PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

CIENCIAS SOCIALES

**TRABAJO GRUPAL**

**Nombres:** Cordova Candiotti Brigitte Carmen, Sánchez Curioso Isaura Yamilé, Torres Policio Luis Alberto, Vargas Cachi Mayra Catalina

**Tipo de evaluación:** Entrega 2

**Curso: Estadística Para el Análisis Político**

**Horario: 690**

**Comisión: 690A**

**Profesor: José Manuel Magallanes**

**Jefe de Práctica: Jair Alva**

**SEMESTRE 2020 -2**

1. **INTRODUCCIÓN**

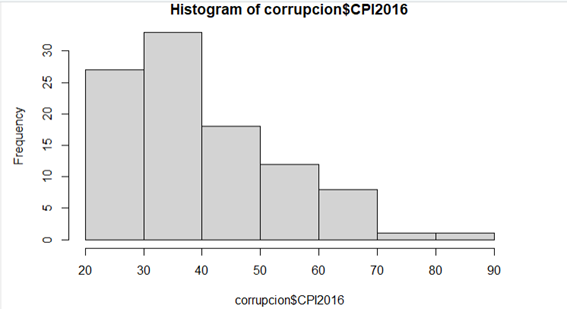
La presente investigación tiene como finalidad demostrar, mediante técnicas estadísticas, como influyen algunos factores en el índice de percepción de la corrupción en los países. De manera específica, se han considerado ocho variables independientes, las cuales se han asociado en cuatro grupos: factores sociales, factores políticos, factores económicos e IDH. Por tal motivo se establece que existe una relación causal entre dichos grupos y la variable de la percepción de corrupción. Para esto, se han establecido cuatro hipótesis. Primero, la percepción de la corrupción es menor cuando los factores políticos (efectividad del gobierno, rule of law, accountability) son mayores. Segundo, la variable independiente es mayor cuando los factores sociales (desempleo, coeficiente de GINI, alfabetización) son mayores. Tercero, el índice de la percepción de la corrupción es menor cuando el gasto público en salud es mayor. Cuarto, la percepción de la corrupción es menor cuando el IDH es mayor.

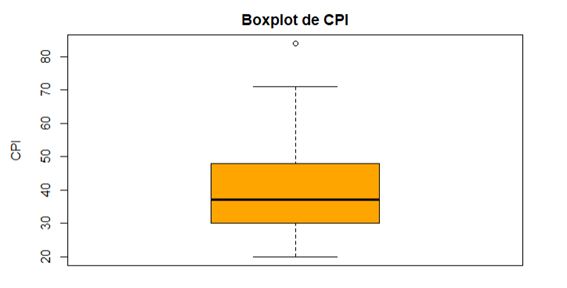
El presente documento, se divide en dos partes. La primera parte presenta el análisis univariado tanto de la variable dependiente como de las independientes. La segunda parte, presenta el análisis bivariado entre la variable dependiente y cada una de las variables independientes. En esta parte se comprobará la influencia de dichas variables en el índice de percepción de la corrupción. Para comprobar y evaluar dichas hipótesis se procedió a realizar regresiones lineales que después permitirán corroborar las hipótesis específicas en relación a los grupos establecidos. Finalmente, se presentan las conclusiones.

1. **ANÁLISIS DE LAS VARIABLES** 
   1. **ANÁLISIS UNIVARIADO DE LAS VARIABLES**
      1. **Dependiente: Índice de Percepción de Corrupción (Isaura)**

La variable dependiente es el índice de los niveles de percepción de corrupción en el sector público. Esta variable está en una escala de 0 a 100, donde 0 es muy corrupto y 100 es muy “limpio” o no corrupto. En el presente trabajo, dicha variable se encuentra en la data *corrupción* como CPI 2016. Para el análisis de dicha variable se empleó el comando *summary,* el cual ha permitido obtener los siguientes datos. Como es una variable numérica; primero, analizamos las medidas de tendencia central, observamos que la mediana de la muestra es de 37; mientras, que el promedio de la percepción de la corrupción de la muestra (100 países) es de 40.41. Aquello permite establecer que la mayoría de países son corruptos. Asimismo, se evidencia que el valor mínimo de esta muestra es de 20, y el valor máximo de 84; esto permite interpretar que hay una significativa diferencia entre los índices de corrupción de los países. Además, al explorar con el comando *skewness*, se observa que la asimetría es de 0.785; esto se comprueba con el histograma, que establece y reafirma una asimetría sesgada a la izquierda. En tal sentido, los índices de percepción de corrupción de la mayoría de países se encuentra en los valores más bajos. De igual manera, con la exploración con el comando *kurtosis* se establece que la curtosis es de 0.04, que señala una distribución leptocúrtica que permite interpretar que hay una gran concentración de los valores en torno a la moda de la variable.

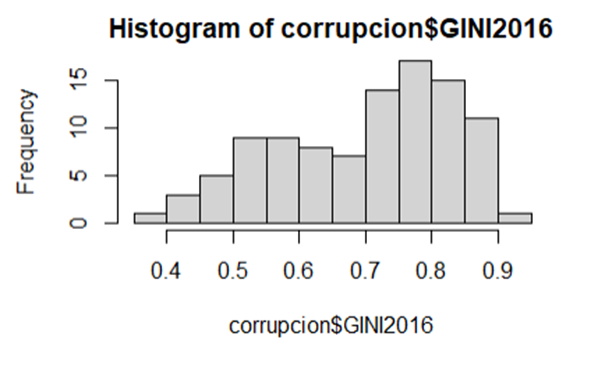
Con respecto a las medidas de dispersión, al utilizar el comando *sd* se obtuvo que la desviación estándar es de 13.34, esto permite analizar la distancia promedio de los datos en la relación al promedio. De esta manera, los datos se encuentran alejados o dispersos entre sí con dicha distancia. Asimismo, el gráfico de boxplot nos permite comprobar lo anteriormente mencionado e incluso dar unas especificaciones. Observamos que el 25% de los países tiene un índice de la percepción de la corrupción de 37; un 50% tiene un índice de 36; mientras, un 75% de los países tiene un índice de corrupción de 48. Todo lo anteriormente mencionado comprueba lo inicialmente mencionado: la mayoría de los países son considerados como corruptos. Por otra parte, al explorar el rango para ver cómo se distribuyen las puntuaciones de las variables de menor a mayor se evidencia que este es de 64. Algo relevante de mencionar es que al explorar con el comando *outliers* se encontró la presencia de un dato atípico, en específico, un país cuenta con un índice de percepción de corrupción de 84, lo cual lo coloca en un país “limpio” o no corrupto.

