo te consideras mayor es el valor en el componente 1 mientras que considerarse Indulgente lo reduce y esto tiene sentido pues las características marcan una personalidad hosca. Frío y Peleonero son las que más pesan.

Cuando observamos el componente 2 este es más fácil de interpretar y parece referirse a personas sociables pues pesan variables como Asertivo, Parlanchín, Tímido y Reservado.

Sobre el componente 3 podemos notar todas las variables que pesan tienen correlación positiva, a mayor valor de cualquiera mayor será el valor del componente, ya que contempla variables como Exhaustivo y Perseverante podemos verlo como una persona optimista y/o dedicada.

4. Identificación de grupos de clientes para focalizar la publicidad de la empresa Oddjob Airways con conglomerados

Identificación de grupos usando el método Kmeans

Asumimos que nuestras variables siguen un comportamiento continuo y obtuvimos algunos grupos convenientes para una mejor interpretación de los datos mediante el método k-means:

```
##  
##   1   2   3  
## 381 118 448
```

Al leer las preguntas que se le hicieron a los clientes potenciales, pudimos notar 4 aspectos del servicio de la compañía: puntualidad, seguridad, mantenimiento e instalaciones, y calidad de servicios ofrecidos. Nosotros creemos que los temas de mantenimiento e instalaciones conllevan a la seguridad del cliente, por lo que dedujimos que la publicidad se puede focalizar en 3 grupos de clientes: los que esperan puntualidad y seguridad, los que esperan seguridad y los que esperan calidad de servicio, puntualidad y seguridad.

Por lo que, procedimos a diseñar modelos, quedándonos con uno para clasificar a los clientes en tres categorías principales en función de sus expectativas.

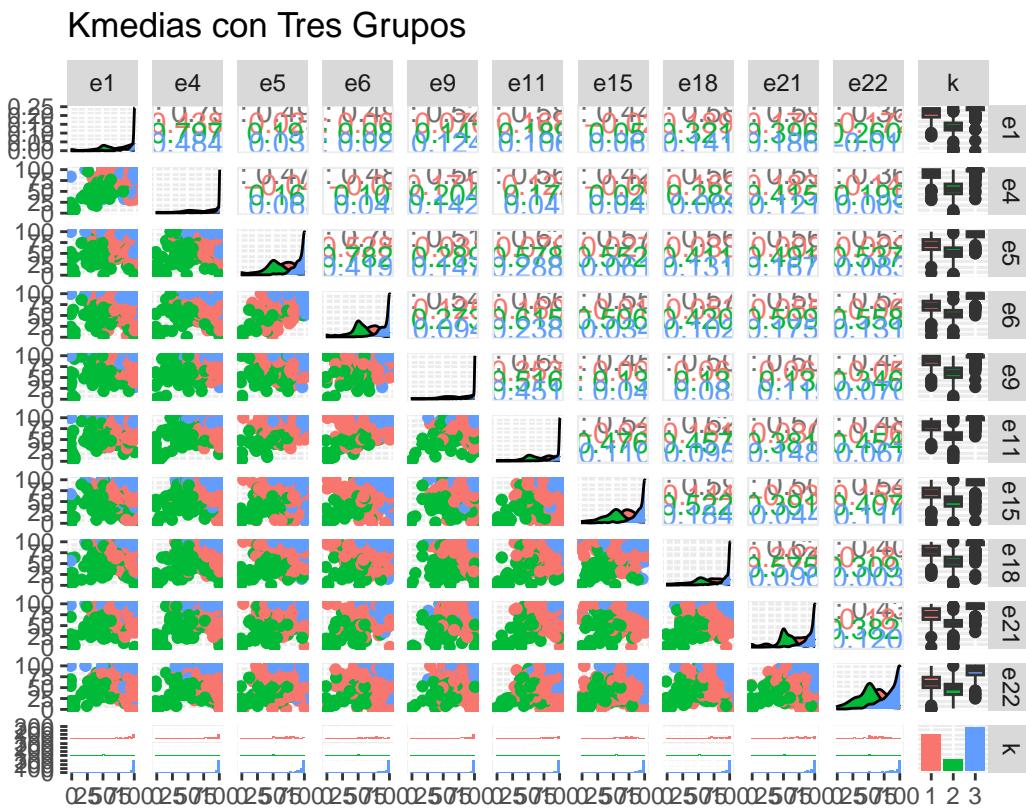


Figura 1: Kmedias con 3 Categorías: Personas a las que les importa la puntualidad y la seguridad, Personas a las que les importa viajar seguras solamente, y Personas a las que les importa la calidad de los servicios, la puntualidad y la seguridad

En la Figura 1 podemos ver cómo los mayores promedios de expectativas se los lleva el grupo 3, siguiéndole el grupo 1 y finalmente el grupo 2.

