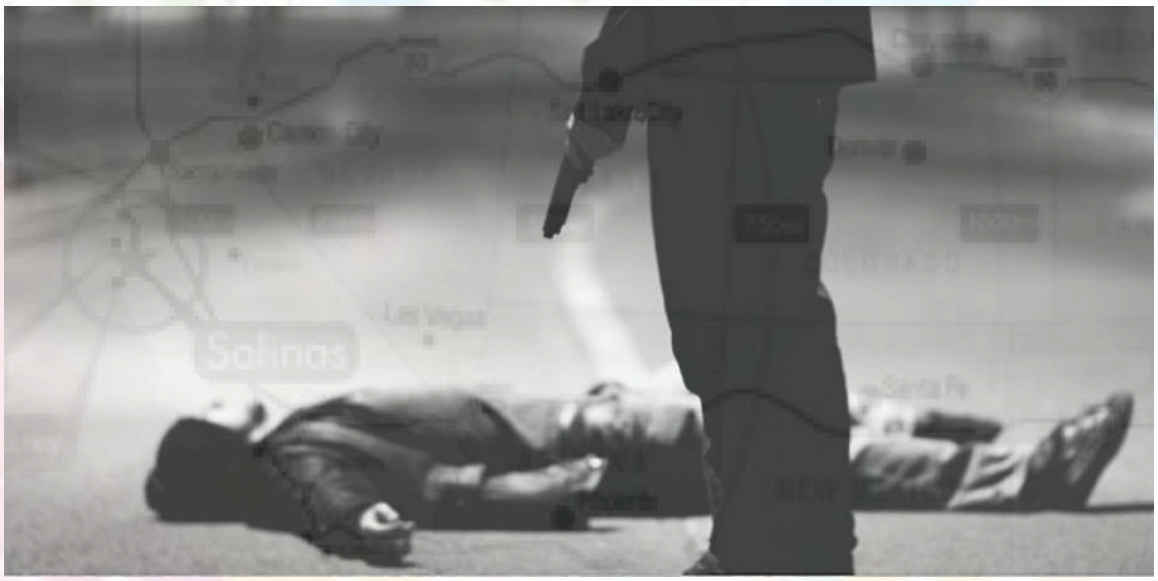
Universidad Autónoma de Yucatán
Facultad de Matemáticas

NIVEL DE VIOLENCIA EN SALINAS, CA.



ANÁLISIS MULTIVARIADO

Reporte Estadístico

Por:

- **Álvarez Samantha**
- **Asencio Alejandra**
- **Sobrino Samantha**

Licenciatura en Actuaría

FACULTAD DE MATEMÁTICAS

Análisis Multivariado

Componentes Principales y Análisis Factorial

Reporte Estadístico

Por:
- Álvarez Herrera Samantha
- Asencio González Alejandra
- Sobrino Bermejo Samantha

Profesora: Rocío Acosta Pech

Licenciatura en Actuaría – Especialización en Estadística.

TABLA DE CONTENIDOS

Página

Abstracto

En este trabajo se presenta un índice para medir los niveles de violencia anual en la comunidad de Salinas, California. Dicho indice permite relacionar la violencia con seis variables importantes y cuyos datos son de fácil alcance para toda persona por medio de las páginas web oficiales de diferentes instituciones gubernamentales. Dichas variables son el número de crímenes, el tamaño de la población, presupuesto (en dólares) del departamento de policía, el número de personas que se encuentran en libertad condicional, el número de personas por conjunto habitacional y la tasa de empleo registradas respectivamente año tras año.

Cabe mencionar que el objetivo inicial del estudio es realizar un análisis exploratorio de las variables para determinar sus características principales puesto que se intuye que son elementos importantes en la causalidad de los crímenes en la comunidad.

Debido al número de variables y puesto que se encontraron niveles altos de correlación entre ellas, se optó por utilizar el método de *componentes principales* para reducir la dimensionalidad. Como se puede ver en el apartado de metodología, este método sugirió la elección de una sola componente que explica suficientemente la variabilidad del modelo, es por esto que se decidió que dicha componente única puede utilizarse como un índice que representa el nivel de violencia en Salinas, California.

Introducción

Desde las primeras sociedades existentes, el crimen ha sido uno de los principales problemas, situación que aún está presente en las sociedades a lo largo del mundo. El crimen causa desequilibrio en todo el sistema social. Es un fenómeno complejo de impactos sociales, políticos y económicos.

Más importante aún, el crimen puede ser visto como un "síntoma indirecto" de fallas en el desarrollo personal, social y educativo de los niños, las cuales representan grandes pérdidas de potencial humano.

