UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO
FACULTAD DE CIENCIAS

LA CORRUPCION, UNA ALTERNATIVA PARA LAS EMPRESAS

T E S I N A

QUE PRESENTA LA **LIC. ANA PATRICIA GUTIERREZ OLIVA**
PARA OBTENER EL DIPLOMA DE LA ESPECIALIZACION EN ESTADISTICA APLICADA

ASESOR: M.EN C. LETICIA GRACIA-MEDRANO

Índice

Introducción

Podemos pasar mucho tiempo tratando de encontrar una solución a los problemas, sin embargo en el momento de implementarlas éstas pueden variar. Por ejemplo, hace algunos años en nuestro país para obtener una licencia de manejo había que presentar y aprobar un examen, pagar los derechos y finalmente obtener la misma. Este sencillo proceso, que en realidad es una forma de que el gobierno coordine a los ciudadanos, se transformó en un verdadero oasis para la corrupción. Nadie hacía el examen de manejo, era más fácil pagar la mordida para omitir este paso y obtener rápidamente la licencia. De igual forma, se daba una pequeña cooperación a los "coyotes" y se omitía el pago de los derechos gubernamentales. Esta buena idea se transformó a la hora de implementarla en algo completamente distinto a lo que deseaba el gobierno; los ciudadanos no comprobaban que sabían manejar lo cual incrementó el índice de accidentes automovilísticos y por otro lado los ingresos gubernamentales por dicho servicio disminuyeron aunque la expedición de licencias aumentó.

Implícitamente el gobierno se comunica con los ciudadanos y las empresas a través de las políticas o leyes que desarrolla para coordinar a los mismos. Supongamos que estamos en San Ángel y deseamos ir a Ciudad Universitaria. Para llegar a este lugar seguimos determinada ruta conformada por algunas calles, en las cuales hay señales y semáforos que coordinan a los automovilistas evitando accidentes. En un mundo ideal el gobierno debería ser como las señales y los semáforos de una avenida. Debería coordinar los esfuerzos de las empresas para que pudiesen lograr sus objetivos a través del marco jurídico y de las políticas establecidas. Supongamos que hay una ruta principal para llegar a Ciudad Universitaria. Si los semáforos están sincronizados y las calles pavimentadas y en términos generales el acceso es rápido y sencillo entonces la mayoría de los automovilistas utilizarán dicha ruta. Pero qué pasa si este camino tiene muchos baches, los semáforos no funcionan o no están sincronizados, si hay que pagar una tarifa altísima por usar esta ruta, entonces los automovilistas buscarán otros caminos alternos para llegar a Ciudad Universitaria. Este ejemplo sirve para mostrar cómo determinadas políticas públicas incentivan a los ciudadanos para apegarse a las normas y reglas diseñadas por el gobierno y cómo estas mismas pueden hacer que los individuos recurran a caminos alternos conocidos como "corrupción". En este sentido la corrupción se ha convertido en una alternativa para los empresarios de manera que puedan obtener servicios o permisos gubernamentales.

¿Qué hace que una empresa opte por un camino o por otro? Haciendo a un lado la parte ética, en primer lugar la comunicación entre el gobierno y las empresas. Para tomar una ruta hay que conocerla. Segundo, si se toman en cuenta las necesidades del usuario y la ruta es diseñada con base en esto entonces habrá más incentivos a utilizarla. Tercero, que la ruta sea fácil, segura y barata. El objetivo de este trabajo es crear una medida que nos indique cuándo el gobierno se comunica con sus empresas, toma en cuenta sus necesidades y les hace el camino fácil, rápido y barato para obtener bienes y servicios gubernamentales y cuándo la corrupción se vuelve una alternativa. A esta variable la llamaremos "viabilidad legal". Retomemos el ejemplo de las licencias de manejo. Hoy en día en el Distrito Federal ya no es

necesario presentar el examen de manejo. Se llena una hoja donde el solicitante declara que sabe manejar. La licencia ya no tiene caducidad lo que reduce su costo a largo plazo y el pago se realiza en las sucursales bancarias para evitar el desvío de recursos. El trámite se ha vuelto fácil, rápido y barato. En este ejemplo el gobierno tiene una buena comunicación con los ciudadanos, la corrupción ha disminuido y en consecuencia la recaudación del gobierno local por este concepto ha aumentado de manera considerable.

No esperamos que todos los países que tengan buena comunicación con sus empresas no tengan corrupción, porque como hemos señalado, del dicho al hecho hay mucho trecho. Puede que dialoguen por horas con los empresarios pero en el momento de implementar los pasos a seguir o los trámites del proceso éstos sean tan lentos, difíciles, caros, o poco entendibles que provoquen la corrupción. De igual forma puede que los gobiernos no tomen en cuenta a las empresas pero que simplifiquen sus trámites de tal manera que las empresas prefieran seguir el camino "correcto".

Para crear este nuevo concepto utilizaremos una serie de variables observadas agrupadas en dos temas: comunicación legal y corrupción. Primero estudiaremos a través de estadísticas descriptivas la estructura de la base de datos. Con el fin de obtener la dimensión del espacio de los mismos aplicaremos el Análisis de Componentes Principales para finalmente utilizar el Análisis de Factores y crear la variable "viabilidad legal". Esto nos permitirá clasificar a los diferentes países dependiendo de los distintos escenarios.

La base de datos con la que trabajaremos fue obtenida del *World Development Report* realizado por el Banco Mundial de 1997. La información fue obtenida a través de una encuesta telefónica aplicada a los dueños o administradores de cientos de empresas en distintos países del mundo[1]. Los países encuestados fueron:

PAISES ENCUESTADOS

[1] La muestra fue obtenida a través de un muestreo bietápico para el cual se aplicó el muestreo aleatorio simple en ambas etapas. El cuestionario de la encuesta se presenta al final de este trabajo.

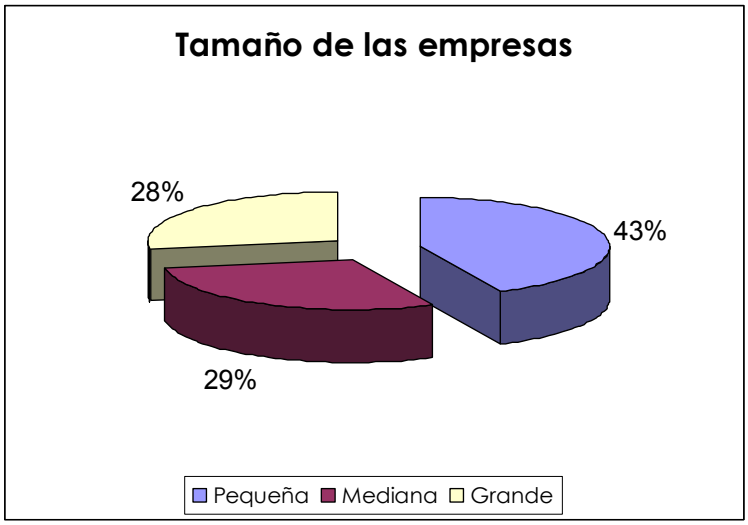
ALBANIA	CZECH REPUBLIC	ITALY	MAURITIUS	UNITED STATES
ARMENIA	GERMANY	JAMAICA	MALAWI	UZBEKISTAN
AUSTRIA	ECUADOR	JORDAN	MALAYSIA	VENEZUELA
AZERBAIJAN	SPAIN	KAZAKHSTAN	NIGERIA	WEST BANK
BENIN	ESTONIA	KENYA	POLAND	SOUTH AFRICA
BULGARIA	FIJI	KYRGYZ REPUBLIC	PORTUGAL	ZAMBIA
BELARUS	FRANCE	LITHUANIA	RUSSIA	ZIMBABWE
BOLIVIA	UNITED KINGDOM	LATVIA	SENEGAL	
CANADA	GEORGIA	MOROCCO	SLOVAK REPUBLIC	
SWITZERLAND	GHANA	MOLDOVA	CHAD	
COTE D'IVOIRE	GUINEA	MADAGASCAR	TOGO	
CAMEROON	GUINEA-BISSAU	MEXICO	TURKEY	
CONGO	HUNGARY	MACEDONIA	TANZANIA	
COLOMBIA	INDIA	MALI	UGANDA	
COSTA RICA	IRELAND	MOZAMBIQUE	UKRAINE	

Podemos observar que los países están distribuidos en todos los continentes. La encuesta realizada originalmente por el Banco Mundial cubría varios aspectos, no solamente marco legal y corrupción, sin embargo para efectos de este trabajo sólo se consideraron las siguientes 6 variables:

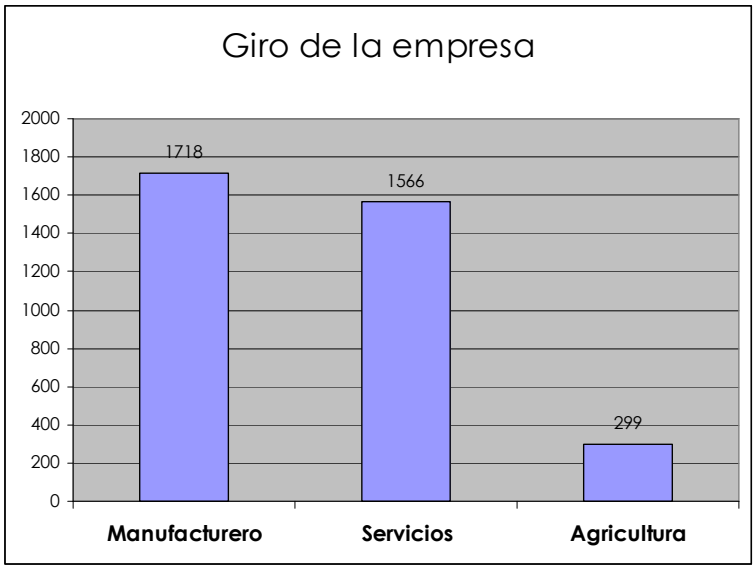
a) Promedio de cambios inesperados en las leyes, reglas o políticas que directamente afectan a las empresas ***(Cambio)***
b) Promedio de ocasiones en que esperan que el gobierno se apegue a las políticas generales anunciadas ***(Seguimiento)***
c) Promedio de veces en que el gobierno toma en cuenta las necesidades y preocupaciones manifestadas por la empresa o por la representación de su industria para la modificación de leyes o políticas que afectan la operatividad de la empresa ***(Consideración)***
d) Promedio en que las empresas de la misma línea de negocios tienen que hacer "pagos adicionales"[2] de forma irregular para lograr que las cosas se hagan ***(Mordida)***
e) Promedio de ocasiones en que las empresas de la misma línea de negocios saben con anticipación "cuanto hay que dar por los "pagos adicionales"***(Monto)***
f) Promedio de veces en las que aun hecho el "pago adicional" la empresa teme que se le pida más dinero, por ejemplo por otro servidor público***(Miedo)***

Todas las variables son cualitativas y fueron obtenidas para 3,583 empresas distribuidas en todo el mundo; para fines de este trabajo utilizaremos los promedios por país pues fue la información que era pública y estaba disponible. La encuesta fue aplicada a empresas de distinto tamaño como se muestra en la gráfica.

[2] Cuando se habla de "pagos adicionales" se refiere a sobornos, regalos, mordidas o cualquier pago fuera de la ley que se debe pagar a los trabajadores del gobierno para poder recibir un servicio público a cambio.



El tamaño de las empresas depende del número de empleados que laboran en ellas. Podemos observar que el 43% de las empresas es de tamaño pequeño. Es decir que en estas empresas laboran menos de 12 personas. Ahora bien, las empresas tienen distintos giros los cuales se presentan a continuación:



La mayoría de las empresas encuestadas fueron del área de manufactura y servicios. Entre estas dos categorías representan el 91% de los datos. En la siguiente sección se analizará la estructura de los mismos utilizando algunas estadísticas descriptivas.

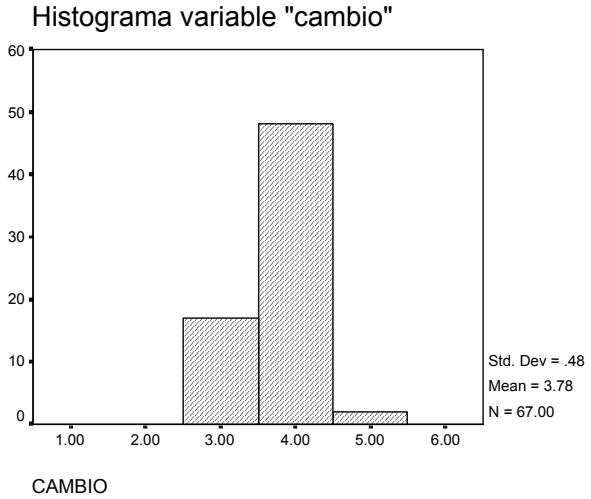
CAPITULO 1. Estadísticas Descriptivas

En este capítulo revisaremos la estructura de las variables elegidas. Primero estudiaremos algunas estadísticas descriptivas como media, moda, varianza a través del uso de histogramas. Después analizaremos la mediana y la presencia de observaciones atípicas con los diagramas de caja y brazos. Finalmente mostraremos el comportamiento en conjunto de las variables aplicando técnicas multivariadas como las Caritas de Chernoff.

Histogramas

Un histograma nos permite describir un conjunto de datos a través de la agrupación en clases. Al número de observaciones en cada clase se le llama frecuencia de clase y el cociente de la frecuencia de clase entre el número total de observaciones se le conoce como frecuencia relativa. Las frecuencias relativas nos permiten observar si existe algún patrón en el conjunto de observaciones. Ahora bien, esperaríamos que una muestra representativa de datos tuviese un comportamiento similar al de la población; por lo que la distribución de frecuencias de la misma nos daría una idea de cómo se comporta ésta.

A continuación se mostrarán los histogramas para cada una de las variables elegidas antecedidos por la pregunta correspondiente de la encuesta aplicada para facilitar la interpretación[1]. Comencemos con el histograma para la variable "**cambio**". La pregunta de la encuesta dice: "¿Tiene usted regularmente que lidiar con cambios inesperados en reglas, leyes o políticas que materialmente afectan su negocio?" Las respuestas van en escala donde 1 corresponde a "*completamente predecible*" y el 6 a "*completamente impredecible*".