En la Figura 2 tenemos a la primera y segunda componente principal, las cuales tienen una proporción de la varianza de 59.1 % y 9.7 % respectivamente, lo cual nos dice que la proporción de la varianza de las demás componentes está por debajo de este último porcentaje. Esta primera componente actúa como un resumen de todas las demás. Es decir, a mayor puntuación en este componente, entonces mayor promedio en general en todas las respuestas del cuestionario.

Variables – PCA

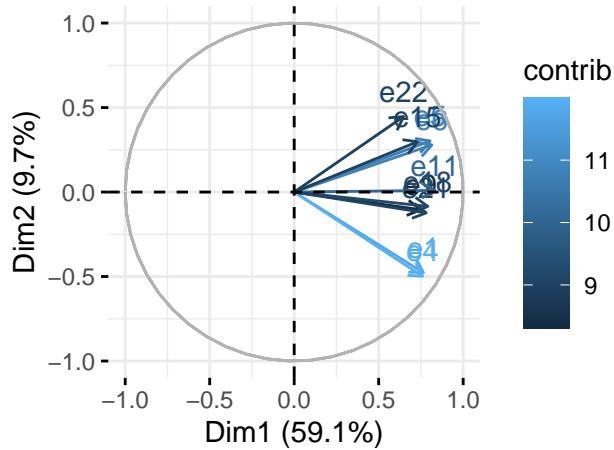


Figura 2: Variables-PCA

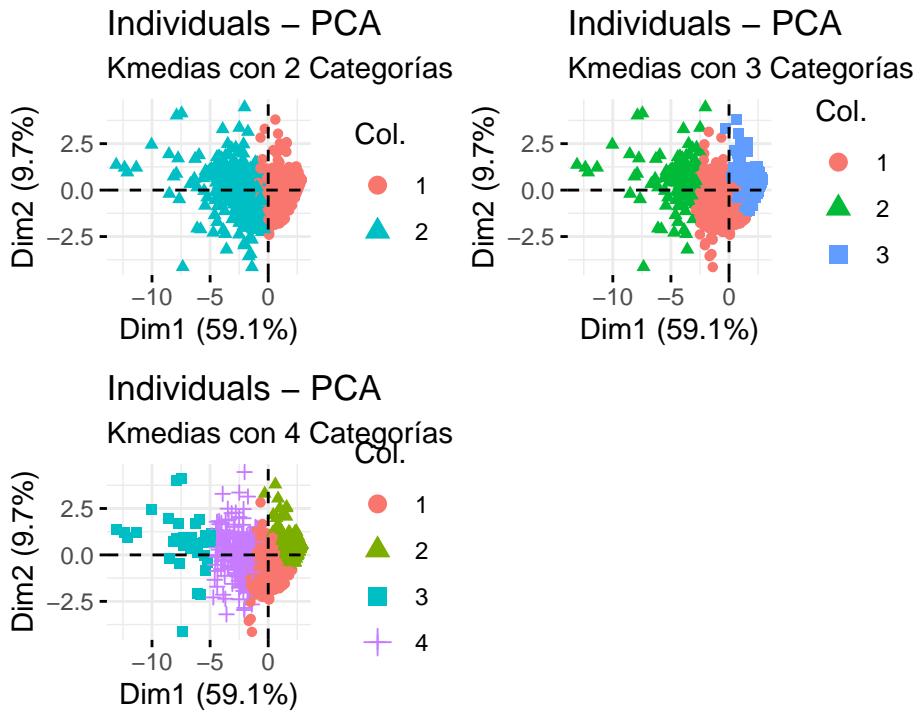


Figura 3: Distribución de datos sin estandarizar

Observemos en la Figura 3 que, en la primer gráfica podemos dividir a los clientes entre los que tienen altas expectativas en todos los aspectos y los que no, lo cual no nos ayuda a focalizar la publicidad. La tercera gráfica nos ayuda más, pero en el chunk “silhouette”, nos guíamos del método con el mismo nombre para afirmar que tener 3 grupos es mejor que tener 4, lo cual también se puede ver en la segunda gráfica. Debido a esto, y observando los componentes principales, podemos decir que es mejor, y hace la diferencia, focalizar la publicidad en 3 grupos de clientes: los que esperan puntualidad y seguridad, los que esperan seguridad y los que esperan calidad de servicio, puntualidad y seguridad, ordenados respectivamente como se ve en la segunda gráfica de la Figura 3.

Identificación de grupos usando el método Jerárquico Aglomerativo

De la misma manera, asumiendo que nuestras variables siguen un comportamiento continuo, hicimos la prueba con el método aglomerativo jerárquico. Para éste también probamos tanto con los datos sin alteraciones como con los datos estandarizados. Hicimos la prueba con todas las combinaciones de disimilitudes entre: euclidea, máxima, canberra, manhattan y minkowski, para los clientes. Y las disimilitudes probadas para los clusters fueron: ward D, ward D2, simple y completa.