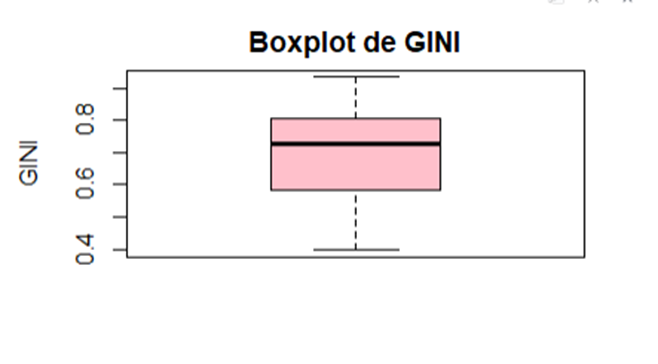




* + 1. **Independientes**
       1. **Variable GINI (Isaura)**

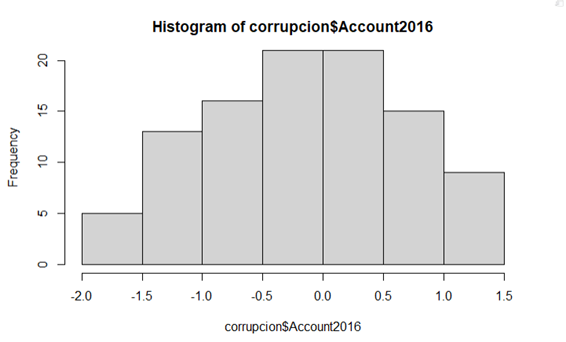
La variable independiente de GINI mide la desigualdad en el ingresos en cada país, esta se mide en una escala de 0 a 1, donde el 0 es que los ingresos y el consumo están distribuidos equitativamente entre toda la población, mientras el 1 representa una situación donde sólo una persona posee toda la riqueza.Dicha variable se encuentra en la data de corrupción como GINI 2016. Para el análisis de dicha variable se empleó el comando *summary,* el cual ha permitido obtener los siguientes datos. Como es una variable numérica analizamos las medidas de tendencia central; primero, la mediana de la muestra es 0.7275; mientras que el promedio de la percepción de la corrupción de los 100 países es de 0.6995 . Esto último permite establecer que la mayoría de países tienen un mayor índice de desigualdad en los ingresos. Asimismo, podemos establecer que el valor mínimo de esta muestra es de -0.3980 , y el valor máximo de 0.9330. Además, encontramos con el comando *skewness*, que la asimetría es de -0.462307, lo cual se comprueba con el histograma que establece una asimetría sesgada a la derecha, donde la mayor cantidad de valores se encuentra en los valores más altos. Por otra parte, con la exploración de la curtosis con el comando *kurtosis* se establece que esta es de -0.8807632, lo cual nos conlleva a señalar que tenemos una distribución platicúrtica que indica que los valores son muy diversos.

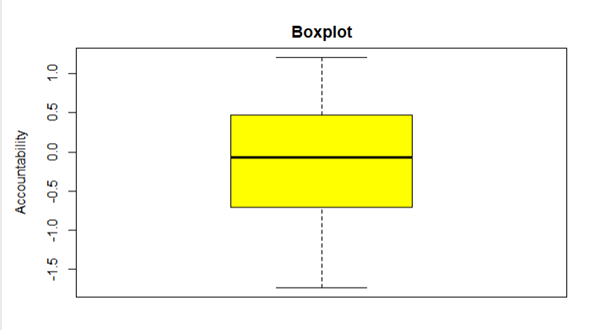
Por otro lado, con respecto a las medidas de dispersión, al utilizar *sd*, obtenemos que la desviación estándar es de 0.1342499. De igual forma, el gráfico de boxplot nos permite señalar que el 25% de los países tiene -0.58525; un 50% tiene 0.72750; mientras, un 75% de los países tiene un índice de corrupción de 0.80650, lo que permite evidenciar que la mayoría de los países tienen una gran desigualdad en los niveles de ingresos. También establecemos que el rango es de 0.535 lo cual nos indica cómo las puntuaciones de la variable se distribuyen de menor a mayor. Asimismo, es relevante indicar que al explorar con el comando *outliers* no se encontró ningún valor atípico. 



* + - 1. **Variable Accountability (Isaura)**

La variable independiente de Voice and Accountability es aquella que captura las percepciones de los ciudadanos en relación a la libertad de expresión, la libertad de asociación y los medios de comunicación libres. Dicha variable se encuentra en la data de corrupción como Account 2016. Para el análisis de dicha variable se empleó el comando *summary,* el cual ha permitido obtener los siguientes datos. Como es una variable numérica; primero, analizamos las medidas de tendencia central, la mediana de la muestra es -0.0657; mientras que el promedio de la variable accountability de los 100 países es de -0.1219. Esto último permite establecer que la mayoría de países tiene un bajo nivel de libertad. Asimismo, podemos establecer que el valor mínimo de esta muestra es de -1.72922, y el valor máximo de 1.20798. Además, encontramos con el comando *skewness*, que la asimetría es de -0.1728429, lo cual se comprueba con el histograma que establece una asimetría sesgada a la derecha; en tal sentido, la mayor cantidad de valores se encuentra en los valores más altos. Por otra parte, con la exploración de la curtosis con el comando *kurtosis* se establece que esta es de -0.93963, lo cual nos conlleva a señalar que tenemos una distribución platicúrtica que indica que los valores están muy dispersos.

Por otro lado, las medidas de dispersión, al utilizar *sd*, obtenemos que la desviación estándar es de 0.7841. De igual forma, el gráfico de boxplot nos permite comprobar lo anteriormente mencionado e incluso dar unas especificaciones. Observamos que el 25% de los países tiene -0.7039; un 50% tiene -0.0657; mientras, un 75% de los países tiene un índice de corrupción de 0.4610; teniendo en cuenta que la escala es de -2.5 a 2.5, se interpreta que la mayoría de los países tiene niveles bajo de accountability. También establecemos que el rango es de 0.535 lo cual indica que la puntuación de la variable no es muy amplia. Como se evidencia en el gráfico del boxplot no hay presencia de datos atípicos.



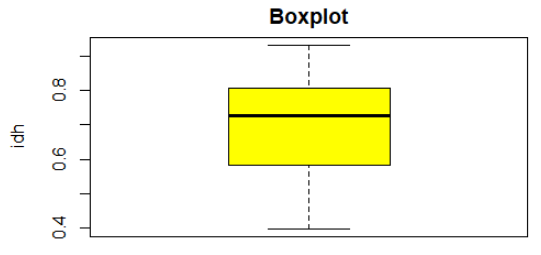
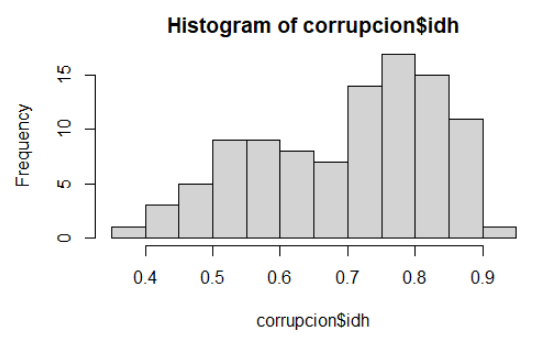
* + - 1. **Variable IDH (Brigitte)**

**Datos:**

La variable IDH se mide de la siguiente manera. Primero, 0-55 a menos es nivel bajo. Segundo, de 0.55 a 0.699 es nivel medio. Tercero, de 0.7 a 0.799 es nivel alto. Cuarto y último,de 0.8 a más, es muy alto.

