Unidad 3: Exploración y Análisis de Datos según Escala de Medición

Distribución de frecuencias

Un tema relevante para la Ciencia Política es el poder que ejercen las instituciones en el dominio de la sociedad. La Encuesta Mundial de Valores analiza el grado de participación y la confianza de los ciudadanos hacia estos organismos. Para presentar el tema de Distribución de Frecuencias se utilizarán algunas variables sobre participación y confianza.

La *distribución de frecuencias* nos va a permitir conocer de qué manera están agrupados los datos según las categorías en que se divide la variable seleccionada. A continuación, para ejemplificar cómo funciona una *distribución de frecuencias*, utilizaremos la Encuesta Mundial de Valores del 2015. Pediremos la *distribución de frecuencias* de la variable **V25**, para esto es necesario utilizar el paquete *Hmisc*.

```
install.packages(“foreign”)          ## solicitamos la instalación del paquete

library (foreign)                    ## pedimos el paquete para empezarlo a usar
```

Abriendo la Encuesta Mundial de Valores del 2015. Se utiliza el paquete *foreign* para importar una data guardad en SPSS

```
Data<- read.spss("WVS2015.sav",use.value.labels=TRUE, max.value.labels=Inf,
to.data.frame=TRUE)

## Nombraremos “Data”  a nuestra base de datos

attach(Data)

## El comando “attach” nos sirve para poder nombrar a cada variable de “Data” sin tener la necesidad de volver a nombrar de donde proviene la variable. (Si no lo hiciéramos tendríamos que escribir así: Data$V25)

Install.packages(“Hmisc”)

## instalamos el paquete Hmisc para poder utilizar el comando describe

library(Hmisc)  ## abrimos el paquete
```

V25- Active/Inactive membership: Church or religious organization

describe(V25) ## el comando “describe” nos genera las principales descriptivas de cada variable

```
V25

     n missing  unique
 85592     680       3
Not a member    (56870, 66%)
Inactive member (12898, 15%)
Active member   (15824, 18%)
```

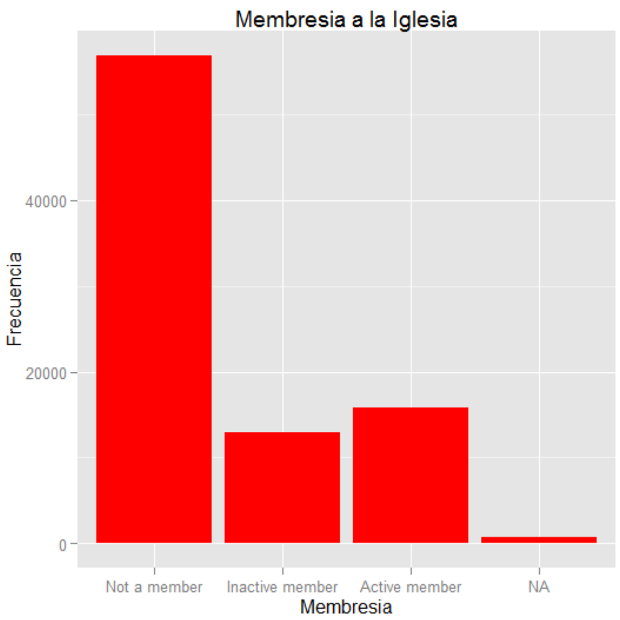
Luego de ejecutar este comando vemos, en primer lugar que hay un total de 85, 592 datos y 680 valores perdidos. Podemos identificar que un 66% de los casos (56, 870 encuestados) se encuentra en el primer grupo, que son las personas que no son parte de una organización religiosa. La segunda opción agrupa al 15% de los casos (12898 encuestados) esto significa que el 15% de la muestra es un miembro inactivo de estas organizaciones a las que pertenece. Por último, 15,824 encuestados son miembros activos de una organización religiosa. La *distribución de frecuencias* nos muestra que este último grupo representa el 18% del total de la muestra.

Un recurso habitual que ayuda a la visualización de los resultados es la utilización de gráficos que ordenan las variables. Determinados gráficos son más útiles dependiendo de la categoría de la variable con la que estamos trabajando. El gráfico circular, el de barras y el histograma son los más usados en estos casos. Tanto el circular- también llamado pie- como el de barras se utilizan para variables nominales y ordinales. El histograma se utiliza para variables escalares. Ejecutamos los siguientes comandos para pedir gráficos.

library(ggplot2)

la librería ggplot2 está diseñada para gráficos estadísticos con mayor estética

ggplot(Data) + geom_bar(aes(V25)) + xlab("Membresia") + ylab("Frecuencia") + ggtitle("Membresia a la Iglesia")



El gráfico de barras corrobora lo que se vio en la tabla de frecuencias". Las barras más alargadas indican mayor frecuencia de la categoría. Los mayor cantidad de personas a nivel mundial no es miembro de alguna variable.