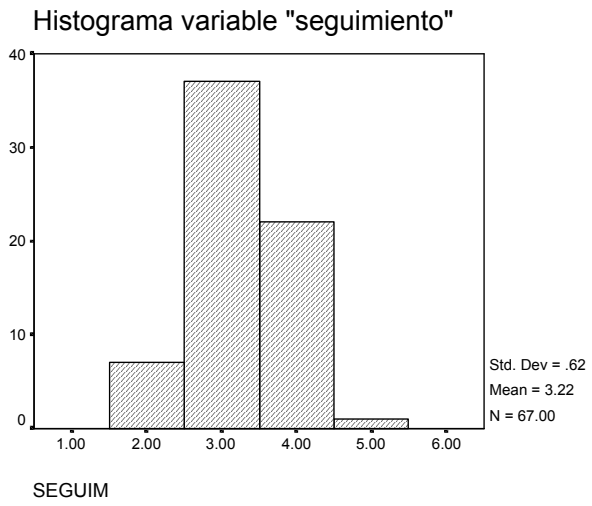


En esta gráfica podemos observar que los datos se concentran en las clases del centro, 3, 4 y 5. La frecuencia relativa más alta está en la categoría 4 (*apenas impredecible*). La media de esta

[1] En el Anexo 1 se muestra el cuestionario que se aplicó para obtener la información

variable[2] es 3.78 y la moda 3.93. Esto quiere decir que en promedio los empresarios piensan que es "*apenas impredecible*" saber si tendrán que lidiar con cambios legislativos que afecten directamente a su negocio. La variabilidad de los datos, medida por la varianza es de 0.2304, no hay una gran dispersión de los datos como se observa en la gráfica. Finalmente si miramos la gráfica podemos ver que los datos tienen una distribución asimétrica hacia la izquierda.

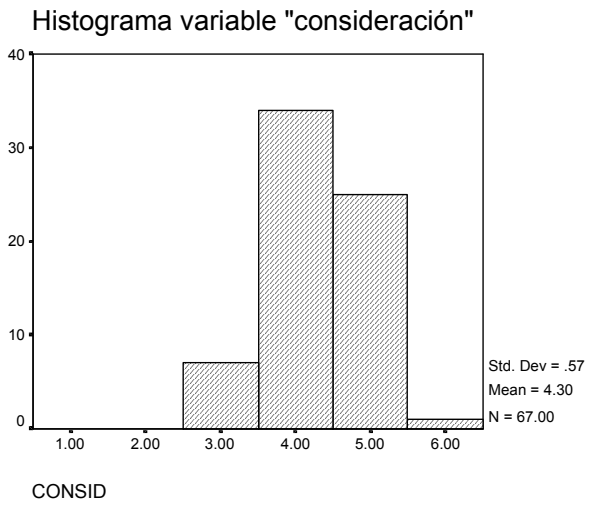
La pregunta para la variable "**seguimiento**" fue: "¿Espera usted que el gobierno se apegue a las grandes políticas anunciadas?" Las respuestas van desde "*siempre*" que es representado por el 1 hasta "*nunca*" ubicado en la categoría 6.



Para la variable "seguimiento" podemos ver que los datos se concentran en dos clases: 3 ("*frecuentemente*"), siendo ésta la que presenta la mayor frecuencia relativa y 4 ("*a veces*"). El promedio de esta variable 3.21 y la moda 3.12 coinciden en la misma clase para este histograma. Es decir que la mayoría de los empresarios dijeron que "*frecuentemente*" esperaban que el gobierno se apegara a las políticas anunciadas. La varianza es 0.382 y nos indica que los datos no están muy dispersos. Los datos tienen una distribución asimétrica hacia la derecha .

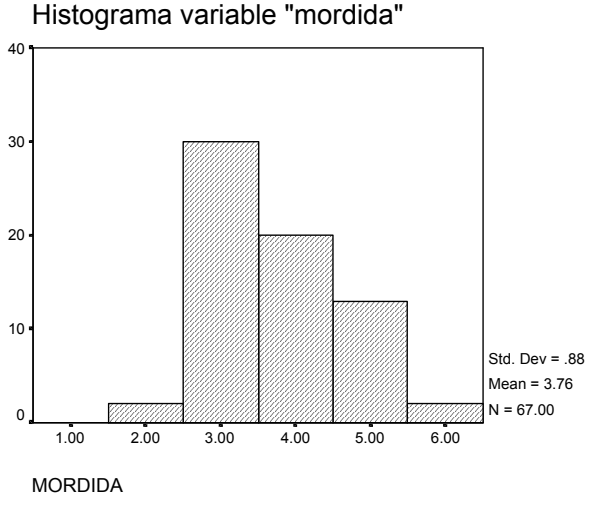
En el caso de la variable "**consideración**" la pregunta hecha fue: "En caso de importantes cambios a la ley o políticas que afectan la operación de mi negocio, el gobierno toma en cuenta mis preocupaciones expresadas por mi persona o por la asociación a la que pertenece mi negocio". Esto es cierto: 1 "siempre" hasta 6 "nunca".

[2] Para poder interpretar los resultados dentro del contexto de los datos redondearemos todos los promedios y modas con el siguiente criterio: cifras de 0 a 0.5 serán consideradas como la unidad menor más próxima. Los datos de 0.51 a 0.99 serán redondeados a la unidad mayor más próxima. Por ejemplo, el promedio de 3.45 será considerado como 3, y una moda de 3.56 será un 4. Sin embargo las gráficas fueron realizadas con los valores originales y únicamente se aplicó el criterio anterior para poder interpretar los resultados de los histogramas.



En el caso de la variable "consideración" podemos observar que los datos se concentran hacia la derecha. La media (4.31) y la moda (4.17) están en la clase 4 *(a veces)*. En términos de la encuesta quiere decir que en promedio *a veces* la autoridad toma en consideración su preocupación sobre cambios importantes en políticas o leyes que afectan la operatividad de sus empresas. La varianza es 0.3249 pues los datos se agrupan en las clases 3, 4 , 5 y 6. Tienen una distribución asimétrica cargada hacia la derecha.

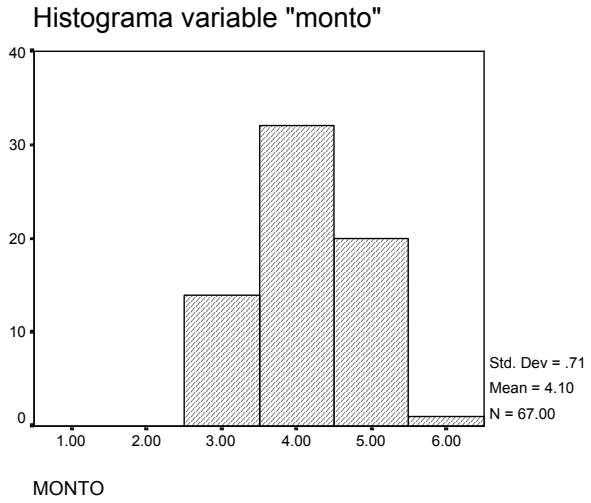
En cuanto a las variables que tienen que ver con corrupción, la pregunta para la variable "**mordida**" fue: "Es común para las empresas de mi línea de negocios que tengamos que realizar algunos pagos adicionales irregulares al gobierno para que las cosas se hagan". Las respuestas van desde 1 "siempre" hasta 6 "nunca".



La variable "**mordida**" tiene una distribución asimétrica a la derecha. La moda (3.3) está en la clase 3, es decir que la mayoría de los empresarios respondió que "frecuentemente" tienen que pagar mordida para que las cosas se hagan. Sin embargo, el promedio (3.75) está en la clase 4; es decir que en promedio los empresarios "a veces" tienen que pagar mordida. La clase 2 representa los países que en promedio "la mayor parte de las veces" hay que pagar mordida,

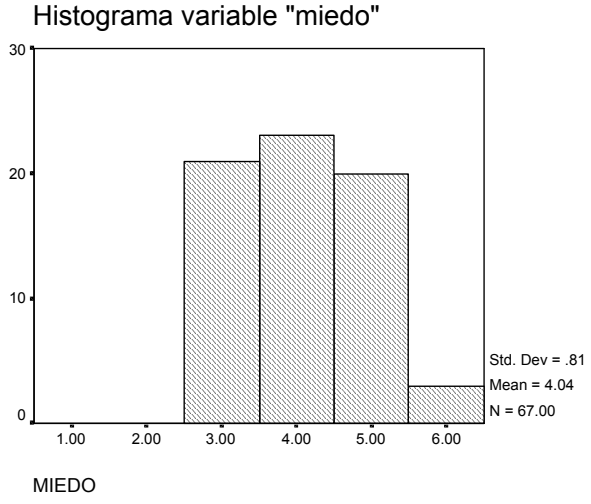
sin embargo también hay países donde "nunca" hay que pagar extra para que las cosas se hagan como lo muestra la clase 6. Esta variabilidad la representa la varianza que es 0.778

Para la variable "monto" la pregunta es: "Empresas de mi línea de negocios usualmente saben de antemano cuánto es el pago adicional". Las respuestas corresponden a la misma escala de la variable "mordida".



Para la variable "monto" los datos están concentrados hacia la derecha. La moda (2.84) está en la clase 3 y el promedio (4.04) está en la clase 4. La mayoría de los empresarios respondieron que "a veces" saben con anticipación el monto de la mordida. Como podemos observar en la clase 6, al menos en promedio en algún país "nunca" saben cuánto será el monto, lo cual indica una enorme inestabilidad legal.

Finalmente para "**miedo**" la pregunta es: "Incluso si mi empresa ha realizado los pagos adicionales siempre tengo miedo que me sea pedido más dinero por ejemplo por otro servidor público". La escala es igual que la variable "mordida".



Los datos para la variable "miedo" se concentran hacia la derecha. La media y la moda coinciden en 4, es decir los empresarios respondieron que "a veces" aunque hayan pagado mordida tienen miedo de que les pidan otro pago adicional. La varianza es de 0.65 y la distribución de los datos podríamos decir que es casi simétrica y muy parecida a una distribución uniforme.

En la siguiente tabla se resumen los estadísticos descriptivos analizados en esta sección.

RESUMEN ESTADISTICAS DESCRIPTIVAS

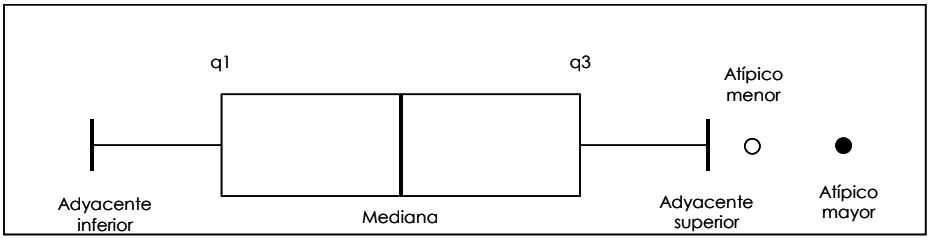
	CAMBIO	SEGUIM	INFORMA	CONSID	MORDIDA	MONTO	MIEDO
Media	3.7846	3.2173	3.9421	4.3046	3.7585	4.0954	4.0427
Moda	3.93	2.56	3.89	4.17	3.33	3.89	2.84
Varianza	.23452	.38174	.36513	.32495	.77831	.50676	.65088

Ahora nos interesa ver si existen observaciones atípicas o sospechosas en nuestra base de datos, para lo cual utilizaremos los diagramas de caja y brazos.

Diagramas de caja y brazos

El diagrama de caja y brazos es una forma conveniente de graficar los siguientes 5 datos: el valor más pequeño y el más alto que no sean observaciones atípicas, el primer y tercer cuartil así como la mediana. Podemos mostrar con esta herramienta la localización, simetría y dispersión de los datos así como la existencia de observaciones atípicas. En el siguiente gráfico se muestran los elementos de un diagrama de caja y brazos.

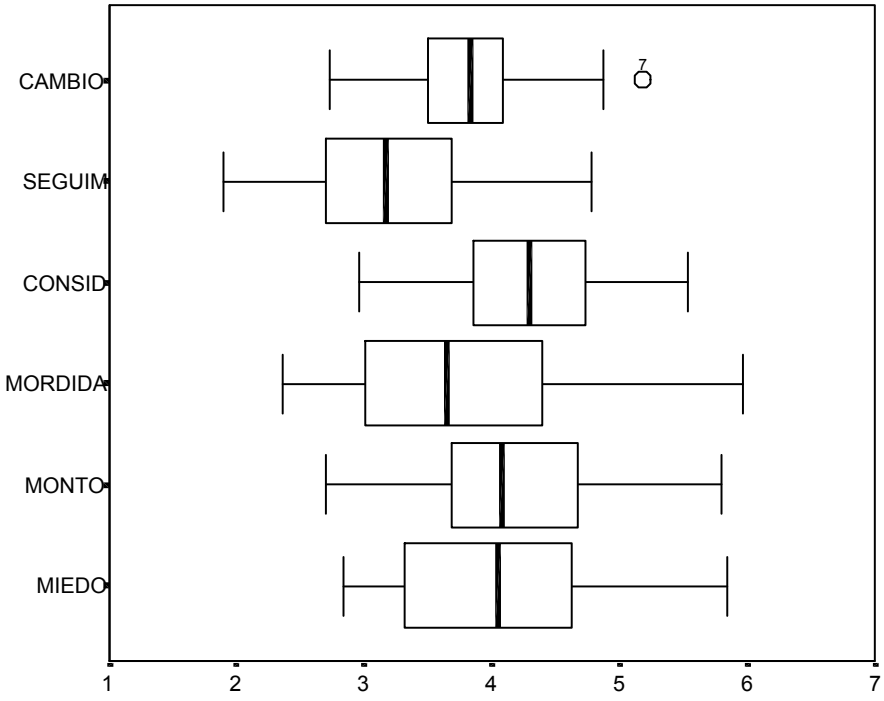
ELEMENTOS DE UN DIGARAMA DE CAJA Y BRAZOS



La caja contiene el 50% de los datos, el borde de la caja de lado izquierdo marca el cuartil 25 (q1) y el borde derecho el cuartil 75 (q3). Al rango que existe entre q1 y q3 se le conoce como rango intercuartílico. Ahora bien, la raya dentro de la caja marca la mediana de los datos. Cuando la línea no está en el centro de la caja indica que la distribución de los datos es asimétrica. Las rayas horizontes de los brazos muestran el valor más pequeño y el más grande de los datos (adyacente inferior y superior).

