



UNIVERSIDAD DE QUINTANA ROO
DIVISIÓN DE CIENCIAS E INGENIERÍA

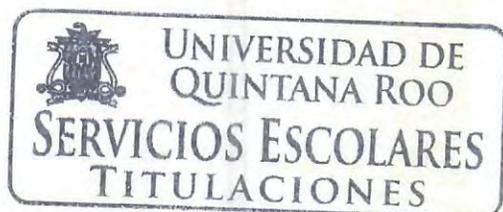
SECUENCIA DE APRENDIZAJE PARA LOS NIVELES DE ENTENDIMIENTO EN LA REDUCCIÓN DE DATOS MEDIANTE COMPONENTES PRINCIPALES

TESIS
PARA OBTENER EL GRADO DE
MAESTRO EN ENSEÑANZA EN LAS MATEMÁTICAS

PRESENTA
Q.F.B. JOSÉ LUIS GONZÁLEZ BUCIO

DIRECTOR
DR. JAIME DIONISIO CUEVAS DOMÍNGUEZ

ASESORES
DR. VÍCTOR HUGO DE JESÚS SOBERANIS CRUZ
M.E.M. WALTER MAGAÑA LANDERO
M.E.M. LORENA ANTONIA GUADALUPE PUC ORTIZ
M.E.M. GLORIA MARÍA ALVARADO FARÍAS



CHETUMAL QUINTANA ROO, MÉXICO, MARZO DE 2017



UNIVERSIDAD DE QUINTANA ROO
DIVISIÓN DE CIENCIAS E INGENIERÍA

TRABAJO DE TESIS ELABORADO BAJO LA SUPERVISIÓN DEL
COMITÉ DEL PROGRAMA DE MAESTRÍA Y APROBADA COMO
REQUISITO PARA OBTENER EL GRADO DE:

MAESTRO EN ENSEÑANZA DE LAS MATEMÁTICAS

COMITÉ DE TESIS

DIRECTOR:

DR. JAIME DIONISIO CUEVAS DOMÍNGUEZ

NOMBRE Y APELLIDOS

ASESOR:

DR. VÍCTOR HUGO DE JESÚS ROBERANIS CRUZ

NOMBRE Y APELLIDOS

ASESOR:

M.E.M. WALTER MAGAÑA LANDERO

NOMBRE Y APELLIDOS

ASESOR:

M.E.M. LORENA ANTONIA GUADALUPE PUC ORTÍZ

NOMBRE Y APELLIDOS

ASESOR:

M.E.M. GLORIA MARÍA ALVARADO FARIAS

NOMBRE Y APELLIDOS



DCI DIVISIÓN DE
CIENCIAS E
INGENIERÍA



CHETUMAL, QUINTANA ROO, MÉXICO, MAZO DE 2011

ÍNDICE

1.	INTRODUCCIÓN	1
1.1	Planteamiento del problema	5
1.2	Objetivo general	7
1.2.1	Objetivos particulares	7
2.	MARCO TEÓRICO	8
2.1	Técnicas de reducción de dimensión	8
2.2	Técnicas multivariadas	9
2.2.1	Análisis de componentes principales (ACP)	9
2.2.2	Generación de las componentes principales desde los datos	10
2.3	Desarrollo de una comprensión de la reducción de datos	11
3.	METODOLOGÍA	14
3.1	Criterios para el diseño de la secuencia de aprendizaje	16
3.1.1	Actividades de la secuencia de aprendizaje	16
3.2	Secuencia de aprendizaje	17
3.2.1	Caso de estudio: Factores de riesgo de enfermedad coronaria, mediante el análisis de componentes principales (ACP)	17
3.2.2	Objetivos de la secuencia didáctica	18
3.2.3	Nivel de comprensión 1	18
3.2.4	Desarrollo intuitivo de la de la reducción de datos	18

3.2.4.1 Actividad 1.- Apertura	18
3.2.4.2 Desarrollo del nivel de comprensión intuitivo de la reducción de datos	21
3.2.4.3 Evaluación	22
3.2.5 Nivel de comprensión 2	24
3.2.5.1 Implementación de la reducción de datos	24
3.2.5.2 Actividad 2.- Apertura	24
3.2.5.3 Desarrollo	24
3.2.5.4 Evaluación	25
3.2.6 Nivel de comprensión 3	27
3.2.6.1 Comprensión global de la reducción de datos y el ajuste de modelos	27
3.2.6.2 Actividad 3.- Apertura	27
3.2.6.3 Desarrollo	27
3.2.6.4 Evaluación	28
3.3 Evaluación del razonamiento de los estudiantes acerca de la reducción de datos por CP	30
3.3.1 Evaluación de los resultados de la secuencia de aprendizaje	30
4. RESULTADOS	34
4.1 Informe del desarrollo de la secuencia	34
4.1.1 Nivel de comprensión 1	34
4.1.2 Nivel de comprensión 2	38
4.1.3 Nivel de comprensión 3	41

4.2	Resultados de las actividades del grupo I	45
4.3	Resultados de las actividades del grupo II	55
5.	CONCLUSIONES	61
6.	CITAS BIBLIOGRÁFICAS	63

CONTENIDO

Figura 1. Histogramas de las variables

Figura 2 Varianza.

Figura 3. Gráfica de dispersión de presión arterial con peso y edad.

Figura 4. Matriz de dispersión de las siete variables.

Figura 5. Vistas de casa.

Figura 6. Figuras de componentes principales.

Figura 7. Diferentes proyecciones en el componente principal CP-I.

Figura 8. Componente principal CP-I.

Figura 9. Componentes principales 1 y 2.

Figura 10. Componentes principales con tres variables.

Figura 11. Componentes principales con siete variables.

CONTENIDO

Tabla 1. Tabla de datos.

Tabla 2. Estadísticos descriptivos.

Tabla 3. Correlaciones lineales de Pearson.

Tabla 4. Nivel de comprensión uno.

Tabla 5. Nivel de comprensión dos.

Tabla 6. Nivel de comprensión tres.

ABREVIATURAS

ACP Análisis de Componentes Principales

SIMBOLOGÍA

a	Eigenvector
Σ	Sumatoria
n	Tamaño de la Muestra
N	Tamaño de la Población
S_{ij}	Covarianza de la Muestra
X	Matriz de Datos
\bar{x}	Media de la Muestra
%	Porcentaje
Z	Matriz de Datos Estandarizados

1. INTRODUCCION

El volumen de datos generados en todos los aspectos, ya sean laborales y no laborales, complican su interpretación y ha traído como consecuencia la necesidad de utilización de métodos especiales del procesamiento de los datos.

Por ejemplo en el análisis ambiental que está conformada por una serie de métodos matemáticos, estadísticos y químicos, permiten el reconocimiento de relaciones complejas y patrones de comportamiento ocultos en grandes cantidades de datos, en una amplia variedad de situaciones en diversos campos (Einax, 1992).

Los resultados de las mediciones medioambientales se caracterizan usualmente por su elevada variabilidad, en la que se incluye, además de su variabilidad natural, la incertidumbre resultante de los procesos analíticos como el muestreo, la conservación de las muestras, las mediciones analíticas, etc., de manera que al emplear métodos univariados para su estudio se pierde a menudo información (Grupo de Quimiometría, 2001)

Una de las principales herramientas para esta complejidad de datos es el Análisis por Componentes Principales (ACP). Esta ha sido empleada en los últimos 20 años, en la interpretación de datos numéricos resultantes de los diseños de experimentos y de la medición de numerosas variables simultáneamente.

El análisis de componentes principales (ACP) es una técnica multivariante diseñada en 1901 por Karl Pearson con la finalidad de encontrar líneas y planos que mejor se ajusten a una nube de puntos en el espacio. La técnica de Pearson ha sido modificada a través de los años (Cela, 1994) empleándose actualmente en muchos tipos de problemas multidimensionales (Quintana., 1994; Carlosena, 1999; Cal, 2001).

La idea básica del ACP es encontrar un pequeño número de combinaciones lineales no correlacionadas a partir de las n variables originales, las cuales expliquen el mayor por ciento de la variabilidad total de los datos originales. Es decir, pasar el conjunto de variables primitivas a un nuevo espacio, trasladando la máxima información original contenida en el espacio

multidimensional analizado, a un espacio de dimensionalidad reducida, como por ejemplo tri o bidimensional (Massart D.L., 1998)

Los estudios multivariados, en cualquier campo al que se apliquen, comienzan frecuentemente a partir de la construcción de una matriz de correlación. Esta constituye una tabla simétrica de coeficientes de correlación de cada variable respecto a cada una de las otras; de ellas pueden emerger patrones y estructuras que a menudo no son distinguibles por simple inspección de los datos. El análisis de correlación sirve como técnica primaria descriptiva estimadora del grado de asociación entre las variables involucradas en el estudio (Sokal, 2012).

Las variables analizadas deben presentar cierto grado de correlación, o sea, que aporten relativamente el mismo tipo de información (por ejemplo, un conjunto de distintos parámetros de composición metálica de sedimentos y organismos) para que pueda llevarse a cabo la reducción del número de variables, manteniendo un porcentaje elevado de la información que aportaban las variables originales (Quintana, 1994).

Por ello, primeramente debe comprobarse la adecuación de los datos para la realización del ACP mediante la prueba de esfericidad de Bartlett (Cela, 1994) que revela cuándo la matriz de correlación es una matriz de identidad, lo cual podría indicar que las variables no estuvieran correlacionadas. Un nivel de significación menor de 0.05 indica que probablemente existen relaciones significativas entre las variables. La prueba de Kaiser-Meyer-Olkin (KMO) también ofrece una medida de la adecuación muestral para la realización del ACP e indica la proporción de varianza de las variables que es varianza común, y que puede ser causada por ejemplo, por factores subyacentes. Valores del estadígrafo KMO mayores que 0.7 indican que los datos estudiados son adecuados para realizar sobre ellos un análisis de componentes principales, dado que las variables están suficientemente correlacionadas como para llevar a cabo un análisis de reducción de dimensionalidad.

Una vez confirmada la adecuación muestral, la matriz de datos formada por n filas (muestras) y p columnas (parámetros evaluados) se considera como un “enjambre” de dimensionalidad $n \times p$ puntos de un espacio p -dimensional, y se constituyen p nuevas variables, llamadas componentes principales. Se denomina componente principal 1 (CP1) a la recta de mínimos cuadrados que pasa por los puntos medios (\bar{x}) de las variables originales. Al trazar con igual criterio una segunda recta

que pasa por (\bar{x}) pero perpendicular a CP1, se obtiene una segunda componente principal (CP2) y así sucesivamente (Cela, 1994). De modo que, las componentes principales significan buscar los ejes sobre los que se maximiza la varianza de los datos estudiados a partir de la combinación lineal de los datos primarios.

Consecutivamente se obtienen tantas componentes principales como variables originales. Cada una de ellas explica una parte de la varianza total de los datos expresada como porcentaje de varianza explicada, sucesivamente menor, hasta explicar el 100% de la varianza total (información original) que estaba contenida en los datos primarios. La suma de cada uno de los porcentajes de varianza explicada por cada componente con la consecutiva, se denomina tanto por ciento de varianza acumulativa y ayuda a decidir cuántas componentes principales retener para realizar el estudio.

En la literatura especializada se proponen diversos criterios para la selección del número de componentes a retener. Uno de ellos es el llamado “Criterio del autovalor medio” que consiste en escoger las componentes cuyo autovalor sea mayor que 1. Otro criterio es el del “Bastón roto” en el que se calcula el parámetro b_k para las CP y se compara con el porcentaje de varianza explicada por cada una. Se consideran significativas las componentes en las que el porcentaje de varianza explicado sea mayor que el valor del parámetro b_k . Adicionalmente existe el criterio del “porcentaje de varianza explicada” mediante el que se seleccionan aquellas componentes que expliquen más de la tercera parte de la varianza original (Cela, 1994). Este último criterio fue el seleccionado para el trabajo desarrollado.

La matriz de “loadings” o peso de las variables primitivas en las nuevas variables (CP) representa la correlación encontrada entre las variables originales y el componente principal p , por lo que esta matriz resulta muy útil en la interpretación de los resultados del análisis.

Los pesos más cercanos a 1 corresponden a variables altamente correlacionadas con la componente principal sobre la que presentan este valor. No obstante, es común obtener pesos que tengan valores muy parecidos para dos componentes, lo cual puede dificultar la interpretación de los resultados. En estos casos puede ser útil practicar una “rotación” a esta matriz.

La matriz inicial de coeficientes o pesos se somete frecuentemente a una rotación para facilitar su interpretación, maximizando los pesos de unas variables en algunas componentes y

minimizándolos en otras. Se han definido diferentes tipos de rotaciones como la quartimax (maximiza la varianza del cuadrado de cada fila de pesos), varimax (maximiza la varianza del cuadrado de cada columna de pesos), equimax (una combinación o compromiso entre varimax y quartimax), entre otras (Cela, 1994). Con todas se consigue mejorar la interpretación de la significación de las variables en las componentes principales, aunque la más comúnmente utilizada es la varimax.

Si adicionalmente sobre el gráfico de dispersión de los pesos se une el punto representado por una variable y el centro de los ejes coordenados, se obtiene un vector. Si se hace esto mismo con la variable, se obtiene un segundo vector. El coseno del ángulo formado por estos dos vectores será el coeficiente de correlación entre ambas variables. Adicionalmente, cuando mayor sea la magnitud de estos vectores y mayor el coseno cuadrado el ángulo formado por ellos con el eje coordenado (CP), mayor será la contribución de la variable en cuestión a la varianza explicada por dicha componente (Cela, 1994).

Los “scores” o puntuaciones de los objetos sobre las componentes constituyen las proyecciones (coordenadas) de las muestras originales sobre los nuevos ejes formados por las componentes principales (Cela, 1994). Este tipo de diagrama se analiza simultáneamente con los gráficos pasos. Las muestras con altas puntuaciones sobre una componente principal se identifican con aquellas que poseen valores mayores de las variables representadas por dicha componente.

Después del concepto de variabilidad, que está en el corazón de la Estadística y es un elemento fundamental de la cultura general en esta rama de las matemáticas (Pfannkuch y Wild, 2004), la reducción de datos multivariados es otro elemento importante del razonamiento estadístico. La reducción de los datos, y particularmente la reducción del número de variables o el análisis de la estructura interna de un grupo de variables, son ideas esenciales para comprender cómo se relaciona un conjunto de variables que determinan el comportamiento de algún fenómeno, ya que los datos crudos se nos presentan ininteligibles, al no tener elementos para extraer información alguna de ellos. Resumir los datos mediante un pequeño conjunto de nuevas componentes, construidas como transformaciones de las originales, con una pérdida mínima de información es uno de los objetivos del Análisis de Datos Multivariados (Peña, 2002).

En este trabajo distinguiremos entre el Análisis Exploratorio de Datos (AED) para datos univariados (AEDU), y el AED para datos multivariados (AEDM). Esta distinción se basa en el hecho de que la implementación de técnicas de Análisis Exploratorio de Datos Multivariados requiere de llevar a cabo una ruptura epistemológica con el AEDU, desde el punto de vista de los objetivos, el tipo de datos tratados (número, naturaleza, variedad), los sujetos de análisis (variables o individuos), el proceso (de los datos al modelo y no a la inversa), los métodos matemáticos empleados y los conceptos implícitos en los mismos (Batanero, 2001). En el AEDM, por *entendimiento global* nos referiremos a la habilidad para investigar, reconocer, describir, clasificar, y explicar patrones generales en el conjunto de datos multivariados por medio de distribuciones, matrices de parámetros u otras técnicas. El analizar con detenimiento una matriz de varianzas y covarianzas para el discernimiento de patrones y generalidades incluyendo la producción de explicaciones, comparaciones, y predicciones basadas en la dispersión y asociación entre las variables son actividades básicas del AED y se incluyen como elementos relevantes del razonamiento estadístico.

1.1 Planteamiento del problema

Se requiere de **nuevas propuestas didácticas** para un entendimiento sólido de la reducción de la dimensión de un conjunto de datos, que significa desarrollar un modelo cognitivo que incluya varios elementos y sus conexiones, y usar este modelo para razonar acerca de la reducción de datos en diferentes contextos (Garfield y Ben-Zvi, 2005). En esta sección se describe un modelo epistemológico que busca capturar los elementos esenciales de los diferentes niveles de entendimiento de la reducción de datos. Este modelo está basado en la propuesta de Garfield y Ben-Zvi (2005), Ben-Zvi, D., (2004) para un desarrollo profundo de la comprensión de la variabilidad, e incluye algunas de las ideas centrales del modelo de Van Hiele sobre los niveles de entendimiento en Geometría (Burger y Shaughnessy, 1986). El modelo de Van Hiele está formado por dos partes: una descriptiva que identifica diversas etapas de aprendizaje denominadas “niveles de razonamiento geométrico”, la otra está compuesta por directrices sobre cómo organizar el proceso de instrucción para ayudar a los estudiantes a alcanzar un nivel superior de razonamiento; estas se conocen como “fases de aprendizaje” (Jaime y Gutiérrez, 1990). Los 5 niveles de razonamiento del modelo Van Hiele son: Reconocimiento, Análisis, Clasificación, Deducción

Formal y Rigor, cada uno de ellos se identifica con un número del cero al cuatro, respectivamente (Jaime y Gutiérrez, 1990). Las fases de aprendizaje del modelo son: Información, Orientación dirigida, Explicitación, Orientación libre e Integración, las cuales se denominan fase 1, fase 2..., fase 5, respectivamente (Jaime y Gutiérrez, 1990). Los descriptores del modelo de van Hiele se refieren al aprendizaje y la enseñanza de la geometría. En nuestro trabajo se generan descriptores análogos para diversos niveles de entendimiento de la Reducción de Datos por Componentes Principales que pueden desarrollar estudiantes de Ingeniería (Licenciatura) que han aprobado un curso universitario de Estadística. El modelo se centra en el Análisis descriptivo de las Componentes Principales y no al análisis inferencial, ya que la instrucción en el concepto de prueba de hipótesis y en el de intervalo de confiabilidad representa una investigación diferente en la enseñanza de la estadística.

Nuestro modelo epistemológico consiste en un conjunto de categorías que no son mutuamente excluyentes. Estas categorías las consideramos como las piedras de toque para la construcción de los niveles de pensamiento para la comprensión del concepto de reducción de datos por Componentes Principales (CPs), análogos a los niveles de entendimiento del Modelo de Van Hiele. Desde nuestra perspectiva, el que un estudiante complete los niveles antes mencionados será equivalente a decir que tiene una *sólida comprensión* de lo que es la reducción de datos por CPs; es decir, significa que además de conocer el concepto es capaz de relacionar a éste con sus conocimientos previos, que entiende cuáles son los contextos en los que se puede llevar a cabo la reducción de datos, así como las limitaciones del método. Que sea capaz de producir un nuevo conocimiento atreves de descubrir relaciones entre las variable u objetos, resolver problemas reales, construir explicaciones, y sacar conclusiones. Una sólida comprensión de la reducción de datos incluye el conocimiento de las condiciones o restricciones para usar la reducción de datos, un conocimiento de sus elementos tales como preservación de la estructura, dirección optima, dispersión y las relaciones entre éstas. Una comprensión solida incluye el discernimiento de aquellos aspectos que son importantes, respecto de los que son accesorios o irrelevantes. Que el estudiante sea capaz de aplicar y adaptar el concepto a situaciones nuevas y diferentes en vez de seguir los procedimientos aprendidos en situaciones familiares.

1.2 Objetivo general

Diseño, implementación y evaluación de una secuencia de aprendizaje, para el Análisis de Componentes Principales (ACP).

1.2.1 Objetivos particulares

Desarrollar en el estudiante una comprensión intuitiva de la reducción de datos para el Análisis de Componentes Principales.

Desarrollar en el estudiante la comprensión del concepto de Componente Principal.

Aplicación e interpretación del estudiante del Análisis de Componentes Principales.

2. MARCO TEÓRICO

2.1 Técnicas de reducción de dimensión

Condensar una gran cantidad de datos en unos cuantos factores que capturen las características esenciales o representativas de un fenómeno es un elemento importante del razonamiento estadístico. El proceso de reducción de datos permite identificar patrones en el comportamiento de variables, posiblemente correlacionadas, con la finalidad de construir conglomerados de éstas que faciliten y hagan más eficiente el manejo e interpretación de la información.

Con los métodos de reducción de dimensión se pretende transformar al conjunto original de variables en un conjunto menor de combinaciones lineales que mantenga la máxima variabilidad respecto del conjunto original de variables. Mediante el estudio de las proyecciones podemos deducir ciertas propiedades de la estructura global de la información, al igual que mediante una fotografía, podemos reconocer los objetos tridimensionales. Estas técnicas, permiten utilizar las facultades de percepción que usamos a diario: sobre los gráficos del análisis factorial es posible visualizar reagrupamientos, oposiciones, tendencias imposibles de discernir directamente sobre una gran tabla de números, incluso después de un examen prolongado (Batanero, 2001).