Con las variables que se presentan líneas abajo, vamos a generar distribuciones de frecuencias y gráficos de barras. Trata de replicar cada ejemplo por tu cuenta

V109- Confidence: The armed forces

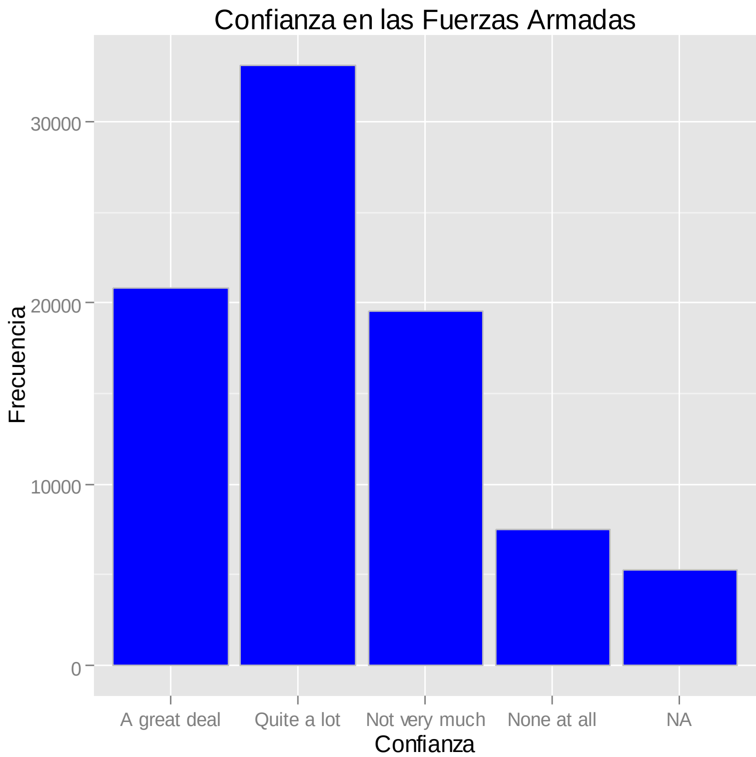
describe(V109)

V109

```
     n missing  unique 
 80964    5308       4 
A great deal (20792, 26%), Quite a lot (33098, 41%)
Not very much (19538, 24%), None at all (7536, 9%)
```

ggplot(Data) + geom_bar(aes(V109), fill="blue") + xlab("Membresia") + ylab("Frecuencia") + ggtitle("Confianza en las Fuerzas Armadas")

#En este caso especificamos que deseamos un gráfico de barras de color azul. Es por eso que en la parte de "fill" escribimos "blue".



V117- Confidence: The police

describe(V117)

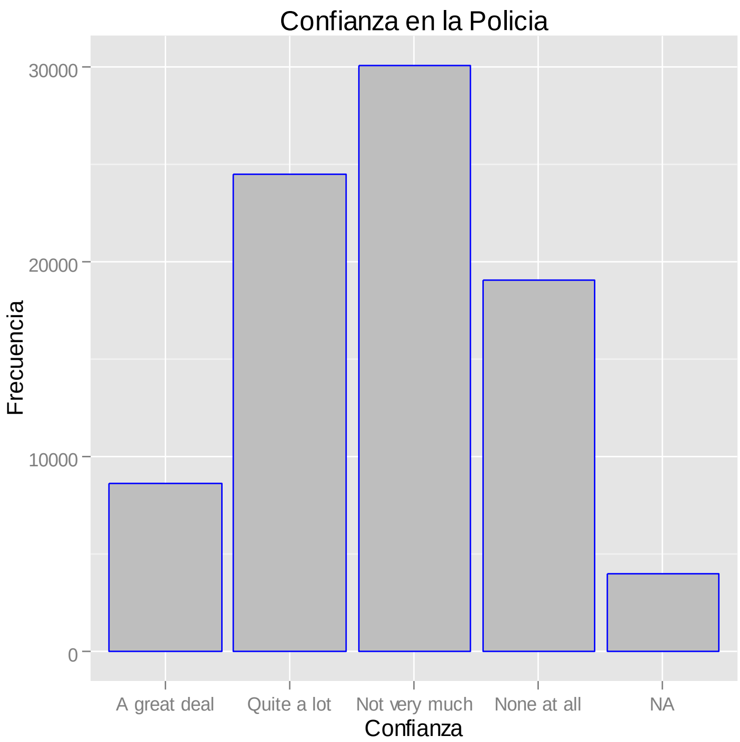
V117

```
      n missing  unique 
  82308    3964       4 
```

A great deal (8658, 11%), Quite a lot (24470, 30%)
Not very much (30078, 37%), None at all (19102, 23%)

ggplot(Data) + geom_bar(aes(V117), fill=”gray”, color=”blue”) + xlab("Membresia") + ylab("Frecuencia") + ggtitle("Confianza en la Policia")

#En este caso generamos un gráfico de barras de color gris (“gray”).