En el caso que se presentan observaciones atípicas los brazos se extienden hasta el 1.5 del rango intercuartílico. Estos deben terminar en un valor observado de la base de datos. Los puntos fuera de los brazos son considerados observaciones sospechosas. Cuando un dato está a menos de 3 veces el rango intercuartílico, a partir de la mediana, es considerado un atípico menor y cuanto el dato excede este límite es considerado un atípico mayor.

DIAGRAMA DE CAJA Y BRAZOS POR VARIABLE



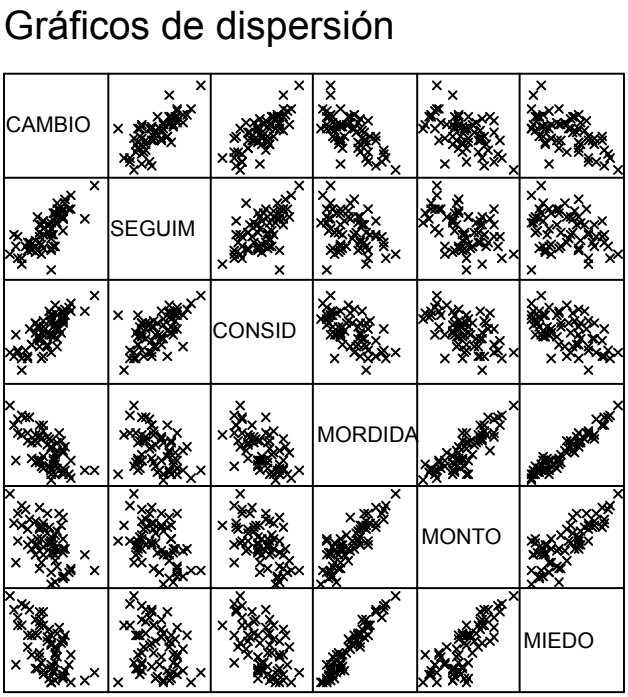
Todas las variables, excepto consideración, tienen una mediana similar pues este valor se localiza entre el rango entre 3 y 4 aproximadamente[3]. La variable con mayor variabilidad es la de "mordida". En los histogramas pudimos constatar que había tanto países en los que nunca se pagaba mordida así como aquellos donde la mayoría de las veces había que recurrir a dicha práctica para que las cosas fueran hechas. Por otro lado, la variable con menor dispersión es la de "cambio" donde todos los datos son cercanos a 4; los empresarios opinaron que era "apenas impredecible" saber si tendrían que lidiar con cambios no esperados en las leyes que afectaban directamente su negocio. En esta misma variable observamos un dato atípico correspondiente a Bielorrusia. En dicho país los empresarios contestaron que era "altamente impredecible saber si lidiarían con cambios legales no esperados". Para 1997 la U.R.S.S. acababa de fragmentarse y Bielorrusia fue uno de los países nuevos que se formaron. La vorágine de formar parte de una potencia mundial y repentinamente convertirse en un pequeño país puede justificar la poca estabilidad del marco jurídico así como falta de aplicación de la ley. Baste decir que hoy día gran parte de la mafia rusa que comercializa con armamento, drogas y prostitución se ha reubicado en este país. Por esta razón continuaremos el análisis sin eliminar a Bielorrusia del estudio.

[3] En esta tabla se resumen los valores de los estadísticos descriptivos:

		CAMBIO	SEGUIM	INFORMA	CONSID	MORDIDA	MONTO	MIEDO
Mediana		3.8300	3.1700	3.9800	4.3000	3.6500	4.0700	4.0400
Minimum		2.72	1.89	2.40	2.96	2.36	2.70	2.84
Maximum		5.18	4.78	5.06	5.54	5.96	5.80	5.84
Cuartiles	**q1(25)**	3.4900	2.6900	3.5200	3.8400	3.0000	3.6800	3.3100
	q3 (75)	4.0900	3.7000	4.3900	4.7400	4.4100	4.6700	4.6700

Matriz de diagramas de dispersión

Para saber si existe una correlación entre las variables utilizaremos la matriz de diagramas de dispersión. Este gráfico resume las correlaciones entre las diferentes variables.



MATRIZ DE CORRELACIONES

	Cambio	Seguimiento	Consideracion	Mordida	Monto	Miedo
Cambio	1	0.7348612	0.738714	-0.6368237	-0.5143172	-0.5784629
Seguimiento	0.7348612	1	0.5805847	-0.3516951	-0.418938	-0.278515
Consideracion	0.738714	0.5805847	1	-0.6006639	-0.5341112	-0.525509
Mordida	-0.6368237	-0.3516951	-0.6006639	1	0.8533771	0.9625132
Monto	-0.5143172	-0.418938	-0.5341112	0.8533771	1	0.7844273
Miedo	-0.5784629	-0.278515	-0.525509	0.9625132	0.7844273	1

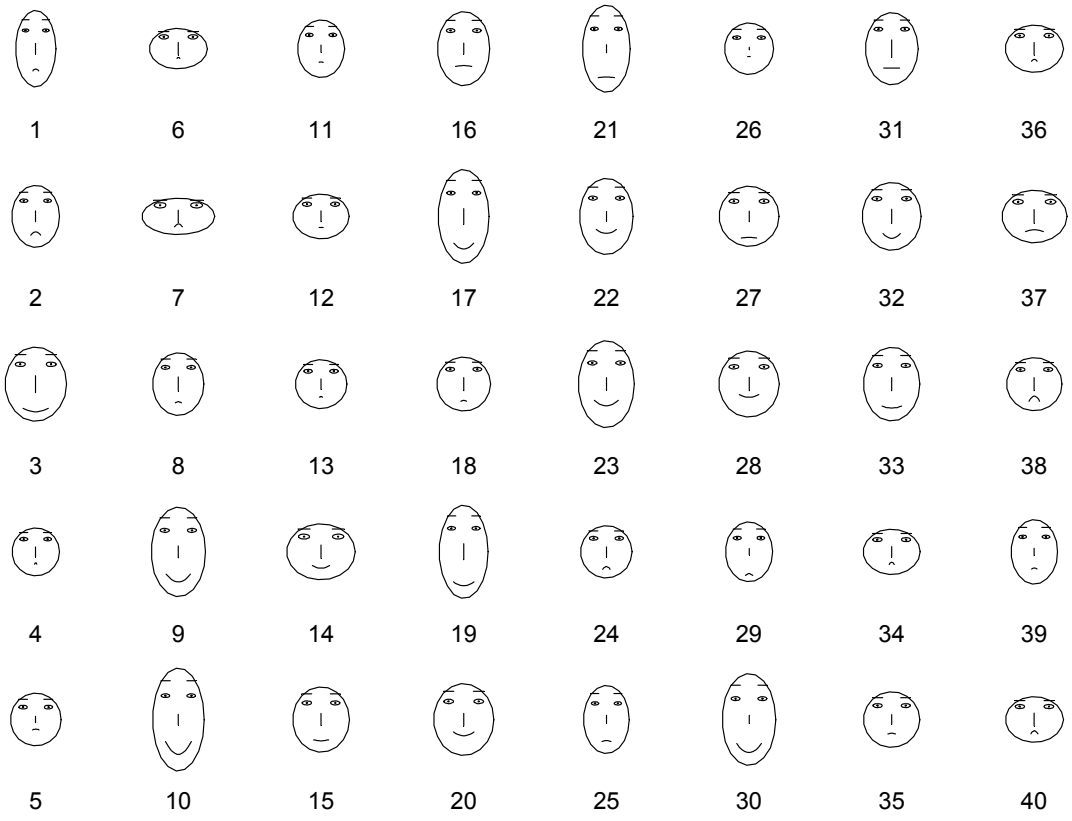
En el gráfico anterior podemos observar que las variables están asociadas linealmente entre sí. En el caso de "mordida" con "miedo" los datos forman casi una recta perfecta, lo cual indica una alta correlación y ya que la recta tiene pendiente positiva, la asociación de los datos también es positiva. En otras palabras a medida que aumenta el pago de "mordidas" se espera que una vez pagado este monto algún otro servidor público les pida a los empresarios más dinero. La tabla nos muestra las cifras para las correlaciones entre las variables. Esta medida de asociación lineal entre dos variables va de -1 a 1, donde 1 es una correlación positiva perfecta entre 2 variables y -1 indica una correlación negativa perfecta entre las variables. La correlación entre "mordida" y "cambio" es de -0.6368, lo cual quiere decir que a medida que las empresas no tienen que lidiar con cambios legales que afectan directamente a su negocio la corrupción disminuye. A pesar de que esta técnica nos permite observar todas las variables no sabemos cómo se comportan en conjunto por lo cual utilizaremos las Caritas de Chernoff.

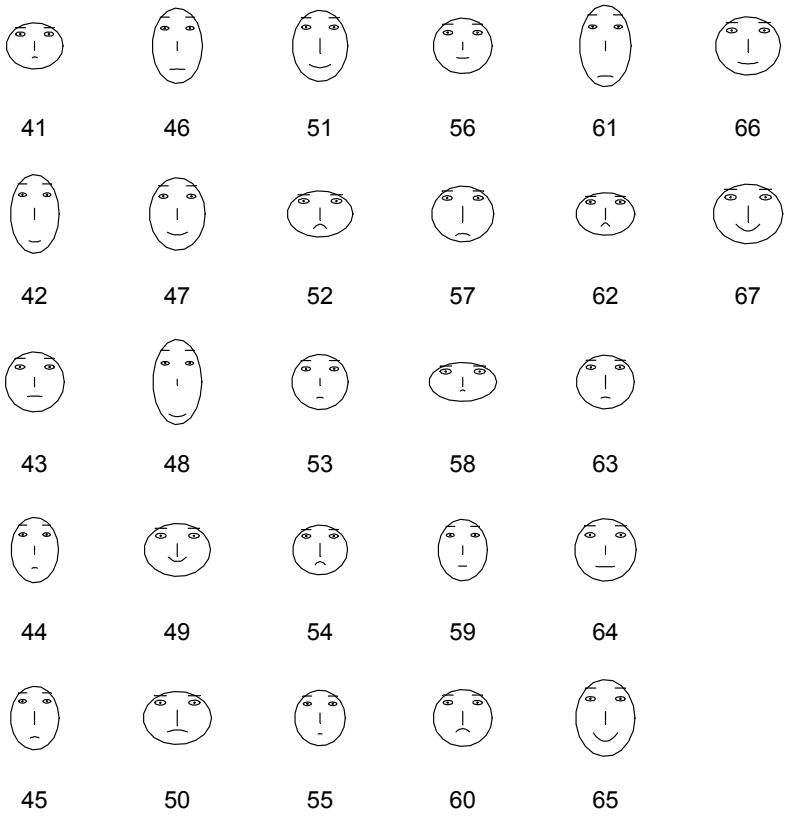
Caritas de Chernoff

En esta técnica cada unidad de estudio corresponde a una carita. Las características de la carita están dadas por los valores particulares de las variables. En el siguiente cuadro se presentan las variables que corresponden a cada rasgo facial y después las caritas para cada uno de los 67 países.

Variable	Mordida	Cambio	Seguimiento	Consideración	Monto	Miedo
Rasgo facial	Área cara	Forma cara	Tamaño nariz	Ubicación boca	Curva sonrisa	Ancho boca

CARITAS DE CHERNOFF





Caritas de Chernoff y país que representan

1	ALB	16	CZE	31	ITA	46	MUS	61	USA
2	ARM	17	DEU	32	JAM	47	MWI	62	UZB
3	AUT	18	ECU	33	JOR	48	MYS	63	VEN
4	AZE	19	ESP	34	KAZ	49	NGA	64	WTB
5	BEN	20	EST	35	KEN	50	POL	65	ZAF
6	BGR	21	FJI	36	KGZ	51	PRT	66	ZMB
7	BLR	22	FRA	37	LTU	52	RUS	67	ZWE
8	BOL	23	GBR	38	LVA	53	SEN		
9	CAN	24	GEO	39	MAR	54	SVK		
10	CHE	25	GHA	40	MDA	55	TCD		
11	CIV	26	GIN	41	MDG	56	TGO		
12	CMR	27	GNB	42	MEX	57	TUR		
13	COG	28	HUN	43	MKD	58	TZA		
14	COL	29	IND	44	MLI	59	UGA		
15	CRI	30	IRL	45	MOZ	60	UKR		

En la gráfica podemos observar que entre más gorda y cachetona es la carita, mayor es la corrupción y menor la comunicación entre el gobierno y las empresas. La curvatura de la boca indica si las empresas saben por adelantado cuánto hay que pagar de mordida: si están tristes

saben cuánto hay que pagar, si sonríen quiere decir que no saben cuánto hay que pagar[4]. Por ejemplo la carita 37 Lituania. La cara redonda indica que hay mucha corrupción, la curvatura de la sonrisa hacia abajo indica que saben cuánto hay que pagar de mordida, la nariz larga indica que el gobierno no considera a las empresas cuando lleva a cabo los cambios en la legislación. Es decir que las caritas gordas y tristes representan países donde hay mucha corrupción y mala comunicación entre el gobierno y las empresas. Ahora miremos la carita rotulada con el número 9, representa a Canadá. Es una carita delgada y sonriente, lo cual indica que hay poca corrupción y la comunicación entre el gobierno y las empresas es buena.

[4] No saben cuánto hay que pagar porque en su país la corrupción es muy baja, entonces al no existir este fenómeno no saben el monto del mismo.

CAPITULO 2. Análisis de Componentes Principales

El objetivo de este trabajo es crear una nueva medida denominada "viabilidad legal" a través del Análisis de Factores, sin embargo para conocer la dimensión de los datos utilizaremos primero el Análisis de Componentes Principales. En esta sección se explicará teóricamente cómo funciona y al final de la sección se correrá con la base de datos.

Algunas técnicas estadísticas como la regresión lineal piden que se cumplan algunos supuestos como la independencia lineal entre las variables para evitar la multicolinealidad. Cuando las variables originales están correlacionadas podemos transformarlas en un nuevo conjunto de variables no correlacionadas llamadas Componentes Principales. Estas "nuevas" variables son combinaciones lineales de las variables originales, de tal forma que la primera componente principal explica tanta variación en los datos como sea posible y así sucesivamente[1]. Es una técnica dirigida por las variables y funciona cuando éstas tienen el mismo peso para el análisis; es decir no hay una variable dependiente y varias independientes como el caso de la regresión lineal.

Se recomienda que el Análisis de Componentes Principales se utilice como técnica exploratoria de manera que el investigador tenga una imagen general del conjunto de datos. Los objetivos principales de esta herramienta son reducir la dimensionalidad del conjunto de datos e identificar nuevas variables subyacentes.