No tuvimos éxito con la disimilaridad simple y completa. Para todas las demás combinaciones de disimilaridades en general no se tuvo un mal desempeño de los modelos.

Pudimos diseñar un modelo que nos pudiera separar los datos en tres grupos de una manera satisfactoria. Usando clusters jerárquicos aglomerativos, el modelo que consideramos como el mejor para dos clusters es usando la disimilaridad de Manhattan entre clientes y ward D2 para clusters. El modelo fue muy parecido al de kmedias y se puede visualizar en las Figuras 5 y 4, siendo la segunda gráfica de esta última.

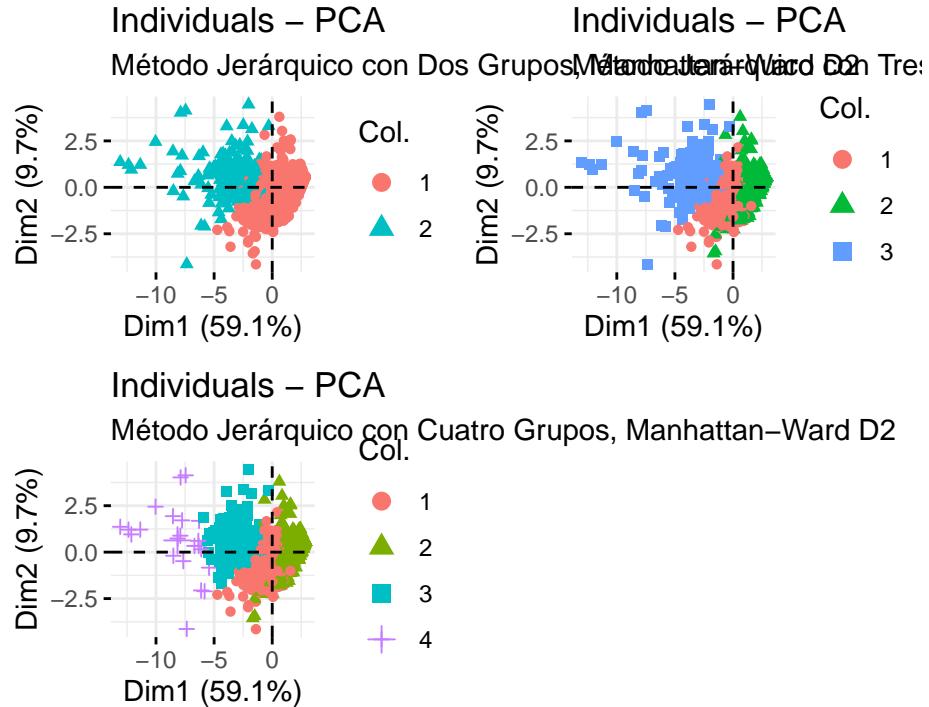


Figura 4: Distribución de los datos: Método Jerárquico usando Manhattan–Ward D2

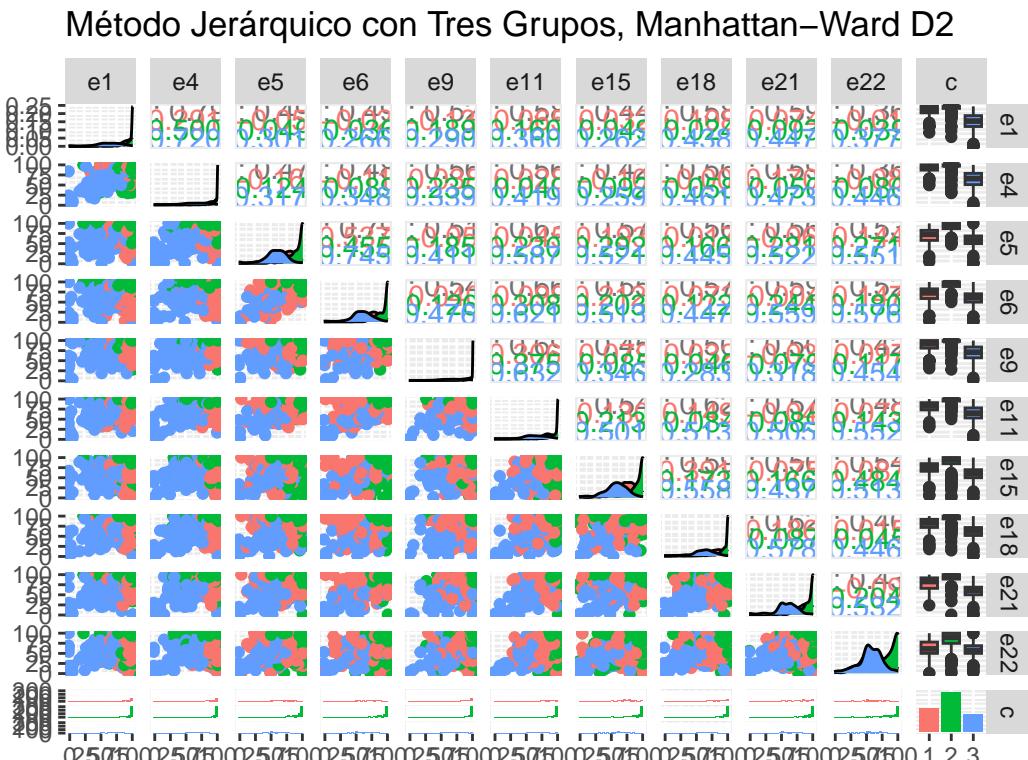


Figura 5: Método Jerárquico con Tres Grupos, Manhattan–Ward D2

En la Figura 4 podemos ver nuevamente la comparación de la distribución de 2, 3 y 4 grupos. Notamos que la gráfica que tiene 3 grupos es muy similar a la obtenida con el método Kmeans. En esta ocasión los 3 grupos de clientes son: los que esperan puntualidad y seguridad, los que esperan calidad de servicio, puntualidad y seguridad, y los que esperan seguridad solamente, respectivamente ordenados como se ve en la gráfica. De igual forma, vemos un cambio relevante, y es que los promedios de expectativas de los clientes sobre seguridad solamente ya están más cerca de los promedios de expectativas de los clientes sobre puntualidad y seguridad, siendo más relevantes los promedios de esperar todo en los servicios.

Identificación de grupos usando ambos métodos con Componentes Principales

Probando en kmedias con las componentes principales, comenzamos con tres clusters y consideramos que la separación es más limpia y fue mejor a la anterior.