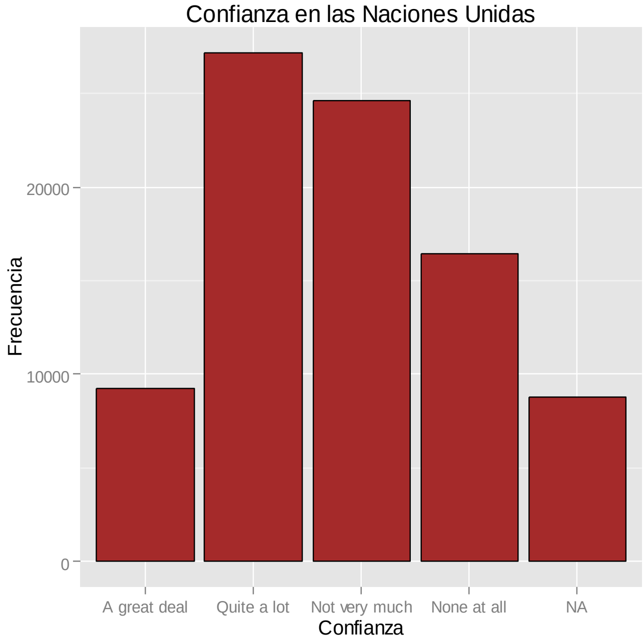
V126- Confidence: The United Nations

describe(V126)

V126

```
      n missing  unique 
  77490    8782       4 
```

A great deal (9217, 12%), Quite a lot (27183, 35%)
Not very much (24613, 32%), None at all (16477, 21%)



Tomando estas tres instituciones en conjunto, se pueden establecer una serie de características con respecto a la expectativa que generan en los ciudadanos. Por ejemplo, se puede señalar que los ciudadanos tienen una mayor confianza hacia las Fuerzas Armadas que hacia los Departamentos de Policías de sus países. Con respecto a las Naciones Unidas se puede señalar que un porcentaje considerable de la muestra no tiene nada de confianza en esta institución.

```
#Para dejar de utilizar "Data$VARIABLE" utilice: attach (Data)
```

Estadísticos descriptivos

La Encuesta Mundial de Valores ha diseñado una sección titulada *Democracy*. En esta se presentan 9 casos que inspeccionan si para el encuestado la función que se le presenta debe ser abordada por un estado democrático. Si en un estado democrático las mujeres tienen igualdad de derechos frente a los hombres o si el Estado debe preocuparse por los salarios de todas las personas, son enunciados que nos permiten conocer el nivel de entendimiento que existe sobre las reglas no escritas de la democracia. En esta sección se presentará el tema de *Estadísticos Descriptivos* haciendo uso de estos enunciados.

Los *estadísticos descriptivos* permiten observar las diversas características que presentan nuestras variables. Básicamente, son las medidas de centralidad, de dispersión, orden y sesgo.

Para entender la relevancia de estas mediadas veamos el siguiente ejemplo con la variable **V131**:

V131- Democracy: Governments tax the rich and subsidize the poor.

describe(V131A)

V131A

n	missing	unique	Info	Mean	.05	.10	.25	.50	.75	.90	.95
82591	3681	10	0.98	6.227	1	1	4	7	9	10	10

Podemos observar algunas cosas de la variable V131A, n como el número de casas, missing los valores perdidos, mean como la media, finalmente, la ubicación de los déciles.

Medidas de centralidad: media, mediana y moda

La primera es el promedio de los valores cuantitativos que hemos recogido en la variable. La mediana nos indica el valor que resulta como punto medio en la distribución de la variable escalar, mientras que la moda es el valor que más veces ha sido recogido por los encuestadores.

Aclaremos, antes de continuar con la explicación, que las medidas de centralidad no son exclusividad de las variables escalares. ¿Qué nos impide establecer la moda de una variable nominal o la de una variable ordinal? En una variable nominal se puede establecer qué variable ha sido recogida en mayor cantidad de oportunidades, pero no podemos determinar un punto medio ya que no están alineadas bajo ningún orden. En una variable ordinal, es posible definir qué valor se repite con más frecuencia y establecer una etiqueta como punto medio de la variable, mas no podemos establecer un valor promedio. Por último, para una variable numérica se pueden utilizar las tres medidas siempre considerando que la media nos otorga una mayor precisión en nuestro análisis. Pidamos ahora la variable **SACSECVAL** *con* el comando *summary.*