Al reducir la dimensión en que se representan los datos, en general tendremos el problema de la no unicidad de solución. No todas las soluciones posibles nos proporcionan la misma información sobre el fenómeno original, al igual que no todas las perspectivas desde las que se fotografía un objeto son igualmente buenas para lograr un reconocimiento adecuado. De modo que se necesitarán ciertos criterios para lograr resultados óptimos. Por ejemplo, al representar gráficamente la relación entre peso y altura de un conjunto de personas, vemos que este conjunto de puntos bidimensional tiene una estructura aproximadamente lineal, aunque para elegir una recta que represente la tendencia de los datos, se añaden ciertas condiciones, como la de minimización de la varianza residual, en el método de mínimos cuadrados, o equivalentemente, la selección de la recta que tome en cuenta la máxima variabilidad que presentan los puntos. En general, este principio - minimización de la varianza residual o maximización de la varianza explicada- se extiende al realizar una reducción de la dimensión en un conjunto de datos. Podemos considerar que con las técnicas de reducción de dimensión estamos haciendo un cambio de base, ya que las observaciones originales de los individuos están expresadas en la base formada por las variables, que pueden o

no estar correlacionadas y que, generalmente, no es una base ortogonal. Al transformar el conjunto inicial de variables en otro no correlacionado, cambiamos la base original por una nueva base ortogonal, en la que la interpretación de las interrelaciones existentes resulta más clara, ya que cada elemento de la nueva base se interpreta independientemente del resto. Entre los métodos que podemos incluir en este apartado se hallan el análisis factorial, análisis de componentes principales, análisis factorial de correspondencias y las técnicas de producción de escalogramas (Batanero, 2001).

2.2 Técnicas Multivariadas

Las técnicas multivariadas son un conjunto de herramientas que sirve para el análisis de datos. Estas técnicas permiten analizar el efecto de varios factores en conjunto. Aquí se transforma una serie de variables en nuevas variables denominadas componentes no correlacionadas, las cuales absorben la varianza (variabilidad) total de los datos (<https://www.dspace.espol.edu.ec/retrieve/94469/D-71989.pdf>), Anderson, C.A, (2000).

2.2.1 Análisis de componentes principales (ACP).

Las aplicaciones del ACP son numerosas y entre ellas podemos citar la clasificación de individuos, la comparación de poblaciones, la estratificación multivariada, etc.

El Análisis de Componentes Principales (ACP) es una técnica cuyo objetivo principal es hallar combinaciones lineales de variables representativas de ciertos fenómenos multidimensionales, con la propiedad de que exhiban varianza mínima y que a la vez no estén correlacionadas entre sí. Para obtener tales combinaciones es necesario construir la matriz de varianzas y covarianzas de esas variables (Dallas. E.J., 2000).

Permite reducir la dimensionalidad de los datos, transformando el conjunto de p variables originales en otro conjunto de q variables no correlacionadas ($q \leq p$) llamadas componentes

principales. Las p variables son medidas sobre cada uno de los n individuos, obteniéndose una tabla de datos o matriz de datos de orden np ($p < n$).

La varianza de la primera componente mientras mayor sea su varianza, mayor será la cantidad de información en dicha componente. Por ello las sucesivas combinaciones o variantes de las componentes se ordenan en forma descendente de acuerdo a la proporción de la varianza total presente en el problema, que cada una de ellas explica (Pérez C., 2000).

La primer componente es por lo tanto, la combinación de máxima varianza; la segunda es otra combinación de variables originarias que obedece a la restricción de ser ortogonal a la primera y de máxima varianza, la tercer componente es aún otra combinación de máxima varianza, con la propiedad de ser ortogonal a las dos primeras; y así sucesivamente.

Por sus propiedades de ortogonalidad, las sucesivas componentes después de la primera se pueden interpretar como las combinaciones lineales de las variables originarias que mayor varianza residual explican, después que el efecto de las precedentes ha sido ya removido y así sucesivamente hasta que el total de varianza ha sido explicado.

(<https://www.dspace.espol.edu.ec/retrieve/94469/D-71989.pdf>)

2.2.2 Generación de las Componentes Principales desde los datos

La reducción de datos mediante componentes principales esencialmente consiste en organizar los datos en una matriz \mathbf{X} de tamaño $n \times p$ donde p es número de variables en el estudio y n es el número de observaciones. Las variables se denotan como X_1, X_2, \dots, X_p , de modo que:

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

La matriz de varianzas-covarianzas muestral sería la matriz $S = (S_{jk})$ de tamaño $p \times p$ donde

$$S_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \quad j, k = 1, 2, \dots, p$$

La primera componente principal estaría en la dirección que capture la máxima variabilidad de los datos originales. Esto se consigue resolviendo la ecuación

$$(S - l_{(1)}I)\underline{a}_{(1)} = \underline{0}$$

Donde $l_{(1)}$ es el multiplicador de Lagrange. El valor de $l_{(1)}$ se escoge de modo que

$$|S - l_{(1)}I| = 0$$

Aquí $l_{(1)}$ es el máximo eigenvalor de la matriz de varianzas-covarianzas S , y $\underline{a}_{(1)}$ es su correspondiente eigenvector. El procedimiento para la obtención de las demás direcciones es directo (Dillon y Goldstain, 1984).

2.3 Desarrollo de una comprensión de la reducción de datos

Un entendimiento sólido de la reducción de la dimensión de un conjunto de datos significa desarrollar un modelo cognitivo que incluya varios elementos y sus conexiones, y usar este modelo para razonar acerca de la reducción de datos en diferentes contextos (Garfield y Ben-Zvi, 2005). En esta sección se describe un modelo epistemológico que busca capturar los elementos esenciales de los diferentes niveles de entendimiento de la reducción de datos. Este modelo está basado en la propuesta de Garfield y Ben-Zvi (2005) para un desarrollo profundo de la comprensión de la variabilidad, e incluye algunas de las ideas centrales del modelo de Van Hiele sobre los niveles de entendimiento en Geometría (Burger y Shaughnessy, 1986). El modelo de Van Hiele está formado por dos partes: una descriptiva que identifica diversas etapas de aprendizaje denominadas “niveles de razonamiento geométrico”, la otra está compuesta por directrices sobre cómo organizar el proceso de instrucción para ayudar a los estudiantes a alcanzar un nivel superior de razonamiento; estas se conocen como “fases de aprendizaje” (Jaime y Gutiérrez, 1990). Los 5 niveles de razonamiento del modelo van Hiele son: Reconocimiento, Análisis, Clasificación, Deducción Formal, Rigor, cada uno de ellos se identifica con un número del cero al cuatro, respectivamente

(Jaime y Gutiérrez, 1990). Las fases de aprendizaje del modelo son: Información, Orientación dirigida, Explicitación, Orientación libre, Integración, las cuales se denominan fase 1, fase 2, ..., fase 5, respectivamente (Jaime y Gutiérrez, 1990). Los descriptores del modelo de van Hiele se refieren al aprendizaje y la enseñanza de la geometría. En nuestro trabajo se generan descriptores análogos para diversos niveles de entendimiento de la Reducción de Datos por Componentes Principales que pueden desarrollar estudiantes de Ingeniería (Licenciatura) que han aprobado un curso universitario de Estadística. El modelo se centra en el Análisis descriptivo de las Componentes Principales y no al análisis inferencial, ya que la instrucción en el concepto de prueba de hipótesis y en el de intervalo de confiabilidad representa una investigación diferente en la enseñanza de la estadística.

Nuestro modelo epistemológico consiste en un conjunto de categorías que no son mutuamente excluyentes. Estas categorías las consideramos como las piedras de toque para la construcción de los niveles de pensamiento para la comprensión del concepto de reducción de datos por Componentes Principales (CPs), análogos a los niveles de entendimiento del Modelo de Van Hiele. Desde nuestra perspectiva, el que un estudiante complete los niveles antes mencionados será equivalente a decir que tiene una *sólida comprensión* de lo que es la reducción de datos por CPs; es decir, significa que además de conocer el concepto es capaz de relacionar a éste con sus conocimientos previos, que entiende cuáles son los contextos en los que se puede llevar a cabo la reducción de datos, así como las limitaciones del método. Que sea capaz de producir un nuevo conocimiento a través de descubrir relaciones entre las variables u objetos, resolver problemas reales, construir explicaciones, y sacar conclusiones. Una sólida comprensión de la reducción de datos incluye el conocimiento de las condiciones o restricciones para usar la reducción de datos, un conocimiento de sus elementos tales como preservación de la estructura, dirección óptima, dispersión y las relaciones entre éstas. Una comprensión sólida incluye el discernimiento de aquellos aspectos que son importantes, respecto de los que son accesorios o irrelevantes. Que el estudiante sea capaz de aplicar y adaptar el concepto a situaciones nuevas y diferentes en vez de seguir los procedimientos aprendidos en situaciones familiares.

Para efectos de adaptar la propuesta de Garfield y Ben-Zvi (2005) y las de Van Hiele en la construcción de nuestro modelo, primeramente describimos los postulados centrales del modelo de Van Hiele (1957).

- (1) Se pueden encontrar varios niveles de perfección en el razonamiento de los estudiantes de Matemáticas (Estadística).
- (2) Un estudiante solo podrá comprender realmente aquellas partes de las Matemáticas (Estadística) que el profesor le presente de manera adecuada a su nivel de razonamiento.
- (3) Si una relación matemática (concepto estadístico) no puede ser expresada en el nivel actual de razonamiento de los estudiantes, será necesario esperar a que éstos alcancen un nivel superior de razonamiento para presentársela.
- (4) No se puede enseñar a razonar a una persona de una determinada forma. Pero sí se le puede ayudar, mediante una enseñanza adecuada de las Matemáticas (Estadística), a que llegue a razonar de esa forma.

3. METODOLOGÍA

Para efectos de adaptar la propuesta de Garfield y Ben-Zvi (2005) y las de Van Hiele en la construcción de nuestra secuencia de aprendizaje, primeramente describimos los postulados centrales del modelo de Van Hiele:

1. Se pueden encontrar varios niveles de perfección en el razonamiento de los estudiantes de Matemáticas (Estadística).
2. Un estudiante solo podrá comprender realmente aquellas partes de las Matemáticas (Estadística) que el profesor le presente de manera adecuada a su nivel de razonamiento.
3. Si una relación matemática (concepto estadístico) no puede ser expresada en el nivel actual de razonamiento de los estudiantes, será necesario esperar a que éstos alcancen un nivel superior de razonamiento para presentársela.
4. No se puede enseñar a razonar a una persona de una determinada forma. Pero sí se le puede ayudar, mediante una enseñanza adecuada de las Matemáticas (Estadística), a que llegue a razonar de esa forma.

Con base en los postulados anteriores, se determinan las características de cada uno de los niveles de comprensión de la reducción de datos:

Nivel 1. Desarrollo intuitivo de la Reducción de Datos

- El estudiante es capaz de reconocer que “muchas” variables no son fáciles de manejar y de que no es fácil darles sentido o interpretarlas globalmente.
- El estudiante distingue dentro de un conjunto de variables aquellas que están correlacionadas de las que no lo están.
- El estudiante puede explicar por qué un conjunto de variables están correlacionadas y puede determinar diversos niveles de correlación.
- El estudiante identifica a la reducción de datos como un procedimiento global respecto de un conjunto de variables.

Nivel 2. Implementación de la Reducción de Datos

- El estudiante puede utilizar una matriz de gráficas bivariadas para identificar posibles asociaciones entre dos variables.
- El estudiante cuenta con habilidad para interpretar la información de una matriz de varianzas y covarianzas.

Nivel 3. Interpretación de los resultados de la Reducción de Datos

- El estudiantes puede utilizar tanto las gráficas como la matriz de varianzas y covarianzas para detectar posibles redundancias en las variables.
- El estudiante comprende que el análisis de la matriz de varianzas y covarianzas es útil para construir grupos de variables asociadas, y puede formar grupos de variables.
- Al examinar cuidadosamente éstos grupos de variables los estudiantes pueden extraer información sobre el comportamiento del fenómeno del cual provienen los datos.

Nivel 4. Comprensión global de la Reducción de Datos y el ajuste de modelos

- El estudiante es capaz de construir diagramas de dispersión e interpretar la distancia entre los puntos en un sistema de ejes definido por dos variables.
- El estudiante puede explicar por qué la reducción de datos preserva la estructura de los puntos originales en un espacio de dimensión menor (una recta) en el sentido de que si dos puntos originales están distantes (o cercanos), que también lo estén en la dimensión menor.
- El estudiante sabe elegir entre las diferentes soluciones en función de mayor o menor nivel de preservación de las características de la estructura original de los datos, al medir cuál es la mejor dirección que preserva tal estructura.
- El estudiante posee habilidad para transformar y reorganizar los datos originales y definir nuevos constructos (grupos de variables), los cuales interpreta en términos del fenómeno de referencia.

- El estudiante puede justificar que al ser ortogonales, las componentes principales permiten un mejor ajuste de alguna variable explicada por las variables originales.

Nivel 5. Valoración de la Reducción de Datos como parte del pensamiento estadístico

- El estudiante examina y discute la posible redundancia de algunas variables que forman parte de un conjunto de datos.
- El estudiante es capaz de explicar una variable a través de un conjunto de otras variables relacionadas con la primera.
- El estudiante puede explicar el agrupamiento de variables a través de constructos que subyacen a las mismas.

3.1 Criterios para el diseño de la secuencia de aprendizaje

Se basan en los postulados centrales del modelo de Van Hiele, se determinan las características de cada uno de los niveles de comprensión de la reducción de datos.

3.1.1 Actividades de la secuencia de aprendizaje

Se integraron 5 equipos en binas, seguidamente se desarrollaron 3 actividades con alumnos del cuarto semestre de la carrera de Ingeniería Ambiental de la División de Ciencias e Ingenierías de la Universidad de Quintana Roo, cada grupo de actividades se desarrollaron durante 2 horas en el aula con: Apertura de trabajo, desarrollo y evaluación de la secuencia de aprendizaje.

3.2 Secuencia de aprendizaje

La secuencia didáctica fue organizada a partir de 3 grupos de actividades.

El desarrollo de la primera actividad se fundamentó de la siguiente manera: los fundamentos para una epistemología de lo intuitivo en el desarrollo de la actividad creadora se encuentran en la teoría

del reflejo, como núcleo de la teoría del conocimiento, según la filosofía dialéctico materialista. A cada grupo de estudiantes (binas) se les propició que desarrollaran su nivel de conocimiento y comprensión en esta primera actividad.

3.2.1 Caso de estudio: Factores de riesgo de enfermedad coronaria, mediante el análisis de componentes principales (ACP)

Se quiere estudiar la situación de un grupo de pacientes hipertensos, con relación a los factores de riesgo de enfermedad coronaria en la ciudad de Chetumal, Quintana Roo. Así como, identificar cuáles son las variables que incrementan el riesgo de presentar una enfermedad coronaria y un posible infarto.

Para poder efectuar un análisis de componentes principales (ACP), con el fin de identificar cuáles son los factores que mejor explican los riesgos relacionados con la enfermedad coronaria, se obtuvo una lista de pacientes hipertensos de los hospitales y centros médicos de la ciudad.

Se elaboró una base de datos con los registros de 7 variables de interés disponibles y una explicación breve de cada parámetro.

Se realizó un análisis descriptivo de los datos (muestra) con especial énfasis en la varianza, dispersión de las variables y medidas de asociación, apoyado en una visualización de los datos.

Fuente de los datos: Grupo de pacientes hipertensos de diferentes hospitales de la ciudad de Chetumal, en el estado de Quintana Roo, de los cuales se seleccionaron 20 pacientes al azar.

Las variables son las siguientes:

Presión _ arterial (mm Hg)

Edad (años)

Peso (Kg)

Superficie corporal (m²)

Duración de la hipertensión (años)

Pulso (pulsaciones/minuto)

Medida del estrés

3.2.2 Objetivos de la secuencia didáctica:

Al finalizar las actividades de la secuencia de aprendizaje el estudiante será capaz de comprender e interpretar los resultados del Análisis de Componentes Principales.

3.2.3 Nivel de comprensión 1:

Primera sesión

3.2.4 Desarrollo intuitivo de la reducción de datos

3.2.4.1 Actividad 1.- Apertura

Objetivos de la actividad 1

- Introducir un problema de estudio (Batanero, 2001) que aporte datos y permita desarrollar la secuencia.
- Recordar conceptos básicos sobre las medidas: media, varianza, covarianza y correlación.

Introducción al problema de estudio.

Se quiere estudiar la situación de un grupo de pacientes hipertensos, con relación a las variables de riesgo de enfermedad coronaria en la ciudad de Chetumal, Quintana Roo. Así como, identificar cuáles son las variables que incrementan el riesgo de presentar una enfermedad coronaria y un posible infarto.

Con el fin de identificar cuáles son las variables que mejor explican los riesgos relacionados con la enfermedad coronaria, se obtuvo una lista de pacientes hipertensos de los hospitales y centros

médicos de la ciudad de Chetumal, Quintana Roo. Del mismo modo se identificaron con un panel de expertos que variables pueden aportar la información requerida.

Se identificaron siete (7) variables numéricas de interés como a continuación se detalla:

Presión_ arterial. - es la presión que ejerce la sangre contra la pared de las arterias. Esta presión es imprescindible para que circule la sangre por los vasos sanguíneos y aporte el oxígeno y los nutrientes a todos los órganos del cuerpo para que puedan funcionar correctamente. Se mide en mm de Hg.

Edad. - Tiempo que ha vivido una persona u otro ser vivo contando desde su nacimiento y se mide en años.

Peso. - El peso corporal es la masa del cuerpo en kilogramos. También se le llama masa corporal.

Superficie corporal. - En fisiología y medicina el área de superficie corporal (ASC) es la medida o cálculo de la superficie del cuerpo humano y se mide en metros cuadrados.

D. Hipertensión. - La Duración de la hipertensión arterial es una patología crónica que consiste en el aumento de la presión arterial y se mide en años.

Pulso. - El pulso de una persona, es la pulsación provocada por la expansión de sus arterias como consecuencia de la circulación de sangre bombeada por el corazón y se mide en pulsaciones por minuto.

Stress. - Es una reacción fisiológica del organismo en el que entran en juego diversos mecanismos de defensa para afrontar una situación que se percibe como amenazante o de demanda incrementada y se considera adimensional.

Instrucción.

I.- Observa los datos de la Tabla 1, sobre medidas clínicas de una muestra de pacientes que te proporcionó el profesor.

¿Cuántas son las variables estudiadas y qué significado pueden darle a cada una de ellas?

¿Qué valores crees que pueda tomar cada dato?

¿Cuántos individuos están en la muestra?

¿Con qué objetivo creen que se midieron estos datos?

¿Qué porcentaje de los pacientes tiene más de 100 kg de peso?

Intuitivamente, ¿crees que es más difícil manejar y analizar los datos con menos o con más variables?

II.- Observe la Figura 1.

¿Cuántos pacientes tienen una presión arterial mayor de 115 mm de Hg?

¿Cuántos pacientes tienen edad entre 40 y 60 años?

Si se considera que una persona sana debe tener la presión arterial entre 90 y 110 mm de Hg ¿Aproximadamente cuántos pacientes cumplen con este nivel?

Realizar una estimación de la media de la presión arterial de los pacientes y ubícala en el histograma correspondiente.

Comparar la media estimada de la presión arterial de la pregunta anterior con la media calculada de la Tabla 2. ¿Qué observaste?

3.2.4.2 Desarrollo del nivel de comprensión intuitivo de la reducción de datos

I.- Observa la media, el valor mínimo y máximo de cada variable en la Tabla 2 y ubíquelo en los Histogramas de la Figura 1.

¿Cuál es la variable que tiene la media más alta y cual la más baja? ¿Crees que es correcto comparar las medias entre las variables de distinta unidad de medida? ¿Por qué?

En la Tabla 2, observa la edad, presión arterial y peso.

¿Qué variables tienen mayor dispersión?

¿Cual tiene varianza más baja y más alta?

Observa la siguiente formula de la varianza muestral y la Figura 2.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \hat{x})^2}{n - 1}$$

¿Están de acuerdo que la varianza muestral es el promedio de la diferencia al cuadrado de cada dato respecto de su media? ¿Por qué?

Señala en la Figura 2, la diferencia entre la edad del paciente número 10 y la media de edad
¿Qué valor obtuviste?

¿Estás de acuerdo que la varianza puede considerarse como el promedio de los errores al cuadrado respecto de la media? ¿Por qué?

Con las discusiones anteriores, responder lo siguiente:

Intuitivamente, ¿crees que es más difícil interpretar y comparar, las medias y varianza conforme aumenta el número de variables? ¿Por qué? ¿Cómo influye la unidad de medida de cada dato?

