**Exploración de datos**

**Información general:**

**Tema**: Desigualdad de género

**Objetivo**: Determinar los factores que explican e influyen en la desigualdad de género a nivel mundial.

**Unidad de análisis**: 196 países

**Link del repositorio:**

**Tabla resumen de variables**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tipo** | **Latente** | **Observable (Operacionalización** | **Fuente** |
| Variable dependiente | Desigualdad de género | Índice de Desigualdad de Género. Tiene un rango que va de 0 a 1: más cerca a 0, hay más igualdad entre mujeres y hombres, más cerca a 1,hay poca o nula igualdad en las dimensiones medidas | PNUD - Informes de Desarrollo Humano |
| Variable independiente | Salud | Rango de 0 a 1. Mide el grado de implementación de un marco legal que proteja los derechos y la salud reproductiva de las mujeres | OCDE - Base de datos sobre género, instituciones y desarrollo (2019) |
| Social | Rango de 0 a 1. Mide el grado de implementación de un marco legal que proteja a las mujeres de violencia incluyendo violencia por parte de su pareja sentimental, violación, y acoso sexual, sin excepciones legales y con un enfoque extensivo |
| Libertades civiles | Porcentaje de mujeres del número total de representantes en el Parlamento. |
| Porcentaje de mujeres del número total de personas que declaran no sentirse seguras caminando solas en la noche en la ciudad o área en la que viven. |
| Porcentaje de mujeres del número total de personas que declaran no tener confianza en el sistema judicial y las cortes de su país. |
| Educación | Porcentaje de mujeres de la población de 25 años a más que ha llegado (pero que no necesariamente ha terminado) a un nivel de educación secundaria. |
| Economía | Proporción de mujeres a hombres que están desempleadas |
| Porcentaje de la población femenina de 15 años o más que tiene una cuenta en una institución financiera o en alguna banca móvil |

**Análisis descriptivo**

El presente estudio analizará los posibles factores relacionados con la desigualdad de género. Como variable dependiente tenemos el Índice de Desigualdad de Género y como independientes tenemos Autonomía reproductiva, Violencia contra la mujer, Voz política, Libertad de movilización, Acceso a la justicia, Población con secundaria completa, Tasa de mujeres en desempleo, Mujeres con acceso a una cuenta en una institución financiera o un servicio móvil de transferencias.

**Análisis univariado:**

**Variable dependiente**: El Índice de Desigualdad de Género mide la desigualdad de género en tres aspectos importantes del desarrollo humano: salud reproductiva, empoderamiento y situación económica. Mide los costos de desarrollo humano de la desigualdad de género. Tiene un rango de 0 a 1. Cuanto más cerca a 1, significa que hay mayores disparidades entre mujeres y hombres; por lo tanto, mayor pérdida de desarrollo humano. Según los estadísticos descriptivos, el valor mínimo es 0.04. Esto quiere decir que hay un grupo de países que sí están logrando cerrar brechas de género, como Dinamarca, Suecia y Suiza. También observamos que el valor máximo es 0.83 que corresponde al caso de Yemen y nos da a entender que es el territorio en el que hay más desigualdad de acuerdo con los parámetro del índice. Además el valor de la media (0.35) indica que a nivel general se están produciendo ciertos avances, pero todavía hay disparidades entre mujeres y hombres.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mínimo | Máximo | Media | Mediana | Desviación Estándar |
| *0.04* | *0.83* | *0.35* | *0.37* | *0.19* |

**Variables independientes**:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variables | 0 | 0.5 | 1 | NA |
| *Autonomía Reproductiva* | *77* | *25* | *67* | *27* |
| *Violencia contra la mujer* | *30* | *40* | *99* | *27* |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variables | Mínimo | Máximo | Media | Mediana | Desviación Estándar |
| *Voz Política* | *0* | *61.3* | *23.54* | *21.9* | *11.31* |
| *Libertad de Movimiento* | *29.7* | *91.1* | *63.49* | *65.4* | *10.82* |
| *Desempleo en mujeres* | *0.38* | *11.67* | *1.48* | *1.17* | *1.23* |
| *Cuenta Financiera o servicio de transferencias móviles* | *1.67* | *100* | *55.11* | *52.71* | *29.07* |
| *Acceso a la Justicia* | *23.20* | *68.1* | *51.56* | *52.35* | *8.18* |
| *Secundaria Completa* | *1.74* | *100* | *61.49* | *66.92* | *29.50* |