Finalmente probamos con los datos de las componentes principales para hacer clusters jerárquicos. Usamos todas las mismas disimilaridades que en el caso anterior.

Los mejores modelos para tres clusters fueron usando ward D2 entre clusters y euclidea o minkowski para clientes. Mostramos a continuación la euclidea:

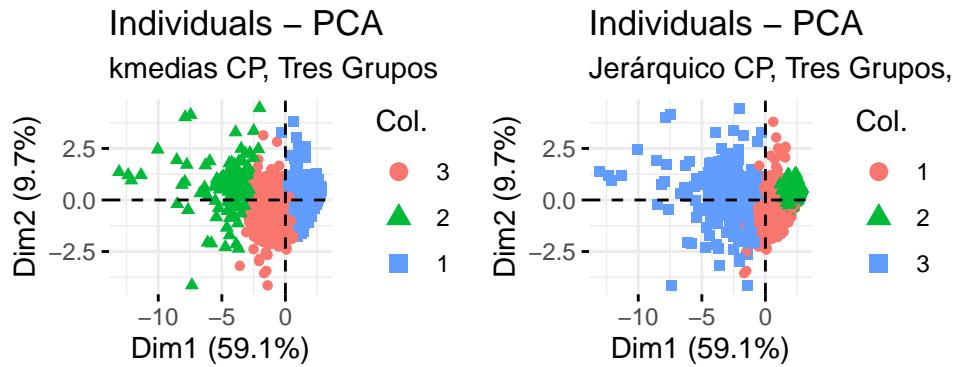


Figura 6: Métodos Kmeans y Jerárquico usando Componentes Principales, con 3 Categorías (#fig:Gráfica 49)

Conclusiones Finales

Finalmente, para tres grupos elegimos el último modelo visto, usando las componentes principales y aglomeramientos jerárquicos con métrica euclidea entre clientes y ward D2 entre clusters. Esto debido a que la separación de las medias en la primera componente principal es la deseada y en la segunda componente, aunque es menos relevante, también es más adecuada la separación. Es decir, en general el grupo 1 tiene un promedio más alto en todas sus respuestas, el grupo 2 un promedio medio y el tercer grupo un promedio de respuestas más bajo que el resto.

La elección de estos modelos dependerá de la publicidad que se quiera usar. Creemos que tiene más sentido enviar tres tipos diferentes de publicidad que cuatro o dos, ya que podríamos dar las mayores ofertas a los de bajas expectativas, como lo son las clientes que sólo buscan seguridad en sus vuelos, enviar más publicidad a los de medias expectativas, como lo son las clientes que buscan seguridad y puntualidad en sus viajes, y a los de altas expectativas, como lo son los clientes que esperan lo mejor en sus vuelos, es decir, la calidad, seguridad y puntualidad, seguir tratándolos con excelencia para seguir con su preferencia al elegir viajar en Oddjob Airways.