CAPITULO 2

2. MARCO TEÓRICO

En este capítulo trataremos las técnicas estadísticas que se aplican de manera amplia, con la recopilación, organización, presentación, análisis e interpretación de los datos con el fin de realizar una toma de decisiones más efectiva.

Partiendo de la estadística descriptiva que se refiere a los procedimientos empleados para la organización presentación y análisis de datos numéricos. También encontraremos la estadística inferencial, que son los métodos empleados para determinar algo acerca de una población, esta población es un conjunto de todos los posibles objetos o mediciones de interés.

Luego de los procedimientos univariados se encontrara los conceptos del análisis multivariado, análisis bivariado, y el análisis no lineal de componentes principales.

2.1. CONCEPTOS BASICOS.

Lo que se desea obtener son cantidades que se definen como medidas numéricas descriptivas de un conjunto de datos. Se buscan números que describan la distribución de frecuencias para cualquier conjunto de mediciones. Se concentrará la obtención en dos tipos de números descriptivos, las medidas de tendencia central y las medidas de dispersión o variabilidad.

La medida de tendencia central más común que se utiliza en la estadística es la media.

Definición.- La *media muestral* de un conjunto de mediciones $y_1, y_2, \ldots y_n$ esta dada por:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

La medida más común de variabilidad usada en la estadística es la varianza que es una función de las desviaciones (o distancias) de las mediciones muestrales con respecto a su media.

Definición.- La *varianza muestral* de un conjunto de mediciones, es la media del cuadrado de las desviaciones de las mediciones con respecto a su media y esta dada por:

La varianza es útil en la comparación de variación de dos conjuntos de mediciones, pero sólo aporta información con respecto a la variación en un solo conjunto cuando se interpreta en términos de la desviación estándar.

Definición.- La *desviación estándar* de un conjunto de mediciones es $y_1, y_2, \ldots y_n$ la raíz positiva de la varianza.

2.1.1 REPRESENTACIONES GRAFICAS

La importancia para comprender, captar y hacer un análisis simple visual que es fácil obtenerlo mediante programas estadísticos.

Al obtener gráficos de los datos de variables individuales, se puede fácilmente de manera rápida una interpretación de cada variable. Para obtener una perspectiva del análisis de los datos es importante determinarlos por medio de gráficas.

Se pueden hacer gráficos de variables estadísticas en forma individual, pero también pueden ser de mucha utilidad y brindar buena información los gráficos de pares de variables.

2.1.1.1 HISTOGRAMA

El histograma es uno de los medios gráficos de más fácil interpretación. Para elaborar un histograma, las frecuencias de clase se marcan en la escala de un eje vertical (eje Y), y uno horizontal (eje X) que marcan los límites declarados, los límites verdaderos o los puntos medios. Se utilizarán los límites declarados y se mostrará sólo el límite inferior de cada clase en el eje X.

2.1.2 MEDIDAS DE ASIMETRIA O SESGO

Una característica que puede medirse es el grado de asimetría de una distribución. Recordando que si una distribución de frecuencias

es simétrica, no tiene sesgo, es decir el *sesgo es nulo*. Si una o más observaciones son muy grandes, la media de la distribución se vuelve mayor que la mediana o la moda. En tales casos se dice que la distribución tiene *sesgo positivo*. Por lo contrario, si una o más observaciones muy pequeñas se encuentran presentes, la media es la menor de los tres promedios y se dice que la distribución tiene sesgo negativo. Este se lo calcula:

$$\gamma_1 = \left\{ \frac{n\left[\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^3\right]^2}{\left[\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right]^3} \right\}^{1/2}$$