**Variable Autonomía reproductiva (“Autonomía”):** Tiene un rango de 0 a 1. La variable original tenía cinco niveles (0, 0.25, 0.5, 0.75, 1), pero para el presente análisis consideramos que sería más eficiente hacer una recodificación y quedarnos con 3 valores (0, 0.5, 1). El valor 0 corresponde a un marco que protege la salud reproductiva en caso de embarazo no deseado que no solicita ninguna justificación y si las solicita, son mínimas (77 países). El valor 0.5 corresponde a un marco que protege la salud reproductiva en caso de embarazo no deseado solo en situaciones justificadas (25 países). El valor 1 corresponde a un marco que protege la salud reproductiva en caso de embarazo no deseado con justificaciones muy estrictas o no brinda ningún tipo de protección en absoluto (67 países). No se tiene información para 27 países.

**Variable Violencia contra la mujer (“Violencia”):** Tiene un rango de 0 a 1. Esta variable mide el grado de implementación de un marco legal que protege a las mujeres de violencia incluyendo violencia por parte de su pareja sentimental, violación, y acoso sexual, sin excepciones legales y con un enfoque extensivo. La variable original tenía cinco niveles (0, 0.25, 0.5, 0.75, 1), pero para el presente análisis consideramos que sería más eficiente hacer una recodificación y quedarnos con 3 valores (0, 0.5, 1). El valor 0 corresponde a un marco legal extensivo que sí brinda protección. El valor 0.5 corresponde a un marco que protege pero con excepciones legales. El valor 1 corresponde a un marco que no brinda protección o en ciertos casos es mínima. No se tiene información para 27 países.

**Variable Voz Política (“VozPolitica”):** Variable numérica en donde el valor máximo corresponde a 61.3. Este indica que hay países que cuentan con un alto número de mujeres que ocupan escaños en el parlamento. Pero el valor de la media (23.54) indica que en general todavía no hay una amplia presencia de mujeres en el congreso lo que sería un reflejo de la poca apertura que tienen para participar en la vida política.

**Variable Libertad de Movimiento (“LibertadMov”):** El valor máximo resalta en esta variable numérica. Esto indica que hay un alto porcentaje de mujeres (91.1%) que no se sienten seguras caminando solas en la noche en la ciudad o área en la que viven y corresponde al país de Finlandia. El valor de la media nos permite deducir que en promedio un poco más de la mitad de mujeres de los países analizados comparten esta percepción de inseguridad.

**Variable Mujeres con cuenta en una institución financiera o servicio móvil de transferencias (“CuentaF”):** Resalta el valor máximo de 100% que indica que en el caso de Dinamarca, todas las mujeres tienen acceso a servicios financieros. Por otro lado, el valor mínimo de 1.67% en el caso de Yemen nos indica que solo un pequeño número de mujeres tienen una cuenta o servicio bancario a su nombre. Con la media (55.11%) podemos deducir que solo un poco más de la mitad de mujeres a nivel general tienen acceso a estos servicios. Estos valores no permiten deducir que todavía un gran porcentaje no tiene independencia económica o si la tiene, es mínima.

**Variable Desempleo de mujeres (“Desempleo”):** En este caso, el valor máximo corresponde a Baréin donde 11.67 mujeres por cada 100 hombres no tiene empleo, y el valor mínimo corresponde a Papúa Nueva Guinea donde 0.38%. Al observar la data, notamos que en países desarrollados como Finlandia, Francia o Alemania la proporción es igual o menor a 1. Con la media observamos que a nivel general hay una proporción baja de mujeres que no tienen empleo, sin embargo no podríamos asumir que aquellas mujeres que sí tienen trabajo cuentan con condiciones laborales justas. Esto solo nos da una idea del acceso al mercado laboral.

**Variable Acceso a la justicia (“AccesoJ”):** En este caso, observamos el porcentaje de mujeres del número total de personas que no confía en el sistema judicial. El valor mínimo corresponde a Burkina Faso y el máximo a Gambia. Al analizar la data vemos que en países desarrollados sorprende que la cantidad de mujeres que no confían en sus sistemas judiciales llegue también casi al 50%, como es el caso de Finlandia e Islandia. Se esperaría que en países con mejores condiciones de vida haya mayor confianza en la justicia.