Primero, explicaré los datos de la medida de tendencia central. Comenzaré explicando los datos encontrados sobre la mediana, media y moda. La mediana, puntuación promedio, es de 0.73. Esto según la clasificación significa que hay un índice de desarrollo humano alto en el promedio mundial. La media, puntuación central en distribución ordenada, es de 0.70, lo que demuestra un índice de desarrollo humano alto. La moda es de 0.572. Esta es la puntuación que ocurre con mayor frecuencia. Esto según los índices significa que la moda es la de “desarrollo humano medio” (se encuentra entre el rango 0.55- 0.69). Segundo explicaré los datos encontrados en las medidas de dispersión. Comenzaré explicando los datos encontrados en asimetría y kurtosis. La asimetría es de – 0.46. Esto significa que la asimetría es positiva (0 es mayor a la asimetría). Es decir, el grado de sesgo o simetría de la curva de distribución de frecuencias determina lo siguiente. Hay una asimetría hacia la derecha. Es decir, la cola de la distribución se alarga (a la derecha) para valores superiores a la media. Esto se observó al notar que la media es mayor que la moda. Puesto que, hay un considerable número de valores “altos”. Por otro lado, la curtosis. Esta es de -0.88. No hay tantos datos cercanos a la media, al parecer; dado que, la curtosis es menor a 0, lo que significa que es de tipo platicúrtica.

Ahora explicaré los datos encontrados en la desviación y varianza. La desviación es de 0.13. Considero que la desviación es baja, porque no sobrepasa ni el 0.20. Esto significa que las puntuaciones de la variable idh no están tan dispersas por la distribución muestral en relación con la puntuación media. Mientras que la varianza es de 0.018. Considero que también es baja. Esto significa que la variación promedio de las puntuaciones en la distribución de esta variable es baja. Esto significa que no hay grandes niveles de dispersión producidos por los datos y outliers. Por otro lado, el rango muestra que la diferencia entre la puntuación mínima y la puntuación máxima de la data es de 0.54. Considero que esto muestra que existe cierta distribución uniforme entre los datos analizados. Sobre los cuartiles, esto demuestra que la mediana es de 0.73(al 53%), y que a los extremos hay datos atípicos tales como (0.39 al 0% y 0.93 al 100%). Para ver con mayor precisión el tema de los datos atípicos, realicé un boxplot en el que revisé los outliers. Al ver el boxplot pude observar que la mayor concentración de outliers se encuentra entre los datos menores de la variable. Cuarto, debo mencionar que al realizar un histograma se pudo observar que, efectivamente, hay una distribución menor entre los valores menores de la variable. Asimismo, que hay una concentración considerable alrededor de la media.

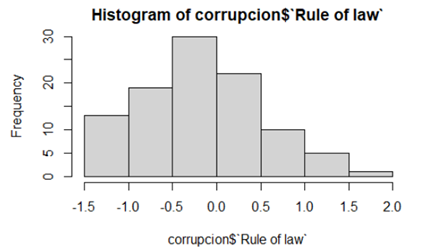


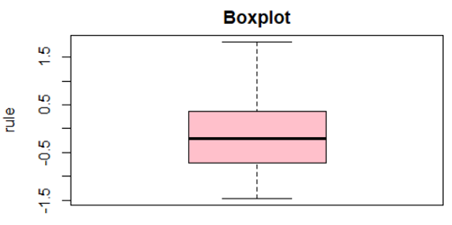
* + - 1. **Variable “rule of law” (Brigitte)**

Dato: Rule of Law se mide de -2.5 a 2.5 (menor a mayor)

Primero, explicaré los datos de la medida de tendencia central. Comenzaré explicando los datos encontrados sobre la mediana, media y moda. La mediana, puntuación promedio, es de -0,21. Esto según la clasificación significa que hay una puntuación promedio relativamente baja sobre el rule of law en los distintos países encuestados. Puesto que, el índice va desde (-2.5 hasta 2.5). La media, puntuación central en distribución ordenada, es de -0.15, lo que demuestra una tendencia a puntaje bajo en esta variable. La moda es de - 0.5 (aproximadamente), un puntaje bajo . Ahora, explicaré los datos encontrados en la asimetría y kurtosis. La asimetría es de 0.33. Esto significa que la asimetría es negativa hacia a la izquierda (0.33 es mayor a 0). Es decir, la cola de la distribución se alarga para valores inferiores a la media. Por otro lado, la curtosis. Esta es de – 0.46. No hay tantos datos cercanos a la media. Asimismo, se puede señalar que es de tipo platicúrtica., al parecer; dado que, la curtosis es mayor a 0, lo que significa que es de tipo platicúrtica.

Tercero, explicaré los datos encontrados sobre las medidas de dispersión. Comenzaré con la desviación y varianza. La desviación es de 0.68. Considero que la desviación es relativamente alta, Esto significa que las puntuaciones de la variable “rule of law” están considerablemente dispersas por la distribución muestral en relación con la puntuación media. Mientras que la varianza o variación promedio de las puntuaciones en la distribución de esta variable es de 0.47. Me parece que es una puntuación a considerar; dado que, existe una variación producida por datos que generan ciertos resultados que pueden afectar la visión real del objeto de estudio. Por otro lado, el rango es de 3.28 y muestra lo anteriormente observado: los datos están dispersos. Sobre los cuartiles, esto demuestra que la mediana es de -0.21(al 50%), y que a los extremos hay datos atípicos tales como (0.39 al 0% y 0.93 al 100%). Para ver con mayor precisión el tema de los datos atípicos, realicé un boxplot en el que revisé los outliers. Al ver el boxplot pude observar que la mayor concentración de outliers entre los datos mayores de la variable. Cuarto, debo mencionar que al realizar un histograma puede observar que efectivamente hay una acumulación en las puntuaciones menores de la variable.





* + - 1. **Variable gasto de salud (Mayra)**

Según el Banco Mundial, Current health expenditure (% of GDP) o ***Gasto actual en salud (CHE por sus siglas en inglés)*** es un estimado de los gastos en salud de cada año considerando los distintos bienes y servicios que son consumidos. Este indicador no evalúa otros gastos relacionados a la salud como son la infraestructura, maquinaria o las existencias en suministros de vacunas.

Al momento de realizar el análisis de esta variable independiente se obtuvieron los siguientes resultados. En primer lugar, se analizaron los estadísticos descriptivos con el comando *summary* que arrojó los valores pertenecientes a el primer y tercer cuartil, el valor mínimo y máximo, la media y la mediana los cuales serán descritos a continuación.

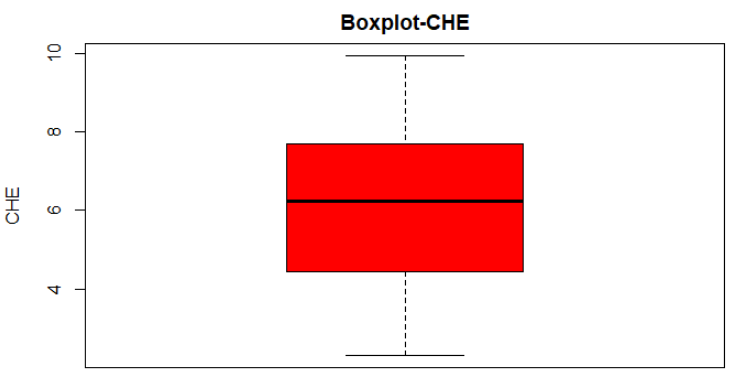
Con relación al valor mínimo y máximo presentes en esta variable, los resultados obtenidos fueron: 2.312 y 9.952 respectivamente. Por parte de los cuartiles, los cuales dividen los datos en partes iguales, el *summary* nos permite visualizar el primer y tercer cuartil, que en el caso de esta variable son: 4.464 y 7.668. Los últimos estadísticos descriptivos que se visualizan son la media y mediana que en este caso arrojan los valores de 6.185 y 6.247 respectivamente.