En el caso del primer objetivo se trata de descubrir la verdadera dimensión del conjunto de datos, si ésta es menor a las "p" variables originales, entonces se pueden sustituir los datos por un número menor de variables subyacentes sin perder información. Para el segundo objetivo, las nuevas variables no siempre tienen significado y por lo tanto solo en algunas ocasiones se pueden interpretar.

Método para obtener las Componentes Principales

Como ya se mencionó deseamos que las variables Componentes Principales no estén correlacionadas y que la primera componente explique tanta variabilidad en los datos como sea posible de tal forma que las subsiguientes Componentes Principales vayan tomando tanta variabilidad restante como les sea posible.

---

[1] Esta es una de las principales aplicaciones del Análisis de Componentes Principales, sin embargo hay otras razones:
a) Exploración de datos: nos permite comprobar la hipótesis establecida sobre el conjunto de datos y para identificar posibles observaciones "sospechosas" en el conjunto de datos.
b) Agrupación : en ocasiones este análisis es útil cuando se quieren formar subgrupos de unidades experimentales.

Supongamos que $\mathbf{X}^T = [X_1, ..., X_p]$ es una variable aleatoria *p*-dimensional con media $\boldsymbol{\mu}$ y matriz de covarianzas $\boldsymbol{\Sigma}$. Lo que necesitamos es encontrar un nuevo conjunto de variables, llamadas $Y_1, Y_2, ...,Y_p$, que no estén correlacionadas y cuyas varianzas sean decrecientes de la primera a la última. Cada Y_j será una combinación lineal de las *X*'s, de manera que

$$Y_j = a_{1j}X_1 + a_{2j}X_2 + ... + a_{pj}X_p \qquad = \mathbf{a}_j^T\mathbf{X} \qquad (1)$$

donde $\mathbf{a}_j^T = [a_{1j,...,}a_{pj}]$ es un vector de constantes y además pedimos[2] que $\mathbf{a}_j^T\mathbf{a}_j = \sum_{k=1}^{p} a_{kj}^2 = 1$.

La primera Componente Principal, Y_1, se encuentra escogiendo $\mathbf{a}_1$ de tal forma que la varianza de Y_1 se maximice. Se selecciona la $\mathbf{a}_1$ que maximice la varianza de $\mathbf{a}_1^T\mathbf{X}$ sujeta a la condición $\mathbf{a}_1^T\mathbf{a}_1$=1. El valor máximo de la varianza de $\mathbf{a}_1^T\mathbf{X}$ entre todos los vectores $\mathbf{a}_1$ que satisfacen $\mathbf{a}_1^T\mathbf{a}_1$=1 es igual al eigenvalor más grande de Σ, llamado λ_1. Este máximo ocurre cuando $\mathbf{a}_1$ es el eigenvector de Σ correspondiente al eigenvalor λ_1. La segunda Componente Principal, Y_2, se encuentra eligiendo $\mathbf{a}_2$ de modo que la varianza de $\mathbf{a}_2^T\mathbf{X}$ sea un máximo entre todas las combinaciones lineales de $\mathbf{X}$ que no están correlacionadas con la primera variable Componente Principal, $\mathbf{a}_1^T\mathbf{a}_2 = 0$ y tenga $\mathbf{a}_2^T\mathbf{a}_2$=1. De igual forma el máximo es igual a λ_2, , el segundo eigenvector de Σ correspondiente al eigenvalor λ_2. Y así sucesivamente se van obteniendo todas las Componentes Principales. En consecuencia $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$ denotan los eigenvalores ordenados de Σ y $\mathbf{a}_1$, $\mathbf{a}_2$, ..., $\mathbf{a}_p$ denotan los correspondientes eigenvectores normalizados[3].

Esperamos que la primera Componente Principal explique tanta variabilidad de los datos originales como sea posible y de forma análoga para las subsiguientes componentes. Por esta razón definiremos la varianza explicada. Sea A la matriz de pxp de eigenvectores para la cual

$$A = [\mathbf{a}_1,...,\mathbf{a}_p]$$

y el vector de *p*×1 de Componentes Principales llamado $\mathbf{Y}$. Entonces,

$$\mathbf{Y} = A^T\mathbf{X} \qquad (2)$$

[2] Esta normalización asegura que las distancias en el *p*-espacio se preserven.

[3] Hay que señalar que los p eigenvalores de Σ deben ser todos no negativos debido a que Σ es definida positiva.

La matriz de covarianzas de **Y** se definirá como Λ y está dada por:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & . & . & . & 0 \\ 0 & \lambda_2 & . & . & . & 0 \\ . & & & & & \\ . & & & & & \\ . & & & & & \\ 0 & & . & . & . & \lambda_p \end{bmatrix}$$

Esta matriz es diagonal pues los componentes elegidos no están correlacionados. Los eigenvalores deben interpretarse como las varianzas de cada Componente Principal. La traza[4] mide la variación total de las variables originales y de igual forma, la suma de las varianzas de los Componentes Principales se obtiene al calcular la traza de Λ.

$$\sum_{i=1}^{p} \mathrm{Var}(Y_i) = \sum_{i=1}^{p} \lambda_i = \mathrm{tr}(\Lambda)$$

Por otro lado sabemos que:

$$\mathrm{tr}(\Lambda) = \mathrm{tr}(\Sigma) = \sum_{i=1}^{p} \mathrm{Var}(X_i)$$

La suma de las varianzas de las variables originales y la suma de varianzas de las Componentes Principales son iguales. La variación total explicada por las variables Componentes Principales es igual a la cantidad total de la variación medida por las variables originales.

Podemos decir que la i-ésima Componente Principal explica una determinada proporción del total de la variación total de los datos originales definida como:

$$\lambda_i / \sum_{j=1}^{p} \lambda_j$$

Con el propósito de utilizar las variables Componentes Principales en otros estudios es necesario calcular las calificaciones de tales componentes para cada unidad de estudio. Dichas calificaciones nos proporcionan la ubicación de cada una de las "n" observaciones con respecto a sus ejes Componentes Principales. Sea x_r el vector de variables medidas para la r-ésima unidad de estudio. Entonces la calificación de la j-ésima variable Componente Principal, para la r-ésima unidad de estudio esta dada por:

---

[4] Recordemos que la traza de una matriz se obtiene al sumar la diagonal de tal forma que:
$\mathrm{tr}(\Sigma) = \sigma_{11} + \sigma_{22} + \ldots + \sigma_{pp}$.

$$y_{rj} = \mathbf{a}_j^T \mathbf{x}_r \qquad \text{para } j = 1, 2, ..., p \text{ y } r = 1, 2, ..., n$$

Hasta ahora hemos supuesto que conocemos la matriz de covarianzas Σ, sin embargo casi nunca contamos con esta información. Por esta razón en lugar de utilizar la matriz de covarianzas utilizaremos la matriz de correlaciones muestral **R**. El procedimiento es el mismo que anteriormente hemos descrito, las Componentes Principales de **X** serán los eigenvectores de **R**. Sean $\hat{\lambda}_1, \hat{\lambda}_2, ..., \hat{\lambda}_p$, los eigenvectores de **R** ordenados de mayor a menor y $\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, ..., \hat{\mathbf{a}}_p$ los eigenvectores correspondientes. Dado que **R** es semidefinida positiva, entonces los eigenvalores son todos no negativos y representan las varianzas estimadas de las distintas componentes.

En el caso que estemos trabajando con una muestra aleatoria, entonces $\{\hat{\lambda}_i\}$ y $\{\hat{\mathbf{a}}_i\}$ pueden considerarse como estimadores de los eigenvalores y eigenvectores de Σ, siendo estos estimadores de las Componentes Principales de **X** [5]. En este caso las calificaciones de las Componentes Principales se estiman por:

$$y_{rj} = \hat{\mathbf{a}}_j \mathbf{x}_r \qquad \text{para } j = 1, 2, ..., p \text{ y } r = 1, 2, ..., n$$

Finalmente necesitamos determinar la dimensión real del espacio en el que se encuentran los datos, es decir el número de Componentes Principales que tienen varianza mayor a cero. Hay dos métodos que explicaremos a continuación:

A) Buscamos el menor de los valores en k, donde k=1,2,..p que por primera vez ***V*** sobrepasa el γ100% de la variabilidad total de las variables originales. Sea $V = \sum_{j=1}^{k} \lambda_j / \sum_{j=1}^{p} \lambda_j$ para valores sucesivos de $k = 1, ..., p$. [6]

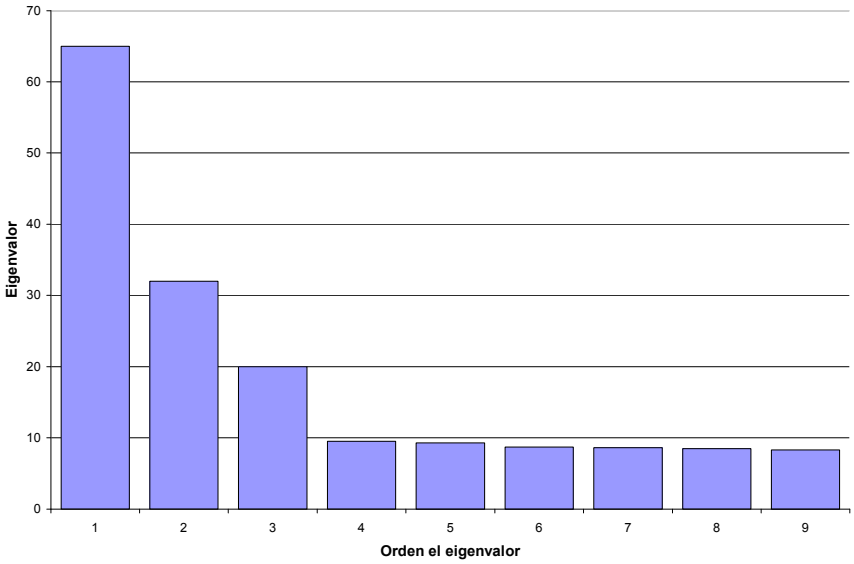
B) Utilizar la gráfica SCREE, la cual se construye graficando los eigenvalores en parejas $(1, \hat{\lambda}_1)$, $(2, \hat{\lambda}_2)$,..., $(p, \hat{\lambda}_p)$. Cuando los puntos de la gráfica se nivelan, entonces estos

[5] Lamentablemente no se establecen supuestos sobre la población subyacente, lo cual hace imposible derivar las propiedades muestrales de los estimadores.

[6] Cuando se trabaja con datos donde todas las variables están controladas como experimentos de laboratorio, puede resultar bastante fácil explicar más de 95% de la variabilidad total con sólo dos o tres componentes principales. Sin embargo cuando trabajamos con datos para "individuos" es posible que se requieran cinco o seis componentes principales para explicar más de 70 al 75% de la variación total. Por desgracia, entre más componentes principales se requieran, menos útil es cada una de ellas.

eigenvalores son tan pequeños que probablemente estén midiendo ruido aleatorio y podemos ignorarlos. A continuación se muestra una gráfica SCREE.

GRAFICA SCREE



En este caso, podemos observar que a partir del cuarto eigenvalor los puntos se nivelan, en este ejemplo el número adecuado de Componentes Principales a utilizar sería tres.

Análisis de Componentes Principales para datos estandarizados

¿Qué hacer si las variables que analizamos no están medidas en las mismas unidades o por lo menos en unidades comparables o las varianzas no tienen tamaños semejantes? Si omitimos este detalle, un cambio de escala o si una de las variables tiene una varianza mucho más grande que las demás el análisis se alterará, pues este rasgo dominará la primera Componente Principal sin importar la estructura de covarianzas. Por eso los investigadores sugieren aplicar el Análisis de Componentes Principales a la matriz de correlación, lo cual es equivalente a utilizar datos estandarizados[7].

En este caso, al igual que los anteriores, las Componentes Principales son los eigenvalores y eigenvectores de **P**, la matriz de correlación. Sean $\lambda_1^+ \geq \lambda_2^+ \geq \ldots \geq \lambda_p^+$ los eigenvalores de **P** y $\mathbf{a}_1^+, \mathbf{a}_2^+, \ldots, \mathbf{a}_p^+$ los eigenvectores respectivos.

[7] Al llevar a cabo una estandarización, estamos haciendo que las variables se midan en unidades comparables

En este caso las calificaciones de las Componentes Principales se calculan a partir de los valores de la variable estandarizada (valores Z). Entonces la calificación de la j-ésima Componente Principal para la r-ésima unidad de estudio es:

$$y_{rj}^{+} = \mathbf{a}_j^{+T}\mathbf{z}_r \qquad \text{para } j = 1,2,...,p \text{ y } r = 1,...,n$$

De igual forma las correlaciones entre las variables originales y la j-ésima Componente Principal están dadas por $\mathbf{c}_j^{+} = \lambda_j^{+1/2}\mathbf{a}_j^{+}$. Y estos vectores se llaman Vectores de Correlaciones de Componentes[8].

Si trabajamos con una muestra y la matriz de correlación **R**; los estimadores serán $\hat{\lambda}_j^{+}$ y $\hat{\mathbf{a}}_j^{+}$,los eigenvalores y eigenvectores. Las calificaciones de las Componentes Principales se calculan a partir de los valores **Z**, utilizando la siguiente fórmula:

$$y_{rj}^{+} = \hat{\mathbf{a}}_j^{+}\mathbf{z}_r \qquad \text{para } j = 1,2,...,p \text{ y } r = 1,...,n$$

Cuando utilizamos la matriz de correlación se pueden utilizar los métodos descritos anteriormente para determinar la dimensión del espacio en el que se encuentran los datos. Además existe en este caso otro método el cual consiste en buscar aquellos eigenvalores mayores a uno. El resultado de esta búsqueda será la dimensión. Se puede utilizar este método porque al trabajar con datos estandarizados, la varianza de cada variable es igual a 1. Ahora bien, si una Componente Principal no puede explicar más variación que una variable original por sí misma, entonces es probable que no sea importante.