2.1.3 COEFICIENTE DE KURTOSIS.

Una medida que mide el grado de agudeza de una distribución. Este permite establecer el grado de apuntamiento o achatamiento de la curva de distribución normal, tenemos distribución *mesocúrtica,* cuando tiene la forma de una normal y su coeficiente es igual a tres, distribución *platicúrtica* cuando es achatada con respecto a una normal y su coeficiente es menor a tres; y por último, distribución

leptocúrtica cuando es mas apuntada que una normal y su coeficiente es mayor a tres. Se calcula a través de la relación entre el cuarto momento central y la varianza al cuadrado se lo indica:

$$\bar{\alpha}_4 = \frac{\bar{\mu}_4}{s^4} = \frac{E\left[(x_i - \bar{x})^4\right]}{s^4}$$

2.1.4 ANALISIS BIVARIADO (Tablas bivariadas)

Con las tablas bivariadas podemos tener un análisis entre dos variables de interés, al obtener este análisis simultaneo, en las tablas encontramos las variables con sus respectivos niveles para determinar el porcentaje que se determinen mediante la combinación.

2.1.5 ANALISIS DE CONTINGENCIA

Como parte importante de los datos enumerativos es la independencia de dos métodos de clasificación de eventos observados. Consta de un arreglo bidimensional en la que se detalla

los factores a ser analizados con igual o diferentes niveles de información que nos permitirá determinar si esos dos factores son independientes al realizar un contraste de hipótesis sobre independencia de los factores.

Una explicación más clara del análisis de contingencia:, como parte del arreglo bidimensional, tiene r filas y c columnas, donde r es el número de niveles del primer factor o de la variable X_i y c el número de niveles del segundo factor o de la variable X_j, cada variable debe tener al menos dos niveles los cuales deben ser exhaustivos y mutuamente excluyentes. Sirven para determinar la dependencia o independencia de dos variable aleatorias X_i y X_j.

Sea A un factor con r niveles y B un factor con c niveles, la tabla de contingencia se muestra en el siguiente cuadro:

TABLA DE CONTINGENCIA

		FACTOR B				
		Nivel 1	***Nivel 2***	***…***	***Nivel c***	$X_{i.}$
FACTOR A	***Nivel 1***	X_{11} E_{11}	X_{12} E_{12}	…	X_{1c} E_{1c}	X_{1} .
	Nivel 2	X_{21} E_{21}	X_{22} E_{22}	…	X_{2c} E_{2c}	X_{2} .
		------- -------	------ -------	----- ------	------- -------	-------- --------
	Nivel r	X_{r1} E_{r1}	X_{r2} E_{r2}	…	X_{rc} E_{rc}	X_{r} .
	$X_{.j}$	$X_{.1}$	$X_{.2}$	…	$x_{.3}$	X.. =n

Donde :

n = es el número de observaciones

X_{ij} = es el número de valores observados que simultáneamente poseen la i-ésima característica del factor A y la característica j-ésima del factor B.

E_{ij} = es el número de observaciones esperadas con la i-ésima característica del factor A y la característica j-ésima del factor B y se lo obtiene:

$$E_{ij} = \frac{X_{i.} * X_{.j}}{n} = \frac{\sum_{j=1}^{c} X_{ij} * \sum_{i=1}^{r} X_{ij}}{n}$$

$X_{i.}$ = al número de observaciones que poseen la característica i-ésima del factor B.

$X_{.j}$ = al número de observaciones que poseen la característica j-ésima del factor A.

Para tener el contraste de hipótesis sobre la independencia tenemos las siguientes postulaciones:

H_o : El factor A es independiente del factor B

vs

H_1 : El factor A es dependiente del factor B

El estadístico de prueba utilizado $\chi^2 = \sum_{i=1}^{h} \sum_{j=1}^{k} (X_{ij} - E_{ij})^2 / E_{ij}$ el cual se distribuye según una ji-cuadrado con (r-1)*(c-1)grados de libertad, se rechaza la hipótesis nula a favor de la hipótesis alterna con (1-α)100% de confianza si:

$$\chi^2 > \chi^2_{\alpha(h-1)(k-1)}.$$

2.1.6 MATRIZ DE DATOS.

Los datos obtenidos, estarán en una matriz de datos, esta matriz estará representada con n filas y p columnas, y los vectores columnas contienen los n valores de cada una de las p variables consideradas.