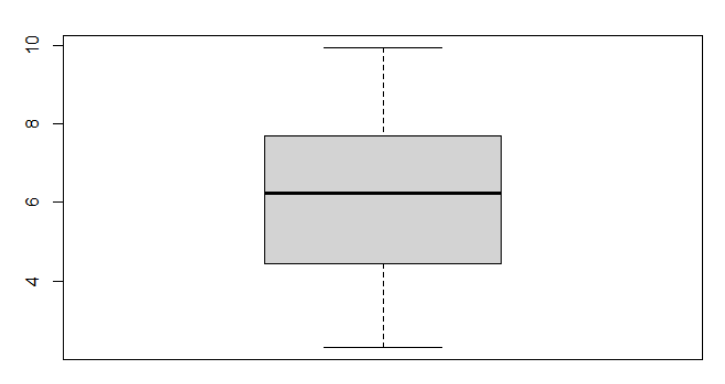
En segundo lugar, se solicitaron los estadísticos de dispersión y se obtuvieron los resultados de la asimetría y la curtosis para lo cual se llamaron a dos comandos del *package* e1071: *skewness* y *kurtosis*. El resultado obtenido para la asimetría en el caso de esta variable fue 0.04373928, mientras que para la curtosis el valor fue -1.089806.

En tercer lugar, se procedió a utilizar los comandos *sd* y *var* para obtener la desviación estándar y la varianza cuyos valores fueron 1.938588 y 3.758124 respectivamente.

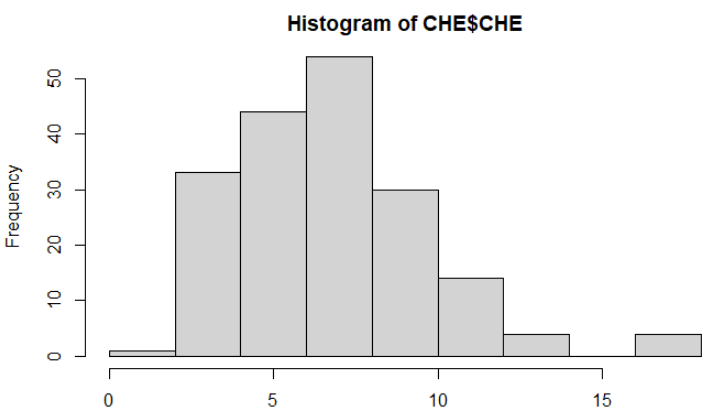
Finalmente se obtuvieron dos gráficas, la primera fue el *boxplot* o gráfico de cajas (Gráfico 1) que muestra los cuartiles, el rango intercuartil y los valores atípicos (Gráfico 2) los cuales también son presentados ya sea de manera numérica en el caso de los dos primeros o gráfica en el caso del tercero. La segunda gráfica obtenida es un histograma (Gráfico 3) el cual nos permite visualizar mayor cantidad de valores los cuales se agrupan en rangos y permite también visualizar la curva de distribución para poder reconocer la presencia de sesgos.

**Gráfico 1**

**Gráfico 2**



**Gráfico 3**



* + - 1. **Variable alfabetización (Mayra)**

La **Tasa de alfabetización de adultos,** según el Banco Mundial, es el porcentaje de personas que pueden leer y escribir. Este indicador permite evaluar el nivel educativo de los países.

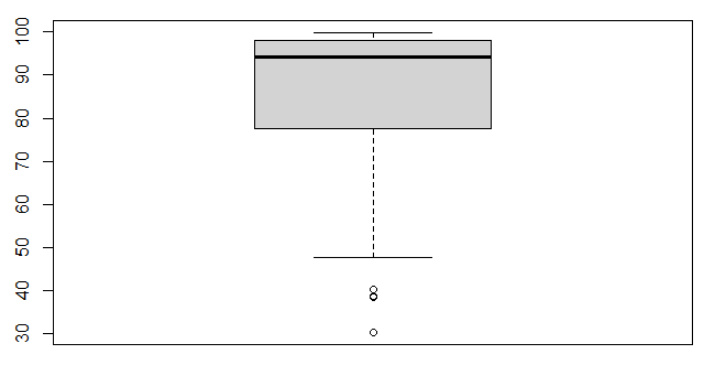
Para analizar esta variable independiente se analizaron los estadísticos descriptivos, los de dispersión, la varianza y la desviación estándar y se solicitaron dos gráficas. A continuación, se procederá a detallar y presentar los resultados obtenidos.

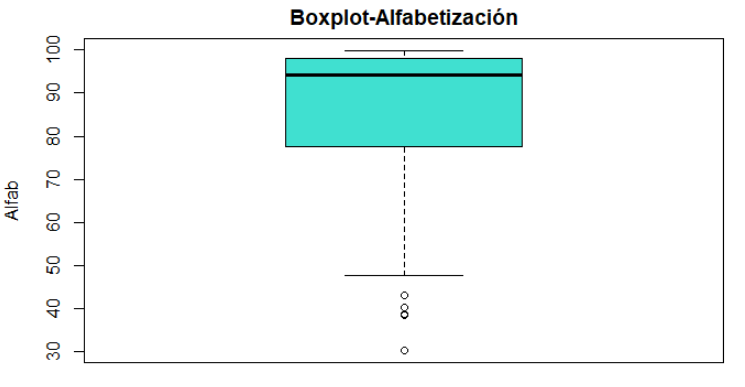
Con relación a los estadísticos descriptivos se utilizó el comando *summary* para obtener los valores del primer y tercer cuartil, el valor mínimo y máximo, la media y la mediana. Para esta variable, los resultados fueron los siguientes: Respecto a los valores mínimos y máximos, los resultados obtenidos fueron 30.41 y 99.89. Por otro los valores de los cuartiles que arrojó el *summary* corresponde al primer y tercer cuartil cuyos resultados son: 77.77 y 97.84. Finalmente, los últimos estadísticos descriptivos considerados son la media y la mediana, cuyos valores son 85.50 y 94.20 respectivamente.