3.2.4.3 Evaluación.

Nuevamente, observa los datos de la Tabla 1.

¿Con qué objetivo se midieron estos datos de cada paciente?

¿Expresa el porcentaje de los pacientes que tiene una presión arterial de 110 mm de Hg?

Observe la Figura 1 y apóyese de la Tabla 1.

¿Cuántos pacientes tienen edad entre 40 y 50 años con una presión arterial entre 120 y 125 mm de Hg? ¿La información de la Figura 1 fue adecuada para poder resolver esta pregunta? ¿Por qué?

Realizar una estimación de la media del peso de los pacientes y ubícala en el histograma de la Figura 1, compara la media estimada de peso con la media calculada de la Tabla 2. ¿Obtuviste el mismo valor? ¿Por qué?

Observa la media, el valor mínimo y máximo de cada variable en la Tabla 2 y ubíquelo en los Histogramas de la Figura 1.

En la Tabla 2, observa: la edad, presión arterial y peso.

¿Qué variables tienen mayor dispersión? ¿Por qué?

Intuitivamente, ¿crees que es más difícil interpretar y comparar, las medias y varianza conforme aumenta el número de variables? ¿Por qué? ¿Cómo influye la unidad de medida de cada dato?

3.2.5 Nivel de comprensión 2:

Segunda sesión

3.2.5.1 Implementación de la reducción de datos

3.2.5.2 Actividad 2.- Apertura.

2. a. Observar las gráficas de dispersión de la Figura 3a. y 3b.

El paciente que tiene presión arterial de 105 mm de Hg ¿Qué edad tiene? ¿Qué peso tiene?

La covarianza entre dos variables es una medida de dispersión entre ellas. De acuerdo a la Figura 3, intuitivamente, ¿Cuál tiene mayor dispersión para?: ¿Presión vs Edad o Presión vs Peso? ¿Por qué?

La correlación entre dos variables es una medida de asociación entre ellas, a mayor dispersión menor asociación y a menor dispersión mayor asociación (correlación). Intuitivamente, ¿cuál tiene mayor correlación para?: ¿Presión vs Edad o Presión vs Peso? ¿Por qué? Compruébalo en la Tabla 3.

3.2.5.3 Desarrollo.

Observar nuevamente la Figura 3.b.

A mayor peso del paciente ¿consideras que tuviera una mayor o menor presión?

Suponte que llega un paciente con un sobrepeso u obesidad, ¿qué esperarías, que tenga una presión alta o una presión baja? ¿Por qué?

Si un paciente tiene un peso de 86 Kg ¿Qué presión esperarías para este paciente? ¿Por qué?

Si un paciente tiene una presión de 112 mm de Hg ¿Qué peso supondrías que tenga? de acuerdo a la Figura 3.b.

Si sabes que una persona tiene una presión baja ¿Esperarías que tenga sobrepeso?

Observa que, a mayor correlación o asociación, una variable explica mejor a otra variable y alguna de ellas podría resultar redundante (es decir se podría omitir). Con esa idea, de acuerdo a la Figura 3, si tuvieras que descartar alguna variable para reducir el número de ellas ¿Cuál variable descartarías edad o peso? ¿Por qué?

Observa la Figura 4 y la Tabla 3, ¿Qué pareja de variables tienen mayor asociación (correlación)?

¿Qué pareja de variables tienen menor correlación?

Con ayuda de la Tabla 3, divide las siete variables en dos grupos, de tal manera que uno de los grupos tenga correlación alta entre ellas.

3.2.5.4 Evaluación.

Observar la gráfica de dispersión de la Figura 3.a y 3.b.

Los pacientes que tienen presión arterial de 110 mm de Hg ¿Qué edad tienen? ¿Qué peso tienen? Comprueba con la Tabla 1.

La correlación entre dos variables es una medida de asociación entre ellas, a mayor dispersión menor asociación y a menor dispersión mayor asociación (correlación). Intuitivamente, ¿cuál tiene mayor correlación peso vs Superficie corporal o Superficie corporal vs D. Hipertensión? ¿Por qué? Comprueba con la Tabla 3.

Observa la Figura 3.b. Si un paciente tiene un peso de 82 Kg ¿Qué presión esperarías para este paciente? ¿Por qué?

Si las siete variables las divides en dos grupos considerando las asociaciones entre ellas ¿Consideras que presión arterial y peso estarían en un mismo grupo? ¿Por qué?

3.2.6 Nivel de comprensión 3:

Tercera sesión

3.2.6.1 Comprensión global de la reducción de datos y el ajuste de modelos

3.2.6.2 Actividad 3.- Apertura.

3. a. Observa la Figura 5 relativo a una casa habitación.

¿Qué proyección te aporta mayor información sobre la distribución de las habitaciones?

¿Por qué?

¿Qué proyección te aporta menor información sobre la distribución de las habitaciones?

¿Por qué?

Observa las líneas azul, roja y morada en la Figura 6.a., 6.b. y 6.c. ¿En qué caso la línea pasa más cercano de la mayoría de los puntos? ¿Por qué?

Observa las líneas azul, roja y morada en la Figura 6.a., 6.b. y 6.c. Señala aquella figura en donde se tenga una menor distancia promedio de los datos con respecto a la línea ¿Por qué?

3.2.6.3 Desarrollo

Suponte que el Componente Principal de la Figura 7.c. es la proyección de cada dato en la línea como se muestra para dos datos (proyección perpendicular a la línea). Proyecta los datos restantes en la componente de la Figura 7.c. Observa que la Componente Principal 1 son los puntos proyectados sobre la línea 1 y es una combinación lineal $CP1 = a * presión_{arterial} + b * edad$. Observa que esta Componente Principal de la Figura 7.c. contiene menos diferencias o errores respecto de los datos que la posible Componente de las Figuras 7.a. y 7.b. compruébalo realizando mediciones con una regla.

Con base a lo anterior una Componente Principal es una proyección de datos sobre una línea o un plano que minimiza las distancias a los datos y por tanto explica mejor su

dispersión y aporta mayor información. Nota que en ejercicio anterior la Componente Principal es una nueva variable que representan las dos variables originales.

Observa la Figura 8. ¿Cuál de las dos Figuras 8.a y 8? b. la línea (Componente Principal) ¿tienen menor error o explica mejor los datos? ¿Por qué? ¿Cuál de las dos figuras tienen mayor correlación? ¿Existe una relación entre la correlación y el menor error en el Componente Principal?

Observa la Figura 9.a. donde se expresan las dos posibles componentes. La Componente 2 es perpendicular a la 1 y en este caso no hay reducción de variables ya que son dos variables originales y dos Componentes Principales. Solo existe una rotación de ejes explicando el 100% de la dispersión.

Observa un reporte de un paquete estadístico del análisis de Componentes Principales en la Figura 9.b. Nota que están reportadas las dos Componentes Principales y las dos variables originales, es similar a la Figura 9.a. pero rotada. La Figura 9.b. indica que el componente uno explica el 97.5 % de la varianza total, como consecuencia de una alta correlación entre la variable de presión arterial y peso corporal. Por tanto, una alta correlación entre dos variables indica que puede generarse o reducirse a una variable (CP1) con pequeñas pérdidas de información.

3.2.6.4 Evaluación.

Observa las líneas azul, roja y morada en la Figura 6.a., 6.b. y 6.c. ¿En qué caso la línea pasa más lejana de la mayoría de los puntos? ¿Por qué?

Observa la Figura 8. ¿Cuál de las dos Figuras 8.a y 8.b. la línea (Componente Principal) ¿tienen mayor error o explica menos los datos? ¿Por qué? ¿Cuál de las dos figuras tienen menor correlación? ¿Existe una relación entre la correlación y el menor error en el Componente Principal?

Observa la Figura 10, un reporte de un paquete estadístico del análisis de Componentes Principales. Se observan tres variables (Presión_ arterial, Peso y Superficie C.) en dos Componentes Principales. ¿Qué Componente explica la mejor varianza?

Si observamos las Figuras 9.b. y 10, ¿Cuál de los dos Componentes explica la mejor varianza? ¿Qué diferencia hay entre las varianzas explicadas? Con esta información arrojada ¿Crees que es más difícil manejar y analizar los datos con menos o con más variables? ¿Por qué? Apóyate de la Tabla 3. Correlaciones lineales de Pearson.

Si observas la Figura 11, en este ACP observamos dos Componentes Principales con las 7 variables ¿Cuál de los dos Componentes explica la mejor varianza? Al compararlo con los Componentes Principales de las Figuras 9.b y 10 ¿Qué sucede con sus varianzas? Con esta información podrías corroborar ¿Qué es más difícil manejar y analizar los datos con menos o con más variables? ¿Por qué? Apóyate de la Tabla 3. Correlaciones lineales de Pearson.

3.3 Evaluación del razonamiento de los estudiantes acerca de la reducción de datos por CP

La mayoría de las evaluaciones tradicionales respecto al conocimiento de la Reducción de Datos se centra en la definición de las CP y el uso de algún software estadístico para obtener las componentes, sin que los estudiantes lleven a cabo interpretaciones contextualizadas de los resultados ni de los procedimientos de reducción. Por otra parte, evaluaciones alternativas se centran en intuiciones, entendimientos conceptuales, y razonamientos acerca de la reducción de datos y su conexión con otros conceptos como variabilidad, asociación, base, dimensión, dirección, transformación, y rotación. Siguiendo nuevamente el trabajo de Garfield y Ben-Zvi (2005) en la siguiente lista se ejemplifican algunas preguntas que pueden ser de utilidad para determinar si un estudiante ha desarrollado un conocimiento profundo e intuición acerca de la Reducción de Datos, en particular de las Componentes Principales, con base en un problema específico que hemos implementado en nuestros cursos de matemáticas.

3.3.1 Evaluación de los resultados de la secuencia de aprendizaje

Después de aplicar la secuencia de aprendizaje, aplicamos los siguientes criterios para evaluar el conocimiento aprendido en las actividades propuestas.

1. Evaluación-Desarrollo intuitivo de la Reducción de Datos

- Reconocimiento de que “muchas” variables no son fáciles de manejar y de que no es fácil darles sentido globalmente: por ejemplo: si existen mediciones de los siguientes contaminantes en un cuerpo de agua, tales como: Mercurio, Cadmio, Cobre, Zinc, Plomo, Fosforo y Antimonio. Si determinamos diez mediciones de cada uno de esos contaminantes en diferentes puntos de muestreo y se le pide al estudiante que dé una descripción de la contaminación puntual, el alumno lo desarrollaría como se propone en las actividades precedentes.
- Cuestionar al alumno, por ejemplo. Un conjunto de variables podrían estar correlacionadas o no correlacionadas: Se sabe que la toma de estas mediciones es

extremadamente cara y difícil, se le pregunta al estudiante ¿Son necesarias las siete mediciones?, ¿Por qué?

- Intentar entender por qué las variables estarían correlacionadas: Se le pregunta al estudiante ¿Podrías agrupar a las variables? ¿Cómo lo harías? ¿Por qué?
- La reducción de datos es un procedimiento global respecto de un conjunto de variables. Nuestros datos están en una matriz 10×7 , se le pregunta al estudiante ¿Sería posible reducir esta matriz a una de $10 \times p$ con $p < 7$ con una pérdida mínima de información? ¿Cómo lo haría?

2. Evaluación-Implementación de la Reducción de Datos

- Una matriz de gráficas bivariadas nos muestran posibles asociaciones de variables dos a dos, ¿se pediría definir grupos de variables correlacionadas dentro de cada grupo y no correlacionadas entre grupos?
- La matriz de varianzas y covarianzas nos dirían más y diferentes cosas, respecto de las variables ¿Se pediría definir baja correlación y alta correlación y de aquí construir grupos de variables con información redundante?

3. Evaluación-Interpretación de los Resultados de la Reducción de Datos

- Tanto las gráficas como la matriz de varianzas y covarianzas nos guiarían hacia la detección de posibles redundancias en las variables: Se le preguntaría al estudiante ¿Con ésta matriz de gráficas bivariadas podrías anticipar la redundancia de alguna de las siete mediciones? ¿Podrías predecir alguna medición de otra o de otras mediciones?
- La matriz de varianzas y covarianzas también nos guiarían para la construcción de grupos de variables asociadas dentro de cada grupo: ¿Podrías anticipar el agrupamiento de variables? ¿Cómo lo harías?

- El examen cuidadoso de éstos grupos de variables nos conducirían a preguntarnos el por qué y cómo de cada uno de los grupos: Se le preguntaría al estudiante ¿Qué es lo que hace que se formen los grupos de variables? ¿Le podrías dar nombre y describirlo o explicarlo?

4. Evaluación-*La Comprensión Global de la Reducción de Datos y el ajuste de modelos*

Diagrama de dispersión y la comprensión de la distancia entre los puntos en un sistema de ejes definido por las dos variables: En uno de los grupos de variables ya definido consideremos a las dos o tres que estén altamente correlacionadas y le pedimos al estudiante que nos de la matriz de las distancias entre los puntos definidos por éstas variables.

- Preservar la estructura de los puntos originales en un espacio de dimensión menor (una recta) en el sentido de que si dos puntos originales están distantes (o cercanos), que también lo estén en la dimensión menor: Le solicitamos al estudiante que obtenga o defina rectas o planos que el considere que preserven la estructura de éstos datos.
- No hay una única solución, hay muchas rectas o planos que lo podrían hacer ¿Cuál preserva más la estructura original?, es decir, ¿Cuál es la mejor dirección o mejores direcciones que preserve tal estructura?: Le pediríamos al estudiante que nos proporcione un criterio para la obtención de la mejor solución. Cada conjunto de posibles soluciones definiría un conjunto de nuevas componentes de nuestros datos originales. La mejor solución definiría nuestro conjunto de componentes principales de nuestros datos originales, en el sentido de que éstas son las que mejor preservan la estructura original.
- La reducción de los datos nos permite transformar y reorganizar los datos originales, lo que a su vez nos permitirá definir nuevos constructos. Por construcción estas nuevas variables (Componentes Principales) serán independientes lo cual permitirá simplificar y entonces mejorar la interpretación de los datos: se solicitó al estudiante que obtenga los datos transformados, es decir, evaluaríamos cómo obtendría el estudiante los datos transformados y su procedimiento para validar los supuestos de los nuevos constructos.

- Los nuevos constructos al ser ortogonales nos permitirán un mejor ajuste de alguna variable explicada por las variables originales: Uno de los objetivos de la obtención de las mediciones de los siete contaminantes considerados es investigar si las concentraciones de Antimonio la podemos relacionar con las otras seis variables originales. Le pedimos al estudiante que exponga que problemas numéricos podrían haber al intentar ajustar un modelo de regresión múltiple y el por qué. Seguidamente le solicitamos que exponga cómo usar los nuevos constructos para el ajuste de un modelo de regresión múltiple y por qué en este nuevo contexto se resuelve el problema numérico. El estudiante contextualizó la solución en términos de las variables originales.

5. Evaluación-Valoración de la Reducción de Datos como parte del pensamiento estadístico

- En la estadística descriptiva, es necesario examinar y discutir la posible redundancia en las variables de nuestros datos.
- La producción de datos es también diseñada pensando en asociación entre variables; es decir, ésta producción es pensada buscando la posible explicación de una variable a través de otras.
- En el análisis estadístico tratamos de explicar posible agrupamientos de variables a través de constructos que subyacen a esas variables.
- Los tres puntos anteriores son parte del razonamiento estadístico, y deben de estar cuando exploremos datos multivariados y resolvamos problemas estadísticos.

4. RESULTADOS

En este capítulo se presenta el informe del desarrollo de la secuencia y el análisis de los resultados obtenidos en la implementación de la secuencia considerando los objetivos basados en los niveles de comprensión planteados.

4.1 Informe del desarrollo de la secuencia

Se aplicó la secuencia de aprendizaje a los alumnos del cuarto semestre de la carrera de Ingeniería Ambiental de la División de Ciencias e Ingeniería de la Universidad de Quintana Roo en el ciclo de primavera en el mes de febrero del 2017. Organizándose en 5 equipos de parejas (binas). Al inicio el profesor dio instrucción de la secuencia de aprendizaje por cada uno de los niveles de comprensión, los cuales son los siguientes:

4.1.1 Nivel de comprensión 1

Desarrollo intuitivo de la reducción de datos

En esta fase, como primera actividad se les presentó a los alumnos la introducción al caso de estudio (el estudio de un grupo de pacientes hipertensos, con relación a las variables de riesgo de enfermedad coronaria en la ciudad de Chetumal, Quintana Roo. Así como, identificar cuáles son las variables que incrementan el riesgo de presentar una enfermedad coronaria y un posible infarto).

Tabla 4. Nivel de comprensión uno.

<p>Actividad 1. Desarrollo intuitivo de la reducción de datos</p>	<p>Descripción de resultados</p>
<p style="text-align: center;">Apertura</p> <p>I. Conteste lo siguiente:</p> <p>Observa los datos de la Tabla 1, sobre medidas clínicas de una muestra de pacientes que te proporcionó el profesor.</p> <p>¿Cuántas son las variables estudiadas y qué significado pueden darle a cada una de ellas?</p> <p>¿Qué valores crees que pueda tomar cada dato?</p> <p>¿Cuántos individuos están en la muestra?</p> <p>¿Con qué objetivo creen que se midieron estos datos?</p> <p>¿Qué porcentaje de los pacientes tiene más de 100 kg de peso?</p> <p>Intuitivamente, ¿crees que es más difícil manejar y analizar los datos con menos o con más variables?</p> <p>II.- Observe la Figura 1.</p> <p>¿Cuántos pacientes tienen una presión arterial mayor de 115 mm de Hg?</p> <p>¿Cuántos pacientes tienen edad entre 40 y 60 años?</p> <p>Si se considera que una persona sana debe tener la presión arterial entre 90 y 110 mm de Hg</p> <p>¿Aproximadamente cuantos pacientes cumplen con este nivel?</p>	<p>Respondieron correctamente todos los estudiantes, diciendo que se trataba de 7 variables y que se estudiaba un caso poblacional de problemas de hipertensión en pacientes adultos.</p> <p>Solamente un equipo (bina) respondió el valor que puede tomar cada dato. Expresaron que cada dato tenía diferente magnitud y unidad de medida (kg, mm de Hg, %, pulsaciones por minuto, etc.).</p> <p>Todos los equipos respondieron correctamente, comentaron que eran 20 pacientes o individuos estudiados.</p> <p>Solamente 3 equipos respondieron correctamente cual era el objetivo de medición de los datos en la muestra. Conocer los riesgos de enfermedades coronarias o cardiovasculares.</p> <p>En la pregunta, donde se pide el porcentaje de los pacientes con peso más de 100 kg. Respondieron correctamente 2 equipos de 5 (3 binas contestaron incorrectamente).</p> <p>Al preguntar, ¿intuitivamente crees que es más difícil manejar y analizar los datos con menos o con más variables? 2 equipos respondieron correctamente, los otros 3 equipos comentaron que es más difícil manejar menos variables,</p>

<p>Realizar una estimación de la media de la presión arterial de los pacientes y ubícala en el histograma correspondiente.</p> <p>Comparar la media estimada de la presión arterial de la pregunta anterior con la media calculada de la Tabla 2. ¿Qué observaste?</p>	<p>aún no tenían los conocimientos necesarios para poder contestar el objetivo principal de esta secuencia de aprendizaje.</p>
<p style="text-align: center;">Desarrollo</p> <p>Desarrollo del nivel de comprensión intuitivo de la reducción de datos</p> <p>Observa la media, el valor mínimo y máximo de cada variable en la Tabla 2 y ubíquelos en los Histogramas de la Figura 1.</p> <p>¿Cuál es la variable que tiene la media más alta y cual la más baja? ¿Crees que es correcto comparar las medias entre las variables de distinta unidad de medida? ¿Por qué?</p> <p>En la Tabla 2, observa la edad, presión arterial y peso. ¿Qué variables tienen mayor dispersión?, ¿Cual tiene varianza más baja y más alta? Observa la siguiente fórmula de la varianza muestral y la Figura 2.</p> $s^2 = \frac{\sum_{i=1}^n (x_i - \hat{x})^2}{n - 1}$ <p>¿Están de acuerdo que la varianza muestral es el promedio de la diferencia al cuadrado de cada dato respecto de su media? ¿Por qué?</p> <p>Señala en la Figura 2, la diferencia entre la edad del paciente número 10 y la media de edad ¿Qué valor obtuviste?</p>	<p>En la sección 2 de preguntas relacionadas con la Figura 1 (Histogramas): los 5 equipos (binas) contestaron correctamente, sin embargo, para algunos alumnos fue difícil interpretar los histogramas, por lo que los alumnos tuvieron que apoyarse de las Tablas 1 y 2.</p> <p>En la tercera sección de la actividad 1, la mayoría de los equipos contestaron correctamente y supieron ubicar la media, el valor mínimo y máximo de cada variable de la Tabla 2 en los Histogramas de la Figura 1.</p>

<p>¿Estás de acuerdo que la varianza puede considerarse como el promedio de los errores al cuadrado respecto de la media? ¿Por qué?</p> <p>Con las discusiones anteriores, responder lo siguiente:</p> <p>Intuitivamente, ¿crees que es más difícil interpretar y comparar, las medias y varianza conforme aumenta el número de variables? ¿Por qué? ¿Cómo influye la unidad de medida de cada dato?</p>	
<p style="text-align: center;">Evaluación</p> <p>Nuevamente, observa los datos de la Tabla 1.</p> <p>¿Con qué objetivo se midieron estos datos de cada paciente?</p> <p>¿Expresa el porcentaje de los pacientes que tiene una presión arterial de 110 mm de Hg? Observe la Figura 1 y apóyese de la Tabla 1.</p> <p>¿Cuántos pacientes tienen edad entre 40 y 50 años con una presión arterial entre 120 y 125 mm de Hg? ¿La información de la Figura 1 fue adecuada para poder resolver esta pregunta? ¿Por qué?</p> <p>Realizar una estimación de la media del peso de los pacientes y ubícala en el histograma de la Figura 1, compara la media estimada de peso con la media calculada de la Tabla 2. ¿Obtuviste el mismo valor? ¿Por qué?</p> <p>Observa la media, el valor mínimo y máximo de cada variable en la Tabla 2 y ubíquelo en los Histogramas de la Figura 1.</p>	<p>Todos los equipos entendieron cuál fue la variable que tiene la media más alta y cual la media más baja.</p> <p>La mayoría de los equipos pudieron contestar que no es correcto comparar las medias entre las variables de distinta unidad de medida e identificaron que variables tienen mayor dispersión y el concepto de varianza.</p> <p>En general, en esta sección retroalimentaron los conceptos de varianza y estuvieron de acuerdo la mayoría de los estudiantes en que se puede considerar a la varianza como el promedio de los errores al cuadrado respecto de la media.</p> <p>También, en esta sección 3 de la primera activada, la mayoría de los alumnos ya pudieron interpretar intuitivamente que: es más difícil interpretar y comparar, las medias y varianzas conforme aumenta el número de variables y el valor de cada unidad de medida.</p>

<p>En la Tabla 2, observa: la edad, presión arterial y peso.</p> <p>¿Qué variables tienen mayor dispersión? ¿Por qué?</p> <p>Intuitivamente, ¿crees que es más difícil interpretar y comparar, las medias y varianza conforme aumenta el número de variables? ¿Por qué? ¿Cómo influye la unidad de medida de cada dato?</p>	
---	--

4.1.2 Nivel de comprensión 2

Implementación de la reducción de datos.