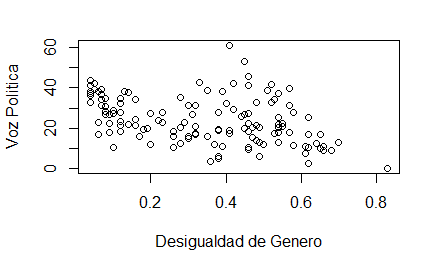
**Variable Secundaria completa (“SecundariaC”):** Podemos observar extremos. Por un lado, tenemos el valor mínimo que corresponde a Chad donde apenas el 1.74% de mujeres mayores de 25 ha llegado a un nivel de educación secundaria y por otro lado, tenemos a países como Finlandia, Austria o Canadá donde todas las mujeres mayores de 25 han recibido educación secundaria. El valor de la media no permite deducir que todavía hay una brecha casi del 40% a nivel mundial en educación secundaria.

**Análisis bivariado**:

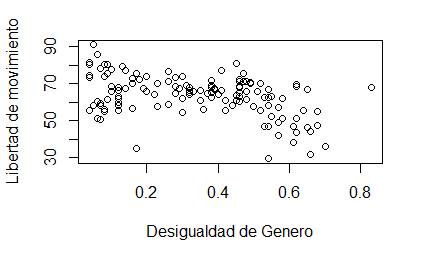
**Variable Autonomía reproductiva (“*Autonomia*”):** De acuerdo a los datos obtenidos de la prueba de hipótesis *Anova de un factor*, al haber obtenido un *P-value* (4.26e-09) menor a 0.05, podemos rechazar la hipótesis nula, la cual postula que la media de la variable “Desigualdad de género” y la media de los grupos de “Autonomía” son iguales. Por ello, usando *Tukey HSD* para la prueba de comparaciones múltiples, obtenemos que en el primer grupo (regular-bueno), el *P-value* es 0.0001; este es menor a 0.05, por ende, se rechaza la Hipótesis nula. En el caso del segundo grupo (malo-bueno), el valor de *P-value* (0.0) también es menor a 0.05; lo que hace que se rechace la hipótesis nula. Por último, el tercer grupo (malo-regular) posee un *P-value* (0.95) mayor a 0.05, por ello, no se rechaza la hipótesis nula. Entonces, según las pruebas de comparaciones múltiples *Tukey*, la media de los grupos “regular-bueno” y “malo-regular” difieren entre sí; mientras que las medias del grupo malo-regular no presentan variación significativa.

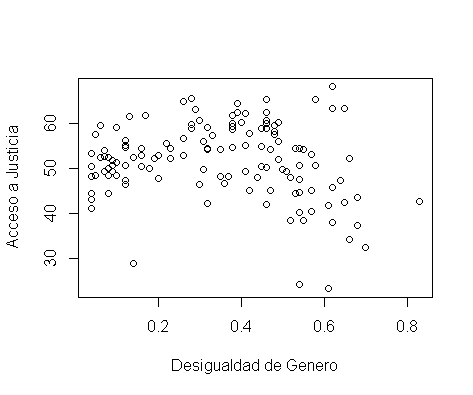
**Variable Violencia contra la mujer (“*Violencia*”)**: El *P-value* (0.006) obtenido de la prueba de hipótesis *Anova de un Factor*, es menor a 0.05, por ende, se rechaza la hipótesis nula. Entonces, es viable decir que la variable “Desigualdad de género” y los grupos de la variable “Violencia”, presentan diferencia de medias en al menos un grupo. Para conocer exactamente qué grupo es el que posee una media distinta a los demás, usaremos el comando TukeyHSD. Con este obtenemos que el primer grupo (regular-buena) posee un *P-value* (0.011) menor a 0.05; por ende, se rechaza la noción de que la media de estos grupos son iguales. Caso similar es el del segundo grupo (malo-bueno), ya que tiene un *P-value* (0.011) menor a 0.05. Por último, el tercer grupo (malo-regular) posee un *P-value* (0.85) mayor a 0.05; entonces no es posible rechazar la hipótesis nula. En conclusión, según las pruebas de comparaciones múltiples *Tukey*, existe diferencia de medias entre los niveles “regular-bueno” y “malo-bueno” de implementación del marco legal que proteja a las mujeres ante casos de violencia.