Finalmente, algunos paquetes estadísticos cuando utilizan el análisis de Componentes Principales, arrojan una gráfica llamada Biplot. Esta es una gráfica en la cual tanto las variables como las observaciones son representadas en un espacio bidimensional definido por las primeras 2 Componentes Principales y es significativa cuando la dimensión de los datos es dos. Entonces en los ejes aparecen los valores de la primera y segunda Componente Principal. Además se presentan los vectores de correlación para cada variable. Si observamos los vectores con respecto a la primera Componente Principal (de izquierda a derecha), aquellos que apuntan hacia la derecha indican que esas variables están asociadas positivamente. De forma análoga, si los vectores apuntan hacia la izquierda, esos vectores están correlacionados de forma negativa con respecto a la primera Componente Principal. Ahora bien, si miramos los vectores con respecto a la segunda Componente Principal (de abajo a arriba), los vectores que apuntan hacia arriba tienen una correlación positiva y los que apuntan hacia abajo una correlación negativa. Para saber qué tan relacionadas están las variables utilizamos el coseno del ángulo formado entre dos vectores. Si el ángulo es pequeño el coseno del mismo es muy grande y positiva, lo cual indica una relación alta y positiva. Si el ángulo es grande, el coseno

[8] Nótese que c_{ij}^{+} es el *i*-ésimo elemento del *j*-ésimo vector de carga de componentes.

del mismo será grande y negativo, de igual forma indica una relación negativa. Finalmente si el ángulo es cercano a 90° quiere decir que las variables no están altamente correlacionadas pues el coseno será cercano a cero. El tamaño de los vectores indica la variabilidad por variable, entre más grande sea la flecha mayor será la varianza. Un ejemplo de esta gráfica se desarrollará más adelante.

Aplicación del Análisis de Componentes Principales

Nuestra base de datos consiste de una muestra de países dentro de los cuales fueron encuestados algunos empresarios y la información disponible fue el promedio por variable. Todas las variables son cualitativas medidas en una escala del 1 al 6; sin embargo sus varianzas no son semejantes como se muestra en la siguiente tabla.

VARIANZAS POR VARIABLE

	Cambio	Seguimiento	Consideracion	Mordida	Monto	Miedo
Varianza	0.2345222	0.3817351	0.3249464	0.7783129	0.5067555	0.6508806

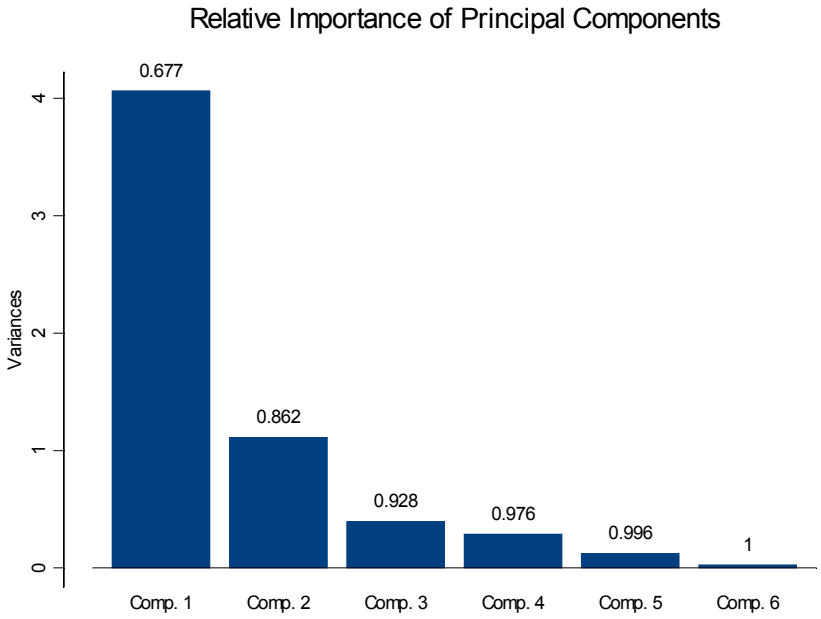
La varianza de "cambio" es muy pequeña en comparación con la de "mordida" por esta razón aplicaremos el Análisis de Componentes Principales a la matriz de correlaciones, es decir estandarizando los datos. Lo primero que obtenemos son los valores de éstas.

COMPONENTES PRINCIPALES

Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6
4.063115024	1.109491449	0.394481848	0.286690425	0.121207517	0.025014174

Para cada variable tenemos un eigenvalor los cuales aparecen ordenados de mayor a menor. La dimensión del espacio la dan aquellos eigenvalores mayores a 1 porque estamos trabajando con datos estandarizados. Las Componentes Principales menores a 1 explican muy poco sobre los datos por lo cual las desecharemos y sólo utilizaremos las primeras dos. Sin embargo, no sabemos cuánta de la variabilidad explican las Componentes Principales por lo que utilizaremos la gráfica Scree.

GRAFICA SCREE



En el eje horizontal se muestran los eigenvalores y en el eje vertical las varianzas. La primera Componente Principal explica el 67% de la variabilidad y entre esta y la segunda Componente Principal explican el 86%. Dado que no estamos trabajando con experimentos controlados nos parece muy razonable seguir utilizando únicamente las primeras dos.

La siguiente tabla muestra las cargas para cada variable.

CARGAS POR COMPONENTE PRINCIPAL

	Comp.1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6
Cambio	-0.418	0.368	0.128	-0.555	-0.6	
Seguimiento	-0.324	0.618	-0.565		0.435	
Consideracion	-0.397	0.29	0.697	0.5	0.146	
Mordida	0.453	0.358		0.151	-0.141	0.79
Monto	0.419	0.304	0.42	-0.55	0.47	-0.179
Miedo	0.426	0.421		0.338	-0.435	-0.579

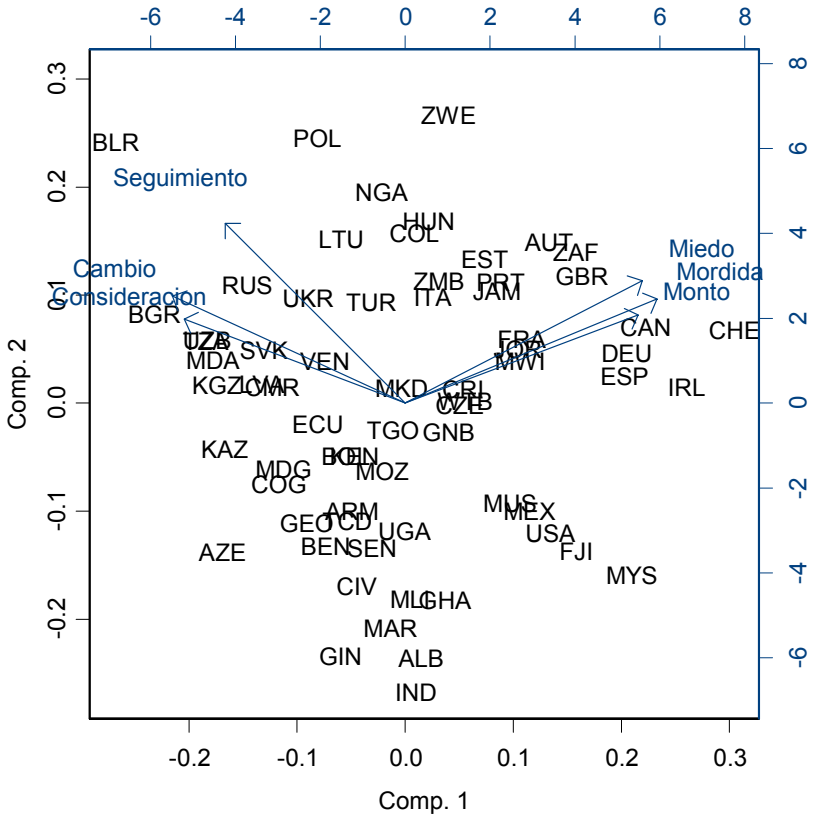
Por un lado, la primera Componente Principal muestra que las variables "cambio", "seguimiento" y "consideración" tienen signos negativos y todas ellas tienen que ver con la comunicación entre el gobierno y las empresas. En la otra mano, las variables "mordidas", "monto" y "miedo" tienen signo positivo. ***Si miramos las preguntas hechas en la encuesta podemos observar que la escala con la cual se miden las variables relacionadas con la comunicación van en sentido contrario a las que miden corrupción.*** Por ejemplo, para "cambio", "seguimiento" y "consideración" la escala de respuestas va de (1) completamente predecible hasta (6) completamente impredecible y en el contexto del tema, entre más alta es la escala peor es la situación para el país. A pesar de que la escala para las variables "monto",

"mordida" y "miedo" es la misma en la interpretación es al revés. Las variables que miden corrupción entre menor sea el valor de la respuesta mejor estará el país.

Los valores de las cargas son muy parecidos porque estamos trabajando con datos estandarizados. La segunda Componente Principal parece un promedio ponderado de la comunicación del gobierno con las empresas en temas que les competen.

La gráfica que se presenta enseguida es conocida como Biplot. Como ya hemos mencionado, resume la posición de las unidades experimentales, los países con respecto a los valores de las dos Componentes Principales en estudio. Además muestra los vectores de correlación de cargas.

BIPLOT PRIMERA Y SEGUNDA COMPONENTE PRINCIPAL



En esta gráfica podemos observar que los vectores de correlación de cargas de las variables "miedo", "monto" y "mordida" respecto a la primera componente tienen una asociación positiva y alta, pues el coseno de sus ángulos es cercano a 1. Los vectores de las variables "consideración", "cambio" y "seguimiento" están relacionadas de forma negativa y dado que el coseno de sus ángulos es pequeño, la asociación es alta. Esto corrobora la información que nos proporcionó las cargas de las primeras dos Componentes Principales. De igual forma podemos ver que la primera Componente Principal mide hacia la derecha las variables que tienen que ver con la corrupción y hacia la izquierda aquellas que tienen que ver con la

comunicación entre el gobierno y las empresas. En cuanto a la segunda Componente Principal, todas las variables apuntan hacia arriba lo que indica que tienen una relación positiva. También se observa que el ángulo entre "seguimiento" y "miedo" es cercano a los 90°, lo que indica una muy baja correlación. Algunos países forman grupos, por ejemplo en el cuadrante derecho en la parte inferior se encuentran Mauritania (MUS), México (MEX), Estados Unidos (USA), Fiji (FJI) y Malasia (MYS). Finalmente determinada la dimensión del espacio para los datos con los que contamos correremos el Análisis de Factores utilizando la información obtenida.

CAPITLO 3. Análisis de Factores

Esta técnica depende de un modelo estadístico razonable y trata de explicar la estructura de covarianza o correlación entre las variables. Al igual que el Análisis de Componentes Principales el Análisis de Factores es una técnica dirigida por las variables. Comúnmente es utilizado cuando deseamos estudiar un concepto que no es posible medir directamente como el coeficiente intelectual. En estos casos, recabamos información sobre algunas variables que puedan ser indicadoras del concepto a estudiar, como las calificaciones y tratamos de descubrir si las relaciones entre el coeficiente intelectual y dichas variables son consistentes.

En otras palabras, el objetivo del Análisis de Factores es revisar si las relaciones entre un conjunto de variables observadas $\mathbf{x}' = [x_1,...,x_p]$ pueden ser explicadas a través de un número pequeño de variables latentes no observadas llamadas factores[1] $f_1,...,f_k$, donde k < p. Al final del análisis lo que se desea es dar una interpretación a estos factores.

El modelo general para el Análisis de Factores[2] se expresa de la siguiente forma:

$$
\begin{aligned}
x_1 &= \mu_{11} + \lambda_{11} f_1 + \lambda_{12} f_2 + ... + \lambda_{1k} f_k + u_1 \\
x_2 &= \mu_{12} + \lambda_{21} f_1 + \lambda_{22} f_2 + ... + \lambda_{2k} f_k + u_2 \\
&... \\
x_p &= \mu_{1p} + \lambda_{p1} f_1 + \lambda_{p2} f_2 + ... + \lambda_{pk} f_k + u_p
\end{aligned}
$$

Donde las x´s son las variables observadas, μ es la media, λ son las **cargas de los factores**, las f´s son una componente aleatoria común a todas las variables medidas llamado **factor común** y las u´s son las componentes aleatorias específicas para cada variable medida conocidas como **factores específicos**. Las cargas de los factores miden la contribución del *k-ésimo* factor a *la j-ésima* variable respuesta.

Para simplificar la notación, podemos expresar el modelo general utilizando matrices y asumiremos que las medias son cero:

$$\mathbf{x} = \Lambda \mathbf{f} + \mathbf{u} \qquad (1)$$

---

[1] Para que el análisis funcione necesitamos que las variables originales estén correlacionadas pues como ya se dijo el Análisis de Factores trata de entender las relaciones entre las variables originales.

[2] En la mayoría de los estudios se aplica el Análisis de Factores a la matriz de correlaciones, utilizamos los valores estandarizados de las variables originales aunque los resultados que serán descritos aplican también para la matriz de covarianzas.

donde:

$$\Lambda = \begin{pmatrix} \lambda_{11} & . & . & . & \lambda_{1k} \\ . & & & & \\ . & & & & \\ . & & & & \\ \lambda_{p1} & . & . & . & \lambda_{pk} \end{pmatrix}$$

$$\mathbf{f} = \begin{pmatrix} f_1 \\ . \\ . \\ . \\ f_k \end{pmatrix}$$

$$\mathbf{u} = \begin{pmatrix} u_1 \\ . \\ . \\ . \\ u_p \end{pmatrix}$$

Supuestos:

a) Los factores comunes f_q son independientes e idénticamente distribuidos (iid) con media cero y varianza uno, para q =1,...,k.

b) Los factores específicos u_j son independientemente distribuidos con media cero y varianza ψ_j, j = 1,2,...,p.

c) Tanto los factores comunes f_q como los factores específicos u_j tienen distribuciones independientes para todas las combinaciones de q y j, q =1,...,k. y j = 1,2,...,p.

Entonces la varianza de x_i es:

$$\text{var}(X_i) = \sigma_i^2 = \sum_{j=1}^{k} \lambda_{ij}^2 + \psi_i \qquad \text{i = 1,...,p.} \qquad (2)$$

donde $\psi_1 = \text{var}(u_i)$.

Ahora bien, si observamos la ecuación (2) podemos ver que la varianza de una variable x_i se puede dividir en dos partes. El primer término es la varianza compartida con las otras variables a través de los factores comunes llamada **comunalidad** y se define como:

$$h_i^2 = \sum_{j=1}^{k} \lambda_{ij}^2$$

El último término es la varianza de xi dada por ψ_i y es la variabilidad de x_i no compartida con el resto de las variables, a esta se le llama **varianza específica o única**.

La covarianza σ_{ij} de las variables x_i y x_j está dada por:

$$\sigma_{ij} = \sum_{l=1}^{k} \lambda_{il}\lambda_{jl} \qquad (3)$$

Esta ecuación indica que la covarianza de dos variables observadas depende exclusivamente de su relación con los factores comunes.