Existen muchas razones por las cuales la gente no respeta la ley al cien por ciento. La delincuencia es tan antigua como la humanidad misma, además, se ha convertido en un fenómeno social común y muchas personas lo ven como si fuera un componente funcional de la organización de agrupaciones humanas (Schafer, 1976).

Sin embargo, se sabe que el comportamiento criminal aumenta constantemente y más de la mitad de la población se preocupa constantemente por la delincuencia y aproximadamente el 75% se siente inseguro mientras está en casa (ONU-Hábitat, 2002).

A pesar de que la literatura que aborda temas relacionados con el crimen mismo, la cual ha logrado identificar componentes causales de este, hasta hoy en día no existe un modelo preciso que logre determinar cuáles son los factores que conducen a que una persona cometa un crimen. Los niveles de violencia o número de crímenes cometidos dentro de una sociedad, si bien pueden mantenerse dentro de un rango, suelen variar en grandes cantidades dependiendo de la zona habitacional que se esté analizando.

En este trabajo se aborda el caso de la comunidad urbana de Salinas, el municipio más grande de Monterey County, ubicado en el estado de California, en los Estados Unidos de Norte América.

El Departamento de Policía de Salinas (SPD, por sus siglas en inglés), junto con las autoridades de la comunidad, ha trabajado incansablemente para mitigar el crimen, relacionado principalmente con las pandillas. Actualmente se están realizando numerosos esfuerzos para reducir la tasa de criminalidad de la ciudad, desde la participación de la comunidad hasta la creación de un grupo de trabajo de pandillas en asociación con las oficinas de policía del condado de los alrededores. Todos estos esfuerzos se derivan de la experiencia y, como es de imaginarse, sin un modelo estadístico que ayude a explicar los niveles de delitos violentos en Salinas. El propósito de este estudio es realizar un análisis exploratorio de las principales variables, que se intuye tienen influencia en los niveles de violencia durante cierto año en la comunidad de Salinas, California y, de ser posible, sugerir un modelo que ayude a medir dichos niveles.

Antecedentes

Se han hecho investigaciones interesadas en averiguar si una mayor cantidad de policías reduce el crimen, una serie de documentos que utilizan una variedad de enfoques diferentes han llegado a una conclusión similar: más policías reducen sustancialmente el crimen (Marvell y Moody, 1968; Levit, 1977; Corman y Mocan, 2000).

Desde que Becker (1968) publicó su artículo seminal sobre el modelo económico del crimen, varios economistas han realizado esfuerzos considerables para determinar su validez empírica, invadiendo el campo utilizando un modelo integral de comportamiento racional individual, en el que una persona se decide por cometer un acto criminal si las ganancias netas, incluyendo el pago de sanciones y otros costos, son mayores que el de las alternativas legales (Erling, 1999).

El estudio del crimen siempre ha sido una actividad multidisciplinar. Aparte de los criminólogos, los sociólogos son quizás el grupo dominante, sin embargo los psicólogos y políticos también han sido destacados durante mucho tiempo. Economistas, econométricos y estadísticos se encuentran entre los participantes más recientes. El estudio de Becker en 1968, “Crime and Punishment: An Economic Approach” (Crimen y castigo: un enfoque económico), sirvió como punto de partida para el trabajo de los economistas modernos sobre el crimen, además de “Estimating the Economic Model of Crime with panel data” (Estimando el modelo económico del crimen con datos de panel), de Cornwell y N. Trumbul (1994) y Marselli (1997). Entre mediados y finales de los años 90, los economistas jóvenes que desarrollaron agendas de investigación se centraron principalmente en el estudio de la delincuencia.

En diciembre de 2009, Jason Clarke y Tracy Onufer realizaron una tesis de NPS (Naval Post-Graduate School) titulada ”Understanding Environmental Factors that Affect Violence in Salinas, California" (Entendiendo los factores ambientales que afectan la violencia en Salinas, California), la cual comparó nueve factores ambientales: nivel económico, población, alojamiento, educación, fuerza policial, influencia carcelaria, rivalidad de pandillas, programas de servicios sociales y la participación de la comunidad, comparados con la tasa de violencia anual en Salinas para determinar qué factores ambientales, de ser el caso, estaban relacionados con los niveles de violencia en Salinas. Clarke y Onufer consideraron la violencia como una combinación de homicidios, robos y agresiones denunciados.