Tabla 5. Nivel de comprensión dos.

Actividad 1. Desarrollo intuitivo de la reducción de datos	Descripción de resultados
<p style="text-align: center;">Apertura</p> <p>2. a. Observar las gráficas de dispersión de la Figura 3a. y 3b.</p> <p>El paciente que tiene presión arterial de 105 mm de Hg ¿Qué edad tiene? ¿Qué peso tiene?</p> <p>La covarianza entre dos variables es una medida de dispersión entre ellas. De acuerdo a la Figura 3, intuitivamente, ¿Cuál tiene mayor dispersión para?: ¿Presión vs Edad o Presión vs Peso? ¿Por qué?</p> <p>La correlación entre dos variables es una medida de asociación entre ellas, a mayor dispersión menor asociación y a menor</p>	<p>La mayoría de los alumnos ya tenían los conocimientos de: el término de dispersión y correlación (asociación), por lo que pudieron proyectar en las gráficas de dispersión correctamente las variables (presión y peso) de cada paciente.</p> <p>La mayoría de los estudiantes, en este nivel de comprensión ya tenían conocimientos de los conceptos de varianza y covarianza (dispersión), y del concepto de dispersión (asociación) de las variables.</p>

<p>dispersión mayor asociación (correlación). Intuitivamente, ¿cuál tiene mayor correlación para?: ¿Presión vs Edad o Presión vs Peso? ¿Por qué? Compruébalo en la Tabla 3.</p>	
<p style="text-align: center;">Desarrollo</p> <p>Observar nuevamente la Figura 3.b.</p> <p>A mayor peso del paciente ¿consideras que tuviera una mayor o menor presión?</p> <p>Suponte que llega un paciente con un sobrepeso u obesidad, ¿qué esperarías, que tenga una presión alta o una presión baja? ¿Por qué?</p> <p>Si un paciente tiene un peso de 86 Kg ¿Qué presión esperarías para este paciente? ¿Por qué?</p> <p>Si un paciente tiene una presión de 112 mm de Hg ¿Qué peso supondrías que tenga? de acuerdo a la Figura 3.b.</p> <p>Si sabes que una persona tiene una presión baja ¿Esperarías que tenga sobrepeso?</p> <p>Observa que, a mayor correlación o asociación, una variable explica mejor a otra variable y alguna de ellas podría resultar redundante (es decir se podría omitir). Con esa idea, de acuerdo a la Figura 3, si tuvieras que descartar alguna variable para reducir el número de ellas ¿Cuál variable descartarías edad o peso? ¿Por qué?</p> <p>Observa la Figura 4 y la Tabla 3, ¿Qué pareja de variables tienen mayor asociación (correlación)?</p>	<p>Los alumnos comenzaron a relacionar las variables, algunos alumnos ya descartaban una variable porque veían que el comportamiento de algunas variables eran similares, Por ejemplo: algunos alumnos comentaban que si la persona presentaba sobrepeso tendría la presión arterial elevada.</p>

<p>¿Qué pareja de variables tienen menor correlación?</p> <p>Con ayuda de la Tabla 3, divide las siete variables en dos grupos, de tal manera que uno de los grupos tenga correlación alta entre ellas.</p>	
<p style="text-align: center;">Evaluación</p> <p>Observar la gráfica de dispersión de la Figura 3.a y 3.b.</p> <p>Los pacientes que tienen presión arterial de 110 mm de Hg ¿Qué edad tienen? ¿Qué peso tienen? Comprueba con la Tabla 1.</p> <p>La correlación entre dos variables es una medida de asociación entre ellas, a mayor dispersión menor asociación y a menor dispersión mayor asociación (correlación). Intuitivamente, ¿cuál tiene mayor correlación peso vs Superficie corporal o Superficie corporal vs D. Hipertensión? ¿Por qué? Comprueba con la Tabla 3.</p> <p>Observa la Figura 3.b. Si un paciente tiene un peso de 82 Kg ¿Qué presión esperarías para este paciente? ¿Por qué?</p> <p>Si las siete variables las divides en dos grupos considerando las asociaciones entre ellas ¿Consideras que presión arterial y peso estarían en un mismo grupo? ¿Por qué?</p>	<p>La mayoría de los alumnos ya tenían los conocimientos de correlación, dispersión, varianza, media, etc. Ya relacionaban las variables cuando estaban correlacionadas y también sabían que variables no tenían buena correlación.</p>

4.1.3 Nivel de comprensión 3:

Comprensión global de la reducción de datos y el ajuste de modelos.

En esta fase, como tercera actividad se estimó la comprensión global de la reducción de los datos.

Tabla 6. Nivel de comprensión tres.

Actividad 1. Desarrollo intuitivo de la reducción de datos	Descripción de resultados
<p style="text-align: center;">Apertura</p> <p>Observa la Figura 5 relativo a una casa habitación.</p> <p>¿Qué proyección te aporta mayor información sobre la distribución de las habitaciones? ¿Por qué?</p> <p>¿Qué proyección te aporta menor información sobre la distribución de las habitaciones? ¿Por qué?</p> <p>Observa las líneas azul, roja y morada en la Figura 6.a., 6.b. y 6.c. ¿En qué caso la línea pasa más cercano de la mayoría de los puntos? ¿Por qué?</p> <p>Observa las líneas azul, roja y morada en la Figura 6.a., 6.b. y 6.c. Señala aquella figura en donde se tenga una menor distancia promedio de los datos con respecto a la línea ¿Por qué?</p>	<p>La mayoría de los alumnos ya tenían los conocimientos de: el término de dispersión y correlación (asociación), por lo que pudieron proyectar en las gráficas de dispersión correctamente las variables (presión y peso) de cada paciente.</p> <p>La mayoría de los estudiantes, en este nivel de comprensión ya tenían conocimientos de los conceptos de varianza y covarianza (dispersión), y del concepto de dispersión (asociación) de las variables.</p>
<p style="text-align: center;">Desarrollo</p> <p>Suponte que el Componente Principal de la Figura 7.c. es la proyección de cada dato en la línea como se muestra para dos datos (proyección perpendicular a la línea). Proyecta</p>	<p>En el desarrollo de esta actividad, generalmente todos los alumnos ya conocían que era un componente principal, supieron proyectar los datos (puntos) en la recta (componente principal).</p>

<p>los datos restantes en la componente de la Figura 7.c. Observa que la Componente Principal 1 son los puntos proyectados sobre la línea 1 y es una combinación lineal $CP1 = a \cdot \text{presión_arterial} + b \cdot \text{edad}$. Observa que esta Componente Principal de la Figura 7.c. contiene menos diferencias o errores respecto de los datos que la posible Componente de las Figuras 7.a. y 7.b. compruébalo realizando mediciones con una regla.</p> <p>Con base a lo anterior una Componente Principal es una proyección de datos sobre una línea o un plano que minimiza las distancias a los datos y por tanto explica mejor su dispersión y aporta mayor información. Nota que en ejercicio anterior la Componente Principal es una nueva variable que representan las dos variables originales.</p> <p>Observa la Figura 8. ¿Cuál de las dos Figuras 8.a y 8? b. la línea (Componente Principal) ¿tienen menor error o explica mejor los datos? ¿Por qué? ¿Cuál de las dos figuras tienen mayor correlación? ¿Existe una relación entre la correlación y el menor error en el Componente Principal?</p> <p>Observa la Figura 9.a. donde se expresan las dos posibles componentes. La Componente 2 es perpendicular a la 1 y en este caso no hay reducción de variables ya que son dos variables originales y dos Componentes Principales.</p>	<p>Al observar las Figuras que ilustraban los componentes principales identificaron la línea con mayor y menor error, también las Figuras que presentaban mayor correlación, además, relacionaron la correlación con el error en el Componente Principal.</p>
---	---

<p>Solo existe una rotación de ejes explicando el 100% de la dispersión.</p> <p>Observa un reporte de un paquete estadístico del análisis de Componentes Principales en la Figura 9.b. Nota que están reportadas las dos Componentes Principales y las dos variables originales, es similar a la Figura 9.a. pero rotada. La Figura 9.b. indica que el componente uno explica el 97.5 % de la varianza total, como consecuencia de una alta correlación entre la variable de presión arterial y peso corporal. Por tanto, una alta correlación entre dos variables indica que puede generarse o reducirse a una variable (CP1) con pequeñas pérdidas de información.</p>	
<p style="text-align: center;">Evaluación</p> <p>Observa las líneas azul, roja y morada en la Figura 6.a., 6.b. y 6.c. ¿En qué caso la línea pasa más lejana de la mayoría de los puntos? ¿Por qué?</p> <p>Observa la Figura 8. ¿Cuál de las dos Figuras 8.a y 8?b. la línea (Componente Principal) ¿tienen mayor error o explica menos los datos? ¿Por qué? ¿Cuál de las dos figuras tienen menor correlación? ¿Existe una relación entre la correlación y el menor error en el Componente Principal?</p> <p>Observa la Figura 10, un reporte de un paquete estadístico del análisis de Componentes Principales. Se observan tres variables (Presión_ arterial, Peso y Superficie C.) en dos</p>	<p>La mayoría de los alumnos contestaron correctamente esta sección, se observó cómo fueron adquiriendo sus conocimientos, como los ordenaron y aplicaron los mismos. Supieron que proyección nos daba mayor información sobre el término de distribución (mejor dimensión) en un plano.</p> <p>La mayoría de los alumnos, reafirmaron los conocimientos aplicados en las actividades 1 y 2, reafirmaron el término de dispersión y correlación en un componente principal. También, proyectaron los puntos dispersos en el componente y midieron sus distancias y así conocer que componente presentaba mayor dispersión.</p>

<p>Componentes Principales. ¿Qué Componente explica la mejor varianza?</p> <p>Si observamos las Figuras 9.b. y 10, ¿Cuál de los dos Componentes explica la mejor varianza? ¿Qué diferencia hay entre las varianzas explicadas? Con esta información arrojada ¿Crees que es más difícil manejar y analizar los datos con menos o con más variables? ¿Por qué? Apóyate de la Tabla 3.</p> <p>Correlaciones lineales de Pearson.</p> <p>Si observas la Figura 11, en este ACP observamos dos Componentes Principales con las 7 variables ¿Cuál de los dos Componentes explica la mejor varianza? Al compararlo con los Componentes Principales de las Figuras 9.b y 10 ¿Qué sucede con sus varianzas? Con esta información podrías corroborar ¿Qué es más difícil manejar y analizar los datos con menos o con más variables? ¿Por qué? Apóyate de la Tabla 3. Correlaciones lineales de Pearson.</p>	<p>En este nivel de entendimiento, la mayoría de los alumnos identificaron que componente principal está mejor proyectado y quién explica el mejor porcentaje de la varianza total. Además, la mayoría de los estudiantes entendieron que es más fácil interpretar y comparar, las medias y varianzas conforme disminuye el número de variables y la reducción de los datos.</p>
--	--

Resumen general:

Con la información generada de este primer análisis de resultados, se tuvo un criterio para clasificar cada equipo de estudiantes en dos grupos: destacados y poco destacados.

En este sentido se identificaron los siguientes grupos:

Grupo I de alumnos destacados: 3 equipos y,

Grupo II de alumnos poco destacados: 2 equipos.

4.2 Resultados de las actividades del grupo I

El grupo I de estudiantes destacados

El grupo de alumnos en el primer nivel de comprensión, identificaron la media, varianza, covarianza y la correlación lineal que existía en las gráficas presentadas, además las variables que especificaban los metadatos y el valor de cada dato, por ejemplo, identificaron que las variables tenían diferentes unidades de medición y no podían ser comparadas una con la otra. Figura 12.

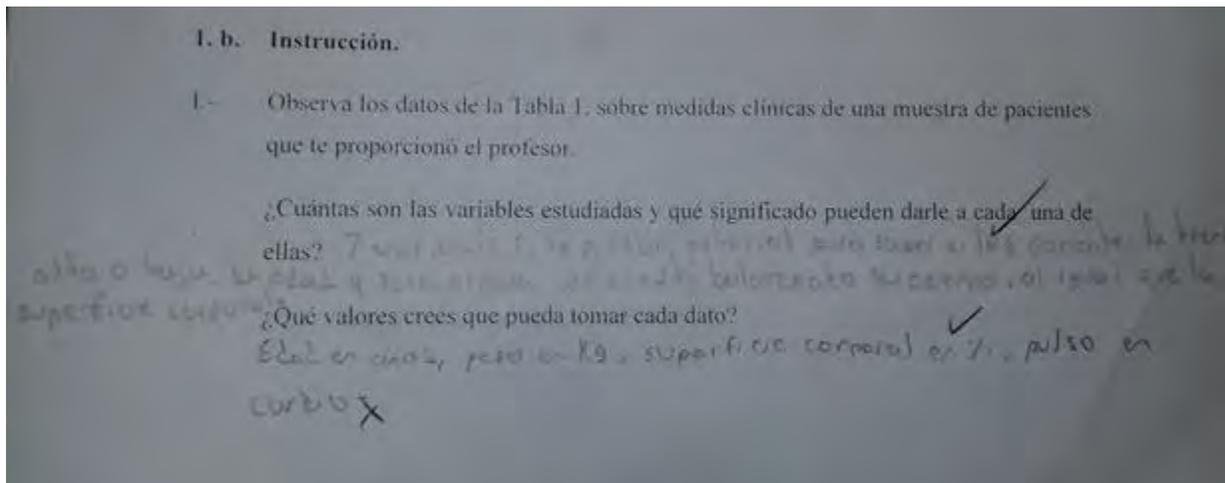


Figura 12. Respuestas del Grupo de estudiantes destacados primera actividad.

También, los alumnos de este grupo identificaron la cantidad de individuos que participaban en la muestra, conocieron el objetivo de la medición de los datos, cuando verificaron el porcentaje de pacientes con más de 100 kg supieron sacar el porcentaje lo que no hicieron los demás. Figura 13.

Sin embargo, aún no podían afirmar correctamente que se podía trabajar y analizar los datos con menos variables, esta pregunta fue incorrecta, estos alumnos destacados contestaron lo contrario, observar respuesta en la Figura 13.

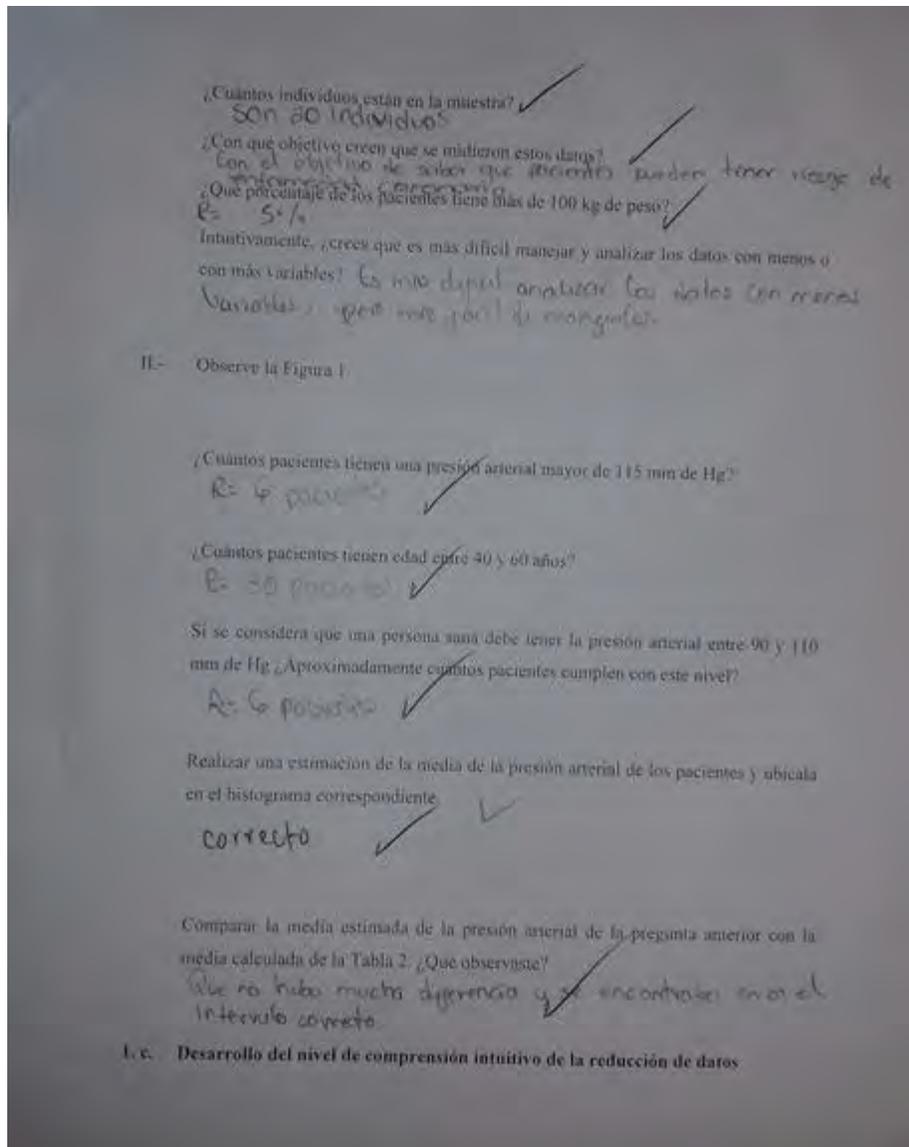


Figura 13. Respuestas del Grupo de estudiantes destacados primera actividad.

Cuando los alumnos de este grupo I, se les pidió trabajar con la Figura 1 (Histogramas) contestaron todo correctamente, lo cual podemos observar en la Figura 14, donde ubicaron correctamente los valores solicitados, tales como, número de pacientes con presión de 115 mm Hg encontrado en edades de 40 a 60 años (respondieron que todos).