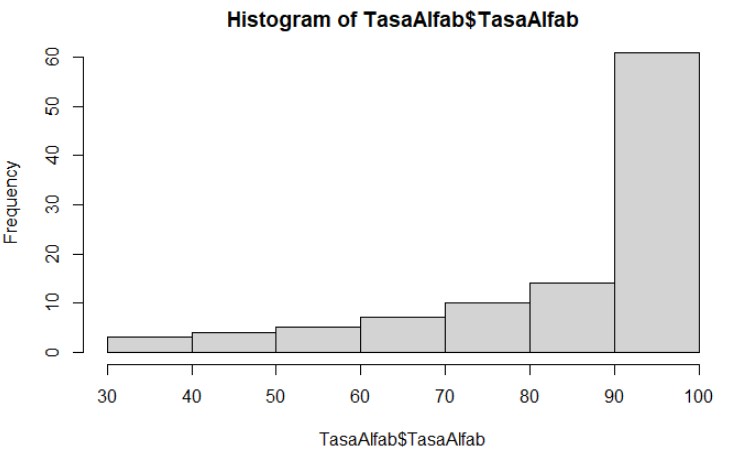
Para el análisis de esta variable también se solicitaron los estadísticos de dispersión a partir comandos del *package* e1071: *skewness y kurtosis* los cuales permitieron visualizar los valoresde la asimetría y la curtosis que para esta variable fueron -1.373262 y 0.9624922 respectivamente.

También se solicitaron los valores de la desviación estándar y varianza a partir de los comandos *sd* y *var* que mostraron como resultado 17.05116 para el primero y 290.742 para el segundo.

Con respecto a los gráficos, se obtuvieron dos: el *boxplot* o gráfico de cajas (Gráfico 1) que muestra los cuartiles,el rango intercuartil y los valores atípicos (Gráfico 2) los cuales también son presentados ya sea de manera numérica en el caso de los dos primeros o gráfica en el caso del tercero. También se solicitó un histograma (Gráfico 3) el cual nos permite visualizar mayor cantidad de valores los cuales se agrupan en rangos y permite visualizar la curva de distribución para poder reconocer la presencia de sesgos.

**Gráfico 1**

**Gráfico 2**

**Gráfico 3**

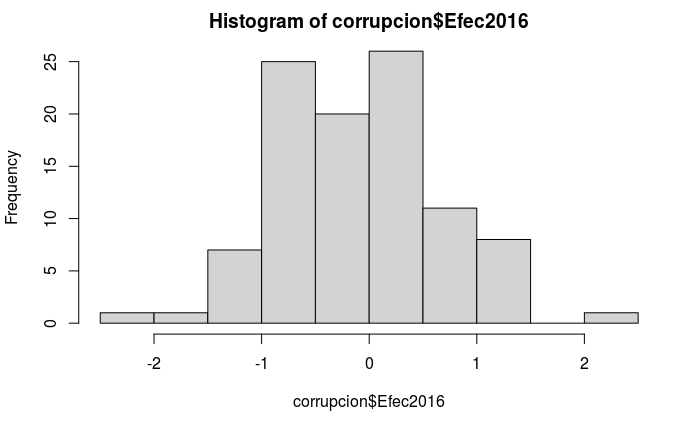
* + - 1. **Variable efectividad (Luis)**

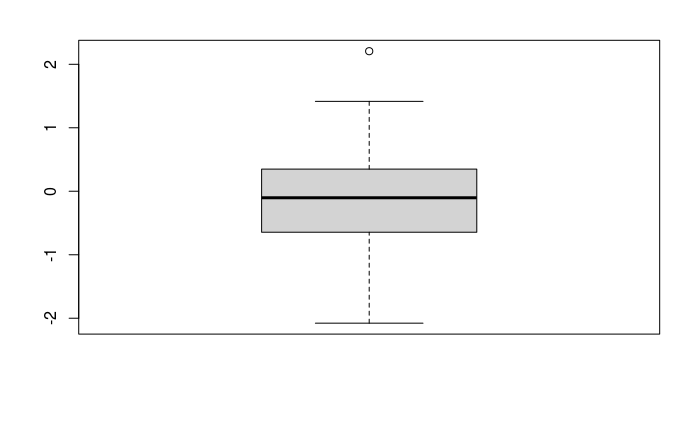
Entendemos por Efectividad del Gobierno a la percepción de la calidad de los servicios públicos, la calidad del servicio civil y el grado de independencia de las presiones políticas además de la calidad de la formulación e implementación de políticas.

Para el análisis de la variable de efectividad(variable independiente) usamos por un lado, estadísticos descriptivos en la cual aplicando *summary* nos da como resultado el valor mínimo (-2,07), el valor máximo (2.20), la mediana (-0,10) y la media (-0,08).

Por otro lado, usamos también *estadísticos de dispersión*, usando los códigos sd(), obtenemos la desviación estándar la cual resulta 0.726, también usamos var() para hallar la varianza el cual resulta 0.52; por su parte usamos también dentro de estos estadísticos la asimetría y curtosis; en el primero mediante el código skewness() obtenemos el valor de 0.21 y por su parte kurtosis() nos arroja un valor de 0.15

Por último, usamos dos gráficos para poder analizar la variable, una es un histograma y un boxplot, en el primero podemos ver la distribución de cada valor en un rango específico, y en segundo lugar podemos ver los cuartiles y tambien el valor atipico que tambien se muestra en el primer grafico.





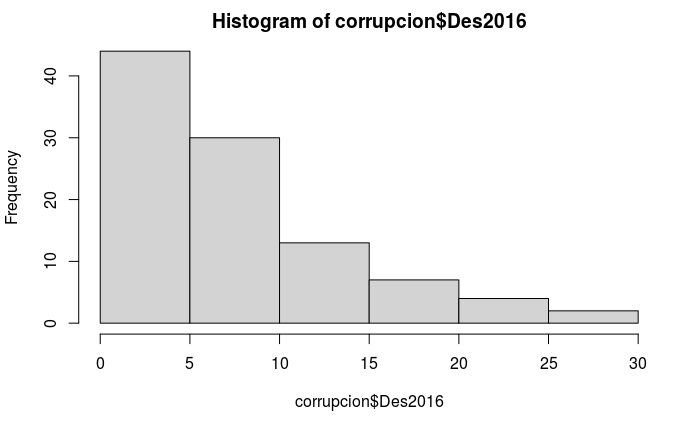
* + - 1. **Variable desempleo (Luis)**

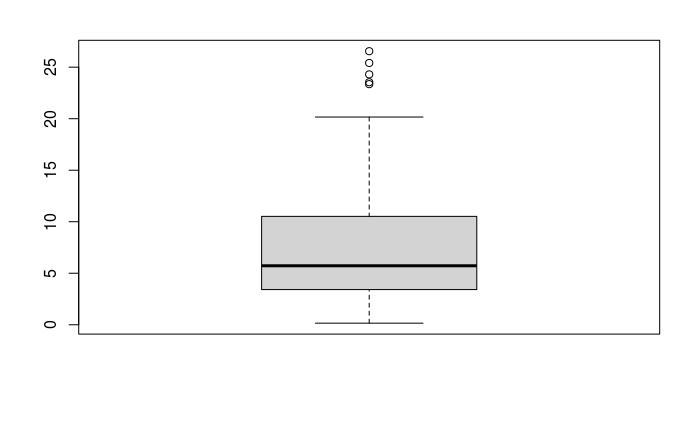
Dentro de lavariable de desempleo se toma en cuenta la proporción de la población activa que no tienen trabajo pero que está en busca de uno y está disponible para realizarlo, aquí también debemos tener en cuenta que las definiciones de población activa y desempleo difieren según el país.

