

Estimación de la esperanza de vida

Maria Alejandra Moncada Agudelo
mamoncadaa@eafit.edu.co

Abelino Sepulveda Estrada
asepulvede@eafit.edu.co

Mayo 2022

Índice

1. Introducción	1
2. Motivación y Justificación	2
3. Revisión de la Literatura	2
4. Marco Teórico	2
5. Análisis descriptivo de los datos	3
5.1. Procesamiento de los datos	3
5.2. Estadísticas descriptivas de los datos	4
5.3. Relación entre factores	4
6. Estimaciones	5
6.1. Regresión Lineal Simple	5
6.2. Regresión Lineal Múltiple	5
6.3. Prueba de Homocedasticidad	6
6.4. Mínimos Cuadrados Generalizados	6
6.4.1. Prueba de Homocedasticidad	6
6.5. No autocorrelación	7
6.6. Prueba de Normalidad	7
6.7. Intervalos de confianza	7
6.8. Análisis de las estimaciones	8
7. Conclusiones	9

1. Introducción

Al nacer, somos etiquetados por un indicador que nos permite saber una estimación de cuántos años viviremos si las condiciones sanitarias prevalecientes en el momento de nuestro nacimiento permanecen iguales durante toda nuestra vida. Dichas estimaciones varía dependiendo del país en el que se nace; este indicador es la *esperanza de vida al nacer*, también conocido como EVN. El EVN es una medida estadística del promedio de tiempo que se espera que un organismo viva. En los humanos el grado de mortalidad se puede ver afectado por factores como el sexo, las condiciones sanitarias, la aplicación de medidas de prevención en salud y el nivel económico de la población.

Hoy en día, se han realizado muchos estudios para explorar y determinar los factores que afectan la esperanza de vida. Estudios como los de [Briceño *et al.*, n.d.] y [Pathirathne & Sooriyarachchi, 2019] evidencian mediante regresiones y análisis de componente principales, que los factores que mayormente afecta dicha medida son variables como las demográficas, la composición del ingreso y las tasas de mortalidad. Estos factores son indicadores clave en el estudio de políticas públicas y la evolución del acceso de la salud ya que le indica a un país qué áreas dar importancia para mejorar de manera

eficiente la esperanza de vida de su población.

Según [Max Roser & Ritchie, 2013] desde 1900 la esperanza de vida mundial se ha más que duplicado y hoy en día superó los 70 años. Adicionalmente, revela que a principios del siglo *XIX* la esperanza de vida comenzó a aumentar en los países industrializados, mientras que se mantuvo baja en el resto del mundo; actualmente la desigualdad de la esperanza de vida sigue siendo muy grande entre los países y dentro de ellos, ya que en el 2019 se reveló la diferencia entre la esperanza de vida de varios países superaba los 30 años.

2. Motivación y Justificación

Como mencionamos anteriormente, muchos estudios motivados por hacer estimaciones sobre la esperanza de vida, consideran y revelan los factores que afectan dicho indicador. Se encontró que en muchos de estos estudios no se tomaba en consideración efectos como la inmunización de algunas enfermedades importantes como el *Hepatitis B* y *Polio* y el índice de desarrollo humano lo que nos motiva a incluir los estos factores a demás de observaciones de la evolución de la esperanza de vida en todos los países en los años 2000 y 2014, todo esto para realizar un regresión lineal múltiple considerando estos datos. Estas estimaciones se centrarán en factores como los de inmunización, de mortalidad, económicos, sociales y otros relacionados con la salud.

Realizar este estudio sobre la estimación de qué factores influyen en la esperanza de vida, ayudaría a sugerir a los países a qué áreas se le debería dar importancia para mejorar de manera eficiente la esperanza de vida de su población. Ya que sería más fácil para un país o ente gubernamental determinar los factores que contribuyen a un menor valor de la esperanza de vida.