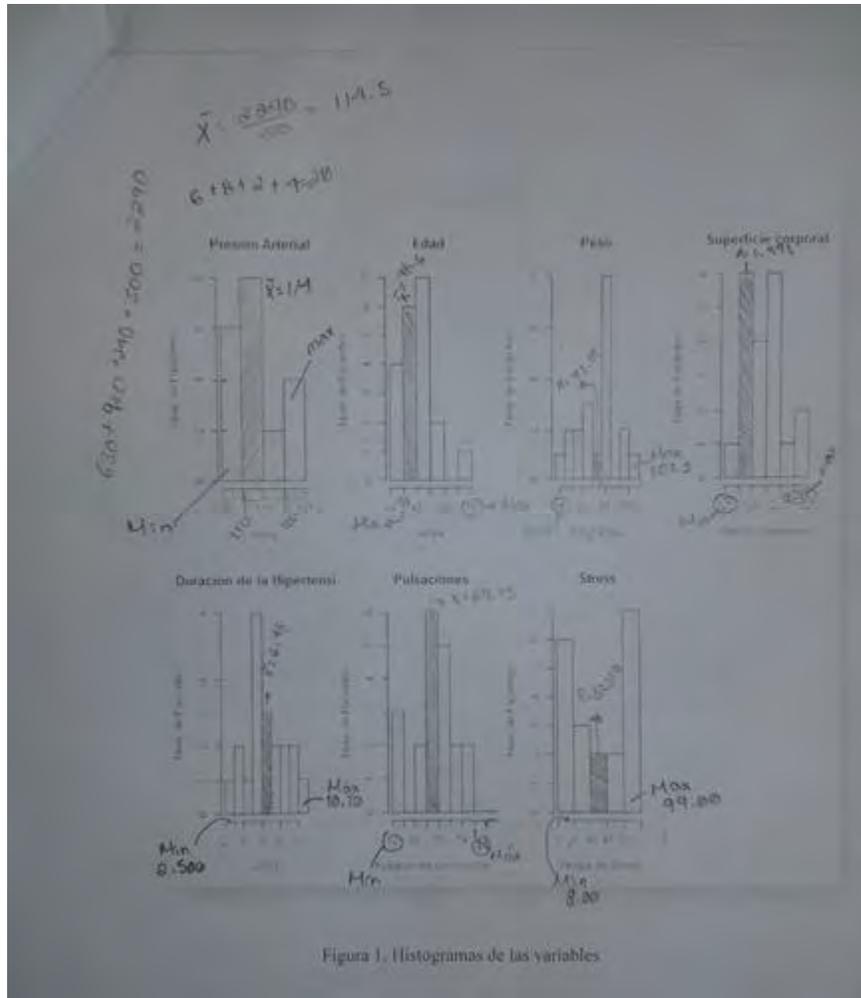


Figura 14. Respuestas del Grupo de estudiantes destacados primera actividad.

Así mismo, los alumnos de este grupo I, ubicaron correctamente la media, máximo y mínimo valor de cada variables (observe Figura 14), identificaron cuál de las variables tenía la media más alta y cuál de las variables tenía la media más baja. Ya tenían conocimiento de cual variable presentaba mayor o menor dispersión y cual variable presentaba mayor o menor varianza. Conocieron el concepto de varianza y lo relacionaron con la dispersión de los datos.

Cuando les preguntan ¿Estás de acuerdo que la varianza puede considerarse como el promedio de los errores al cuadrado respecto de la media? ¿Por qué? No todos los integrantes del grupo I respondieron correctamente, contestaron la mayoría que NO.

Sin embargo, el grupo I de estudiantes contestaron correcto, que es más fácil interpretar y comparar las medias y varianzas conforme disminuye el número de variables y que la unidad de medida de cada variable influye considerablemente. Figura 15.

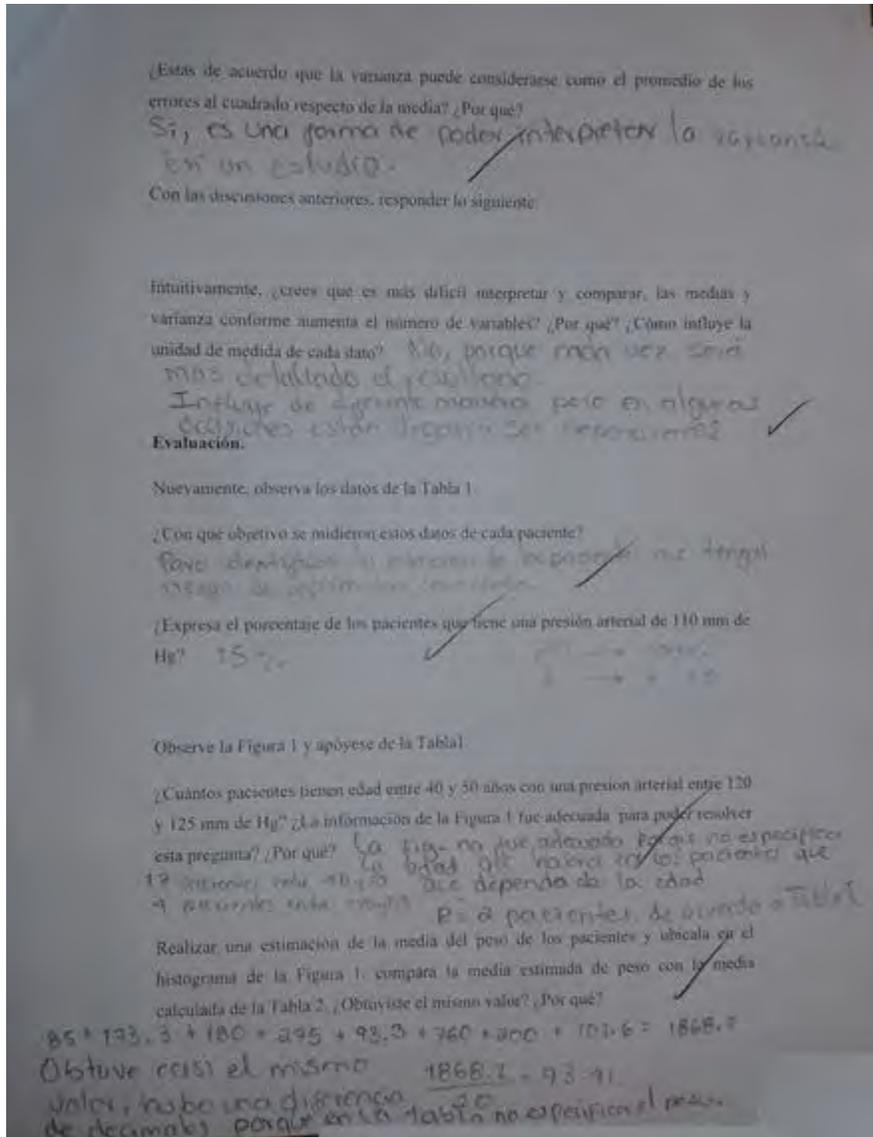


Figura 15. Respuestas del Grupo de estudiantes destacados primera actividad.

En la evaluación de esta primera actividad, los alumnos del grupo I contestaron correctamente e inclusive ya que el manejar menos variables podían interpretar mejor las medidas de varianza y media conforme disminuyen las variables en un estudio, ver Figura 16.

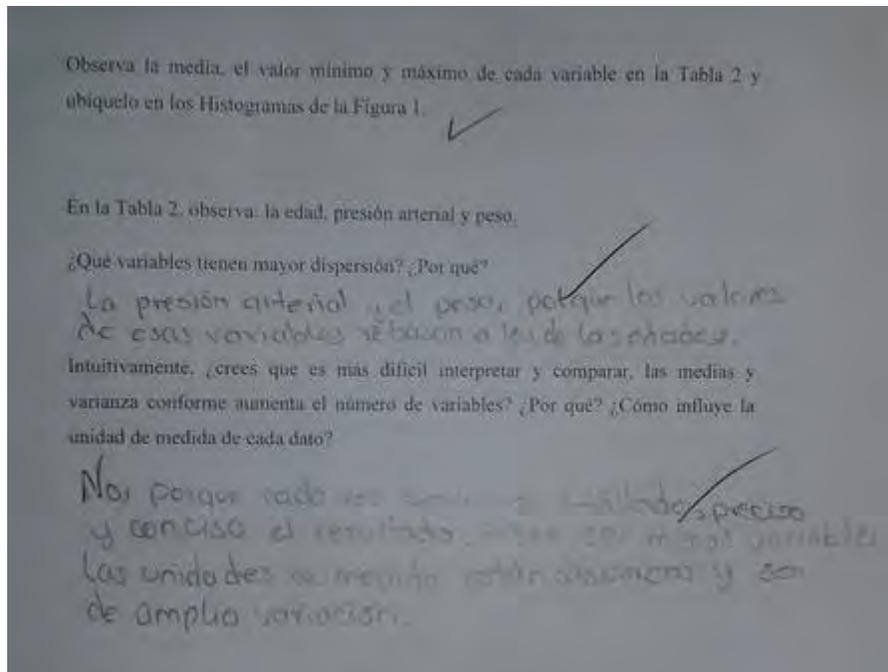


Figura 16. Respuestas del Grupo de estudiantes destacados primera actividad.

El grupo I de alumnos, en el segundo nivel de comprensión o actividad 2, se les preguntó: que observaran las gráficas de dispersión de las Figuras 3.a y 3.b. Se les pidió que encontraran el peso y la edad de acuerdo a la presión arterial (105 mm Hg), por lo eligieron la gráfica correcta (la que tenía menos dispersión) y mayor correlación, proyectaron el valor de la presión arterial y obtuvieron los valores de peso y edad correctamente, observaron que a mayor presión mayor sería su peso y viceversa. Ver Figura 17.

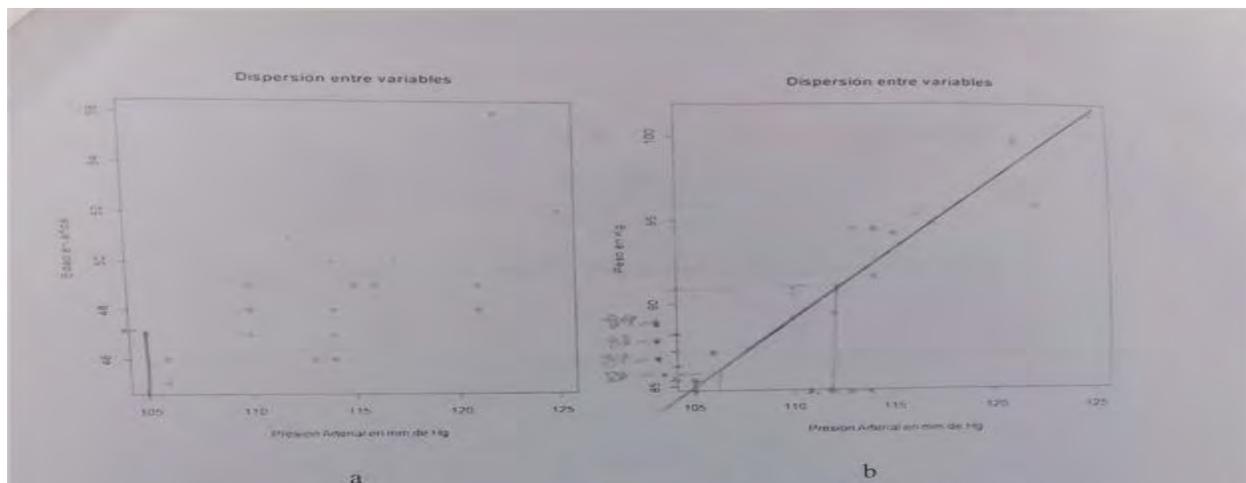


Figura 3. Gráfica de dispersión de Presión arterial con Peso y Edad.



Figura 4. Matriz de dispersión de las siete variables.

Figura 17. Respuestas del Grupo de estudiantes destacados segunda actividad.

El grupo I de estudiantes, algunos no pudieron descartar una variable de tres variables (pero una dependiente de la otra), algunos sí pudieron responder correctamente. En la Figura 17, el grupo I de estudiantes proyectaron los puntos dispersos en el componente correspondiente y supieron que había menos distancia intuitivamente de los puntos dispersos al componente en la Figura 3.b comparada con las Figuras 3.a y 3.b. Ver Figura 17.

Así mismo, se les pidió observar la Figura 4 (Ver Figura 17) y tabla 3, observaron cuales de las parejas de variables tenían mejor correlación (asociación) y cuales tenían menos correlación (más

dispersos los datos) y contestaron correctamente, agruparon las variables con mejor correlación lineal y las variables que no tenían buena correlación.

En la actividad 3, en este nivel de comprensión global de datos y el ajuste de modelos, los alumnos del grupo I ya tenían reafirmados sus conocimientos, se les pidió ver la Figura 5 relativo a una casa habitación y las respuestas a la pregunta fueron correctas, comentaron que la Figura 5.a. tenía mejor distribución de las habitaciones (mejor dimensión). Ver Figura 18.

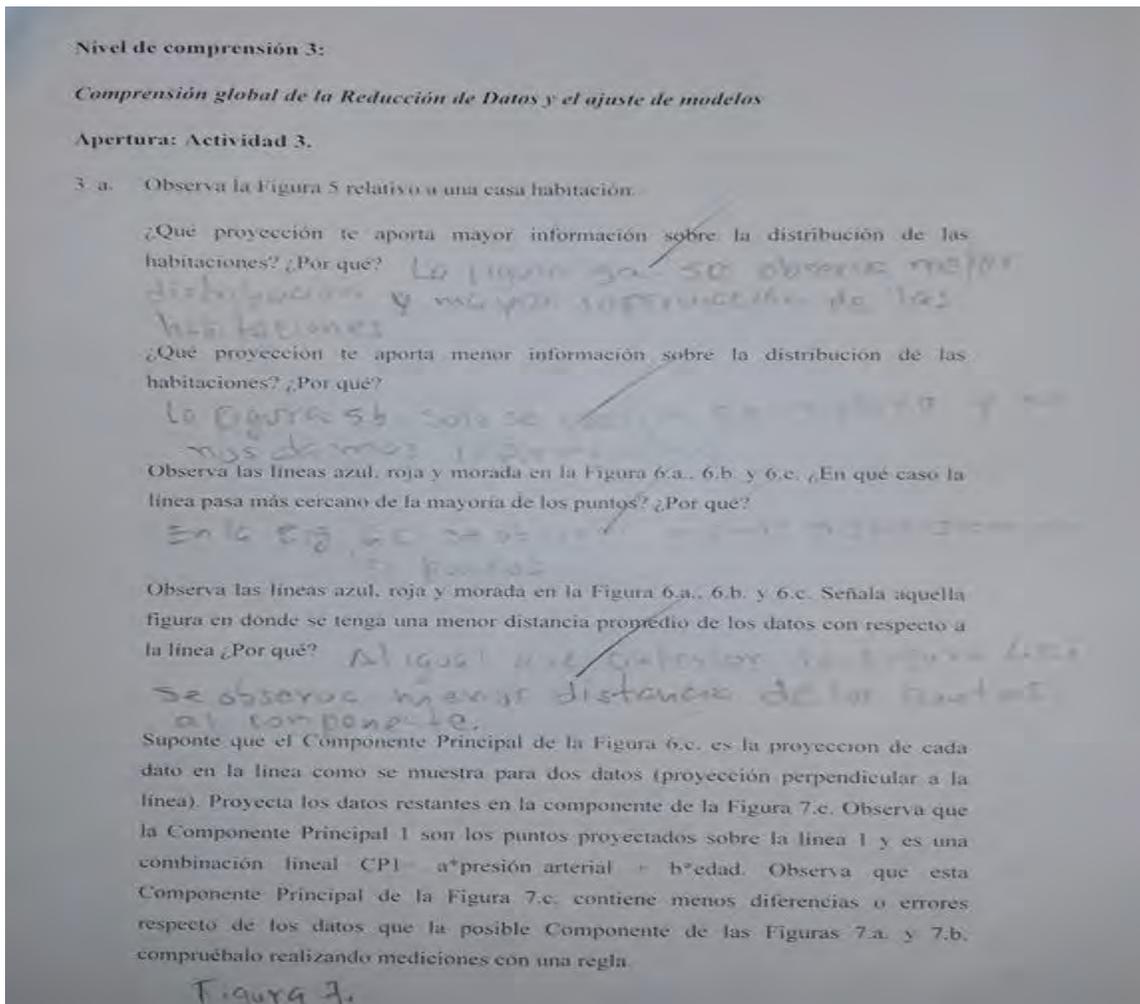


Figura 18. Respuestas del Grupo de estudiantes destacados tercera actividad

En la Figura 18, los alumnos del grupo I comentaron que las Figuras 7.a, 7.b y 7.c de (Figura 19), señalaron correctamente que la Figura 7.b se observaba menos dispersión de los datos y mejor

correlación entre los puntos o datos y viceversa se indicó que las Figuras 7.a y 7.b mostraban mayor dispersión de los datos.

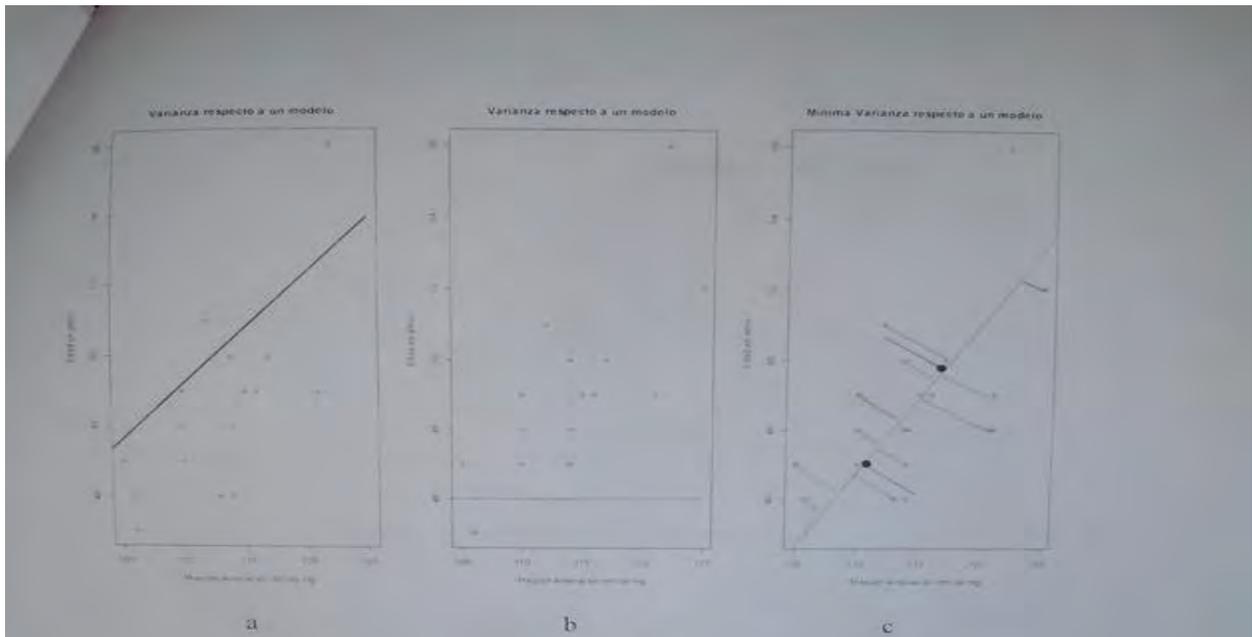


Figura 7. Diferentes proyecciones en el Componente Principal CP-I.

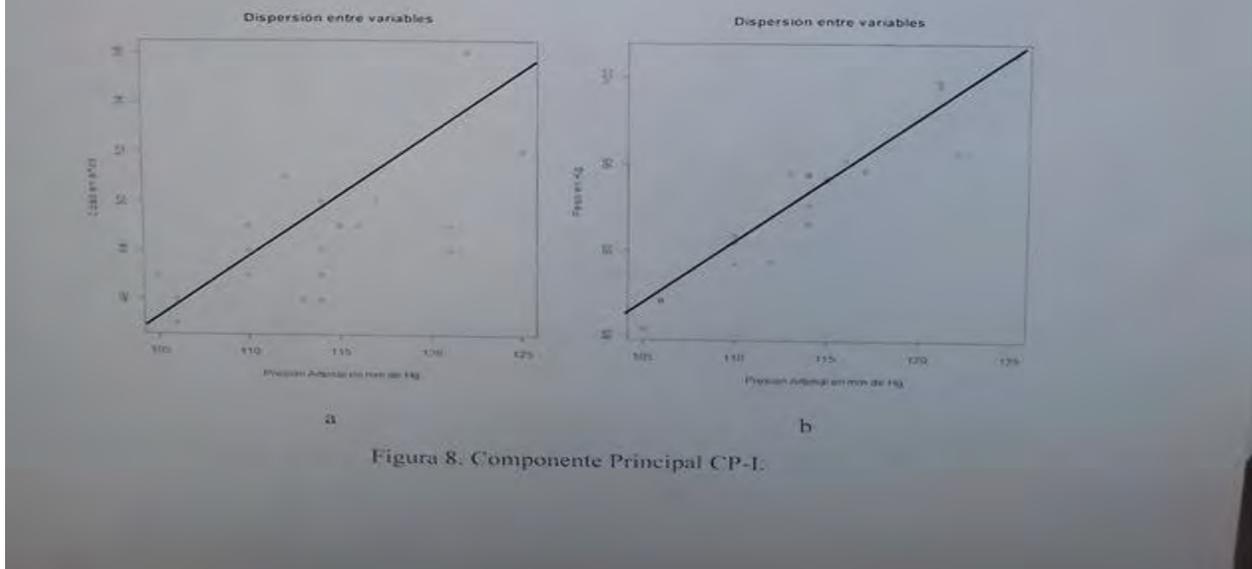


Figura 8. Componente Principal CP-I.