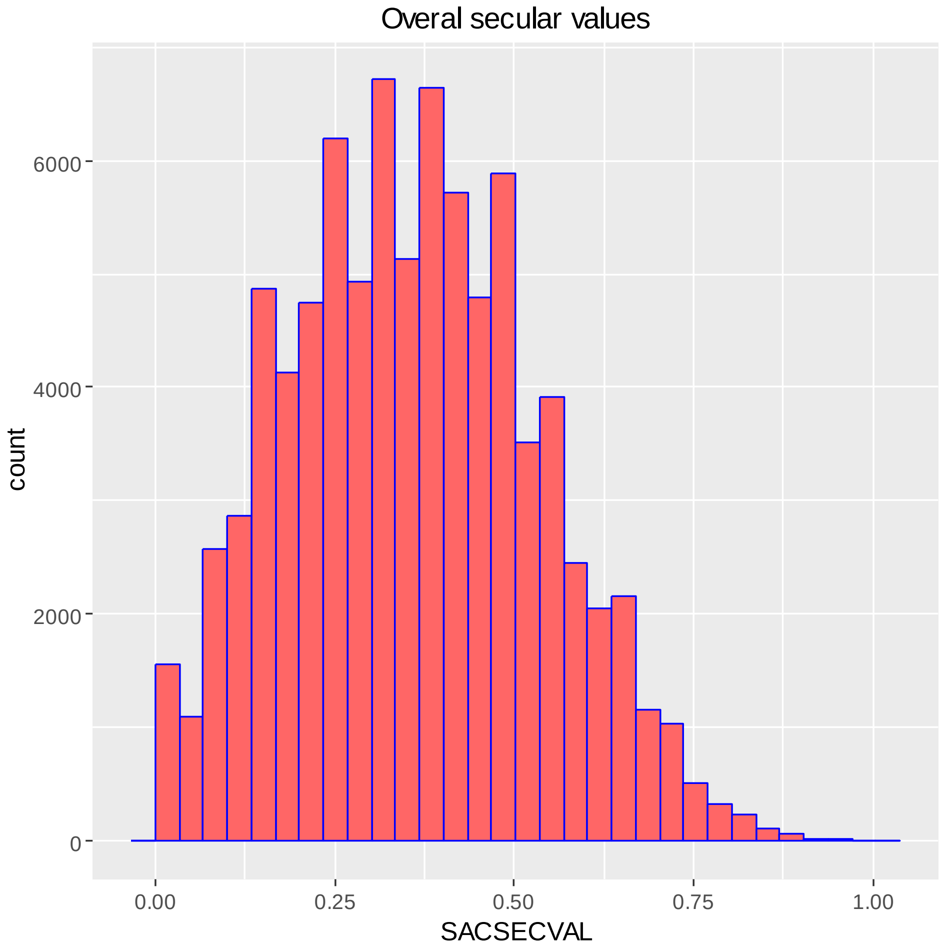
SACSECVAL: Overall Secular Values

summary(SACSECVAL) #para pedir estadísticos descriptivos de centralidad

Min.	1st Qu.	**Median**	**Mean**	3rd Qu.	Max.	NA's
0.0000	0.2274	**0.3592**	**0.3593**	0.4800	1.0050	849

Los resultados me muestran que el valor promedio en la variable es 0.3593 y el dato central; es decir, el valor posicionado sobre el 50% de los casos- el punto medio- es 0,3592. Los recursos analíticos se expanden cuando empezamos a utilizar *estadísticos descriptivos*. Como ya se había mencionado, las variables escalares están conectados a un gráfico especial: el histograma, este se ejecutará con el paquete *ggplot*.

ggplot(Data, aes(SACSECVAL)) + geom_histogram(width=1.5, colour="blue", fill="#FF6666")+ ggtitle("Weight for overal secular values")

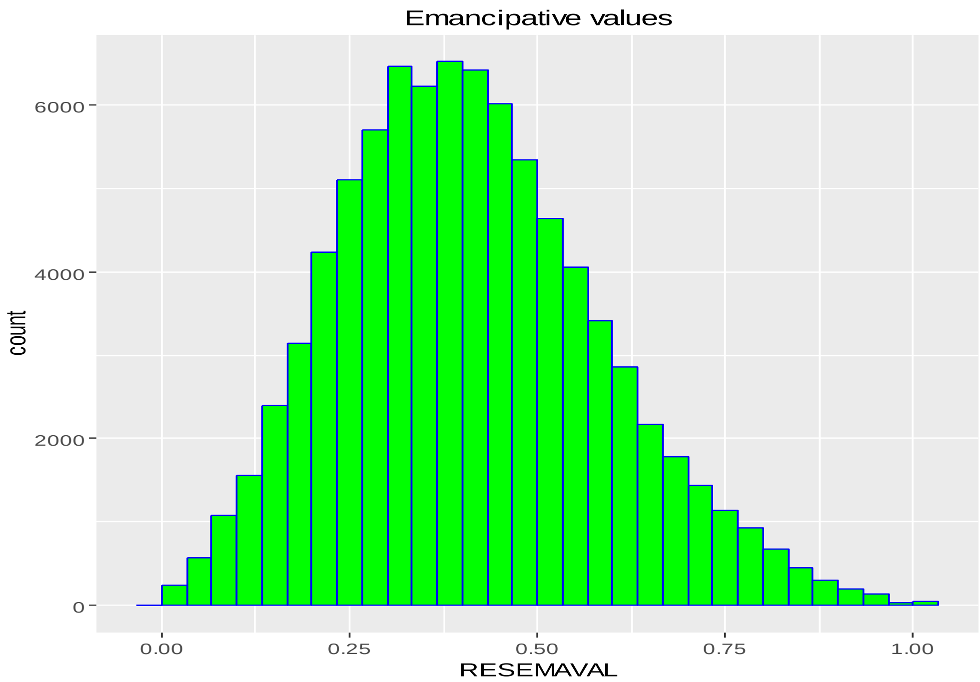


VRESAMAVAL- Emancipative Values

summary(RESEMAVAL)

Min.	1st Qu.	**Median**	**Mean**	3rd Qu.	Max.	NA's
0.0000	0.2844	**0.3966**	**0.4098**	0.5208	1.0000	974

ggplot(Data, aes(RESEMAVAL)) + geom_histogram(width=1.5, colour="blue", fill="green")+ ggtitle("Egmancipative values")



<u>Medidas de dispersión: varianza, desviación estándar y rango</u>

Estos estadísticos nos permiten conocer la distribución de los valores de una variable. Los más utilizados son la varianza, la desviación estándar, el mínimo, el máximo y el rango. La varianza se consigue al calcular la desviación estándar al cuadrado. Ambos se utilizan para determinar que tanto se dispersan los datos desde la media hacia los extremos. Asimismo el rango nos ubica tanto en el primer valor recogido- el mínimo- como en el último valor- el máximo-, mientras mayor sea este, mayor será la dispersión de los datos. Por la naturaleza cuantitativa de los valores, las medidas de dispersión son aplicables solo a las variables escalares.