El problema de las pandillas de Salinas se remonta a la década de 1950. En su informe de 90 días en 2009, el jefe de policía Fetherolf citó la declaración del jefe de policía de 1950, McIntyre: "Las peleas de pandillas no serán toleradas en la ciudad de Salinas" (Fetherolf, 2009, pág. 6). El informe de 90 días continúa mencionando varias instancias de la violencia de Salinas antes de su informe de 2009.

El crimen en Salinas ha aumentado constantemente y ha mantenido un nivel más alto que el promedio nacional. En 2009, Salinas tuvo un récord de 29 homicidios, aproximadamente cuatro veces el promedio nacional, seguido de 19 homicidios en 2010, de nuevo aproximadamente cuatro veces el promedio nacional. En 2009, Salinas ocupó el cuarto lugar en California en homicidios per cápita (Fetherolf, 2010, p. 2). Homicidios, asaltos y robos se sumaron y se utilizaron como una sola variable: violencia.

Metodología.

Para este estudio se cuenta con 6 variables, cada una con 22 observaciones. Se realizó una reducción de la dimensionalidad de variables a través del método de componentes principales, donde el número de componentes fue decido según el criterio del valor propio mínimo, es decir, se seleccionan todas aquellas componentes principales cuyo valor propio asociado sea mayor a 1.

Debido a que los datos se encuentran medidos en diferentes escalas se decidió estandarizar los datos.

NOTA: Al trabajar con las variables estandarizadas, los valores y vectores propios fueron obtenidos utilizando la matriz de correlaciones.

Análisis descriptivo

Procederemos a puntualizar las características de todas las variables involucradas en este estudio, así como la relación existente entre ellas. (*Para mayor detalle, consultar Anexo B)*:

- **Crimen:** Se refiere a el número de crímenes o violaciones de la ley mediante el uso de la violencia anuales en la comunidad de Salinas, California. A saber, el número registrado por año se compone por el total de homicidios, asaltos y robos.

  Los datos recolectados de esta variable cuentan con una media de 1154.73 y la desviación estándar es igual a 124.455. Además, su curtosis estandarizada se encuentra dentro del rango esperado para los datos de una distribución Normal.

- **Población:** Se define como "población" al conjunto de habitantes en un lugar. En el presente estudio se tomó en cuenta los datos oficiales correspondientes al número de habitantes por año desde el 1990 hasta el 2011 en la comunidad de Salinas, en el estado de California. Por naturaleza de esta variable, se espera un aumento año tras año, dado que el periodo estudiado es de solamente un poco más de 20 años.

  Su media y su desviación estándar son de 136225 y 15024.3, respectivamente. La curtosis y el sesgo estandarizados se encuentran dentro del intervalo $(-2,2)$.

- **Presupuesto destinado al Departamento de Policía de Salinas:** Para el estudio, esta variable se midió en dólares y hace referencia a la cantidad de dinero total que el gobierno asignó a la seguridad pública, es decir, al departamento de policía. Según estudios precedentes, se esperaría que, con mayor cantidad de dinero, se pudiera brindar un mejor servicio, elevando la eficiencia de la Policía local y derivando en una disminución de la violencia.

  Esta variable tiene una media de 23614600 USD y su desviación estándar es igual a 9851620 USD. Además, el sesgo y la curtosis se encuentran dentro del intervalo de -2 a 2, indicando que los datos pueden haber provenido de una distribución normal.

- **Personas en libertad condicional:** Se conoce como "libertad condicional" al permiso que se le concede a un condenado en la última parte de la pena por haber mostrado buen comportamiento. Es decir, esta variable representa el número total de ex reos a quienes el gobierno del estado de California (o de los Estados Unidos) les ha concedido su libertad después de haber cumplido parte de su condena, lo que quiere decir que tienen antecedentes penales. Debido a que en muchas ocasiones al momento de solicitar un trabajo se solicita un documento que compruebe no tener antecedentes penales, no es de sorprenderse que esta variable esté relacionada con la tasa de desempleo, variable que también se consideró para el estudio, sin embargo, esta relación no representa un problema para la funcionalidad del modelo ajustado final.

  Su media y su desviación estándar son de 107072 y 14429.2, respectivamente. Su sesgo y curtosis estandarizados se encuentran dentro del intervalo $(-2,2)$ de una distribución normal.

- **Número de personas por conjunto habitacional:** Esta variable hace referencia al número total de personas que viven bajo un mismo techo, según el censo correspondiente a cada año. En estudios anteriores se ha encontrado que padres de familia o responsables del hogar comenten crímenes al sentirse "ahorcados" por los gastos que conlleva la manutención de las personas de quién es responsable, así como personas que viven en grupos comunitarios, que realizan robos con tal de obtener el dinero suficiente para poder independizarse.