Figura 19. Respuestas del Grupo de estudiantes destacados tercera actividad

En la Figura 19. Los alumnos del grupo I comprobaron que la Figura 7.c la Componente Principal 1 son los puntos proyectados sobre la línea 1 y tienen menos distancia que los proyectados en las Figuras 7a. Y 7.b.

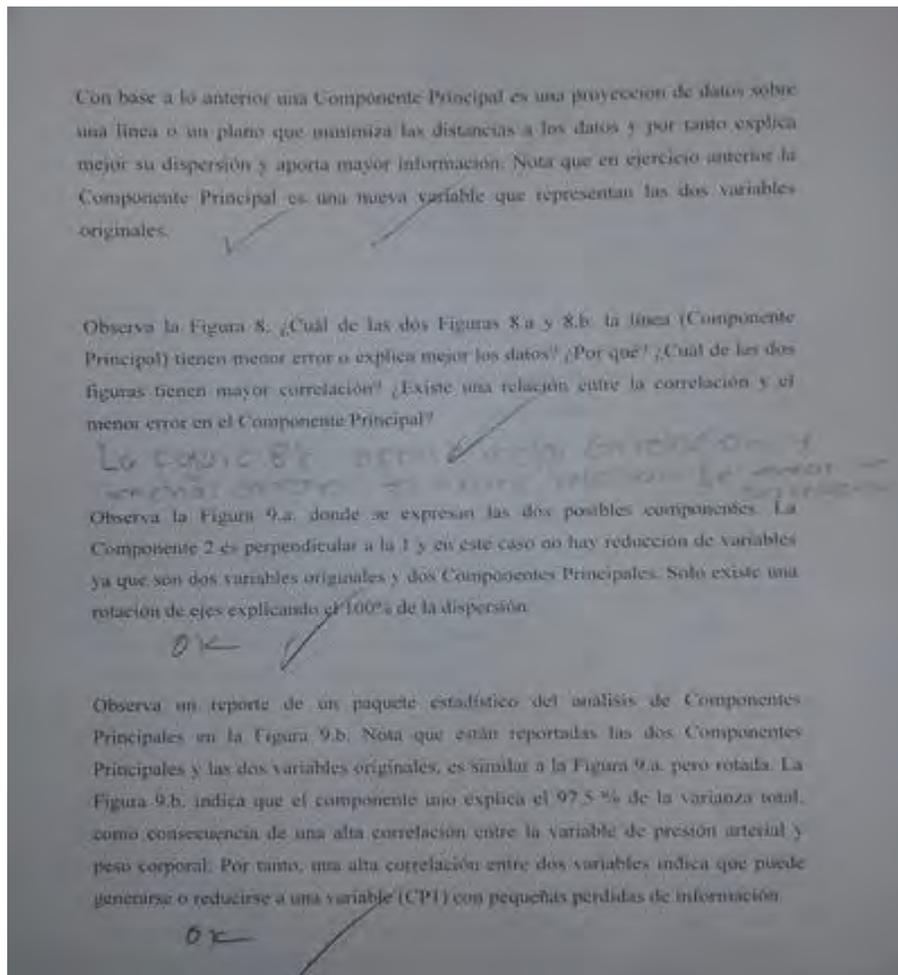


Figura 20. Respuestas del Grupo de estudiantes destacados tercera actividad

En la Figura 20. Los alumnos del grupo I demostraron que la Figura 8.b de (Figura 19) ofrecía mejor correlación lineal y menos error que la Figura 8.a. Además, observaron que en un reporte de un paquete estadístico del análisis de Componentes Principales en la Figura 9.b. estaban reportadas las dos Componentes Principales y las dos variables originales y que era similar a la Figura 9.a. pero rotada.

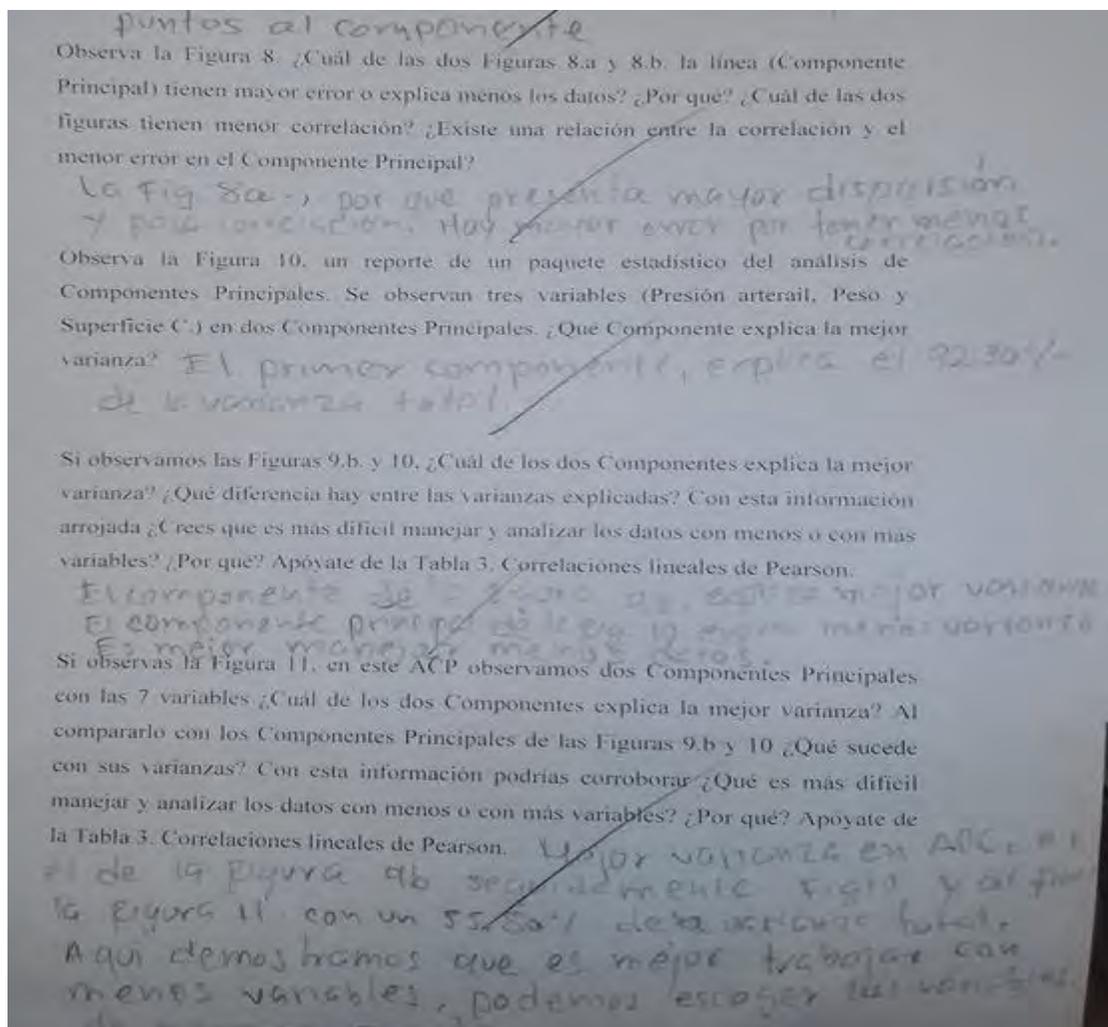


Figura 21. Respuestas del Grupo de estudiantes destacados tercera actividad.

Podemos observar la respuesta de los alumnos del grupo I en la Figura 21, al preguntar ¿Cuál de los dos Componentes explica la mejor varianza? ¿Qué diferencia hay entre las varianzas explicadas? Con esta información arrojada ¿Crees que es más difícil manejar y analizar los datos con menos o con más variables? ¿Por qué? Los alumnos respondieron que el componente de la figura 9.b explica mejor la varianza total en el primer componente comparado con el componente principal de la Figura 10.

Por último, los alumnos más destacados llegaron a la conclusión que la mejor varianza en el componente principal se explicaba en la Figura 9.b seguidamente el componente de la Figura 10 y por último el componente de la Figura 11 (las 7 variables), por lo que concluían que es mejor

trabajar con menos variables y poder escoger las variables que presenten mayor correlación entre ellas. Ver Figura 21.

4.3 Resultados de las actividades del grupo II

El grupo de alumnos en el **primer nivel de comprensión**, identificaron las variables que contenían los metadatos, pero no identificaron el valor de cada dato, por ejemplo, no identificaron que las variables tenían diferentes unidades de medición (kg, presión en mm de Hg, años, etc.), sí pudieron cuantificar cuantos individuos participan en la muestra, identificaron claramente el objetivo de la investigación, este grupo de estudiantes si comentó que es mejor trabajar con un análisis estadístico con menor número de variables el análisis es más fácil. Ver Figura 22.

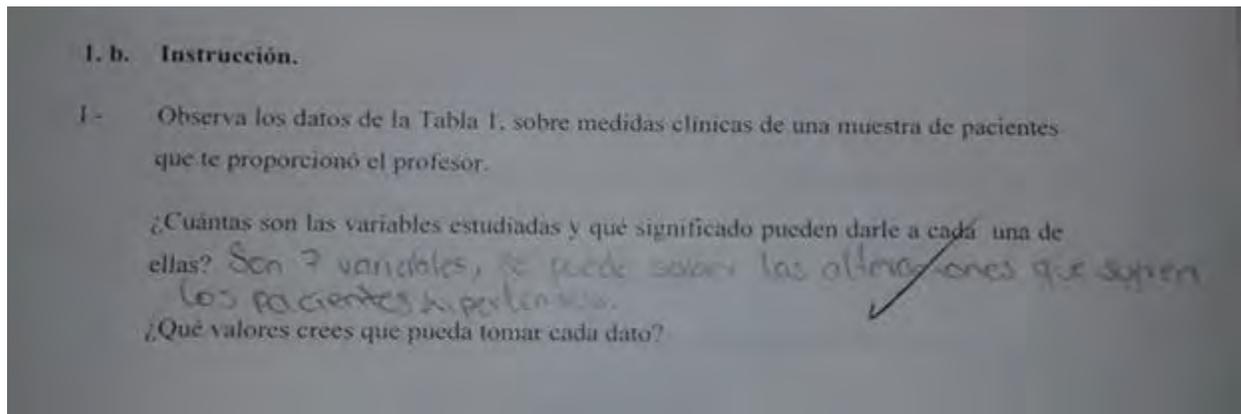


Figura 22. Respuestas del Grupo de estudiantes destacados primera actividad.

En la Figura 23. Los alumnos del grupo II contestaron algunas preguntas correctamente, por ejemplo: la cantidad de individuos que participaban en la muestra, el objetivo del estudio, el porcentaje de individuos con peso mayor de 100Kg. Sin embargo, no supieron utilizar los histogramas de la Figura 1, ni tampoco comprobar las medias, valores máximos y mínimos en los Histogramas de cada variable.

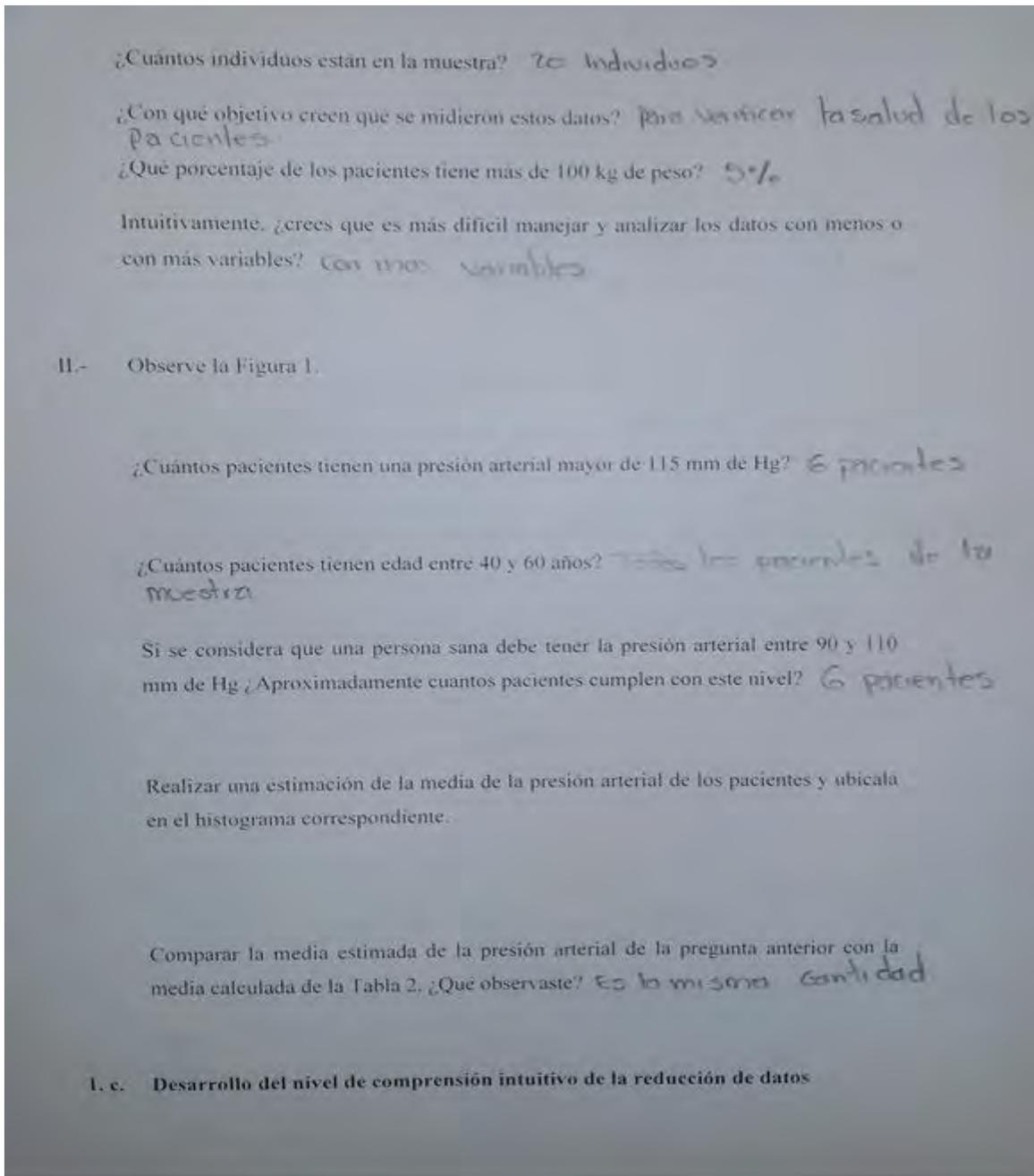


Figura 23. Respuestas del Grupo de estudiantes destacados primera actividad.

En la Figura 24. Podemos observar que este grupo de alumnos presentó dificultad al responder ¿Qué variables tienen mayor dispersión? Cuando observan la presión, edad y peso. No pudieron obtener el valor de la media, utilizando la fórmula de la varianza muestral.

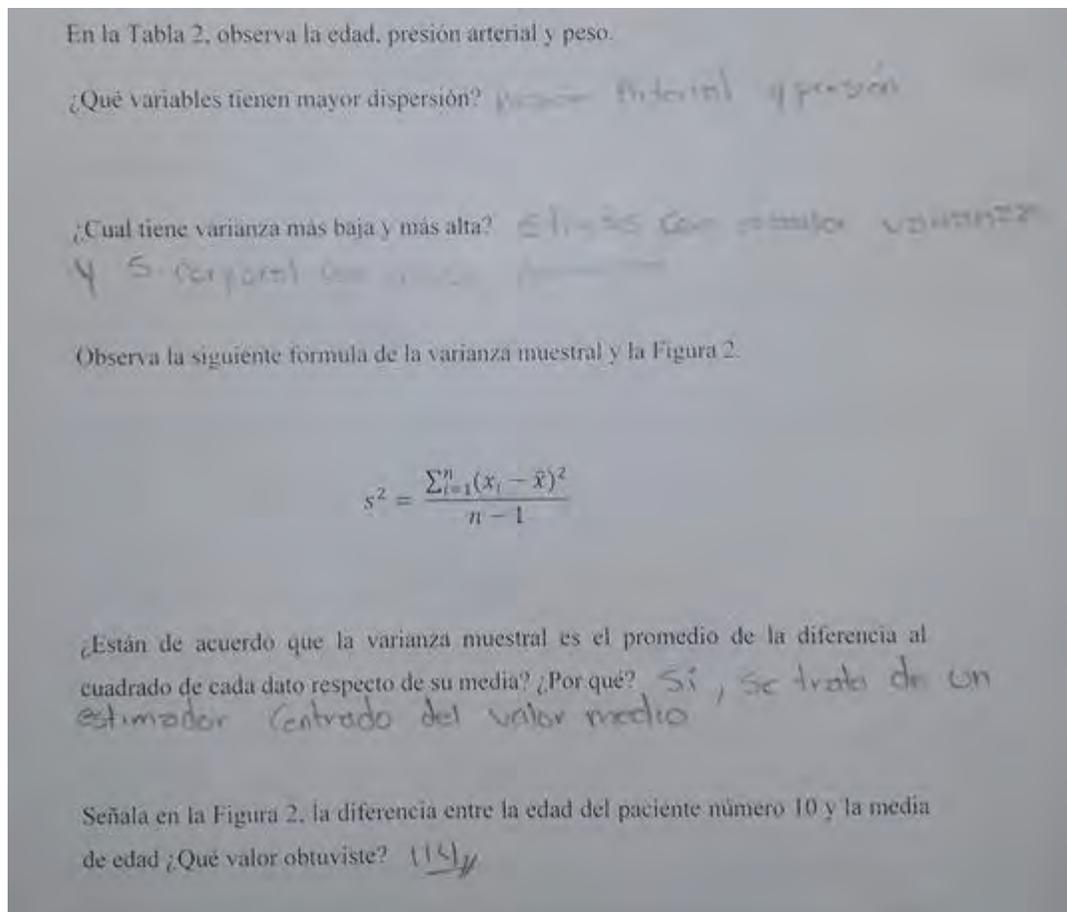


Figura 24. Respuestas del Grupo de estudiantes destacados primera actividad.

Los alumnos del grupo II, en la actividad uno, no desarrollaron conocimientos en los siguientes conceptos: varianza, dispersión, el promedio de los error al cuadrado respecto a la media que sea el término de varianza y no supieron expresar la pregunta ¿crees que es más difícil interpretar y comparar, las medias y varianza conforme aumenta el número de variables?

En la segunda actividad, segundo nivel de comprensión. El grupo II de estudiantes, contestaron correctamente toda esta sección, lo que significa que los alumnos de este nivel retroalimentaron los conceptos y supieron aplicarlos a la actividad 2. Ya sabían el concepto de dispersión (asociación), varianza (menor asociación), proyección de valores en la recta, relación de variables, etc. ver Figura 25.