Ahora juntaremos la ecuación (2) y (3) de la siguiente forma[3]:

$$\Sigma = \Lambda\Lambda^T + \Psi \qquad (4)$$

Estas ecuaciones son conocidas como **Ecuación de Análisis de Factores;** lo que se desea es encontrar Λ y Ψ de tal forma que se cumpla esta ecuación.

Nótese que Λ la matriz de cargas de los factores no está determinada de forma única[4] por lo que es necesario establecer una serie de restricciones sobre los parámetros del modelo. Para lograr este objetivo definiremos a la matriz G como:

$$G = \Lambda^T \Psi^{-1} \Lambda \qquad (5)$$

Necesitamos que la matriz G sea diagonal con sus elementos arreglados en orden descendente de magnitud; lo cual ocasiona factores tales que el primero hace la máxima contribución a la

---

[3] $\Sigma = \text{cov}(\mathbf{x})$ y $\mathbf{x} = \Lambda\mathbf{f} + \mathbf{u}$

$\Sigma = \text{cov}(\Lambda\mathbf{f} + \mathbf{u})$

$\Sigma = \Lambda \text{cov}(\mathbf{f}) \Lambda^T + \Psi$

$\Sigma = \Lambda I \Lambda + \Psi = \Lambda \Lambda^T + \Psi$

[4] Si existen Λ y Ψ de tal forma que:
$\mathrm{P} = \Lambda\Lambda^T + \Psi$,
Rescribiendo $P = \Lambda MM^T\Lambda + \Psi$ para toda matriz ortogonal M (MMT=I). Entonces $\mathrm{P} = (\Lambda M)(\Lambda M)^T + \Psi$; por lo que ΛM es de igual forma una matriz de cargas para toda matriz ortogonal de M y por lo tanto Λ no es única.

varianza común en los elementos de x, el segundo hace la máxima contribución, sujeta a estar no correlacionada con el primero, y así sucesivamente.

¿Cómo sabemos si existe un conjunto de k factores subyacentes? Necesitamos revisar si existen Λ y Ψ tales que:

$$\mathrm{P} = \Lambda\Lambda^{T} + \Psi$$

Primero tenemos que el total de cifras desconocidas en Λ y Ψ es pk+p = p(k+1). El número de cantidades desconocidas en P es p(p+1)/2 [5] . Por lo tanto, las Ecuaciones de Análisis de Factores (4) dan como resultado un sistema de p(p+1)/2 ecuaciones en p(k+1) incógnitas. Hay tres posibles escenarios:

a) Si p(k+1)> p(p+1)/2, o de manera equivalente, si k>(p-1)/2, entonces no existe una solución única al sistema pues existen más ecuaciones que incógnitas.
b) Si k = (p-1)/2, entonces el modelo contiene tantos parámetros como elementos de P. Se puede encontrar una solución única la cual no necesariamente implica una solución con todas las varianzas específicas mayores a cero.
c) Si k<(p-1)/2.entonces hay menos parámetros en el modelo de factores que elementos en P. El Análisis de Factores proporciona una explicación más simple de las relaciones entre las variables observadas que aquella brindada por los elementos de P. Este es el escenario que en realidad nos interesa.

Hay varios métodos para dar soluciones aproximadas a $\mathrm{P} = \hat{\Lambda}\hat{\Lambda}^{T} + \hat{\Psi}$ pues es imposible encontrar una solución exacta. Los tres más conocidos son: **Factores Principales**, **Factores Principales con iteración** y **Máxima Verosimilitud**[6].

Factores Principales

Lo primero que debemos hacer es estimar las comunalidades o varianzas específicas para lo cual corremos un Análisis de Componentes Principales sobre la matriz de correlaciones. Las primeras k Componentes Principales se utilizan para obtener los estimadores de las cargas de factores. Utilizando la siguiente ecuación obtenemos las varianzas específicas o si despejamos las comunalidades.

---

[5] P es una matriz simétrica
[6] Antes de empezar , es importante mencionar que para llevar a cabo los métodos anteriores es necesario tener estimadores preliminares de las comunalidades (varianzas específicas). Algunos estimadores de las comunalidades pueden ser:
(1) El cuadrado del coeficiente de correlación múltiple de la i-ésima variable con todas las otras variables R^2, pues este dato es el porcentaje de la variabilidad en la i-ésima variable que se explica por las otras variables.
(2) El mayor de los valores absolutos de los coeficientes de correlación entre la i-ésima variable y una de las variables restantes.

$$\hat{\psi}_i = s_i^2 - \sum_{j=1}^{k} \hat{\lambda}_{ij}^2 \qquad (6)$$

donde s_i^2 es la varianza muestral de la variable xi.

Factores Principales con Iteración

Este método funciona igual que el de Factores Principales pero una vez que hemos encontrado la solución inicial se utilizan las comunalidades reestimadas con la ecuación (6) como conjetura inicial y se vuelve a repetir todo el procedimiento. Esto se hace hasta que todas las estimaciones convergen o hasta que nos topamos con un resultado absurdo.

Máxima Verosimilitud

Podemos obtener los estimadores de las cargas de factores y varianzas específicas si asumimos que los datos provienen de una Distribución Normal. Una ventaja al utilizar el método de máxima verosimilitud es que es posible probar la hipótesis de que ***k*** factores comunes son suficientes para describir las relaciones observadas en los datos.

Maximizamos la siguiente función de verosimilitud para obtener los estimadores:

$$L = -\frac{1}{2} n \ln|\Sigma| + traza[S\Sigma^{-1}]$$

donde Σ es la matriz de covarianzas predicha por el modelo k-factorial. Σ es función de Λ y Ψ, por lo tanto la función L también lo es. Si rescribimos L será más fácil resolverla. Utilizamos la siguiente transformación: $L = -\frac{1}{2} nF$ +una función de las observaciones. Es más útil minimizar la siguiente función la cual equivale a maximizar L:

$$F = \ln|\Sigma| + traza[S\Sigma^{-1}] - \ln|S| - p \,. \qquad (7)$$

Tenemos que F = 0 si Σ = S y F > 0 .

¿Cómo determinar el número de factores?

Hay varias técnicas para evaluar qué tan bien se ajustan los datos al modelo con un número determinado de factores comunes y a continuación se mencionan:

(1) Utilizar el número de factores cuyos eigenvalores se mayores a uno cuando se utiliza la matriz de correlaciones.

(2) Se grafican los eigenvalores y se elige el número de factores correspondientes al punto donde los eigenvalores empiezan a decrecer para formar una línea horizontal. Esta estrategia se llama Prueba Scree.

(3) Utilizar una prueba Ji-cuadrada para muestras grandes, asociada con la solución por máxima verosimilitud. El estadístico de prueba es:

$$U = n' \min(F)$$

Donde $n' = n - 1 - \frac{1}{6}(2p+5) - \frac{2}{3}k$ y F está dada por la ecuación (7). Si k factores comunes son suficientes para describir los datos, entonces U se distribuye asintóticamente como una Ji-cuadrada con v grados de libertad, donde $\nu = \frac{1}{2}(p-k)^2 - \frac{1}{2}(p+k)$.

En muchos estudios exploratorios k no puede especificarse por adelantado y, por lo tanto, se usa un procedimiento secuencial para determinar k. Iniciando con k = 1 y se estiman los parámetros en el modelo de factores usando el método de máxima verosimilitud. Si la estadística de prueba U no es significativa, aceptamos el modelo con este número de factores. En otro caso, incrementamos k en una unidad y repetimos el proceso hasta que se obtenga una solución aceptable. Si en cualquier etapa, los grados de libertad v se vuelven iguales a cero, entonces o no hay una solución no trivial apropiada o alternativamente el modelo de factores en sí mismo, con su supuesto de relaciones lineales entre variables observadas y latentes, es cuestionable[7].

Rotación de los Factores

Las restricciones que se impusieron a la ecuación[8] (5) ocasionan ciertas propiedades, como arrojar factores ortogonales ordenados de forma descendente en importancia, no son inherentes al Modelo de Factores y eso ocasiona que la interpretación de los mismos se dificulte.

La dificultad radica en que las variables pueden tener cargas altas en más de un factor y las demás sean bajas o cercanas a cero. Por otro lado, puede ser que los factores, excluyendo el

[7] Existen algunos puntos a considerar cuando elegimos el número k de factores: primero, no incluir factores triviales. Las variables que tiene cargas muy altas en un solo factor no están correlacionadas con las demás variables y estas por sí mismas son características subyacentes. Lo mejor es eliminar estas variables y correr de nuevo el análisis. Segundo, algunos paquetes pueden arrojar matrices de diferencias entre las correlaciones observadas entre las variables originales y aquellas generadas por el análisis de factores. Si estas diferencias son pequeñas podemos reducir la cantidad de factores. Cuando estas diferencias son mayores a 0.25 será necesario aumentarlos.

[8] Estas restricciones provocan que los parámetros del modelo de factores fuesen únicos.

primero, tengan cargas positivas y negativas. Por esta razón se ha establecido la rotación de los factores[9].

La rotación de los factores se obtiene al multiplicar la matriz de cargas por una matriz ortogonal, es decir se "rotan los ejes del espacio de factores" obtenidos originalmente. Si la transformación es la correcta, entonces la interpretación de la nueva matriz será más fácil creando una estructura más simple.

Esta nueva matriz tendrá las siguientes características:

a) Cada renglón de Λ debe contener al menos un cero y cada columna de Λ al menos k ceros.
b) Cada par de columnas de Λ debe tener muchas respuestas cuyas cargas tienden a cero en una columna pero no en la otra.
c) Si el número de factores es igual o mayor a cuatro, cada par de columnas contendrá un número grande de respuestas con cargas iguales a cero en ambas columnas.
d) Para cada par de columnas de Λ, sólo un pequeño número de respuestas deberían tener cargas distintas de cero en ambas columnas.

Es importante hacer notar que a medida que se giran los ejes originales hacia los nuevos cambiarán las correlaciones que las variables tienen con cada uno de los factores girados. Existen distintos métodos para obtener la rotación sin embargo el más utilizado es el **Varimax.**

El criterio de **Varimax** consiste en maximizar la siguiente ecuación

$$V^* = \sum_{q=1}^{k} \left(\frac{\left[\sum_{j=1}^{p} b_{jq}^4 - \left(\sum_{j=1}^{p} b_{jq}^2 \right)^2 / p \right]}{p} \right)$$

Donde $B = \Lambda T$, y T es una matriz ortogonal[10]. Lo que deseamos es maximizar la varianza de las cargas elevadas al cuadrado en la *q-ésima* columna de B. Ya que las cargas se encuentran entre 0 y 1, esto es equivalente a forzar el mayor número de cargas hacia 0 y el resto hacia 1.

Sin embargo este método da igual peso a las variables sin considerar el "tamaño" de sus comunalidades. Entonces se sugirió dividir las cargas de los factores para cada variable en dos

[9] En un modelo ideal la interpretación sería más fácil si cada variable tuviese valores altos en un solo factor y el resto de las cargas fuesen positivas o cercanas a cero pues las variables podrían ser separadas en conjuntos disjuntos cada uno asociado a un determinado factor.
[10] Al utilizar una rotación ortogonal no alteramos las comunalidades.

partes: la comunalidad propia de la variable y la comunalidad común. A continuación se maximiza la suma de las varianzas de las razones elevadas al cuadrado, dentro de una columna. En estos términos la ecuación que deseamos maximizar es:

$$V = \frac{1}{p^2} \sum_{q=1}^{k} \left[p \sum_{j=1}^{p} \frac{b_{jq}^4}{h_j^4} - \left(\sum_{j=1}^{p} \frac{b_{jq}^2}{h_j^2} \right)^2 \right]$$

Donde h_j^2 es la comunalidad de la *j-ésima* variable respuesta, $j = 1,2,...,p$. La matriz ortogonal T que maximiza la suma mostrada arriba produce la rotación Varimax. Este ajuste da mayor peso a las variables que tienen las comunalidades más grandes y menor a las que tienen las comunalidades pequeñas.

Puntajes de los Factores

Para obtener las calificaciones de cada unidad de estudio existen dos métodos: Mínimos Cuadrados Ponderados (o Método de Bartlett) y Método de Regresión (o Método de Thompson).

El modelo para cada individuo es $\mathbf{x} = \Lambda\mathbf{f} + \mathbf{u}$, en donde $\mathbf{u}$ no se conoce y Λ se estima. Para un vector de observaciones dado, $\mathbf{x}$, $\mathbf{f}$ no puede ser determinada de forma explícita. Entonces tras resolver el modelo de FA, obtenemos $\mathbf{z} = \Lambda\mathbf{f} + \mathbf{u}$, en donde $\mathbf{u} \sim (\mathbf{0}, \Psi)$.

En el **Método de Mínimos Cuadrados Ponderados** se establece que para encontrar $\mathbf{f}$ hay que minimizar la siguiente ecuación:

$$(\mathbf{z}_r - \hat{\Lambda}\mathbf{f})\hat{\Psi}^{-1}(\mathbf{z}_r - \hat{\Lambda}\mathbf{f})$$

Donde $\mathbf{z}_r$ es el vector de datos estandarizados para el *r-ésimo* individuo. Para un $\mathbf{z}_r$ dado, la expresión anterior se minimiza cuando:

$$\mathbf{f}_r = (\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}\hat{\Psi}^{-1}\mathbf{z}_r$$

Entonces se toma $\mathbf{f}_r$ como el vector de las calificaciones estimadas de los factores para el *r-ésimo* individuo, $r = 1,2,...,N$.

El **Método de Regresión** utiliza la distribución conjunta de **x** estandarizado para datos con distribución Normal y define que f se distribuye como:

$$\begin{bmatrix}\mathbf{z}\\ \mathbf{f}\end{bmatrix} \sim N\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathrm{P} & \Lambda\\ \Lambda' & I\end{bmatrix}\right)$$

Esto implica que la esperanza condicional de $\mathbf{f}$, dado que $\mathbf{z}=\mathbf{z}^*$, es

$$\mathrm{E}[\mathbf{f} \mid \mathbf{z}=\mathbf{z}^*] = \Lambda'\mathrm{P}^{-1}\mathbf{z}^*$$

Por lo tanto, en el método de Regresión estima el vector de calificaciones de los factores, para el *r-ésimo* individuo como $\mathbf{f}_r = \hat{\Lambda}R^{-1}\mathbf{z}_r$. Algunos paquetes pueden usar $\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$ en lugar de R en la fórmula anterior.