Para el análisis de esta variable usaremos los métodos ya especificados anteriormente; por un lado, tenemos los *estadísticos descriptivos*, aplicando el *summary* tenemos los siguientes resultado; el mínimo es de 0,15, el máximo de 26,5, la mediana tiene un resultado de 5.7 y la media toma el valor de 7.5

Por su parte, los *estadísticos de dispersión* nos da como resultado lo siguiente; por un lado,la desviación estándar nos da un valor de 6.0 mientras que la varianza resulta 36.05; por otro lado, usando el comando skewness nos arroja una simetría de 1.3 mientras que curtosis es de 1.2.

Respecto a los gráficos, usamos el histograma y el boxplot para poder analizar la variable y ver de manera gráfica los rangos, cuartiles y los datos atípicos de la variable, en el primer caso el histograma nos muestra la distribución de los valores y los diferentes rangos; en el segundo caso, el boxplot a parte de mostrar los cuartiles nos muestra los valores atípico, que a diferencia de la variable anterior son mucho más.

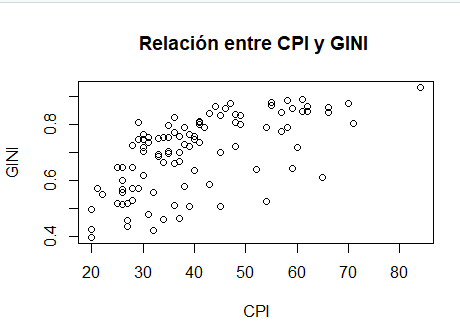




* 1. **ANÁLISIS BIVARIADO** 
     1. **Regresión lineal** 
        1. **Modelo 1: CPI- GINI (Isaura)**

Para el análisis bivariado entre la variable dependiente (CPI) y la variable independiente (GINI), primero se han establecido las hipótesis. Por un lado, tenemos la hipótesis nula, la cual establece que la variable GINI no influye en el la variable de percepción de la corrupción. Por otro lado, tenemos a la hipótesis alternativa señala que la variable GINI influye en la variable de percepción de la corrupción.

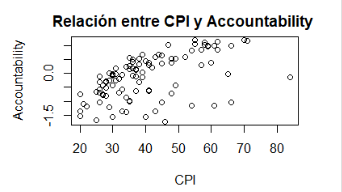
En primer lugar, podemos observar en el diagrama de dispersión una tendencia positiva; sin embargo, se puede evidenciar que existe un valor alejado del resto. Segundo, para enfocarnos si la variable independiente influye en la variable dependiente hemos de recurrir a observar el p-value. El resultado en la prueba nos arroja que este es menor a 0.05, así pues, rechazamos la hipótesis nula. De esta manera, comprobamos la hipótesis alternativa e indicamos que la variable gini si tiene efecto en la variable del índice de percepción de la corrupción. Asimismo,esto se puede comprobar con el nivel de correlación que es de 0.624035. En tercer lugar, para saber cuánto puede explicar dicha variable la incidencia en la variable dependiente, se procedió a realizar un modelo de regresión simple. De esta manera, gracias al R ajustado, se establece que el modelo 1 explica en un 38.32% la incidencia de la variable GINI en el índice de percepción de corrupción. Por último, la ecuación del modelo es la siguiente modelo1= -2.971 + 62.020 \* GINI2016.

****

* + - 1. **Modelo 2: CPI-Accountability (Isaura)**

Para motivos de interpretación de este modelo, hay que tener en cuenta las hipótesis de la investigación. Por un lado, tenemos que establecer una hipótesis nula, esta hace referencia a que la variable independiente (accountability) no influye en el índice de percepción de corrupción. Mientras, que la hipótesis alternativa (la cual es establecida por el investigador) hace referencia a que dicha variable podría explicar dicha percepción de corrupción en los países.

En primer lugar, tenemos que enfocarnos si dicha variable influye en la variable dependiente Por ello, recurrimos a observar el p-value, el resultado en la prueba nos arroja que este es 0.5391, es decir menor a 0.05; así pues, rechazamos la hipótesis nula. De esta manera, comprobamos la hipótesis alternativa e indicamos que si es posible el efecto de la variable independiente en la variabilidad del índice de percepción de la corrupción. Asimismo, esto se puede comprobar con el nivel de correlación que es de 0.5391. En tercer lugar, para saber cuánto puede explicar dicha variable la incidencia en la variable dependiente, se procedió a realizar un modelo de regresión simple. De esta manera, gracias al R ajustado, se establece que el modelo 2 explica en un 28% la incidencia de la variable GINI en el índice de percepción de corrupción. Por último, la ecuación del modelo es la siguiente modelo2= 41.529 + 9.175\* accountability2016.

****

* + - 1. **Modelo 3: CPI-IDH (Brigitte)**

*Ho: La variable idh (índice de desarrollo humano) no influye en la variable de corrupción.*

*Hi: La variable idh (índice de desarrollo humano) influye en la variable corrupción.*

Primero, comencé con un análisis del diagrama de dispersión. Al observar el diagrama se puede notar que existe una tendencia hacia arriba, dirección positiva. Sin embargo, es preciso mencionar que se ven varios outliers en todo la distribución. Segundo, realicé el test de correlación. Los datos obtenidos son los siguientes. El p-value es de 4.026e-12 (menor a 0.05) por lo que se puede aceptar esta correlación como significativa. Mientras que el nivel de correlación es de 0.62 entre las variables presentadas. Es alto, según lo dicho por Cohen (0.5 a más es alto). Tercero, solicité el modelo de regresión simple. Este modelo demostró que existe una influencia significativa del IDH en la corrupción. Puesto que, al observar la tabla se muestra que el p-value de esta variable es de 4.03e-12 menor a 0.05. Entonces, se rechaza la hipótesis nula. Asimismo, se observa que el R ajustado es de 0.38 o 38%. Esto significa que el poder explicativo de este modelo es de 0,38%, un término medio (aunque para las ciencias sociales es realmente bueno). Por último, revise los coeficientes para poder construir la ecuación. El intercepto es de -2.97 y la variable idh tiene coeficiente de 62.02. Por lo tanto, la ecuación sería: modelo 3= -2.97 + 62.02\*estimado de idh). Así, puedo señalar que, por ejemplo, si un país tiene 0.53 en estimado de idh, la ecuación sería la siguiente: modelo 3(corrupción)= -2.97+62.02(0.53). Al realizar la operación se descubre que el resultado es de 29.9. Según el índice de corrupción, esto supondría unos niveles altos de corrupción que corresponden a la existencia de niveles bajos de idh.