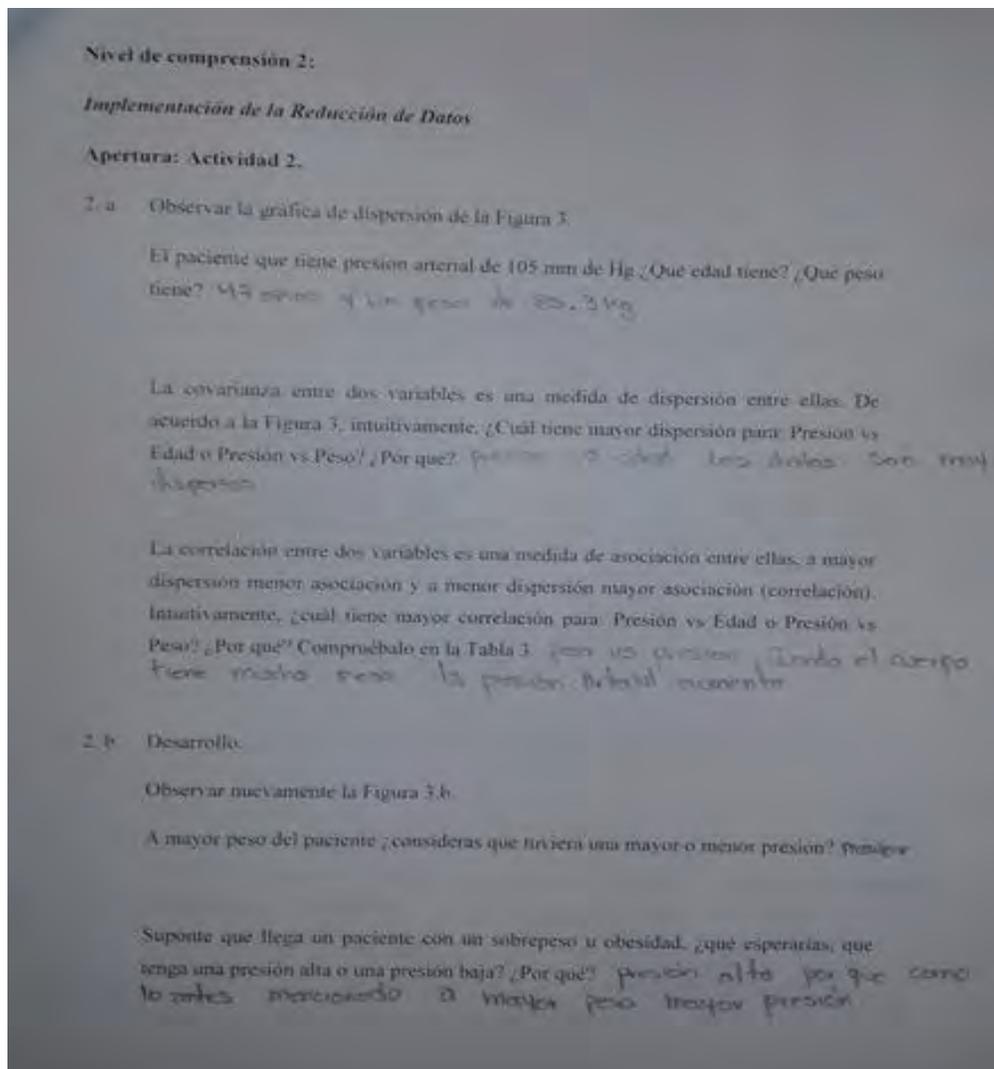


Figura 25. Respuestas del Grupo de estudiantes destacados segunda actividad.

En la tercera actividad, comprensión global de la reducción de datos y el ajuste de modelos.

Los alumnos de este grupo II, se les presento una serie de figuras de una casa habitación, se les preguntó que cual de las figuras aportaba mayor información sobre la distribución de las habitaciones y contestaron correctamente. Ver figura 26.

También, contestaron correcto cuando se les preguntó cuál de las Figuras 6.a, 6.b y 6.c presenta menos distancia de los puntos a la recta.

Los alumnos del grupo II, hicieron las proyecciones de los puntos hacia la recta en la Figura 7.c y las compararon con las otras Figuras 7.a y 7.b y observaron que las distancias de las proyecciones a la recta eran diferentes a la Figura 6.c.

Nivel de comprensión 3:

Comprensión global de la Reducción de Datos y el ajuste de modelos

Apertura: Actividad 3.

3. a. Observa la Figura 5, relativo a una casa habitación.

¿Qué proyección te aporta mayor información sobre la distribución de las habitaciones? ¿Por qué? La figura 6, ya que muestra la mayor distribución y mayor información, las figuras 7.a y 7.b no aportan tanta información.

¿Qué proyección te aporta menor información sobre la distribución de las habitaciones? ¿Por qué? La figura 5, que muestra la menor información.

Observa las líneas azul, roja y morada en la Figura 6.a., 6.b. y 6.c. ¿En qué caso la línea pasa más cercano de la mayoría de los puntos? ¿Por qué? En la figura 6c, porque la línea está más cerca de la mayoría de los puntos.

Observa las líneas azul, roja y morada en la Figura 6.a., 6.b. y 6.c. Señala aquella figura en donde se tenga una menor distancia promedio de los datos con respecto a la línea ¿Por qué? La 6c, porque la línea está más cerca de la mayoría de los puntos.

Suponte que el Componente Principal de la Figura 7.c. es la proyección de cada dato en la línea como se muestra para dos datos (proyección perpendicular a la línea). Proyecta los datos restantes en la componente de la Figura 7.c. Observa que la Componente Principal 1 son los puntos proyectados sobre la línea 1 y es una combinación lineal $CP1 = a \cdot \text{presión arterial} + b \cdot \text{edad}$. Observa que esta Componente Principal de la Figura 7.c. contiene menos diferencias o errores respecto de los datos que la posible Componente de las Figuras 7.a. y 7.b. compruébalo realizando mediciones con una regla.

Relativo a la Figura 7.c.:

Figura 26. Respuestas del Grupo de estudiantes destacados tercera actividad.

En este nivel de comprensión, los alumnos del grupo II, demostraron que ya habían reafirmado sus conocimientos, demostraron que la Figura 8.b de ofrecía mejor correlación lineal y menos error que la Figura 8.a. También, observaron y cotejaron que en un reporte de un paquete estadístico el análisis de Componentes Principales en la Figura 9.b. estaban reportadas las dos Componentes Principales y las dos variables originales y que era similar a la Figura 9.a. pero rotada.

Cuando a los alumnos del grupo II se les preguntó ¿Cuál de los dos Componentes explica la mejor varianza? ¿Qué diferencia hay entre las varianzas explicadas? Con esta información arrojada ¿Crees que es más difícil manejar y analizar los datos con menos o con más variables? ¿Por qué? Los alumnos del grupo II tuvieron semejante respuesta a los alumnos del grupo I: respondiendo que el componente de la figura 9.b explica mejor la varianza total en el primer componente comparado con el componente principal de la Figura 10.

Por último, los alumnos menos destacados alcanzaron el nivel de los alumnos del grupo I y llegaron a la conclusión que la mejor varianza de los componentes principales debe ser cuando exista mayor porcentaje de la varianza total del ACP. Por lo que, en este estudio fue mejor trabajar con menos variables que con el total de las variables.

5. CONCLUSIONES

El modelo de Van Hiele, abarca el aspecto descriptivo, el cual identificamos en el momento que los alumnos experimentaron diferentes formas de razonamiento, partiendo de lo geométrico y fueron valorando el progreso del entendimiento. En el aspecto instructivo, el profesor a partir de la instrucción de la secuencia de aprendizaje dio las pautas para favorecer el razonamiento geométrico de los datos procesados.

En los tres niveles de comprensión aplicados: 1.- Desarrollo Intuitivo de la de la Reducción de Datos: en este nivel los alumnos difícilmente llegaron al entendimiento de cómo y por qué reducir datos en un análisis estadísticos. En el segundo nivel de comprensión: Implementación de la Reducción de Datos, los alumnos alcanzaron a un mejor entendimiento del manejo y reducción de los datos y finalmente, en el tercer nivel de comprensión: Comprensión global de la Reducción de Datos y el ajuste de modelos, los alumnos alcanzaron un mejor entendimiento y desarrollaron sus habilidades, aplicando lo aprendido en las asignaturas algebra lineal y geometría en esta secuencia de aprendizaje, en ese momento pudimos percibir que desarrollaron los alumnos la comprensión, aplicación e interpretación del concepto de Componentes Principales.

El grupo I de estudiantes destacados

En el primer nivel de comprensión, demostraron tener conocimientos para: identificar la varianza, covarianza, media, máximo, mínimo y la correlación lineal que existía en las gráficas presentadas. Así como, este grupo de estudiantes identificaron las variables que especificaban los metadatos y supieron identificar el valor de cada dato y la magnitud de cada una de ellas y que no podían ser comparadas una con la otra por la falta de uniformidad de unidades.

Los alumnos, identificaron claramente el objetivo del estudio y medición de los datos. Sin embargo, aún no podían afirmar correctamente que se podía trabajar y analizar los datos con menos variables.

Los alumnos, en el segundo nivel de comprensión, a pesar que no pudieron descartar una variable de tres variables (pero una dependiente de la otra), tuvieron los conocimientos del concepto de dispersión (no asociación) y del concepto de correlación (asociación).

Los alumnos en el tercer nivel de comprensión, tenían reafirmados sus conocimientos, llegaron a la conclusión que la mejor varianza en el componente principal se explicaba con mejor porcentaje

de la varianza total, por lo que sería mejor trabajar con menos variables y poder escoger las variables que tenga el mejor índice de correlación.

El grupo II de estudiantes poco destacados

Este grupo de alumnos en el primer nivel de comprensión demostraron tener poco conocimiento de los conceptos: de varianza, covarianza y la correlación lineal que existía en las gráficas presentadas, Además, este grupo de alumnos no identificaron las variables que especificaban los metadatos, tampoco supieron identificar el valor de cada dato y la magnitud que representaba, por lo que no se logra el objetivo de esta sesión de la secuencia.

Los alumnos de este grupo, en el segundo nivel de comprensión, pudieron trabajar correctamente con esta sección, supieron descartar una variable de tres variables (cuando son dependientes de la otra), enriquecieron los conocimientos y supieron diferenciar el concepto de dispersión (no asociación) y el concepto de correlación (asociación).

Por último, los alumnos de este grupo, en el tercer nivel de comprensión, ya habían reafirmados conocimientos en: varianza, covarianza, dispersión de puntos, correlación lineal y conocieron discernir que a mejor varianza mejor expresión tendrá el componente principal, y también llegaron a la conclusión que es mejor trabajar con menos variables para obtener mejores rendimientos de varianza en los ACP.

6. CITAS BIBLIOGRÁFICAS

Anderson, C.A, (2000). *Exploratory Multivariate Data Analysis with Applications in Food Technology. PhD Dissertation*. The Royal Veterinary and Agricultural University. Rolighedsrej, Denmark.

Batanero, C., (2001). *Didáctica de la Estadística*. Departamento de Didáctica de las Matemáticas. Universidad de Granada, España.

Ben-Zvi, D., (2004). *Reasoning about Data Analysis. The Challenge of Developing Statistical Literacy, Reasoning and Thinking*. Dordrecht, the Netherlands: Kluwer Academic Publishers, 121-145

Burger W.F. y Shaughnessy, J.M. (1986). *Characterizing the van Hiele Levels of Development in Geometry*. Journal for Research in Mathematics Education, Vol. 17, No.1. 31-48.

Carlosena A. (1999). Clasificación of edible vegetables affected by different traffic intensities usin potencial curves. Talanta, 48, 745.

Cela R. (1994). *Avances en Quimiometría Práctica*. Ed. Univ. Stgo. de Compostela.

Dallas. E. J. (2000). *Métodos multivariados aplicados al análisis de datos*. México: Thompson, pp. 93-396

Dillon, W.R. y Goldstain, M. (1984). *Multivariate A nalysis, Methods and Applications*. Wiley, New York.

Einax J.E. (1992). Multivariate data analysis in environmental analytical Chemistry. GIT-Fachz-Lab, 36(8):815.

Garfiel, J. y Ben-Zvi, D (2005). *A Framework for Teaching and Assessing Reasoning about Variability*. Statistics Education Research Journal, 4(1). 92-99.

Góngora G. J., Cardiel L.N., Zamorano C.J. (2011). *Estadística Básica para estudiantes de Ciencias. Departamento de Astrofísica y Ciencias de la Atmosfera Facultad de Ciencias Físicas Universidad Complutense de Madrid.* 21-27.

Grupo de Quimiometria, i *Qualimetria de Tarragona. Quimiometría una disciplina per al ' analisi química.* Universitat de Rovira i Virgili. 2001.

<https://www.dspace.espol.edu.ec/retrieve/94469/D-71989.pdf>

Jaime, A. y Gutiérrez, A., (1990). “Una Propuesta de Fundamentación para la Enseñanza de la Geometría: El Modelo de Van Hiele” en *Teoría y práctica en educación matemática.* S. Llinares, M.V. Sánchez Eds. Sevilla: Alfar, 295-384.

Lange, J., Burrill, G., y Romberg, T. A. (1993). *Data visualization: Mathematics in context, learning, and testing.* Scotts Valley, CA: Wings for Learning.

Massart D.L. (1998). *Handbook of Chemometrics and Qualimetrics: Part B.* Ed. Elsevier Sc.

Peña, D., (2002). *Análisis de Datos Multivariantes.* McGrawHill, Madrid.

Pérez C. (2000). *Técnicas de análisis multivariante de datos. Aplicaciones con SPSS.* Madrid; Pearson, pp. 121-154.

Pfannkuch, M. y Wild, C., (2004). *Towards an understanding of statistical thinking. The Challenge of Developing Statistical Literacy, Reasoning and Thinking.* Dordrecht, the Netherlands: Kluwer Academic Publishers. 17-46.

Quintana I., Mora G. (1994) *Análisis de suero Humano por Espectroscopía Atómica.* Tesis de Diploma, Facultad de Química, Universidad de la Habana, Cuba.

Sokal, R. (2012). *Biometría: los principios y la práctica de la estadística en la investigación biológica.* 2ª ed. Libro en PDF en la Revista de la Sociedad Estadística Real Seria.

Van Hiele, P. M. (1957): *el problema de la comprensión (en conexión con la comprensión de los escolares en el aprendizaje de la geometría.* Tesis doctoral. Universidad de Utrecht. Utrecht, Holanda.

Anexo: Figuras

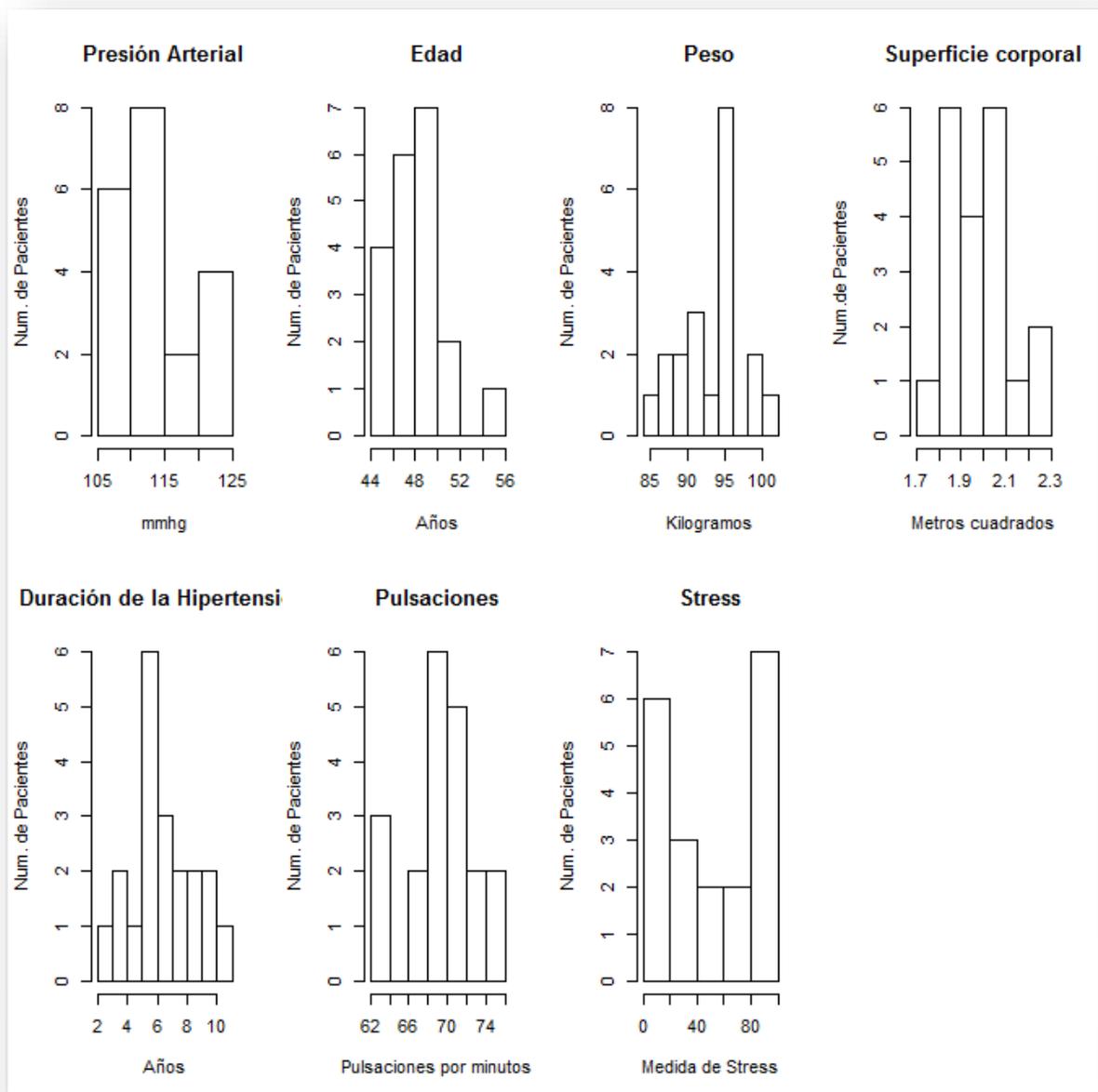


Figura 1. Histogramas de las variables

Variación de la Presión

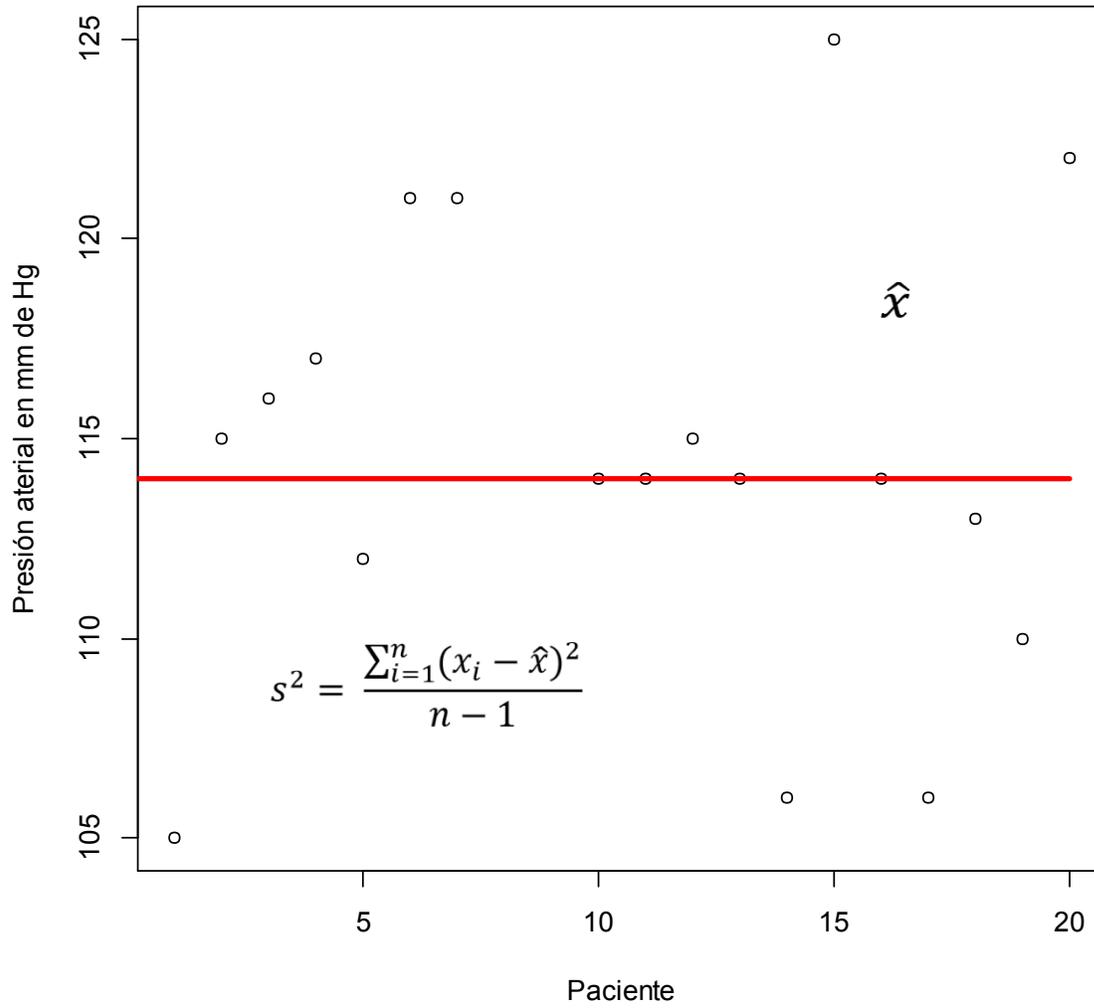
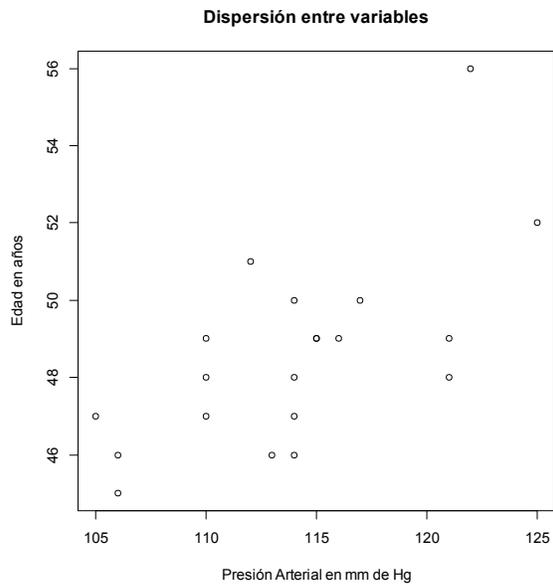
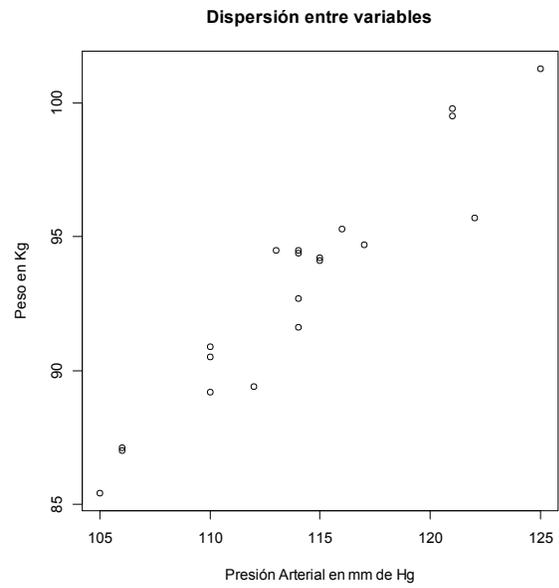


Figura 2. Varianza.



a



b

Figura 3. Gráfica de dispersión de Presión arterial con Peso y Edad.