En donde cada elemento (X_{ij}) de la matriz representa la i-ésima observación de la j-ésima variable:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & . & . & . & x_{1n} \\ x_{21} & x_{22} & . & . & . & x_{2n} \\ . & & . & & & . \\ . & & & . & & . \\ . & & & & . & . \\ x_{p1} & x_{p2} & . & . & . & x_{pn} \end{bmatrix} = \begin{bmatrix} \mathbf{X_1} & \mathbf{X_2} & \ldots & \mathbf{X_n} \end{bmatrix}, \quad \mathbf{X_i} \in R^P$$

2.1.7 MATRIZ DE VARIANZAS Y COVARIANZAS

La matriz de covarianza son un arreglo de **p** filas y **p** columnas, es decir es una matriz cuadrada que tiene la propiedad de ser simétrica. Se nota de la siguiente manera:

$$S_{i\,\mathrm{j}} = \begin{bmatrix} \mathrm{S}_{11} & \mathrm{S}_{12} & -- & \mathrm{S}_{1\mathrm{p}} \\ \mathrm{S}_{21} & \mathrm{S}_{22} & -- & \mathrm{S}_{2\mathrm{p}} \\ -- & -- & -- & -- \\ \mathrm{S}_{\mathrm{p}1} & \mathrm{S}_{\mathrm{p}2} & -- & \mathrm{S}_{\mathrm{pp}} \end{bmatrix}$$

Donde:

$$S_{ij} = \begin{cases} \text{covarianza entre la variable i j, cuando } i \neq j \\ \text{varianza de la variable i , cuando } i = j \end{cases}$$

Se calcula S_{ij} de la siguiente forma:

$$S_{ij} = \frac{1}{n} \sum_{k=1}^{n} (X_{ki} - \overline{X}_i)(X_{kj} - \overline{X}_j)$$

En la matriz, se encuentran en la diagonal de la matriz las varianzas de cada una de las variables de estudio, y en la matriz triangular superior e inferior las covarianzas.

2.1.8 MATRIZ DE CORRELACION.

La matriz de correlación es en forma igual a la matriz de correlación es decir un arreglo de **p** filas y **p** columnas, que es una matriz cuadrada que tiene la propiedad de ser simétrica.

Se define ρ (matriz de correlación) como : $\rho = \mathbf{V}^{-1/2}\,\Sigma\,\mathbf{V}^{-1/2}$, donde Σ es la matriz de varianzas y covarianzas de **?** y V es la matriz cuya diagonal contiene el inverso de la desviación estándar de cada variable, ρ resume los coeficientes de correlación de todos los pares de variables entre las p dadas $X_1, X_2, ..., X_p$.

$$\rho = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \cdots\cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \cdots\cdots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} & \cdots\cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots\cdots & \rho r_{1p} \\ \rho_{21} & 1 & \rho_{22} & \cdots\cdots & \rho_{2p} \\ \rho_{31} & \rho_{32} & 1 & \cdots\cdots & \rho_{3p} \\ . & . & & \cdots\cdots & \\ \rho_{p1} & \rho_{p2} & \rho_{p3} & \cdots\cdots & \rho_{pp} \end{bmatrix}$$

El coeficiente de correlación $\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}$ determina el grado de dependencia lineal entre la i-ésima y la j-ésima variable. Si ρ_{ij}=1 o ρ_{ij}=-1 determina en ambos casos que existe una perfecta

dependencia lineal entre las variables, la diferencia en el signo determina únicamente el sentido de variación si es positivo ambas variables aumentan o disminuyen simultáneamente y si es negativo el sentido de variación es opuesto lo que significa que al aumentar una disminuye la otra. El hecho que el coeficiente de correlación sea igual a cero indica que entre ambas variables no existe relación lineal.