  Esta variable tiene una media de 3.56773 y una desviación estándar igual a 0.178083. Además, el sesgo y la curtosis se encuentran dentro del intervalo de -2 a 2, indicando que los datos pueden haber provenido de una distribución normal.

- **Tasa de desempleo:** Es una medida de la extensión del desempleo, se calcula como un porcentaje dividiendo el número de personas desempleadas por todas las personas que se encuentran en la fuerza laboral, es decir, personas que cumplen con las características necesarias para desempeñar un empleo. Como se mencionó anteriormente, esta variable presenta una relación con el número de personas en libertad condicional.

  Tiene una media de 10.1045 y una desviación estándar de 1.47977. Su sesgo y curtosis se encuentran dentro del intervalo de -2 a 2, por lo que los datos pueden haber provenido de una distribución normal.

Finalmente, para poder aplicar el método de componentes principales debemos comprobar si existe correlación entre las variables del estudio. Después de analizar la matriz de correlaciones concluimos que las variables se encuentran correlacionadas, por lo que podemos continuar con la aplicación del método. *(Consultar Anexo C)*

Nota: Cabe mencionar que para el método de componentes principales no es necesaria la Hipótesis de normalidad multivariada, por lo tanto, no es un requisito que las variables del estudio sigan una distribución normal univariada.

Recolección de los datos

Los datos del proyecto provienen de una tesis, realizada en Junio de 2012, de la escuela Naval de Post-grado de California, cuyo autor es Jarrod S. Shingleton, quién agregó los datos correspondientes al 2010 y 2011 a la tesis realizada anteriormente por Clark y Onufer (2009).

Los autores mencionados recolectaron los datos de las páginas en línea de las Secretarías Federales correspondientes, pues su propósito fue brindar al Departamento de Policía de Salinas una herramienta de fácil acceso para estimar los niveles de delincuencia con datos fácilmente accesibles, por lo que todos los datos utilizados son de acceso público. Todos los datos fueron validados en 2012 por las autoridades de la Institución correspondiente. Lo anterior nos da la certeza de la veracidad y aleatoriedad de la muestra.

Objetivos del estudio

General: Realizar un análisis exploratorio de las variables para describir sus características y relación con el nivel de violencia anual en Salinas, California.

Específicos: Reducir el número de variables mediante el método de Componentes Principales. Analizar los pesos que tienen las variables en las componentes creadas y seleccionadas.

Componentes Principales

Según el problema de este estudio, contamos con seis variables, por lo que tendremos seis componentes principales

La i-ésima componente principal de las variables estandarizadas está dada por:

$$Y_i = \sum_{j=1}^{6} \phi_{ij} Z_j = \sum_{j=1}^{6} \phi_{ij} \left(\frac{X_i - \mu_i}{\sigma_i} \right)$$
$$para\ i = 1, .., 6$$

Donde:

$\phi_1, \dots, \phi_7$: son los vectores propios para **R**.
X_1 es el número de crímenes
X_2 número de habitantes por año
X_3 es el presupuesto del SPD
X_4 es el número de personas en libertad condicional;
X_5 es el número de personas por conjunto habitacional;
X_6 es la tasa de desempleo del al i-ésimo año;

Resultados

Mediante la utilización de un software estadístico obtenemos que sólo una componente principal tiene un valor propio mayor que la unidad. Esta componente explica el 64.576 de la varianza.

La componente está dada como sigue:

$$Y_1 = 0.29056Z_1 - 0.495869Z_2 - 0.420906Z_3 - 0.446013Z_4 - 0.430832Z_5 + 0.328608Z_6$$

Como se obtuvo una única variable nueva (componente) que es combinación lineal de todas las variables del estudio (ponderadas), esta se puede utilizar como índice de violencia, es decir, que es suficiente para obtener valores correspondientes a los niveles de violencia en la comunidad de Salinas, California, a partir de las variables originales del estudio.

Analizando el componente obtenido vemos que las variables *crimen* y la *tasa de desempleo* tienen un peso positivo, es decir, tienen relación directa con el nivel de violencia. Esto nos indica que el índice de violencia se incrementa si aumenta el número de crímenes o la tasa de desempleo.

Por otro lado, las variables *población*, el *presupuesto del SPD*, el *número de personas en libertad condicional* y el *número de personas por conjunto habitacional* tienen un peso negativo, por lo que el índice de violencia disminuye ante el aumento de cualquiera de dichas variables. *(Anexo D)*

Conclusiones

Utilizando el método estadístico multivariado de componentes principales logramos la reducción del número de variables cuantitativas a analizar. Podemos observar que al principio debíamos trabajar con 6 variables y al final obtenemos un único componente.