3. Revisión de la Literatura

La esperanza de vida ha sido un tema de interés para varias áreas del conocimiento. Por lo cual no es extraño que su predicción en años futuros también sea un tema atractivo para la inteligencia artificial o la probabilidad. Lo anterior deriva en varios métodos de predicción que toman en cuenta diferentes factores y buscan superar las dificultades que presentaron métodos anteriores.

[Pascariu *et al.*, 2018] proponen la construcción de un modelo de predicción de la esperanza de vida femenina, basado en un análisis de la brecha entre la esperanza de vida de un país con el resto del mundo. Por otra parte [Kontis *et al.*, 2017] desarrollan un modelo bayesiano promediado (BMA) con un conjunto de 21 modelos, donde cada uno proporciona una distribución de la esperanza de vida en el futuro y la predicción final es una combinación probabilística proveniente de esos modelos individuales. Otro acercamiento es el de [Sormin *et al.*, 2019] que utiliza redes neuronales artificiales para estimar la esperanza de vida mundial es los años siguientes.

4. Marco Teórico

Según [Andreev *et al.*, 2006] si se mantienen las mejores prácticas, la esperanza de vida aumentará de manera lineal y no estocástica, además la brecha actual entre la esperanza de vida de la población y la esperanza de vida de los registros permanecerá constante, por lo que se considera que un análisis de regresión lineal puede ser oportuno para este caso.

[Monsef & Mehrjardi, 2015] mencionan que los factores que afectan a la esperanza de vida se pueden dividir en tres clases: social, económico y ambiental. Por falta de datos en este trabajo no se incluirán factores ambientales. Por otro lado como factor social se decidió incluir la educación por medio de predictores como años de escolaridad e índice de desarrollo humano, ya que según [Hazan, 2011] esta última tiene una correlación positiva con la esperanza de vida. [Balan, 2011] encontró que el número de camas y médicos en los hospitales influyen positivamente en nuestra variable dependiente, por lo que elegimos variables que represente en gasto público en salud.

Finalmente se encontró que no es común incluir factores de inmunización ante enfermedades que normalmente afectan a países en vía de desarrollo como lo son la Hepatitis B o el Polio, por lo que en este estudio se incluirán, con el fin de determinar si realmente tiene un impacto en la esperanza de vida.

5. Análisis descriptivo de los datos

El depósito de datos del Observatorio Mundial de la Salud (GHO) de la Organización Mundial de la Salud (OMS) realiza un seguimiento del estado de salud y de muchos otros factores relacionados de todos los países. El conjunto de datos utilizados en este trabajo se recopiló del mismo sitio web del repositorio de datos de la OMS y cuenta con datos relaciones con la esperanza de vida y los factores de salud de 193 países, sus datos económicos se recopilaron del sitio web de las Naciones Unidas.

5.1. Procesamiento de los datos

Mediante el uso del software Python se hizo una limpieza de los datos, es decir se eliminaron los países que contaban con datos vacío, a demás de eliminar los factores que no se contemplaban en el alcance del trabajo. Posteriormente, se hizo la selección de los años nuestro interés, es decir, el dataset se dividió en 2, los datos del año 2000 y los del 2014.

Los factores predictores a tomar en consideración para las estimaciones son los siguientes:

- **Status:** toma el valor de cero si el país está en desarrollo y 1 si está desarrollado
- **Adult Mortality:** Tasa de mortalidad de adultos de ambos sexos (probabilidad de morir entre los 15 y 60 años por cada 1000 habitantes)
- **infant deaths:** Número de muertes infantiles por cada 1000 habitantes
- **percentage expenditure:** Gasto en salud como porcentaje del PIB per capita (%)
- **Hepatitis B:** Cobertura de vacunación contra Hepatitis B entre los niños de 1 año (%)
- **BMI:** Índice corporal promedio de toda la población
- **Polio:** Cobertura de vacunación contra la poliomielitis entre los niños de 1 año (%)
- **Total expenditure:** Gasto del gobierno general en salud como porcentaje de gasto total del gobierno
- **GDP:** Producto Interno Bruto per capita (USD)
- **Income composition of resources:** Índice de desarrollo humano en términos de composición de ingreso de los recursos
- **Schooling:** Número de años de escolaridad