2.1.9 COMPONENTES PRINCIPALES

El Análisis de Componentes Principales, también conocido como ACP, es utilizado para describir una matriz X (matriz de datos), es decir una matriz de variables continuas del tipo individuos x variables. Las componentes principales explican la estructura de varianza y covarianza de un conjunto de variables a través de pocas combinaciones lineales de ellas, el objetivo de las componentes principales es reducir el número de variables de trabajo y simplificar la interpretación.

Sea $\mathbf{x} \in \Re^P$ un vector aleatorio p variado $X^T = [X_1, X_2, ..., X_P]$, no necesariamente normal p variado con una matriz de varianzas y covarianzas Σ y con un vector de medias μ, se procede a calcular los valores y vectores propios asociados a la matriz de varianzas y covarianzas para formar las combinaciones lineales de acuerdo al siguiente criterio $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_p$

$$
\begin{aligned}
Y_1 &= a_1^T X = a_{11}X_1 + a_{12}X_2 + L + a_{1P}X_P \\
Y_2 &= a_2^T X = a_{21}X_1 + a_{22}X_2 + L + a_{2P}X_P \\
&- - - - - - \\
Y_P &= a_P^T X = a_{P1}X_1 + a_{P2}X_2 + L + a_{PP}X_P
\end{aligned}
$$

donde :

$$\mathbf{Var}(Y_i) = \mathbf{a}_i' \Sigma \mathbf{a}_i = \lambda_i \qquad para \;\; i = 1,2,.....p$$

$$\mathbf{Var}(Y_i) > \mathbf{Var}(Y_{i+1})$$

$$\mathbf{Cov}(Y_i, Y_k) = \mathbf{a}_i' \Sigma \mathbf{a}_k = 0 \quad i \neq k$$

Vale destacar que los vectores propios empleados en las combinaciones lineales son ortogonales $\|\mathbf{a}_i\| = 1$ y que las varianzas de las componentes principales son las más altas posibles.

Para elegir el número de Componentes que se han de retener, se tiene criterios que se utilizan para la elección de las componentes.

Criterio #1.-

Para contener las componentes, con los valores propios debemos aspirar que al menos el 80% de la variabilidad total contenida en las p variables estén presentes en las K componentes principales.

- p variables tienen el 100% de la variabilidad
- k<p debería contener por lo menos el 80% de la variabilidad.

Criterio #2.-

Este criterio indica que se debe retener las componentes para cada λi o valor propio sean mayor que el promedio de λi.

Verificación de la independencia entre las p variables:

Siendo:

S= $\sum$ R= ρ

γ= n-1; donde n es el tamaño de la muestra .

$$u = \frac{\det S}{S11 * S22 * \ldots\ldots\ldots * Spp} = \det R$$

Con (1-α)% de confiaza.

Si $u' > \chi^2 \alpha(n-1)$

Donde
$u' = [-(\gamma - 1/6 (2p + 5)] \ln u$

Si $u' > \chi^2$. Esto quiere decir que alguno de las p variables son correlacionadas, entonces podría aplicarse componentes principales.

2.1.10. COMPONENTES PRINCIPALES NO LINEALES

Un tratamiento no estándar del análisis multivariante es, el denominado análisis no lineal multivariante, el cual trabaja tanto con variables métricas como con variables categóricas o nominales. Este análisis determina una nueva ponderación para la codificación de los datos categóricos. A partir del concepto de homogeneidad utiliza algoritmos que maximizan la homogeneidad de un grupo de variables.

El análisis de componentes principales no lineales, procede de la siguiente forma:

Primero.- se pondera la codificación de las variables cualitativas o categóricas.

Segundo.- a estas variables reponderadas les aplica el ACP clásico.

A continuación se describe uno de los algoritmos de ponderación más utilizada.