Lo anterior que es de gran utilidad ya que lo podemos utilizar como un índice para medir el nivel de violencia anual en la comunidad de Salinas, California. Dicho índice está relacionado con los seis factores, de manera positiva con el número de crímenes y con la tasa de desempleo, es decir, al haber un aumento en el desempleo y en los crímenes habrá un aumento en el índice de violencia.

El nivel de violencia queda inversamente relacionado con el número de habitantes por año, el presupuesto SPD, el número de personas en libertad condicional y el número de personas por conjunto habitacional, es decir, al aumentar estos últimos el índice de violencia anual debe disminuir.

Este tipo de modelos puede ayudar a las autoridades a reducir el índice de violencia, sabiendo que factores deben controlar, o cuales aumentar, para mantener un índice de violencia bajo

Referencias

Becker, G. (1968). *Crime and Punishment: An Economic Approach. The Journal of Political Economy,76/2* , 169-217

C.Cornwell and N.Trumbul. (1994). *Estimating the Economic Model of Crime with panel Data.* The Review of Economics and Statistics,Vol.76.No.2(May,1994), 360-366.

Ehrlich. (1999). *Crime, Punishement and the Market for Offenses*. The Journal of Economic Perspectives,Vol.10.1, 43-67.

Johnson, R. y Wichern, D. (2007) *Applied Multivariate Statistical Analysis*, sexta edición, Pearson, Nueva Jersey.

Schafer. (1976). *Fear of Crime and Criminal Victimizations:Gender-based Contrasts.* Journal of Criminal Justice,Vol.34, 285-301.

Shingleton, Jarrod S. (Junio 2012). *Crime Trend Prediction Using Regression Models For Salinas, California.* Naval Postgraduate School, Thesis. Monterrey, California, E.E.U.U. Recuperado en Septiembre, de https://apps.dtic.mil/dtic/tr/fulltext/u2/a563653.pdf

ANEXO A: TABLA DE DATOS DEL ESTUDIO.

Año		Crimen	Población	Prep_Pol	P_Lib_Cond	Num_PP_Casa	T_Desempleo
1990	1	1051	108777	11456256	73096	3.21	9.7
1991	2	1065	111184	13144292	85470	3.24729	11.4
1992	3	1127	114736	13634881	89453	3.33079	12.5
1993	4	1419	116686	15683718	88858	3.36624	13.1
1994	5	1284	120885	13612478	92958	3.46189	12.4
1995	6	1459	121960	14471238	96110	3.45003	12.3
1996	7	1304	124972	16393545	100934	3.47405	11.3
1997	8	1261	127369	16929407	105449	3.49622	11
1998	9	1118	132449	17575700	111875	3.56309	10.8
1999	10	1096	136797	18852899	117612	3.60341	9.7
2000	11	1195	142685	19288170	121414	3.662	7.4
2001	12	1213	144728	21713995	121820	3.69	7.8
2002	13	1079	146659	22040439	117138	3.702	8.9
2003	14	1143	148117	24224300	114136	3.7	9
2004	15	1147	149838	25241659	113768	3.699	8.3
2005	16	987	149626	29704910	115001	3.654	7.3
2006	17	1073	148707	33356709	121808	3.614	6.9
2007	18	1103	148782	35416564	126906	3.601	7.1
2008	19	992	150898	38380314	123597	3.637	8.4
2009	20	1066	150215	41187794	109026	3.643	11.8
2010	21	1139	150441	37360500	108656	3.685	12.8
2011	22	1083	150441	39852481	100490	4	12.4

Donde:

Prep_Pol: Presupuesto en dólares destinado al SPD.

P_Lib_Cond: Número de personas en libertad condicional.

Num_PP_Casa: Número de personas por conjunto habitacional (casa).

T_Desempleo: Tasa de desempleo.

ANEXO B

ANÁLISIS DESCRIPTIVO DE LAS VARIABLES

Análisis individual de las variables.

El estudio cuenta con seis variables de interés, con 22 observaciones. A cada una se le determinó su media y su desviación estándar y se evaluó si su sesgo y su curtosis estandarizados se encuentran dentro del rango esperado para los datos de una distribución Normal. También se aplicó la prueba de Shapiro-Wilk para normalidad con un nivel de significancia de $\alpha = 5\%$. El no rechazar la hipótesis nula indicaría que la muestra se ajusta a la distribución Normal.