Para utilizar los estadísticos de dispersión vamos a instalar el paquete *psych*, además, para resaltar las características de este nuevo paquete trabajaremos con las variables **SACSECVAL** y **RESAMAVAL.**

install.packages("psych")

library(psych)

V. SACSECVAL

describe(SACSEVAL) #con psych, el comando *describe* incluye medidas de dispersión

	vars	n	mean	**sd**	median	trimmed	**mad**	min	max	**range**	skew	kurtosis	se
1	1	85423	0.36	**0.17**	0.36	0.36	**0.19**	0	1	**1**	0.22	-0.44	0

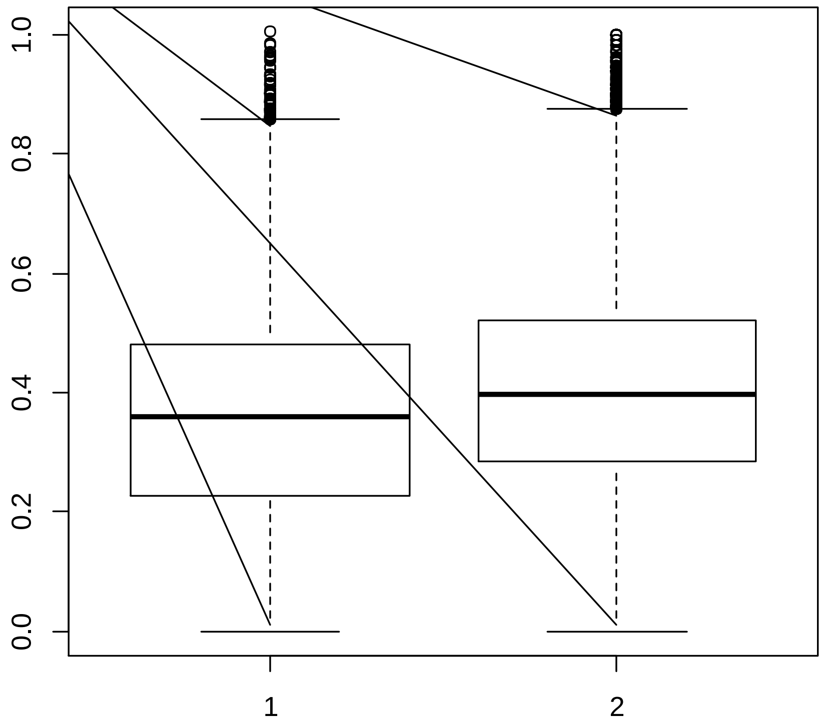
V. RESEMAVAL

describe(RESEMAVAL)

vars	n	mean	**sd**	median	trimmed	**mad**	min	max	**range**	skew	kurtosis	se
1	1 85298	0.41	**0.17**	0.4	0.4	**0.18**	0	1	**1**	0.39	-0.07	0

Para el análisis de dispersión de ambas variables identificamos la desviación estándar y el rango. Las medias de **SACSECVAL** y **RESEMAVAL** son relativamente similares y las desviaciones nos indican que para **SACSECVAL** un amplio número de casos se encuentra entre 0.24 y 0.58 , por la desviación de 0.17; mientras que para **RESAMAVAL** está delimitación se da entre 0.19 y 0.53, por la desviación de 0.17. Para analizar la dispersión de ambas variables en conjunto, podemos utilizar una herramienta extra: el diagrama de cajas (boxplot)

boxplot(SACSECVAL. RESEMAVAL) # pedir boxplot



Este gráfico nos permite visualizar la dispersión e incluso comparar varias variables a la vez. La caja está delimitada por tres líneas horizontales que representan el 25%, 50% y 75% de los datos, el 50% de los datos- mediana- se resalta por lo general en negro. Las líneas superior e inferior, fuera de la caja, señalan el máximo y el mínimo, respectivamente. Las circunferencias fuera de la caja se consideran valores extremos- los más cercanos a la línea inferior y superior- y atípica- los más lejanos al gráfico en su totalidad-. En nuestras variables, se evidencia que la variable **RESEMAVAL** está menos dispersa que la variable **SACSECVAL** ..