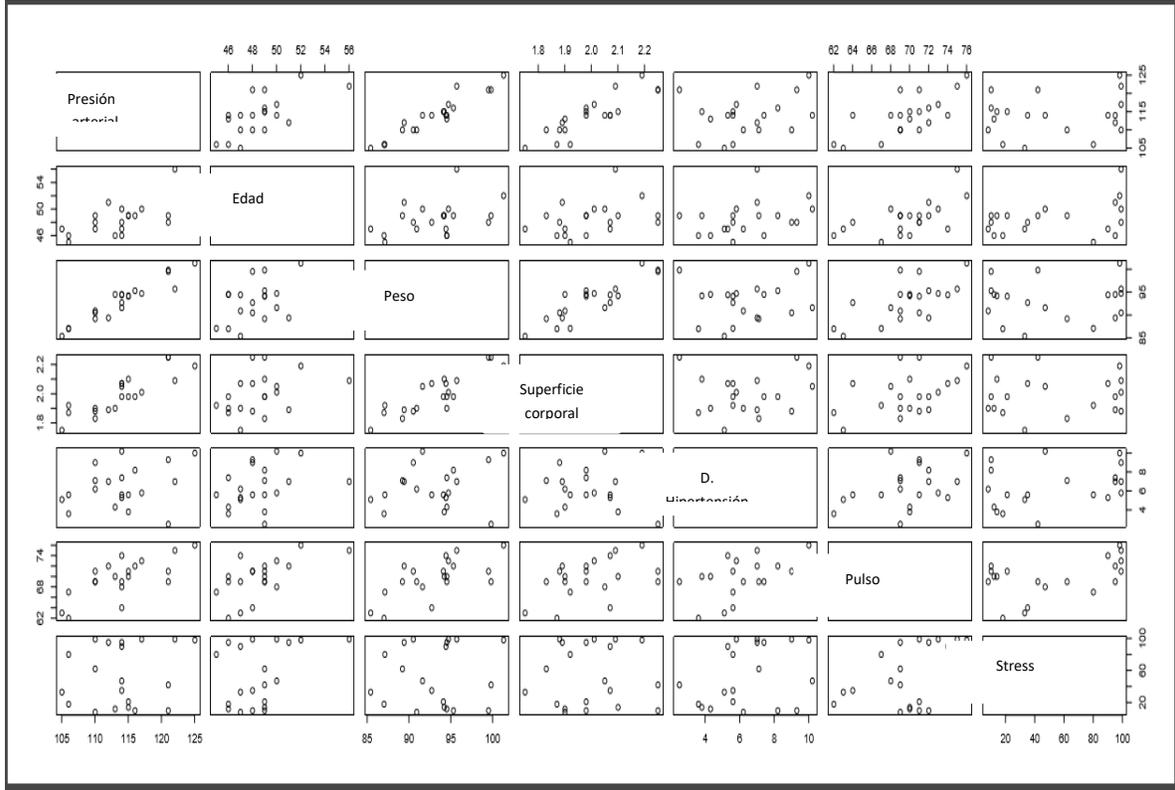


Figura 4. Matriz de dispersión de las siete variables.



Figura 5. Vistas de casa. Fuente: <https://www.google.com.mx/search>

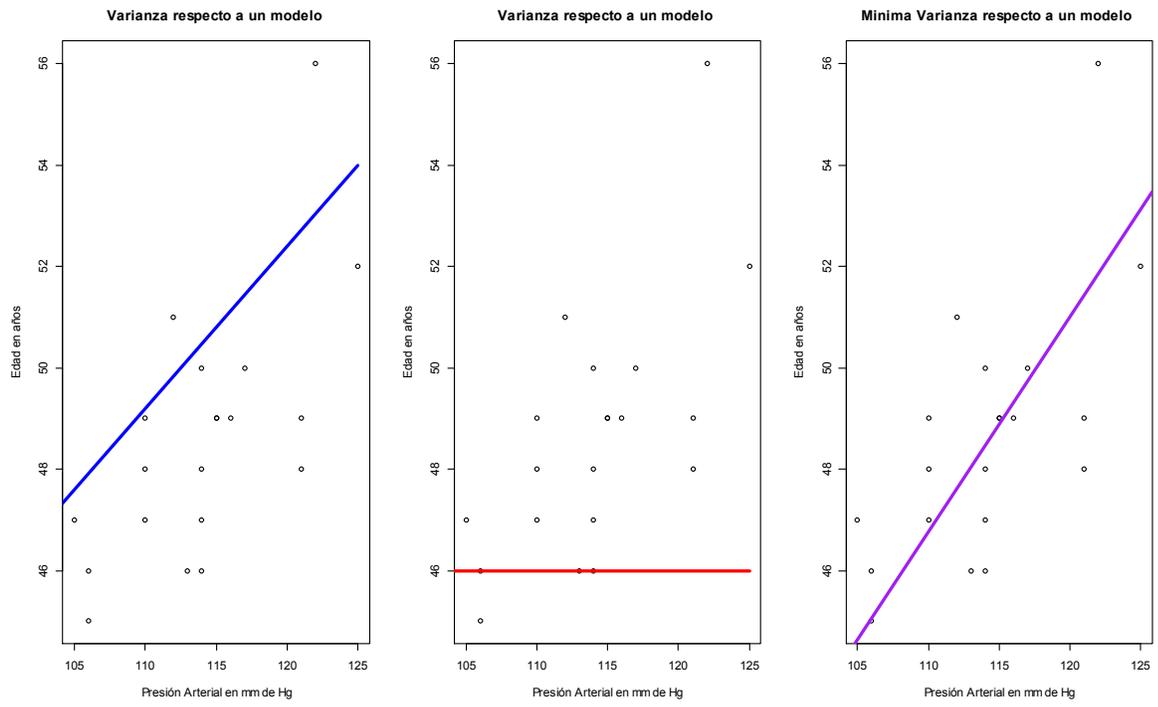


Figura 6. Figuras de componentes principales.

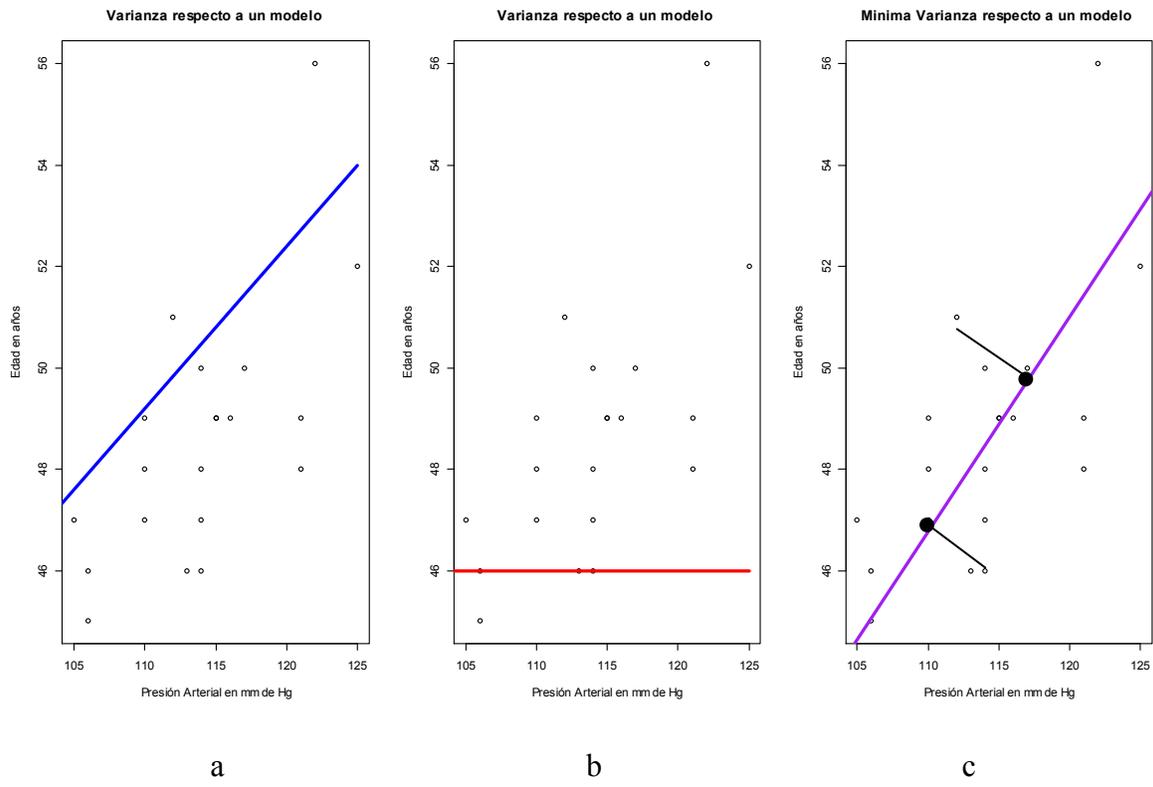
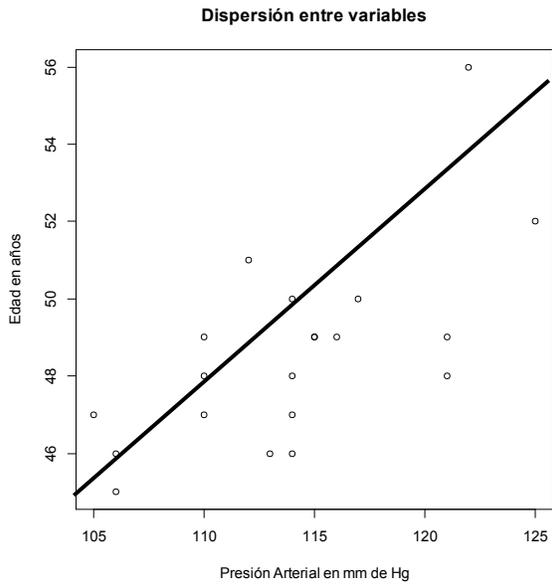
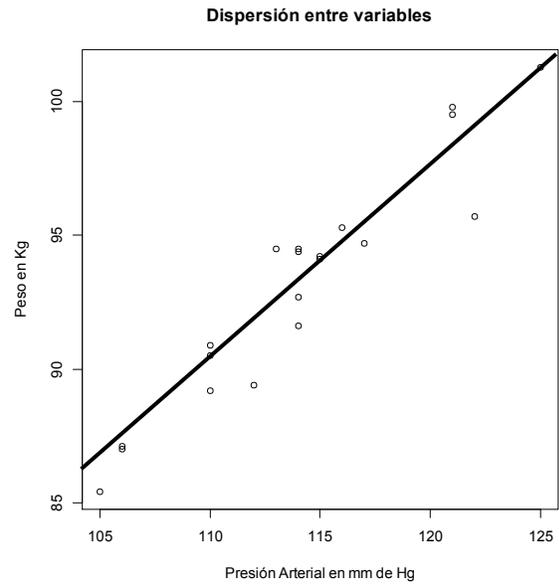


Figura 7. Diferentes proyecciones de los datos en distintos ejes.

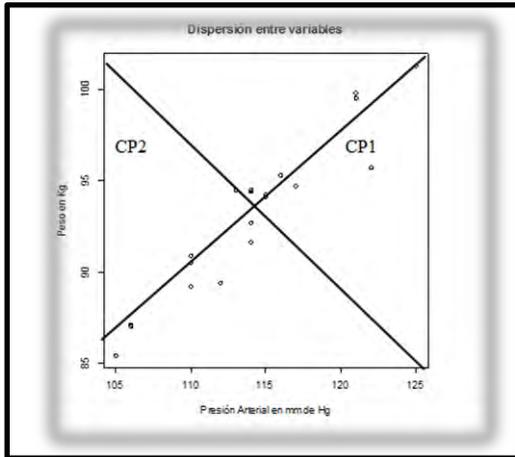


a

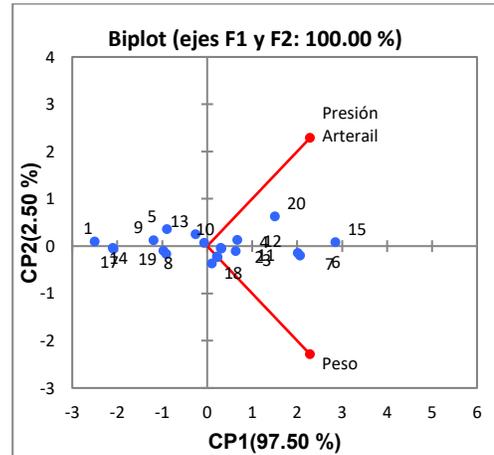


b

Figura 8. Componente principal para dos variables.



a



b

Figura 9. Componentes principales 1 y 2

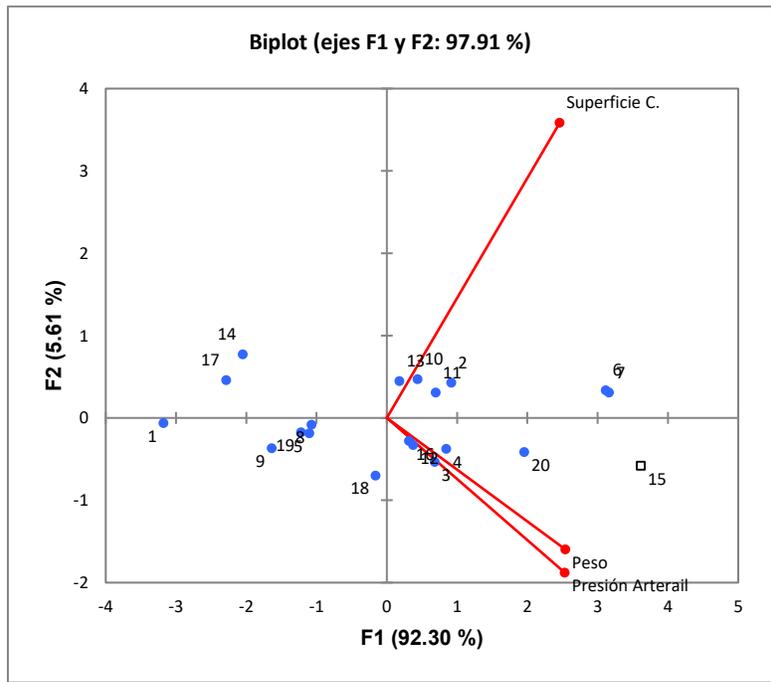


Figura 10. Componentes principales con tres variables.

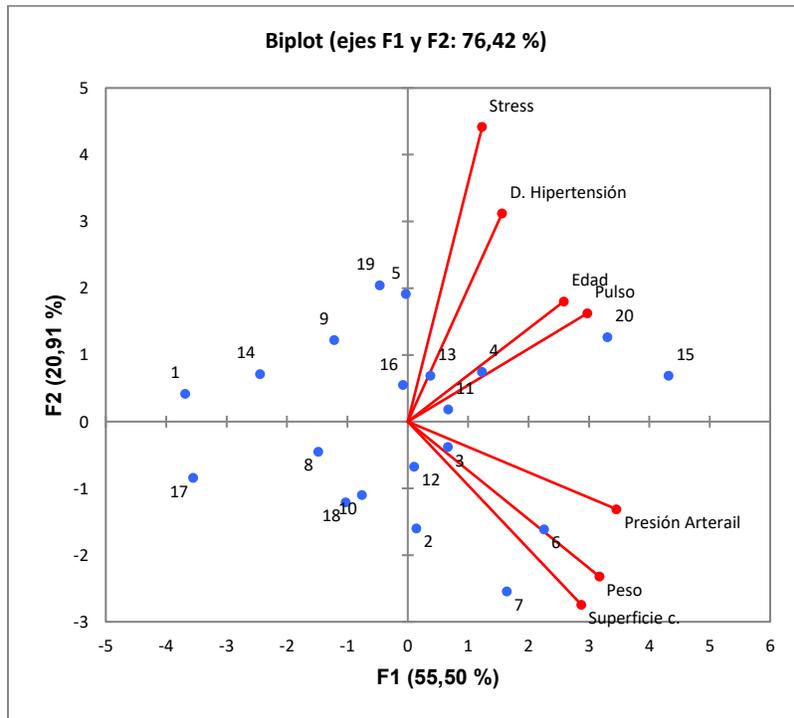


Figura 11. Componentes principales con siete variables.

Anexo: Tablas

Tabla1. Tabla de datos

Número de Paciente	Presión_ arterial	Edad	Peso	Superficie corporal	D. Hipertensión	Pulso	Stress
1	105	47	85	1.75	5.1	63	33
2	115	49	94	2.1	3.8	70	14
3	116	49	95	1.98	8.2	72	10
4	117	50	95	2.01	5.8	73	99
5	112	51	89	1.89	7	72	95
6	121	48	100	2.25	9.3	71	10
7	121	49	100	2.25	2.5	69	42
8	110	47	91	1.9	6.2	69	8
9	110	49	89	1.83	7.1	69	62
10	114	48	93	2.07	5.6	64	35
11	114	47	94	2.07	5.3	74	90
12	115	49	94	1.98	5.6	71	21
13	114	50	92	2.05	10	68	47
14	106	45	87	1.92	5.6	67	80
15	125	52	101	2.19	10	76	98
16	114	46	95	1.98	7.4	69	95
17	106	46	87	1.87	3.6	62	18
18	113	46	95	1.9	4.3	70	12
19	110	48	91	1.88	9	71	99
20	122	56	96	2.09	7	75	99

Tabla 2. Estadísticos descriptivos

Variable	Observaciones	Obs. con datos perdidos	Obs. sin datos perdidos	Mínimo	Máximo	Media	Desv. típica
Presión arterial	20	0	20	105.000	125.000	114.000	5.429
Edad	20	0	20	45.000	56.000	48.600	2.501
Peso	20	0	20	85.400	101.300	93.090	4.295
Superficie corporal	20	0	20	1.750	2.250	1.998	0.136
D. Hipertensión	20	0	20	2.500	10.200	6.430	2.145
Pulso	20	0	20	62.000	76.000	69.750	3.712
Stress	20	0	20	8.000	99.000	53.350	37.086

Tabla 3. Correlaciones lineales de Pearson.

Matriz de correlaciones (Pearson (n)):							
Variables	Presión_ arterial	Edad	Peso	Superficie corporal	D. Hipertensión	Pulso	Stress
Presión_ arterial	1	0.659	0.950	0.866	0.293	0.708	0.164
Edad	0.659	1	0.407	0.378	0.344	0.607	0.368
Peso	0.950	0.407	1	0.875	0.201	0.654	0.034
Superficie	0.866	0.378	0.875	1	0.131	0.446	0.018
D. Hipertensión	0.293	0.344	0.201	0.131	1	0.407	0.312
Pulso	0.708	0.607	0.654	0.446	0.407	1	0.467
Stress	0.164	0.368	0.034	0.018	0.312	0.467	1