- Crimen

Resumen estadístico para Violencia

Recuento	22
Promedio	1154.73
Desviación Estándar	124.455
Coeficiente de Variación	10.7778%
Mínimo	987.0
Máximo	1459.0
Rango	472.0
Sesgo Estandarizado	2.09407
Curtosis Estandarizada	0.772963

Tests de Normalidad para Violencia

Test	*Estadístico*	*P-Valor*
Shapiro-Wilk W	0.899872	0.0273312

En el Resumen Estadístico notamos que la media de los datos de violencia es de 1154.73 y la desviación estándar es igual a 124.455. Además, la curtosis estandarizada se encuentra dentro del intervalo de -2 a 2, es decir, está dentro del rango esperado para los datos de una distribución normal. Esto no se cumple para el sesgo.

Efectuando una prueba de Shapiro-Wilk vemos que los datos de violencia no se ajustan a la distribución Normal. $(P = 0.0273312 < .05 = \alpha)$

- Población

Resumen estadístico para población

Recuento	22
Promedio	136225.
Desviación Estándar	15024.3
Coeficiente de Variación	11.0291%
Mínimo	108777.
Máximo	150898.
Rango	42121.0
Sesgo Estandarizado	-1.13128
Curtosis Estandarizada	-1.23659

Tests de Normalidad para Población

Test	*Estadístico*	*P-Valor*
Shapiro-Wilk W	0.840276	0.00167289

En el Resumen Estadístico notamos que la media de la población es de 136225 y la desviación estándar es igual a 15024.3. Además, el sesgo y la curtosis estandarizados se encuentran dentro del intervalo de -2 a 2, es decir, están dentro del rango esperado para los datos de una distribución normal.

Utilizando una prueba de Shapiro-Wilk vemos que los datos de población no se ajustan a la distribución Normal. $(P = 0.00167289 < 0.05 = \alpha)$

- Prep_Pol: Presupuesto destinado a la Policía

Resumen estadístico para Prep_Pol

Recuento	22
Promedio	2.36146E7
Desviación Estándar	9.85162E6
Coeficiente de Variación	41.7183%
Mínimo	1.14563E7
Máximo	4.11878E7
Rango	2.97315E7
Sesgo Estandarizado	1.15659
Curtosis Estandarizada	-1.08026

En el Resumen Estadístico notamos que la media del Presupuesto destinado a la Policía es de 2.36146E7 y la desviación estándar es igual a 9.85162E6. Además, el sesgo y la curtosis estandarizados se encuentran dentro del intervalo de -2 a 2, es decir, están dentro del rango esperado para los datos de una distribución normal.

Realizando una prueba de Shapiro-Wilk vemos que la variable no se ajusta a la distribución Normal.
$(P = 0.0159512 > 0.05 = \alpha)$

Tests de normalidad para Prep_Pol

Test	*Estadístico*	*P-Valor*
Shapiro-Wilk W	0.8888	0.0159512

- P_Lib_Cond: Número de personas en libertad condicional

Resumen estadístico para P_Lib_Cond

Recuento	22
Promedio	107072.
Desviación Estándar	14429.5
Coeficiente de Variación	13.4765%
Mínimo	73096.0
Máximo	126906.
Rango	53810.0
Sesgo Estandarizado	-1.34611
Curtosis Estandarizada	-0.235959

En el Resumen Estadístico notamos que la media del Número de personas en libertad condicional es de 107072 y la desviación estándar es igual a 14429.2. Además, el sesgo y la curtosis se encuentran dentro del intervalo de -2 a 2, es decir, están dentro del rango esperado para los datos de una distribución normal.

Aplicando la prueba de Shapiro-Wilk vemos que se ajusta a la distribución Normal.$(P = 0.202774 > 0.05 = \alpha)$

Tests de normalidad para P_Lib_Cond

Test	*Estadístico*	*P-Valor*
Shapiro-Wilk W	0.940661	0.202774

- Num_PP_Casa: Número de personas por casa

Resumen estadístico para Num_PP_Casa

Recuento	22
Promedio	3.56773
Desviación Estándar	0.178083
Coeficiente de Variación	4.99151%
Mínimo	3.21
Máximo	4.0
Rango	0.79
Sesgo Estandarizado	-0.187077
Curtosis Estandarizada	0.807067

En el Resumen Estadístico notamos que la media del Número de personas por casa es de 3.56773 y la desviación estándar es igual a 0.178083. Además, el sesgo y la curtosis se encuentran dentro del intervalo de -2 a 2, es decir, están dentro del rango esperado para los datos de una distribución normal.