* + - 1. **Modelo 4 : CPI- Rule of law (Brigitte)**

*Ho: La variable rule of law no influye en la variable de corrupción.*

*Hi: La variable rule of law influye en la variable corrupción.*

Primero, comencé con un análisis del diagrama de dispersión. Al observar el diagrama se puede notar que existe una tendencia hacia arriba, dirección positiva. Considero que existen menos outliers, y los que hay, se encuentran en los niveles mayores. Segundo, realicé el test de correlación. Los datos obtenidos son los siguientes. El p-value es de 2.2e-16 (menor a 0.05) por lo que se puede aceptar esta correlación como significativa con un 95% de confianza. Mientras que el nivel de correlación es de 0.92 entre las variables presentadas. Es muy alto, según lo dicho por Cohen (0.5 a más es alto). Tercero, solicité el modelo de regresión simple. Este modelo demostró que existe una influencia significativa del IDH en la corrupción. Puesto que, al observar la tabla se muestra que el p-value de esta variable es de 2e-16 (menor a 0.05). Entonces, se puede señalar que esta variable tiene influencia significativa. Asimismo, se observa que el R ajustado es de 0.84 o 84%. Esto significa que el poder explicativo de este modelo es bastante alto. Por último, revise los coeficientes para poder construir la ecuación. El intercepto es de 43-13 y la variable Rule of law tiene coeficiente de 17.84. Por lo tanto, la ecuación sería: modelo 4= 43.14+ 17.84\*estimado de Rule of Law). Así, puedo señalar que, por ejemplo, si un país tiene -1.5 (bajo) en estimado de Rule of Law, la ecuación sería la siguiente: modelo 4(corrupción)= 43.14 + 17.84 (0.53). Al realizar la operación se descubre que el resultado es de 16.37. Según el índice de corrupción, esto supondría unos niveles altos de corrupción que corresponden a niveles muy bajos de rule of law. Así, se descubre la influencia de la variable “Rule of Law” frente a los índices de corrupción.

* + - 1. **Modelo 5 : CPI- CHE**

En este caso se procederá a detallar los resultados obtenidos del análisis bivariado entre la variable dependiente***Índice de percepción de corrupción (CPI por sus siglas en inglés)*** y la variable independiente ***Gasto Actual en salud (CHE)***para lo cual se formularon dos hipótesis: la primera es la hipótesis nula (H0) bajo la cual se supone que la variable independiente ***Gasto Actual en salud (CHE)***no explica a la variable dependiente ***Índice de percepción de corrupción (CPI por sus siglas en inglés).***La segunda hipótesis planteada para este análisis será la alternativa (Ha) bajo la cual se supone que la variable independiente si explicaría a la variable dependiente.

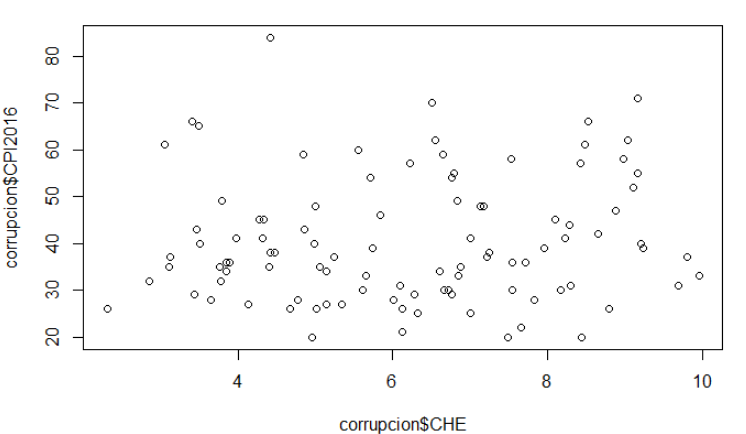
Para realizar el análisis en primer lugar se solicitó un diagrama de dispersión (Gráfico 1) de ambas variables para ver la relación existente. En este caso, se puede apreciar que el gráfico tiene los puntos muy dispersos lo cual dificulta determinar la relación.

En segundo lugar, se procedió a realizar el test de correlación de ambas variables utilizando el comando *cor.test*. El resultado que se visualizó fue el *p value* cuyo valor es 0.3609. Al ser mayor a 0.05 no se rechaza la hipótesis nula (H0). Por lo tanto, no se puede asumir que la variable independiente ***Gasto Actual en salud (CHE)***explique a la variable dependiente ***Índice de percepción de corrupción (CPI por sus siglas en inglés)****.* Además, el *cor.test* arrojó el valor de la fuerza de correlación *(cor)* que en este caso es 0.09232716 lo cual muestra que es baja.

En tercer lugar, se procedió a realizar un modelo de regresión simple para lo cual se utilizó el comando *lm* y se solicitó el *summary.* El resultado que se visualizó fue el R cuadrado ajustado, cuyo valor es -0.001593 lo que permite entender que la variable independiente ***Gasto Actual en salud (CHE)*** no explicaría a la variable dependiente ***Índice de percepción de corrupción (CPI por sus siglas en inglés).***

Finalmente, se procedió a utilizar el comando *coefficients* para elaborar la ecuación de este modelo a pesar de su bajo resultado. Los resultados que se visualizaron fueron el intercepto *(Intercept)* que es 36.4798387 y el valor de la pendiente que sería 0.6354493. A partir de estos valores, la ecuación del modelo sería CPI = 36.479 + 0.635\*CHE

**Gráfico 1**



* + - 1. **Modelo 6 : CPI- alfabetización**

Para el análisis bivariado entre la variable dependiente***Índice de percepción de corrupción (CPI por sus siglas en inglés)*** y la variable independiente ***Tasa de alfabetización de adultos (%)*** se formularon dos hipótesis: la primera es la hipótesis nula (H0) bajo la cual la variable independiente ***Tasa de alfabetización de adultos (%)*** no explica a la variable dependiente ***Índice de percepción de corrupción (CPI por sus siglas en inglés)****.* La segunda hipótesis planteada para este análisis será la alternativa (Ha) que supone que la variable independiente si explicaría a la variable dependiente.

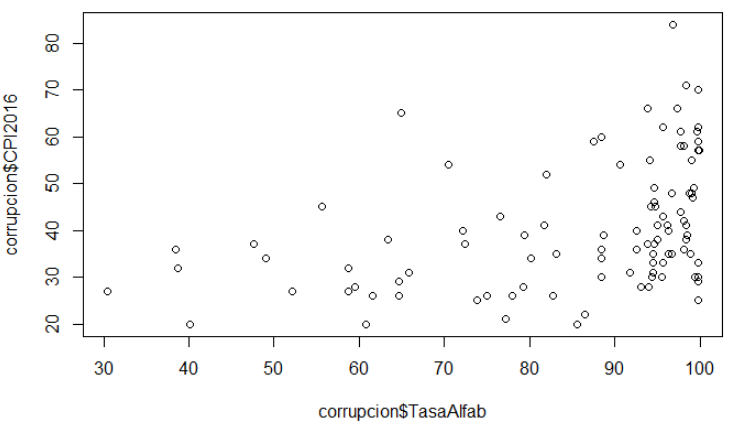
Para realizar el análisis bivariado entre estas variables en primer lugar se solicitó un diagrama de dispersión (Gráfico 1) para ver la relación existente. En este caso, el gráfico presenta los puntos muy dispersos, pero poco a poco estos se van agrupando. Sin embargo, aún no se puede determinar la existencia de una relación.