**Variable Voz Política (“*VozPolitica*”)**:

La forma de la nube de puntos, del gráfico de dispersión (aquel que nos dará un primer vistazo de la asociación entre las variables), nos informa que existe una relación inversa entre las variables; es decir, mientras la variable dependiente “Desigualdad de Género” aumenta, la variable independiente “Voz Política” disminuye. Sin embargo, por la dispersión de los puntos de la nube, debemos de realizar la prueba *R de Pearson*. En esta obtuvimos un *P-value* de *6.529e-06*, el cual, al ser menor a 0.05 nos permite rechazar la hipótesis nula; por ende, podemos sostener que las variables índice de desigualdad de género y el porcentaje de mujeres del total de políticos en el Parlamento son estadísticamente dependientes. Asimismo, con un coeficiente de correlación de -0.38, se sostiene que esta asociación entre ambas variables es inversa y cuenta con una intensidad media.

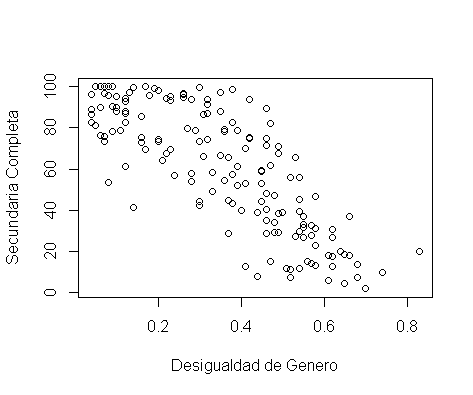
**Variable Libertad de Movimiento (“*LibertadMov*”)**:

Explorando visualmente este gráfico de dispersión, podemos notar que existe asociación entre las variables Desigualdad de Género y Libertad de Movimiento. También es notable la existencia de una relación inversa entre ellas; es decir, mientras la variable dependiente aumenta, la independiente disminuye. Para brindar validez a nuestra deducción, realizamos la prueba *R de Pearson*, de la cual obtuvimos un *P-value* (2.41e-07) menor a 0.05, lo cual nos permite rechazar la hipótesis nula. Entonces, ante la existencia de correlación lineal entre el índice de desigualdad de Género y el porcentaje de mujeres que han declarado sentirse inseguras al caminar por la ciudad en las noches. Asimismo, el coeficiente de correlación (-0.43) nos indica que esta asociación presenta una intensidad media y un sentido inverso (la variable dependiente aumenta, la independiente disminuye).

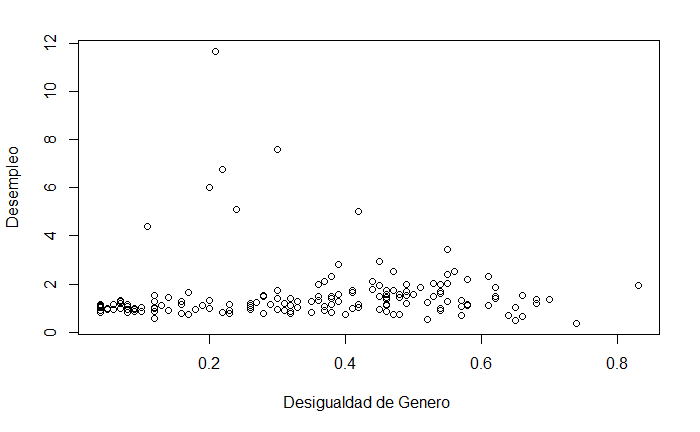
**Variable Acceso a la justicia (“*AccesoJ*”)**:

En el gráfico de dispersión de la variable Desigualdad de Género y Acceso a la Justicia (independiente) notamos una nube de puntos un poco dispersa. Al ser complicado determinar a simple vista las características de la asociación entre ambas variables, hemos realizado la prueba *R de Pearson* con la cual hemos obtenido un *P-valor* de 0.06. Este es mayor a 0.05, por ello, es viable sostener que no se puede rechazar la hipótesis nula; en otras palabras, los índices de desigualdad de género y el porcentaje de mujeres que no le tienen confianza al sistema de justicia de su país son estadísticamente independientes.