Aplicando la prueba de Shapiro-Wilk vemos que se ajusta a la distribución Normal.$(P = 0.14135 > .05 = \alpha)$

.

Tests de normalidad para Num_PP_Casa

Test	*Estadístico*	*P-Valor*
Shapiro-Wilk W	0.933205	0.14135

- T. Desempleo: Tasa de Desempleo

Resumen estadístico para T_Desempleo

Recuento	22
Promedio	10.1045
Desviación Estándar	2.11063
Coeficiente de Variación	20.8879%
Mínimo	6.9
Máximo	13.1
Rango	6.2
Sesgo Estandarizado	-0.249399
Curtosis Estandarizada	-1.46052

En el Resumen Estadístico notamos que la media de la Tasa de Desempleo es de 10.1045 y la desviación estándar es igual a 2.11063. Además, el sesgo y la curtosis se encuentran dentro del intervalo de -2 a 2, es decir, están dentro del rango esperado para los datos de una distribución normal.

Aplicando la prueba de Shapiro-Wilk vemos que se ajusta a la distribución Normal.($P = 0.0561688 > .05 = \alpha$)

Tests de normalidad para T_Desempleo

Test	*Estadístico*	*P-Valor*
Shapiro-Wilk W	0.914497	0.0561688

ANEXO C

ANÁLISIS DE LAS CORRELACIONES

Presentamos la matriz de correlaciones entre las siete variables:

Correlaciones

	Crimen	Población	Prep_Pol	P_Lib_Cond	Num_PP_Casa	T_Desempleo
Crimen		**-0.4427**	**-0.4904**	**-0.3007**	**-0.2716**	**0.4659**
		(22)	(20)	(22)	(22)	(22)
		0.0391	0.0282	0.1738	0.2215	0.0289
Población	-0.4427		**0.9019**	**0.8541**	**0.8960**	**-0.5332**
	(22)		(20)	(22)	(22)	(22)
	0.0391		0.0000	0.0000	0.0000	0.0106
Prep_Pol	-0.4904	0.9019		**0.7677**	**0.7405**	**-0.6507**
	(20)	(20)		(20)	(20)	(20)
	0.0282	0.0000		0.0001	0.0002	0.0019
P_Lib_Cond	-0.3007	0.8541	0.7677		**0.6992**	**-0.6985**
	(22)	(22)	(20)		(22)	(22)
	0.1738	0.0000	0.0001		0.0003	0.0003
Num_PP_Casa	-0.2716	0.8960	0.7405	0.6992		**-0.3028**
	(22)	(22)	(20)	(22)		(22)
	0.2215	0.0000	0.0002	0.0003		0.1707
T_Desempleo	0.4659	-0.5332	-0.6507	-0.6985	-0.3028	
	(22)	(22)	(20)	(22)	(22)	
	0.0289	0.0106	0.0019	0.0003	0.1707	

Correlation
(Sample Size)
P-Value

Cuando las variables están altamente correlacionadas, pocos de los primeros componentes pueden ser suficientes para describir la mayor parte de la variabilidad presente. En este caso, observamos correlaciones altas entre las variables, por lo que el método de componentes principales resulta útil para reducir la dimensionalidad del estudio.

ANEXO D

ANÁLISIS DE COMPONENTES PRINCIPALES

Puesto que las variables originales del estudio fueron medidas en diferentes escalas y unidades por su naturaleza, se decidió trabajar con los datos estandarizados, de esta manera se evita también que las varianzas de las variables afecten en la ponderación de las variables originales al construir las combinaciones lineales de estas (componentes). Por lo anterior, el método de componentes principales utilizó la matriz de correlaciones muestrales, así como los datos y variables estandarizadas.

Con el paquete estadístico obtenemos que:

Principal Components Analysis

Component Number	*Eigenvalue*	*Percent of Variance*	*Cumulative Percentage*
1	4.43973	73.996	73.996
2	0.916015	15.267	89.262
3	0.359978	6.000	95.262
4	0.204065	3.401	98.663
5	0.0781885	1.303	99.966
6	0.0020196	0.034	100.000

Donde:

Los *valores propios d*e la matriz de correlaciones **R** están dados por las soluciones de la ecuación:

$$|\boldsymbol{R} - \lambda \mathbf{I}| = 0$$

Donde $\boldsymbol{R}_{\mathbf{6x6}}$ es la matriz de correlaciones de la muestra:

$$\boldsymbol{R} = \begin{bmatrix} 1 & -0.4427 & -0.4904 & -0.3007 & -0.2716 & 0.4659 \\ -0.4427 & 1 & 0.9019 & 0.8541 & 0.8960 & -0.5332 \\ -0.4904 & 0.9019 & 1 & 0.7677 & 0.7405 & -0.6507 \\ -0.3007 & 0.8541 & 0.7677 & 1 & 0.6992 & -0.6985 \\ -0.2716 & 0.8960 & 0.7405 & 0.6992 & 1 & -0.3028 \\ 0.4659 & -0.5332 & -0.6507 & -0.6985 & -0.3028 & 1 \end{bmatrix},$$

$\mathbf{I}_{\mathbf{6x6}}$ es la matriz identidad.