El diagrama de cajas o Boxplot es un gráfico que puede ser usado para variables numéricas y categóricas. En el siguiente ejemplo observaremos su utilidad para variables categóricas :

V135- Democracy: The army takes over when government is incompetent.

```
V135A<-as.numeric(V135)
describe(V135A)
```

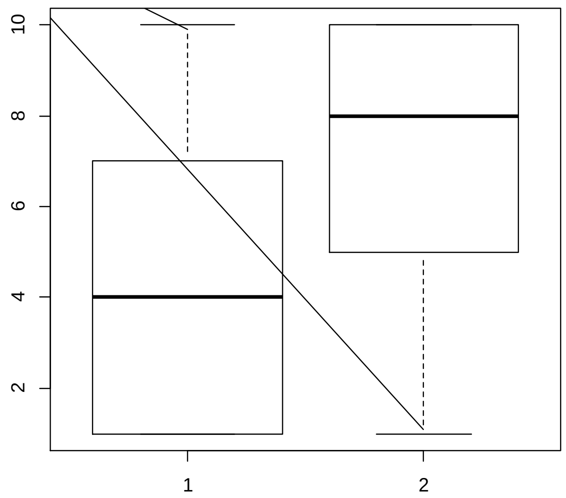
vars	n	mean	**sd**	median	trimmed	mad	min	max	**range**	skew	kurtosis	se
1	1 79078	4.48	**3.08**	4	4.23	4.45	1	10	**9**	0.4	-1.14	0.01

V136- Democracy: Civil rights protect people's liberty from state oppression

```
V136A<-as.numeric(V136)
describe(V136A)
```

vars	n	mean	**sd**	median	trimmed	mad	min	max	**range**	skew	kurtosis	se
1	1 81148	7.36	**2.57**	8	7.69	2.97	1	10	**9**	-0.79	-0.29	0.01

```
boxplot(V135A, V136A)
```



<u>Medidas de orden y sesgo: cuartiles, n-tiles, asimetría y kurtosis</u>

Las medidas de orden y sesgo son el último grupo de estadísticos descriptivos que vamos a revisar en esta sección. Los estadísticos de orden, en primer lugar, nos permiten segmentar la información de una variable. Estas funciones son muy útiles cuando intentamos identificar rendimientos y escalas para generar comparaciones entre grupos. La medidas de sesgo nos ayudan a determinar la tendencia de una variable: si sus valores se concentrar sobre el punto medio o si están muy pegados al valor más bajo de la variable escalar o si se concentran en la sobre el valor más alto de la variable escalar.

Primero, veamos las medidas de orden. La división más utilizada es la de los cuartiles, esto nos permiten dividir la muestra en cuatro grandes grupos: 1° hasta el 25% de los datos, luego hasta el 50% de los datos, después hasta el 75% y, finalmente, del 75% al 100% de los datos. Sin embargo, otras medidas de orden también son comunes, como los n-tiles. Esta nomenclatura puede tomar la forma que nosotros deseemos: deciles, sixtiles, quintiles, etc. Para entender estos conceptos utilicemos la variable **V137**.

V137- Democracy: The state makes people's incomes equal

```
V137A<-as.numeric(V137)
quantile(V137A,na.rm="true") #para pedir medidas  de orden
```

0%	25%	50%	75%	100%
1	**4**	**6**	**9**	**10**

El primer cuartil (25%) nos indica que el valor que divide la primera mitad de las variables es 4. El segundo cuartil (50%) es también la mediana de la variable, en este caso el valor es 6. El tercer cuartil representa el 75% de los valores y cae sobre 9. El cuarto cuartil es también el máximo valor de la variable, en este caso 10.

Al sacar los estadísticos descriptivos de la variable **V137A**, comprobamos que la mediana y Q2- segundo cuartil- son el mismo valor y que el máximo y Q4- cuarto cuartil- son también el mismo valor.

V137A

```
describe(V137A)
```

	vars	n	mean	sd	**median**	trimmed	mad	min	max	range	skew	kurtosis	se
1	1	82297	5.96	2.99	**6**	6.07	2.97	1	10	9	-0.21	-1.15	0.01

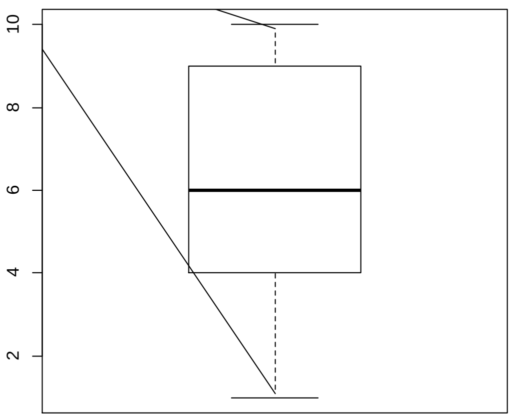
Una herramienta adicional es el **Rango Intercuartil**, esta es la diferencia entre Q3 y Q1. Esta es una medida que se utilizar a la par de la mediana y que puede derivar en otros estadísticos como la desviación cuartil. Para la variable **EDAD**, veamos el RIQ.

V137- Democracy: The state makes people's incomes equal

```
IQR(V137A, na.rm=”true”) #para pedir Rango Intercuartil
```

[1] 5

```
boxplot(V137A)
```



El gráfico que se adecúa muy bien a los estadísticos de orden es el boxplot, este diagrama genera una línea por cada cuartil en la caja. De esta forma, la primera línea de la caja es el Q1, la línea sombreada en negro es la mediana: Q2, y la línea que cierra la caja es el Q3. La lectura de las demás figuras del gráfico se explicó líneas arriba cuando se presentó el boxplot.