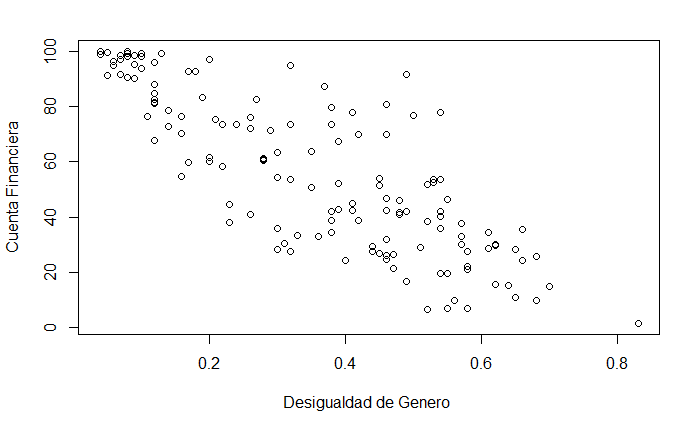
**Variable Secundaria completa (“*SecundariaC*”)**:

Nuestro análisis visual al gráfico de dispersión de las variables “Desigualdad de Género” (dependiente) y “Secundaria Completa” (independiente) nos muestra la existencia de una asociación entre estas. Podemos observar la relación siguiente: mientras que la variable dependiente aumenta, la independiente disminuye. Ante ello, obtuvimos —gracias a la prueba *R de Pearson*— un *P-value* (2.2e-16) menor a 0.05 y un coeficiente de correlación de -0.79. Esto nos permite concluir que los índices de desigualdad de género y el porcentaje de mujeres de 25 años a más, y que cuentan con secundaria completa, presentan una correlación lineal. Asimismo, podemos decir que esta asociación entre ambas es de intensidad alta y de sentido inverso (una aumenta mientras la otra disminuye).

**Variable Desempleo total en mujeres (“*Desempleo*”)**:

De acuerdo al gráfico de dispersión, se observa que si bien hay una nube de puntos fácilmente identificable esta está muy cerca al eje X, es decir, a la variable Desigualdad de Género; sin embargo, es complicado establecer una relación entre ellas a simple vista. Debido a ello, para poseer una noción más clara, aplicamos la prueba *R de Pearson*; la cual nos genera un *P-valor* (0.82) mayor a 0.05. Por ello, no podemos rechazar la hipótesis nula, la cual versa en que la variable dependiente e independiente son estadísticamente independientes. Por ende, es viable concluir que no hay una correlación lineal entre los índices de desigualdad de género y la proporción de mujeres a hombres que se encuentran desempleadas.

**Variable Mujeres con cuenta en una institución financiera o con dinero móvil (“*CuentaF*”)**:

En este gráfico de dispersión percibimos la existencia de una nube de puntos un poco dispersa pero esto no es impedimento para visualizar que mientras la variable dependiente (“Desigualdad de Género”) aumenta, la independiente (“Cuenta Financiera”) disminuye. Además, con la prueba *R* de Pearson obtuvimos un *P valor* de *2.2e-16* y como este es menor a *0.05*, podemos rechazar la hipótesis nula. Debido a ello, es viable concluir que las variables Desigualdad de género y Cuenta Financiera están correlacionadas de forma lineal; es decir, son estadísticamente dependientes. Asimismo, el coeficiente *R de Pearson* es *-0.83*, entonces podemos decir que existe una correlación inversa, con una fuerza de asociación alta entre los índices de desigualdad de género y el porcentaje de mujeres de 15 años a más que poseen una cuenta en alguna institución financiera.

**Bibliografía:**

Fernández, A. (2006). Mujeres y política en América Latina: dificultades y aceptación social. Argumentos (México, D.F.), 19(51), 117-142. Recuperado de <http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-57952006000200005&lng=es&tlng=es>

INEI.(2018). Perú: Brecha de género. INEI. Recuperado de: <http://www2.congreso.gob.pe/sicr/cendocbib/con5_uibd.nsf/C9B86E6376B3547D052583910073BDCF/$FILE/7.libro.pdf>

MIMP. (4 de abril de 2019) Decreto Supremo.Política Nacional de Igualdad de Género. El Peruano. Recuperado de <http://www.consulado.pe/Documents/Igualdad-Genero/politica-nacional-igualdad-degenero.pdf>