Al resolver la ecuación obtenemos que los 6 valores propios son:

$\lambda_1 = 4.43973, \quad \lambda_2 = 0.916015, \quad \lambda_3 = 0.359978,$

$\lambda_4 = 0.204065, \quad \lambda_5 = 0.0781885, \quad \lambda_6 = 0.0020196$

La *proporción de varianza (estandarizada) total explicada por cada componente* es:

Primera componente:

$$\frac{\lambda_1}{\sum_{i=1}^{6} \lambda_i} = \frac{4.43973}{5.9999961} = 0.74$$

Segunda componente:

$$\frac{\lambda_2}{\sum_{i=1}^{6}\lambda_i} = \frac{0.916015}{5.9999961} = 0.1527$$

Tercera componente:

$$\frac{\lambda_3}{\sum_{i=1}^{6}\lambda_i} = \frac{0.359978}{5.9999961} = 0.06$$

Cuarta componente:

$$\frac{\lambda_4}{\sum_{i=1}^{6}\lambda_i} = \frac{0.204065}{5.9999961} = 0.034$$

Quinta componente:

$$\frac{\lambda_5}{\sum_{i=1}^{6}\lambda_i} = \frac{0.0781885}{5.9999961} = 0.013$$

Sexta componente:

$$\frac{\lambda_6}{\sum_{i=1}^{6}\lambda_i} = \frac{0.0020196}{5.9999961} = 0.0003$$

Utilizando el criterio del mínimo valor propio, donde este es igual a 1, se decide que la primera componente puede reemplazar a las seis variables originales con una pequeña perdida de información. Esto es debido a que la primera componente es la única cuyo valor propio es mayor a la unidad ($\lambda_1 = 4.43973$), y explica un 74% de la varianza (estandarizada) total.

Para determinar los "pesos" de cada variable original en la primera componente necesitamos los valores del primer vector propio, obtenido con la matriz de correlaciones. Podemos observar los resultados del paquete estadístico:

Table of Component Weights

	Component 1
Crimen	0.253954
Población	-0.465707
Prep_Pol	-0.424735
P_Lib_Cond	-0.442906
Num_PP_Casa	-0.429753
T_Desempleo	0.3967

El *vector propio* es obtenido resolviendo:

$$(\boldsymbol{R} - \lambda_1 \mathbf{I})\boldsymbol{e_1} = (\boldsymbol{R} - 4.43973\,\mathbf{I})\boldsymbol{e_1} = 0$$

Donde $\boldsymbol{R}$ es la matriz de correlaciones y $\lambda_1 = 4.43973$ es el primer valor propio.

De donde:

$$\boldsymbol{e_1^t} = (0.253954 \quad -0.465707 \quad -0.424735 \quad -0.442906 \quad -0.429753 \quad 0.3967)$$

Calculamos el *componente principal* con el que trabajaremos:

$$Y_1 = \boldsymbol{e_1^t} * \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_5 \\ Z_6 \end{pmatrix}$$

Donde las Z_i son las variables estandarizadas del problema, es decir $\mathbb{E}(Z_i) = 0$ y $Var(Z_i) = 1$, para $i = 1,2,..,6$.

Por lo que la componente 1 nos queda:

$$Y_1 = 0.29056Z_1 - 0.495869Z_2 - 0.420906Z_3 - 0.446013Z_4 - 0.430832Z_5 + 0.328608Z_6$$

Podemos observar que las variables (estandarizadas) crimen (Z_1) y la tasa de desempleo (Z_6) tienen un peso positivo. Esto nos indica que el índice de violencia se incrementa si aumenta el número de crímenes o la tasa de desempleo.

Por otro lado, las variables (estandarizadas) población (Z_2), el presupuesto del SPD (Z_3), el número de personas en libertad condicional (Z_4) y el número de personas por conjunto habitacional (Z_5) tienen un peso negativo, por lo que el índice de violencia disminuye ante el aumento de cualquiera de dichas variables.