También se puede aplicar otro estadístico descriptivo de orden: los n-tiles. Esto es establecer puntos de corte igual entre el rango de frecuencia de la variable; es decir, de 0% a 100% la variable se cortará en la cantidad de veces que el usuario le pida al comando. Esto se puede realizar de dos maneras: por vectores y por n-tiles específicos. Utilizando las variables **V137, e**l vector se pide la siguiente manera:

V137- Democracy: The state makes people's incomes equal

decil <- seq(0, 1, 0.1)

decil

[1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

quantile(V137A, decil, na.rm=”true”)

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

1 1 3 4 5 6 7 8 9 10 10

Lo primero que se hace es señalar la secuencia de corte y nombrarla, en este caso el corte es **0.1** y el nombre *decil*. Luego, verificamos los puntos de corte corriendo *decil*. Por último, obtenemos los resultados con el comando *quantile*(variable, decil). A continuación, programemos un quintil para la variable **V138.**

V138- Democracy: People obey their rulers

```
V138A<-as.numeric(V138)

quintil <- seq(0,1, 0.2)

quintil

 [1] 0.0 0.2 0.4 0.6 0.8 1.0

quantile(V138A, quintil, na.rm=”true”)

0% 20% 40% 60% 80% 100%

1    3     5     7     9    10
```

La segunda forma de pedir los n-tiles se configura de la siguiente manera y es útil cuando se quiere encontrar un cuantil en específico; por ejemplo, con la variable **V139**.

V139- Democracy: Women have the same rights as men.

```
V139A<-as.numeric(V139)

quantile(V139A, probs=c(.33, .44, .66, .855))

33%   44%   66% 85.5%

7       8      10      10
```

Por último, para señalar los valores de la variable desde el percentil, se procede de la siguiente manera. Esta acción nos permite identificar un valor y descubrir, posteriormente, sobre qué percentil se encuentra. En el ejemplo siguiente, el valor 7 se encuentra sobre el 34% de los datos.

```
Per <- ecdf(V139A)

Per(7)

 [1] 0.3482148
```

El siguiente grupo de estadísticos es el de sesgo: asimetría y kurtosis. Estos estadísticos ya los hemos pedido para las medidas de tendencia central que se estudió líneas arriba y nos sirven para entender cómo se distribuyen los valores de una variable, cuál es su tendencia y sobre qué puntos está acumulados.

Una variable puede ser simétrica cuando el coeficiente de asimetría es igual a 0, esto significa que la distribución de los valores hacia la izquierda y la derecha de la media es la misma. También puede ser de asimetría negativa si el coeficiente resultante es menor a 0 y de asimetría positiva si el coeficiente es mayor a 0. Para el primer caso, los valores estarán a la derecha de la media y para el segundo caso, los valores se reúnen a la izquierda de la media.

La kurtosis nos indica qué tan concentrados están los valores de una variable. Cuando estos se encuentran sobre la media de manera regular, la variable es mesocúrtica. Los otros dos tipos son la variable leptocúrtica y la variable platicúrtica. La primera nos indica que un alto número de valores se posiciona sobre la media, esto tendrá un coeficiente positivo. La segunda nos indica que los valores no están concentrados sobre el punto medio, en este caso el coeficiente será negativo.

V133- Democracy: People choose their leaders in free elections

describe(V133A) # con psych, el comando describe incluye medidas de sesgo

	vars	n	mean	sd	median	trimmed	mad	min	max	range	**skew**	**kurtosis**	se
1	1	83019	7.95	2.5	9	8.39	1.48	1	10	9	**-1.2**	**0.52**	0.01

La asimetría para la variable **V133** es negativa porque el coeficiente que R nos da es menor a cero. Esto significa, que los valores de esta variable se concentran a la derecha de la media. Asimismo, el coeficiente de kurtosis es positivo pero bajo, se puede complementar con un gráfico para confirmar que esta es una variable mesocúrtica.

Desarrollemos el mismo análisis con la variable **V134**

V134- Democracy: People receive state aid for unemployment

describe(V134A)

	vars	n	mean	sd	median	trimmed	mad	min	max	range	**skew**	**kurtosis**	se
1	1	83029	6.98	2.73	8	7.28	2.97	1	10	9	**-0.64**	**-0.61**	0.01

Distribución Normal

Este concepto hace alusión a la distribución de los valores de una variable escalar. Se caracteriza por ser simétrica y mesocúrtica. Los valores se concentran en la parte media de la distribución y se extienden de igual manera hacia la zona baja de la muestra, como a la zona alta de la muestra. Su relevancia radica en que es el requisito previo para una serie de técnicas que se utilizarán más adelante. Por eso, una serie de pruebas se han desarrollado con el objetivo de determinar si una variable es normal o no. En este capítulo solo se utilizará las de Shapiro-Wilk y Kolmogorov-Smirnov.