Aplicación del Análisis de Factores a los datos

El objetivo principal de este trabajo es crear la variable "viabilidad legal". Este concepto no es observable directamente, es una variable subyacente, por lo cual decidimos utilizar variables observadas relacionadas con la corrupción y la comunicación entre el gobierno y los ciudadanos para crearla. Además en el capítulo de Estadísticas Descriptivas pudimos observar cómo todas las variables están correlacionadas entre sí.

En la sección anterior al correr el Análisis de Componentes Principales observamos que sólo dos eigenvalores eran mayores a uno por lo cual asumimos que la dimensión real de los datos es ésta. Dicho número será la cantidad de factores que utilizaremos al aplicar el Análisis de Factores. Seremos muy conservadores en la distribución de los datos y dado que no tenemos evidencia contundente para decir que los datos de la muestra provienen de una distribución Normal utilizaremos Factores Principales y de entrada no rotaremos los datos. Los resultados se presentan a continuación:

ESTRUCTURA DE CARGAS PARA FACTORES

LOADINGS	Factor1	Factor2
Cambio	-0.829	0.424
Seguimiento	-0.605	0.550
Consideracion	-0.738	0.267
Mordida	0.950	0.382
Monto	0.801	0.218
Miedo	0.864	0.398

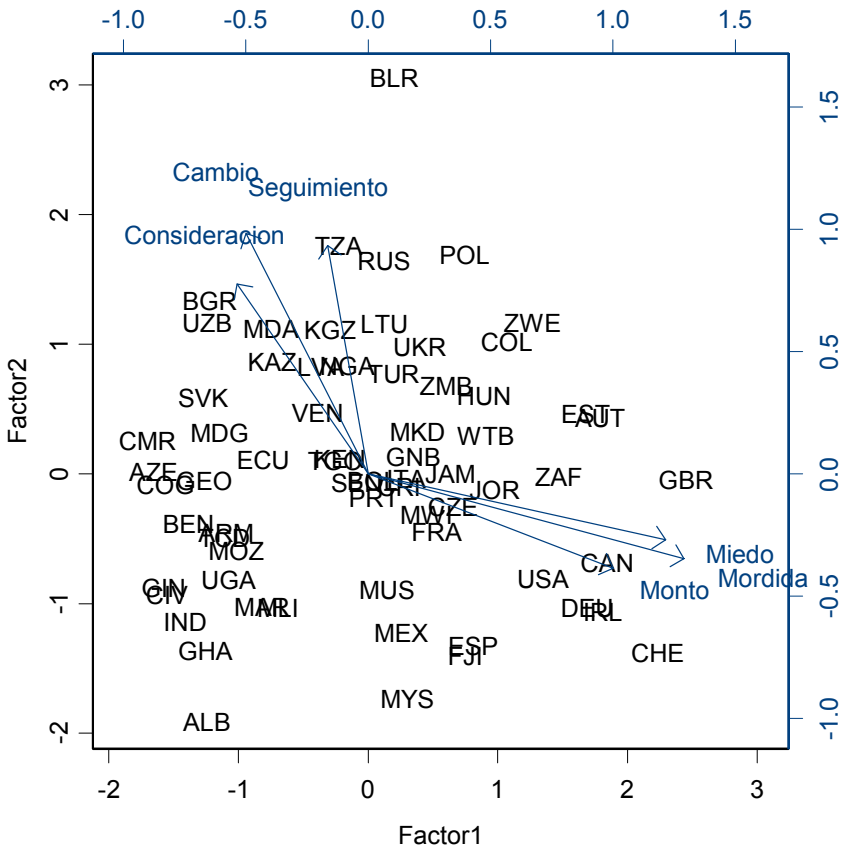
Buscamos que para cada variable haya una carga alta únicamente en alguno de los factores. Sin embargo en el factor 2 sólo hay una carga alta correspondiente a la variable "seguimiento". Esto es un problema porque dicho factor se considera trivial. Entonces buscamos una estructura de datos más fácil de interpretar y rotaremos los datos utilizando el método de Varimax.

ESTRUCTURA DE CARGAS PARA FACTORES UTILIZANDO ROTACION VARIMAX

LOADINGS	Factor1	Factor2
Cambio	-0.379	0.851
Seguimiento	-0.125	0.808
Consideracion	-0.406	0.672
Mordida	0.979	-0.300
Monto	0.760	-0.334
Miedo	0.922	-0.234

Se observa que la estructura realmente mejoró pues para el factor 2 hay varias variables con carga alta, erradicando el problema que teníamos. Además se mantiene la condición de que para cada variable sólo un factor tenga carga mayor a 0.5. Al igual que en el análisis anterior, para el primer factor las variables "cambio", "seguimiento" y "consideración" presentan signos negativos y "mordida", "monto" y "miedo" positivos. Esto ocurre de forma inversa para el segundo factor. Con el fin de interpretar estos datos utilizaremos la gráfica Biplot.

BIPLOT PARA DOS FACTORES



Mirando la gráfica y utilizando la información de las cargas de los factores, podemos decir que el primer factor mide la existencia de corrupción en los países. Aquellos con valores positivos tienen un nivel muy bajo de corrupción, mientras que los países con valores negativos recurrentemente optan por este camino para obtener lo que necesitan del gobierno. Por ejemplo, Suiza (CHE) está en el extremo derecho de la gráfica mientras que Albania (ALB) está en el extremo opuesto[11]. Como ya mencionamos en cuanto al factor 1, las variables "cambio", "seguimiento" y "consideración" tienen una asociación negativa pues sus vectores apuntan hacia la izquierda. De forma análoga, "miedo", "mordida" y " monto" muestran una relación positiva pues los vectores apuntan hacia la derecha.

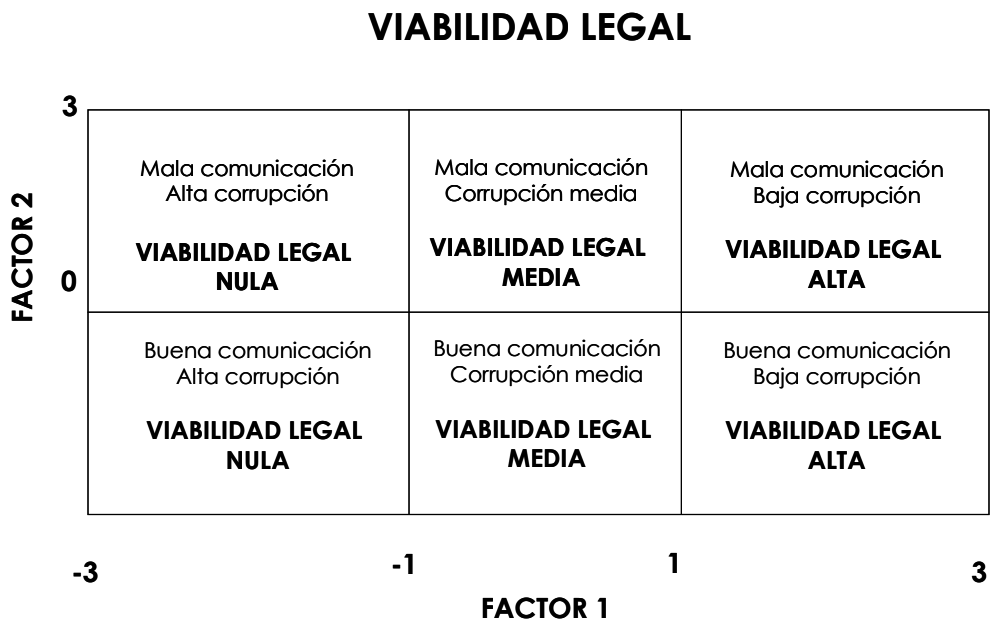
En cuanto al segundo factor tienen mayor peso las variables de comunicación entre el gobierno y las empresas en cuanto a los cambios jurídicos. Los países con mejor comunicación tienen valores negativos y de forma inversa para aquellos cuyos gobiernos no los consideran en temas legales. Bielorrusia (BLR) tiene mala comunicación con su gobierno, basta recordar que para los 90´s se independizó de Rusia, proceso que explica la inestabilidad política del país recién formado. Albania (ALB), localizado hasta debajo de la gráfica manifestó que sus empresarios

[11] En los anexos se muestran los valores de los dos factores para cada país.

son generalmente considerados por el gobierno en los temas legales que les atañen[12]. Si miramos los vectores en cuanto al factor 2, "miedo", "monto" y "mordida" tienen relación negativa y en contraste "cambio", "seguimiento" y "consideración" presentan asociación positiva.

En cuanto a los ángulos que forman los vectores entre sí, podemos observar que "cambio", "seguimiento" y "consideración" están altamente relacionados y de forma positiva; de igual forma "monto", "mordida" y "miedo". Ahora bien, los ángulos que se forman entre estos dos grupos de variables y sus múltiples combinaciones presentan ángulos grandes y si recordamos el coseno de estos indican relaciones altas y negativas. El tamaño de los vectores mide la varianza de las variables. Podemos observar que "mordida" es aquella variable cuya varianza es la mayor. Esto corrobora los datos que muestra el histograma de esta variable, el cual muestra que hay países donde la corrupción es muy baja y aquellos donde ésta es muy alta.

Los dos factores en conjunto nos permiten crear el concepto de "viabilidad legal". Esta noción surge al observar que en la práctica cuando las leyes de un país son muy complicadas y difíciles de cumplir o no se hacen valer, las empresas desarrollan mecanismos alternos denominados actos de corrupción para conseguir la aprobación legal de su actuar. Aunado a esto, la comunicación entre el gobierno y los ciudadanos en ocasiones no existe, lo cual provoca que las leyes no se sigan. Por ejemplo en nuestro país abrir una empresa puede ser un proceso muy largo y engorroso por lo cual muchos empresarios están dispuestos a tomar un "atajo" pagando mordidas y sobornando a los funcionarios públicos. Utilizando esta información podemos asignar a cada país un escenario para la "viabilidad legal" de la siguiente forma:



[12] En el ANEXO 2 se muestra una tabla en la que se encuentran las abreviaturas para cada país

El que la comunicación entre el gobierno y el sector empresarial sea buena por sí misma no garantiza la inexistencia de la corrupción. Podría ser que los trámites sean difíciles de llevar a cabo, poco transparentes, muy caros o que sean muchos, lo cual incentiva a los empresarios a pagar mordidas o sobornar a los empleados gubernamentales para obtener el trámite o servicio deseado. Ante este panorama y para aquellos países cuya comunicación entre empresas y gobierno es mala aunado a la presencia de alta corrupción establecemos que la "viabilidad legal" es NULA.

Para aquellos países cuya comunicación es buena o mala pero presentan corrupción media, entonces asignamos una "viabilidad legal" MEDIA. Como dice el dicho: "Del plato a la boca se cae la sopa", es decir que los gobiernos pueden dialogar con sus empresarios o no hacerlo, pero al final del día los trámites legales que establecen hacen que los negocios opten en algunas ocasiones por la corrupción.

En el otro extremo puede ser que haya muy mala comunicación entre el gobierno y las empresas pero que los trámites sean simples, rápidos, baratos por lo cual las mismas se apegan al marco de la ley. O que la comunicación sea buena y además la corrupción baja. A estos países les asignamos una "viabilidad legal" ALTA.

La siguiente tabla muestra los países que integran cada escenario dependiendo de la "viabilidad legal".

CLASIFICACIÓN DE LOS PAÍSES SEGÚN LA VIABILIDAD LEGAL

	Corrupción alta				Corrupción media				Corrupción baja	
Mala comunicación	6	AZE			46	TGO	47	LVA	3	AUT
	41	CMR			57	ECU	22	NGA	20	EST
	55	MDG			7	KEN	66	KAZ	13	COL
	2	SVK			14	GNB	32	UKR	67	ZWE
	59	UZB			42	WTB	15	KGZ		
	45	BGR			31	MKD	50	MDA		
					48	VEN	21	LTU		
					27	HUN	19	RUS		
					43	ZMB	28	POL		
					60	TUR	64	TZA		
							33	BLR		
	VIABILIDAD LEGAL NULA				VIABILIDAD LEGAL MEDIA				VIABILIDAD LEGAL ALTA	
Buena comunicación	11	ALB	5	MOZ	39	MYS	56	MWI	16	CHE
	4	GHA	54	TCD	18	FJI	58	CZE	30	IRL
	26	IND	24	ARM	40	ESP	35	PRT	17	DEU
	12	CIV	25	BEN	34	MEX	49	JOR	61	USA
	10	GIN	1	COG	44	MLI	53	CRI	9	CAN
	29	UGA	62	GEO	63	MAR	8	SEN	23	GBR
					38	MUS	51	BOL	65	ZAF
					36	FRA	52	ITA		
							37	JAM		
	VIABILIDAD LEGAL NULA				VIABILIDAD LEGAL MEDIA				VIABILIDAD LEGAL ALTA	
	Corrupción alta				Corrupción media				Corrupción baja	

Los países con mala o buena comunicación pero con alta corrupción fueron catalogados con una viabilidad legal nula. En el primer cuadrante, podemos encontrar a Azerbaiyán. En 1997 este país se acababa de independizar liderado por Heidar Aliyev, tras un largo periodo de inestabilidad económica y política. Durante este año se estableció su constitución así como su política exterior. Por esta razón hace sentido que esté clasificado en un escenario de viabilidad nula pues podemos pensar que había mala comunicación entre el gobierno y las empresas así como una alta corrupción.

En la parte inferior izquierda de la tabla está India, este país también tiene una viabilidad legal nula. Para 1996, el 50% de la población estaba en condiciones de pobreza. Entre 1996 y 1998 hubo una gran inestabilidad política, primero hubo un gobierno de centro derecha, a través del Partido Baratilla Janata. A éste le sucedió un régimen de izquierda encabezado por la Coalición Frente Unido. Nos podemos imaginar la gran corrupción que se presentó cuando en un período de tiempo tan corto ocuparon el gobierno partidos de ideas tan distintas. Ambos necesitaban el apoyo de los empresarios para ocupar la presidencia, por lo cual la comunicación con el gobierno era buena; sin embargo la corrupción era altísima.