En segundo lugar, se procedió a realizar el test de correlación de ambas variables con el comando *cor.test*. El resultado que se visualizó fue el *p value* cuyo valor es 6.847e-05. Al ser menor a 0.05 se rechaza la hipótesis nula (H0), con lo cual se puede asumir que la variable independiente ***Tasa de alfabetización de adultos (%)*** explica la variable dependiente ***Índice de percepción de corrupción (CPI por sus siglas en inglés).***Además, el *cor.test* arrojó el valor de la fuerza de correlación (*cor)* que en este caso es 0.3873634, lo cual muestra que la correlación tiene una fuerza media.

En tercer lugar, se realizó un modelo de regresión simple con el comando *lm* y se solicitó el *summary* con el cual se visualizó el R cuadrado ajustado, cuyo valor es 0.1414 lo que permite entender que la variable independiente ***Tasa de alfabetización de adultos (%)*** explicaría a la variable dependiente ***Índice de percepción de corrupción (CPI por sus siglas en inglés)*** en un 14.14%

Finalmente, se procedió a utilizar el comando *coefficients* para elaborar la ecuación de este modelo. Los resultados que se visualizaron fueron el intercepto *(Intercept)* que es 14.493740 y el valor de la pendiente que sería 0.303111. Es así que la ecuación del modelo sería CPI = 14.49+ 0.303\*Tasa de alfabetización.

**Gráfico 1**



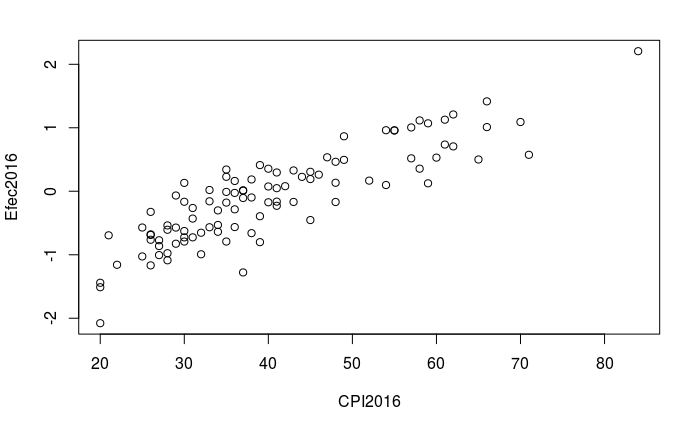
* + - 1. **Modelo 7 : CPI- efectividad (Luis)**

Partiremos del supuesto que existen dos hipótesis; la hipótesis nula la cual nos indica que la variable independiente (efectividad) no influye en nuestra variable dependiente (percepción de corrupción), mientras que la hipótesis alternativa nos indica que la variable efectividad si influye en la variable de percepción de corrupción.

Comenzamos comprobando que nuestro modelo es válido o no, luego de hacer la regresión nos dirigimos al valor del p-value el cual toma el valor de <2.2e-16 siendo este menor que 0.05, por lo tanto con este resultado rechazamos la hipótesis nula y podemos concluir que nuestro modelo es válido.

Ahora nos dirigimos al valor del R2 ajustado para saber que tanto nuestro modelo puede explicar la variabilidad de la dependiente, en este caso el valor del R2 ajustado es de 0.7501, es decir explica el 75% del modelo.

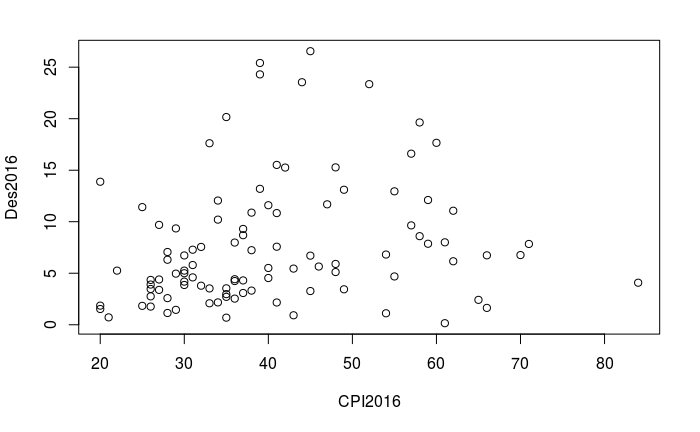
Por último, debemos construir la ecuación y la identificación de los coeficientes, al ser una regresión lineal simple, solo tenemos dos variables el intercepto y el coeficiente de efectividad, en este caso el intercepto tiene un valor de 41.8 mientras que el de efectividad un valor positivo de 15.9; con estas dos variables la adaptamos a la ecuación “y=a+bx” dando como resultado “CPI=41.83+15.92(efectividad)”

**8.Modelo 8 : CPI- desempleo (Luis)**

Al igual que los modelos anteriores se parte del supuesto de la existencia de dos hipótesis, por un lado la hipótesis nula, la cual nos indica que la variable independiente desempleo no afecta o influye en la variable dependiente percepción de la corrupción; y por otro lado, la hipótesis alternativa en la cual la variable desempleo su influye en la dependiente percepción de la corrupción.

Comenzamos con probar si es que el modelo es válido o no, aplicando una regresión simple obtenemos que nuestro p-valor tiene una valor de 0.04 el cual es menor a 0.05, dando como resultado que nuestro modelo si es válido rechazando así la hipótesis nula.

Seguimos con la capacidad explicativa de nuestro modelo, en este caso tenemos un valor de 0.031 en el R2 ajustado, teniendo así que el modelo explica solo el 3% de la variabilidad de la dependiente.

Por último, construiremos la ecuación “*y=a+bx”* a partir de los valores del intercepto y el coeficiente de desempleo, en este caso la ecuación tiene dos valores a reemplazar, el intercepto (a) con un valor de 36.96 y el coeficiente de desempleo (b) con un valor de 0.45, teniendo como resultado la ecuación: “CPI= 36.95+0.45(desempleo)

1. **CONCLUSIONES**

Tras aplicar los métodos estadísticos antes expuestos, se concluye que por ahora las hipótesis planteadas son correctas y que algunos de los cuatro grupos de variables influyen de manera significativa en el índice de la percepción de corrupción (CPI). Con respecto a la primera hipótesis, sobre los factores políticos, encontramos los siguientes resultados:La variable ***efectividad del gobierno*** si influye en el la variable dependiente ***Índice de Percepción de Corrupción*** ya que esta explica el 75% de la variabilidad de esta última. ***Rule of law*** por su parte, tiene un R2 ajustado de 0.84 es decir que explica el 84% de la variabilidad del CPI. Finalmente, la variable ***accountability*** a diferencia de los dos factores anteriores solo explica un 28%, siendo el más bajo de este grupo. Con respecto a la segunda hipótesis, la cual abarca los factores sociales, tenemos los siguientes resultado: ***el coeficiente de GINI*** nos arroja un valor explicativo del 38%, mientras que la variable ***Tasa de alfabetización*** posee un valor de 14%. En el caso de la variable ***Desempleo*** tiene un valor de 3% por lo cual no podemos decir en este caso exista influencia en la variable dependiente (no de manera significativa) y por último, en la tercera tenemos que ***IDH*** posee un valor explicativo de 38%.

Finalmente, es necesario mencionar que la variable ***Gasto público en salud*** (cuarta hipótesis)tiene un valor explicativo negativo de -0.15% lo cual indica que no posee relevancia al momento de explicar la variabilidad de la dependiente.