OBJETOS Y CATEGORIAS

Sea un conjunto de *n* objetos o individuos. Una variable h_j hace corresponder al conjunto de los individuos un conjunto finito de k_j categorías, este conjunto de categorías se denomina el rango de h_j. Vamos a asumir que existe un número finito de m variables h_j ($j=1,...m$). El producto cartesiano de todas estas categorías se denomina rango multivariante, sus elementos son todas las posibles combinaciones de las m categorías, y se denominan perfiles. La matriz de datos *H* es una matriz *nxm* con elementos h_{ij} que nos indican la categoría de la variable h_j para el individuo *i*. Estos elementos no necesariamente son números.

VARIABLES INDICATRICES

La matriz H puede ser codificada utilizando variables indicatríces:

Para cada variable h_j se define una matriz binaria G_j nxk_j, de la siguiente forma:

$$g_{(j)ir} = \begin{cases} 1 & si\ el\ i-ésimo\ individuo\ está\ en\ la\ r-ésima\ categoría\ de\ h_j \\ 0 & caso\ contrario \end{cases}$$

G_j se denomina matriz indicatriz de h_j.

Tales matrices pueden ser reunidas en una matriz particionada $G=(G_1,...,G_j,...,G_m)$ de dimensión $nx\Sigma k_j$ también conocida como matriz indicatriz.

La matriz indicatriz G_j se dice completa si cada fila de G_j tiene sólo un elemento igual a *1* y el resto iguales a *0*, es decir que la suma de cada fila de G_j es igual a *1*. En lo que sigue, se puede escribir $G_ju=u$, donde u es un vector de unos. Si todas las matrices G_j son completas, su matriz combinada G también se dice completa, y se tiene que

$Gu=mu$, donde m es el número de variables y u es un vector de unos.

Sea d_j el vector de totales por columna de G_j. Su r-ésimo elemento $d_{(j)r}$ corresponde a la frecuencia marginal de la categoría r-ésima de h_j. La suma de los elementos de d_j es igual a $u^t d_j = n$.

Sea $D_j = G_j^t G_j$, esta matriz es diagonal puesto que las columnas de G_j son ortogonales y los elementos de su diagonal son los mismos que las frecuencias marginales dados por d_j.

Definimos $C_{jl} = G_j^t G_l$, que es un cruce de las variables h_j y h_l. Sus elementos corresponden a la frecuencia de individuos caracterizados por una particular combinación de una categoría en h_j y una en h_l.

Sea $C = G^t G$, esta matriz combina todas las C_{jl} y su diagonal esta formada por las matrices $C_{jj}=D_j$. C es una matriz de marginales bivariantes (también conocida en la literatura francesa como tabla de Burt).

D es una matriz que consta de las submatrices de la diagonal de C y el resto de sus elementos es cero. D es una matriz de marginales univariantes.

CUANTIFICACION

Las categorías de las variables pueden ser valores numéricos, como puntos medios de intervalos de alguna variable continua. En este caso la matriz H_{nxm} es una matriz de datos clásica y puede ser manejada con las técnicas clásicas del análisis multivariante. En el presente documento no se va a asumir tal cuantificación a priori. Incluso en el caso donde exista tal cuantificación a priori, ésta debe ser ignorada y reemplazada por una categorización nominal. Por ejemplo, si se dispone de la variable edad, esta debe ser dividida en intervalos y a cada uno de éstos asignarle una etiqueta (por ejemplo los puntos medios de cada intervalo):

Supongamos que tenemos una variable “edad” que asigna a los individuos a 15 grupos de edad, cada grupo representado por el punto medio del intervalo en la escala edad. La matriz de datos *H* estará formada por una columna con 15 valores. Su correspondiente matriz indicatriz *G* en cambio tendrá 15 columnas una por cada grupo de edad. Bien se podría olvidar el origen métrico de estas 15 categorías y pensarlas como 15 categorías nominales.

La cuantificación de categorías sigue ciertas reglas, con la intención de optimizar algún criterio, generalmente este criterio es una función de pérdida. Por el momento no se discutirá tal función, sin embargo se indicará en forma global como la cuantificación de una matriz indicatriz es factible:

La cuantificación de las categorías de la variable h_j implica que sus k_j categorías son asignadas como los k_j valores numéricos de un vector y_j. Entonces la variable cuantificada $q_j=G_j y_j$ viene a ser un vector (en $\mathbb{R}^n$) que nos proporciona un resultado numérico para cada individuo con respecto a h_j.