Para un escenario de corrupción media y mala comunicación con las empresas, es decir viabilidad legal media, se encuentran Venezuela y Ecuador. En esa época Venezuela estaba gobernada por Caldera, el cual estableció un régimen de corte populista. Una de sus principales políticas fue establecer el control de precios de recursos energéticos. Incluso la gasolina tuvo el precio más bajo a nivel mundial. Esto arruinó la economía del país y fomentó que el General Hugo Chávez fuese subiendo en popularidad. Por otro lado, en Ecuador durante

1995 ganó las elecciones Sixto Durán Ballén. Dicho presidente fomentó la privatización de empresas públicas provocando serios disturbios sociales. Durante 1996, Abdalá Bucaram toma el gobierno aunque su mandato sólo duró 6 meses. A pesar de la inestabilidad política del país, la unidad se mantiene. Estos acontecimientos nos dan una idea de por qué estos países presentan mala comunicación y corrupción a nivel medio.

México se encuentra en la categoría de buena comunicación y corrupción media, de igual forma que Ecuador, obtuvo una calificación de viabilidad legal media. Durante el sexenio del Dr. Ernesto Zedillo, el país percibía una mayor estabilidad económica después de la crisis del 94. En este lapso se gestó lo que actualmente conocemos como la Secretaría de la Función Pública, organismo encargado de combatir la corrupción a través del servicio civil de carrera para los funcionarios públicos y la promoción de la capacitación. A partir de este sexenio se eliminó a contratación directa en los puestos altos de gobierno y se sustituyó por las convocatorias abiertas. A grandes rasgos los empresarios percibieron que había una buena relación con el gobierno aunque aun persiste la corrupción.

Colombia se encuentra en el extremo superior de lado derecho de la tabla, clasificado con los países cuya viabilidad legal es alta. Durante 1991 en este país se modificó la Constitución, en este proyecto participaron indígenas, grupos de izquierda, guerrilleros, minorías religiosas y representantes de los poderes tradicionales. A partir de ahí el gobierno ha mejorado muchísimo su economía. Para 1997, época donde se realiza la encuesta cuyos datos estamos utilizando, el gobierno tenía una mala comunicación con las empresas, no las tomaba en cuenta cuando establecía sus políticas sin embargo la implementación de las mismas ha sido excelente pues la corrupción ha bajado. Hoy día el gobierno colombiano tiene una página de Internet a través de la cual se pueden realizar cientos de trámites gubernamentales lo cual ha hecho muchísimo más fácil y rápido obtener servicios tanto para las empresas como los ciudadanos.

Finalmente, en el extremo inferior derecho de la tabla de viabilidad encontramos a los países "estrellas", países donde la viabilidad legal sin duda es alta. Estos son ejemplos de gobiernos que se comunican con sus empresarios a la par que implementan sus políticas incentivando a los mismos a seguir el camino legal. Dicho grupo está conformado por Suiza, Irlanda, Alemania, Estados Unidos, Canadá, Gran Bretaña y Sudáfrica. Para el caso de Canadá a partir de los 90´s existen las "Oficinas únicas". Estos sitios son dependencias gubernamentales donde pueden realizarse todos los trámites. No es necesario desplazarse de una secretaría a otra. Por ejemplo, para abrir una empresa los trámites se pueden realizar en menos de 3 días lo cual muestra la gran eficiencia.

Otro dato que arroja el Análisis de Factores son las varianzas específicas las cuales se muestran a continuación:

VARIANZAS ESPECIFICAS

Cambio	Seguimiento	Consideracion	Mordida	Monto	Miedo
0.13281440	0.331763	0.38338730	-0.049148	0.311446	0.09510055

Como se explicó en la parte teórica, el Análisis de Factores se compone de dos varianzas: la varianza compartida o comunalidad y la varianza específica o no compartida. El método de evaluación para éstas últimas, es considerar que las varianzas altas no están correlacionadas con las otras variables. La tabla muestra que todas las varianzas únicas son pequeñas, la más alta aún está por debajo de 0.5, entonces quiere decir que todas las variables están correlacionadas entre sí. La varianza única más pequeña corresponde a la variable "mordida" lo cual corrobora su gran relación con el resto.

CONCLUSIONES

El objetivo de este estudio descriptivo fue crear una variable llamada "viabilidad legal". Este nuevo concepto mide las decisiones que toman las empresas para obtener un bien o servicio gubernamental. Frente al camino planteado por las autoridades pueden optar por seguir el marco jurídico o tomar ciertos atajos conocidos como corrupción. Para que una empresa decida la "ruta" a seguir primero debe conocerla. Además el camino debe estar diseñado para los usuarios, por ejemplo, si se establece una tramitología para empresas que operan en Internet, difícilmente será aplicable a las armadoras de autos porque su problemática es completamente distinta. Finalmente, el camino debe ser rápido, fácil y barato de seguir.

Esta variable difícilmente puede ser observada como tal en la realidad, es una medida subyacente, por lo cual recurrimos a variables medibles. Dichas variables se agruparon en dos clases: aquellas que miden la comunicación sobre temas legales entre las empresas y el gobierno; las que miden los atajos legales o "corrupción". Con esta información decidimos utilizar el Análisis de Factores para obtener la viabilidad legal para los 67 países en estudio.

Con el fin de conocer la estructura de la base de datos se utilizaron estadísticas descriptivas como: histogramas, matriz de correlaciones, diagramas de dispersión y Caritas de Chernoff. Tras mostrar que las variables estaban linealmente correlacionadas utilizamos el Análisis de Componentes Principales para saber cuántos factores utilizar. Con esta técnica observamos que era necesario usar dos factores. Al aplicar el Análisis de Factores obtuvimos resultados que no eran fácilmente interpretables por lo que se rotaron los datos utilizando el método de Varimax.

El primer factor mostró que medía el nivel de corrupción en los países y el segundo factor la comunicación entre empresas y gobierno. Al ser sólo dos factores pudimos crear un mapa a través del cual clasificar a los países según su viabilidad legal. Se plantearon 3 categorías: nula, media y alta. Para respaldar este categorización se tomaron algunos países y se analizaron brevemente sus circunstancias políticas y económicas. Por ejemplo, México tiene una viabilidad legal media porque a pesar de que la comunicación entre el gobierno y los negocios es buena, a la hora de implementar las reglas a seguir, los empresarios a veces se apegan al marco legal y a veces recurren a la corrupción.

La viabilidad legal puede ser utilizada para estudios posteriores con el fin de lograr que los gobiernos no sólo establezcan buenas políticas gubernamentales sino que durante la implementación de las mismas en realidad logren los objetivos. En términos coloquiales, que del plato a la boca no se caiga la sopa.

ANEXO 1. Encuesta

1. Do you regularly have to COPE with unexpected changes in rules, laws or policies which materially affect your business?

Changes in laws and policies are

(1) completely predictable
(2) highly predictable
(3) fairly predictable
(4) fairly unpredictable
(5) highly unpredictable
(6) completely unpredictable

2. "The process of developing new rules or policies is usually such that affected businesses are informed"

This is true

(1) always
(2) mostly
(3) frequently
(4) sometimes
(5) seldom
(6) never

3. "In case of important changes in laws or policies affecting my business operation the government takes into account concerns voiced either by me or by my business association"

This is true

(1) always
(2) mostly
(3) frequently
(4) sometimes
(5) seldom
(6) never

4. "It is common for firms in my line of business to have to pay some irregular "additional payments" to get things done"

This is true

(1) always
(2) mostly
(3) frequently
(4) sometimes
(5) seldom

(6) never

5. "Firms in my line of business usually know much this "additional payment" is."

This is true

(1) always
(2) mostly
(3) frequently
(4) sometimes
(5) seldom
(6) never

6. "Even if a firm has tom make an "additional payment" it always has to fear that it will be asked for more, e.g. by another official."

This is true

(1) always
(2) mostly
(3) frequently
(4) sometimes
(5) seldom
(6) never

ANEXO 2. VALOR DE LOS FACTORES PARA CADA PAIS

Pais	FACTOR 1	FACTOR 2
ALB	-1.243054455	-1.922376281
ARM	-1.096666613	-0.467632973
AUT	1.788235137	0.416391362
AZE	-1.655713923	0.000617414
BEN	-1.384552446	-0.399112711
BGR	-1.224268345	1.322073291
BLR	0.201671681	3.042168132
BOL	0.03200883	-0.066263165
CAN	1.841417849	-0.699311794
CHE	2.229505998	-1.394341576
CIV	-1.553430042	-0.946364301
CMR	-1.703445373	0.237471276
COG	-1.564253936	-0.102164143
COL	1.066292905	1.00711228
CRI	0.23766769	-0.115935256
CZE	0.653117494	-0.270787223
DEU	1.686499952	-1.043541056
ECU	-0.810455294	0.08965997
ESP	0.807373123	-1.342615614
EST	1.677257102	0.448032776
FJI	0.745591992	-1.415232309
FRA	0.526076434	-0.460497746
GBR	2.449178441	-0.060205017
GEO	-1.26418437	-0.064323436
GHA	-1.256036582	-1.37875575
GIN	-1.576529691	-0.893317319
GNB	0.346953531	0.111686813
HUN	0.89163293	0.583306733
IND	-1.411356753	-1.157813854
IRL	1.806064268	-1.075982574
ITA	0.292408376	-0.053034655
JAM	0.633002667	-0.01779348
JOR	0.973645924	-0.14251361
KAZ	-0.739666574	0.846241983
KEN	-0.216935448	0.098352157
KGZ	-0.293989672	1.093026078
LTU	0.125455573	1.139597824
LVA	-0.366366994	0.810233065
MAR	-0.820131104	-1.041467832
MDA	-0.749599793	1.10354373
MDG	-1.147095284	0.301454985
MEX	0.252506052	-1.241983722
MKD	0.379041616	0.310384804
MLI	-0.696970973	-1.045871109
MOZ	-1.017522064	-0.605983016
MUS	0.143175319	-0.909154584
MWI	0.447245747	-0.327495144
MYS	0.30029873	-1.745925385
NGA	-0.162014223	0.816980749
POL	0.742010068	1.672732636
PRT	0.04259531	-0.199544691
RUS	0.121381	1.625254439
SEN	-0.087939566	-0.084463555
SVK	-1.273225664	0.571344961
TCD	-1.108144313	-0.509396543
TGO	-0.25507738	0.084854249
TUR	0.196103656	0.754405452
TZA	-0.228118027	1.747430934
UGA	-1.076681042	-0.833817313
UKR	0.395589094	0.962750057
USA	1.345155467	-0.822146947
UZB	-1.241280723	1.150407572
VEN	-0.390145131	0.458737797
WTB	0.906216236	0.278501191
ZAF	1.467511231	-0.038404672
ZMB	0.600731852	0.660781703
ZWE	1.264232521	1.150033945

ANEXO 3. Abreviaturas

Abreviatura	País	Abreviatura	País
ALB	ALBANIA	**KEN**	KENYA
ARM	ARMENIA	**KGZ**	KYRGYZ REPUBLIC
AUT	AUSTRIA	**LTU**	LITHUANIA
AZE	AZERBAIJAN	**LVA**	LATVIA
BEN	BENIN	**MAR**	MOROCCO
BGR	BULGARIA	**MDA**	MOLDOVA
BLR	BELARUS	**MDG**	MADAGASCAR
BOL	BOLIVIA	**MEX**	MEXICO
CAN	CANADA	**MKD**	MACEDONIA, FYR
CHE	SWITZERLAND	**MLI**	MALI
CIV	COTE D'IVOIRE	**MOZ**	MOZAMBIQUE
CMR	CAMEROON	**MUS**	MAURITIUS
COG	CONGO	**MWI**	MALAWI
COL	COLOMBIA	**MYS**	MALAYSIA
CRI	COSTA RICA	**NGA**	NIGERIA
CZE	CZECH REPUBLIC	**POL**	POLAND
DEU	GERMANY	**PRT**	PORTUGAL
ECU	ECUADOR	**RUS**	RUSSIA
ESP	SPAIN	**SEN**	SENEGAL
EST	ESTONIA	**SVK**	SLOVAK REPUBLIC
FJI	FIJI	**TCD**	CHAD
FRA	FRANCE	**TGO**	TOGO
GBR	UNITED KINGDOM	**TUR**	TURKEY
GEO	GEORGIA	**TZA**	TANZANIA
GHA	GHANA	**UGA**	UGANDA
GIN	GUINEA	**UKR**	UKRAINE
GNB	GUINEA-BISSAU	**USA**	UNITED STATES
HUN	HUNGARY	**UZB**	UZBEKISTAN
IND	INDIA	**VEN**	VENEZUELA
IRL	IRELAND	**WTB**	WEST BANK
ITA	ITALY	**ZAF**	SOUTH AFRICA
JAM	JAMAICA	**ZMB**	ZAMBIA
JOR	JORDAN	**ZWE**	ZIMBABWE
KAZ	KAZAKHSTAN		

BIBLIOGRAFIA

Anderson, T.W., 2003, An introduction to multivariate statistical analysis, Wiley Interscience, Third Edition

Bennet S. y Bowers D., 1976, An introduction to multivariate techniques for social and behavioural sciences, London:Macmillan

Cooley, W. y Lohnes P. 1971, Multivariate Data Analysis, New Cork: Wiley

Darlington, Richard B., Sharon Weinberg, and Herbert Walberg. 1973. Canonical variate analysis and related techniques. Review of Educational Research.

Giri, Narayan C.,2004,Multivariate Statistical Analysis, Marcel Dekker Inc., Second Edition

Green, P. 1976, Analyizing Multivariate Data, Illinois:Dryden

Jambu, Michael, 1991, Exploratoy and multivariate data analysis, Academic Press Inc.

Johnson, Dallas E. 1998, Applied Multivariate Methods for Data Analysts, Brooks Cole Publishing Company

Jobson, J.D., 1992, Applied multivariate data analysis, Volume II: Categorial and multivariate methods, Springer-Verlang

Johnson, Richard A. y Wichern, Dean W. 1988, Applied multivariate statistical analysis, Second Edition, Prentice Hall

Kaiser, H. 1958, The Varimax criterion for analytic rotation in factor analyisis, Psychometrika, vol.23, 187-200

Lam, Longhow, 1999, An introduction to S-Plus for Windows, Mathsoft

Lawley, D. 1940, The estimation of factor loading by the method of maximum likelihood, Proceedings of the royal society of Edinburg 60, p.p. 64-82

Tabachnick, Barbara y Fidell, Linda S. 1989, Using multivariate statistics, Second Edition, Harper Collins Publishers

Tacq, Jaques, 1997, Multivariate analysis techniques in social science research, Sage Publications

http://www.statsoft.com/textbook/stfacan.html