La primera se utiliza cuando el grado de libertad de una variable es menor a 50; por lo tanto, la segunda prueba se utiliza cuando la población es mayor a 50. La prueba de normalidad utiliza una H0: existe normalidad, y una H1: no existe normalidad, de tal manera que cuando la significancia sea mayor a 0.05 se debe aceptar la H0 y cuando esta sea rechazada, se debe aceptar la H1. Veamos el desarrollo de este concepto con las variables **RESEMAVAL (Emancipative Values)** y **SACSECVAL(Overall Secular Values)** , para esto se debe instalar el paquete *nortest* y utilizar los comandos *shapiro.test(variable)p.value* y *lillie.test(variable)$p.value*.

```
install.packages(nortest)

library(nortest)

data1<-Data[Data$V2=="Peru",] #segmentamos la data con las casos del Perú

describe(data1$RESEMAVAL)
```

data1$RESEMAVAL

n	missing	unique	Info	Mean	.05	.10	.25	.50	.75	.90	.95
1194	16	940	1	0.43	0.22	0.26	0.34	0.43	0.51	0.61	0.67

lowest : 0.05528 0.08250 0.09667 0.11524 0.11650

highest: 0.81944 0.83333 0.84218 0.86133 0.87417

```
lillie.test(data1$RESEMAVAL)$p.value #para pedir test de Shapiro-Wilk
```

[1] 0.1502064

```
shapiro.test(data1$RESEMAVAL)$p.value #para pedir test de Kolmogorov-Smirnov
```

[1] 0.002400692

```
describe(data1$SACSECVAL)
```

data1$SACSECVAL

n	missing	unique	Info	Mean	.05	.10	.25	.50	.75	.90	.95
1196	14	560	1	0.40	0.16	0.22	0.29	0.40	0.49	0.58	0.63

lowest : 0.02778 0.04139 0.05528 0.06889 0.08250

highest: 0.83278 0.84583 0.85779 0.86083 0.87333 lillie.test(data1$SACSECVAL)$p.value

[1] 0.001391838

shapiro.test(data1$SACSECVAL)$p.value

[1] 0.003510007

Primero, se ha descrito la variable **RESEMAVAL** para ver si se debe utilizar la prueba de Shapiro-Wilk o la de Kolmogorov-Smirnov. El **n** de la variable (n>50) indica que se debe utilizar la prueba de Kolmogorov-Smirnov. La significancia de esta prueba: 0.1502064, es mayor a 0.05; por lo tanto, se acepta la H0: la existencia de normalidad en la variable **RESEMEVAL**. Ahora es momento de analizar la variable **SACSECVAL** con las pruebas de normalidad. La significancia de esta prueba: 0.0013, es menor a 0.05; por lo tanto, se rechaza la H0: la existencia de normalidad en la variable **SACSECVAL.**

Luego de haber realizado las pruebas de normalidad, hemos determinado si estas variables pueden ser analizadas con determinado grupo de herramientas que describiremos a lo largo de las unidades siguientes.

Ahora es tu turno…

Ejercicio 1

En el libro "El desborde Popular", José Matos Mar refiere que para 1940, el 17% de la población vivía en la ciudad, teniendo un alto porcentaje de personas que residían en las zonas rurales. Sin embargo, esta cifra sufrió grandes cambios para 1977, ya que el 65 % de la población, ahora, se encontraba residiendo en zonas urbanas. Este drástico cambio de movilidad social da pie al argumento central de Matos, las altas tasas de migración que se realizaron de manera desordenada generó una sociedad con dos esquemas económicos: un esquema formal y uno informal, ante el cual el Estado no supo, ni pudo hacer frente.

En la base de datos "Ejercicio 3", encontrarás las cifras de población urbana, población rural e inmigración reciente a nivel nacional y regional de los censos nacionales de 1981, 1993 y 2007, con la que podrás observar cómo ha cambiado el argumento de Matos en el Perú, resolviendo algunas de las siguientes preguntas, que deberán utilizar las medidas de tendencia central y dispersión; así como los gráficos más adecuados que puedan explicar mejor tu argumento.

Identifica las tres regiones que hayan sufrido los cambios más grandes respecto al crecimiento de su población urbana, y que a su vez hayan obtenido las mayores cifras de dispersión, durante estos tres censos. Para ello es necesario que compares estas cifras en términos de porcentajes, por lo que tendrás que sumar la cantidad de población urbana y rural, lo que te brindará la población total por cada año, para realizar el promedio con la cantidad de población urbana. Finalmente, realiza las medidas de dispersión que consideres adecuadas para observar en cuáles de las regiones este crecimiento se ha dado de manera menos sostenida. Es necesario que presente los gráficos adecuados para que pueda sustentar sus resultados

Realiza este mismo proceso con la cantidad de población inmigrante recientemente y compara como se ha dado esta cifra respecto al crecimiento de población urbana en los departamentos y a nivel nacional.

Ejercicio 2

El estudio que realiza la Universidad de Vanderbilt en la encuesta de LAPOP 2014 para Perú incluye algunas preguntas para medir la confianza en algunas instituciones entre las que incluye el sistema judicial, las fuerzas armadas, la Policía Nacional, los partidos políticos y la Iglesia Católica. Por otro lado, contamos con la base de datos "Lima como vamos 2014" que realiza el Instituto de Opinión Público de la PUCP en la cual incluye una pregunta similar a la de LAPOP sobre confianza en las instituciones. Si bien ambas preguntas no son iguales, mantienen ciertas similitudes en sus categorías de respuesta.

A continuación, encontrarás las bases de datos de ambos estudios para que realices una comparación entre los resultados en la ciudad de Lima. Para ello es necesario que realices los gráficos descriptivos que mejor pueda ejemplificar tu respuesta.