Definimos x como el vector promedio de todos los q_j:

$$x = \frac{1}{m}\sum q_j$$

El vector $x \in \mathbb{R}^n$ contendrá la cuantificación de los individuos y diremos que para alguna cuantificación directa y_j de categorías, " x es el puntaje inducido de los individuos".

Por otro lado, si *x* es alguna cuantificación directa de los individuos, se puede definir una categorización inducida por *x* como el promedio de los puntajes de aquellos objetos que asignados en dicha categoría:

$$y_j = D_j^{-1} G_j^t x$$

En lo que sigue, se asume que D_j tiene inversa, lo que significa que no hay categorías con frecuencia cero. Si se este fuese el caso, se debe quitar a esta columna de la matriz indicatriz.

Ambos procedimientos se pueden unir de la siguiente forma:

Sea y_j una cuantificación directa de las categorías de la *j*-ésima variable. Sea *y* un vector que esté compuesto por todos los vectores y_j, es decir tiene Σk_j componentes. Los puntajes inducidos de los individuos son: Gy/m.

Se requiere que una solución para la cuantificación directa de los individuos, *x*, sea proporcional a los puntajes inducidos de los individuos y viceversa, que la cuantificación directa de las categorías, y_j, sea proporcional a la cuantificación inducida de las categorías $D_j^{-1} G_j^t x$. En la parte 3 veremos dos métodos para obtener soluciones a este problema.

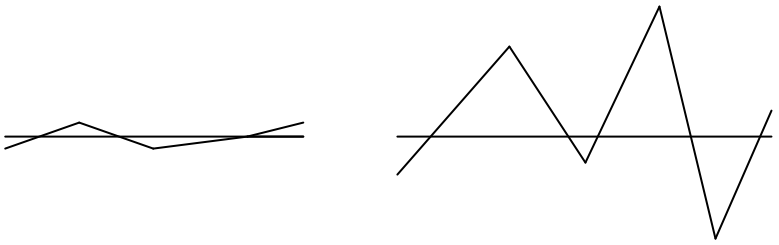
ANÁLISIS DE HOMOGENEIDAD

El término análisis de homogeneidad puede usarse en sentido estricto y amplio. En sentido estricto, denota una técnica para el análisis de datos puramente categóricos con determinada función de pérdida. En sentido amplio, se refiere a una clase de criterio para analizar datos multivariantes en general, compartiendo las características que ayudan a optimizar la homogeneidad de las variables bajo varias formas de manipulación y simplificación. Esta clase de criterio será utilizado en las próximas secciones

HOMOGENEIDAD DE VARIABLES

Históricamente, la idea de homogeneidad es cercana a la idea de que diferentes variables pueden medir la “misma cosa”. Si lo último fuese cierto, la matriz de datos (asumiendo que las variables no son otra cosa que desviaciones de su media, es decir que son estandarizadas) podrían dar valores idénticos en cada fila o, si dibujamos las

observaciones como perfiles, cada uno de ellos sería una recta. Si la idea de “medir la misma cosa” fuese no fuese muy exacta (las variables miden la “misma cosa” pero con un error aleatorio), las filas de la matriz de datos pueden tener elementos que varían un tanto. Un gráfico de los perfiles sería una línea quebrada.



Si reemplazamos tales perfiles por líneas rectas (la media) tendremos una pérdida de información. Las variables son homogéneas si la pérdida de información es relativamente pequeña

MAXIMIZANDO LA HOMOGENEIDAD POR COMBINACIONES LINEALES DE PESOS.

La correlación promedio de las variables nos brinda un estimado de cuan bien ellas pueden ser reducidas a un vector de puntajes, si las mantenemos en su forma original (en este caso en forma estandarizada). Supongamos que es permitido re-escalar a las variables antes de promediarlas, i.e asignar pesos a h_j, en un intento de incrementar la homogeneidad.

Sea *x* un vector de puntajes arbitrario (en I^n) y con media cero. Sea *a* un vector de *m* pesos. Re-escalar las columnas de *H* es equivalente a reemplazar h_j por $a_j h_j$. El problema que tenemos es el de elegir *x* y *a* de tal forma que se maximize la homogeneidad. Más explícitamente, minimizar la pérdida de homogeneidad. Por función de pérdida consideramos a la función:

$$\sigma(x,a) \equiv \frac{1}{m}\sum_j SSQ(x - a_j h_j)$$

Evidentemente, esta función de pérdida tiene un mínimo absoluto en *x=0* y *a=0*. Para excluir esta solución trivial es necesario normalizar *x*

así que $x^t x=c$ donde c es una constante dada distinta de 0 (generalmente igual a 1).

El objetivo de elegir puntajes y pesos así como de maximizar la homogeneidad o de minimizar la función de pérdida es una de las posibles definiciones que se puede utilizar para describir a la (primera) componente principal de H. Esto involucra a combinaciones lineales de pesos, puesto que $a_j h_j$ pueden ser vistas como una transformación lineal de h_j. Realmente calcular x y a es más complicado que hacer un simple promedio. En la próxima sección veremos que una sucesión de promedios ponderados es suficiente para aproximar la solución tanto como se desee.

ALGORITMO DE PESOS NORMALIZADOS:

En este algoritmo los pesos satisfacen la restricción $a^t a=1$. Se requiere un valor inicial arbitrario de $x^0 \neq 0$:

1. Actualización de pesos: $a^0 \leftarrow H^t x^0$
2. Normalización: $a^+ \leftarrow \frac{a^0}{\left\| a^0 \right\|}$
3. Actualización de puntajes: $x^+ \leftarrow \frac{1}{m} H a^+$

4. Test de convergencia: Regresar a (1), hacer que $x^0 \leftarrow x^+$, mientras los valores de x^+ y a^+ no estén los suficientemente estabilizados (de acuerdo a algún criterio de convergencia previamente definido).

ALGORITMO DE PUNTAJES NORMALIZADOS.

En este algoritmo los puntajes de los individuos se sujetan a la restricción $x^t x=1$. El algoritmo requiere un vector de pesos inicial y arbitrario $a^0 \neq 0$:

1. Actualización de puntajes: $x^0 \leftarrow \frac{1}{m} H a^0$
2. Normalización: $x^+ \leftarrow \frac{x^0}{\|x^0\|}$
3. Actualización de pesos: $a^+ \leftarrow H^t x^+$
4. Test de convergencia: Regresar a (1), hacer que $a^0 \leftarrow a^+$ mientras los valores de x^+ y a^+ no estén suficientemente estabilizados (de acuerdo a algún criterio de exactitud previamente establecido).

Descripción del algoritmo:

1. Corresponde al mínimo condicional no restringido de la función de pérdida (2.1) para un a^0 fijo. Notemos que Ha^0/m es un vector que

contiene los promedios de las filas reescaladas por a_j^0. Los puntajes actualizados x^0 por consiguiente también minimizan la perdida relativa *W/T* para *H* re-escalada con pesos fijos a^0.

2. Es la proyección de x^0 sobre la hiper-esfera de todos los *x* normalizados, lo que transfiere la restricción de minimización a una región factible (la región que contiene todas las soluciones que satisfacen la restricción).
3. Corresponde al mínimo condicional no restringido de la función de pérdida (2.1) para un x^+ fijo. Puesto que x^+ y las columnas de *H* son centradas y normalizadas, a^+ es un vector de correlaciones.
4. El algoritmo converge monótonamente, puesto que los pasos 1 y 2 conjuntamente y el paso 3, siempre dan un pequeño valor de la función de perdida, la cual está acotada inferiormente por *0*.