5.2. Estadísticas descriptivas de los datos

Variable	Min	Mediana	Media	Maximo
Life expectancy	39	71	66.75	81.10
Adult Mortality	2	155	181	665
infant deaths	0	4	37.54	1800
percentage expenditure	0	35.661	473.851	8246.130
Hepatitis B	4	84	73.67	99
BMI	1.40	37.9	34.70	67.9
Polio	3	88	76.4	99
Total expenditure	1.1	5.430	5.612	13.7
GDP	3.69	1276.29	5138.77	48736
Income composition of resources	0	0.6238	0.5274	0.9204
Schooling	0	11.50	10.74	20.40

Cuadro 1: Estadísticas descriptivas datos 2000

Variable	Min	Mediana	Media	Maximo
Life expectancy	48.1	73.70	71.54	89
Adult Mortality	1	135	148.7	522
infant deaths	0	2	24.56	957
percentage expenditure	0	151.10	1001.91	19479.91
Hepatitis B	2	92.71	82.58	99
BMI	2	47	40.75	77.1
Polio	8	94	84.73	99
Total expenditure	1.210	5.840	6.194	17.140
GDP	12.28	3154.51	9631.28	119172.74
Income composition of resources	0.3125	0.7270	0.6907	0.9459
Schooling	4.90	13.10	12.87	20.40

Cuadro 2: Estadísticas descriptivas datos 2014

Podemos notar en Cuadro 2 y 1 que dichos factores presentan cambios evidenciabiles dentro este periodo de tiempo. Por ejemplo nuestra variable dependiente, es decir Life expentancy pasa de tener un promedio de 66.75 años en 2000 a 71.54 en 2014.

5.3. Relación entre factores

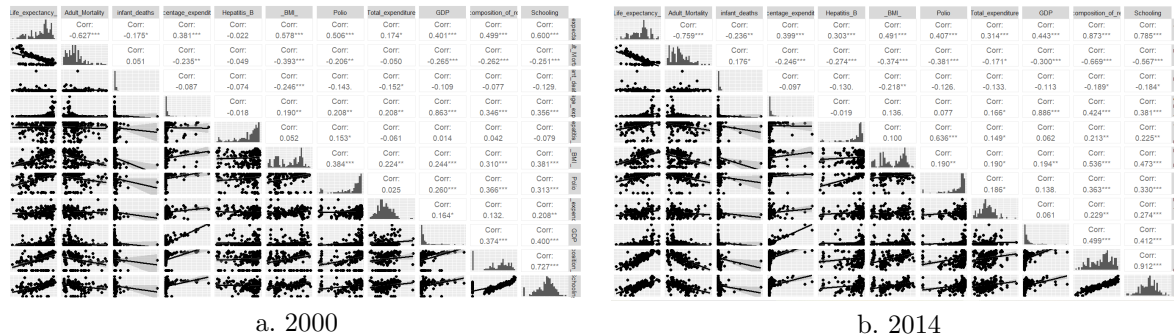


Figura 1: Correlación entre los factores

En la Figura 1 podemos observar que para ambas observaciones los factores *Income composition of resources* y *Schooling* esto podría ser explicado porque el *indice de desarrollo humano* también toma en cuenta la educación, lo cual justificar la correlación entre ambas, por lo cual *Schooling* podría ser redudante para la regresión. Análogamente con las variables *Total expenditure* y *GDP*.

6. Estimaciones

6.1. Regresión Lineal Simple

Como mencionó anteriormente contamos con el factor *Status* el cual mide si un país está en desarrollo o está desarrollado, por lo cual decidimos hacer una *regresión lineal simple* para observar si éste era un predictor relevante para la esperanza de vida, en ambos años de interés

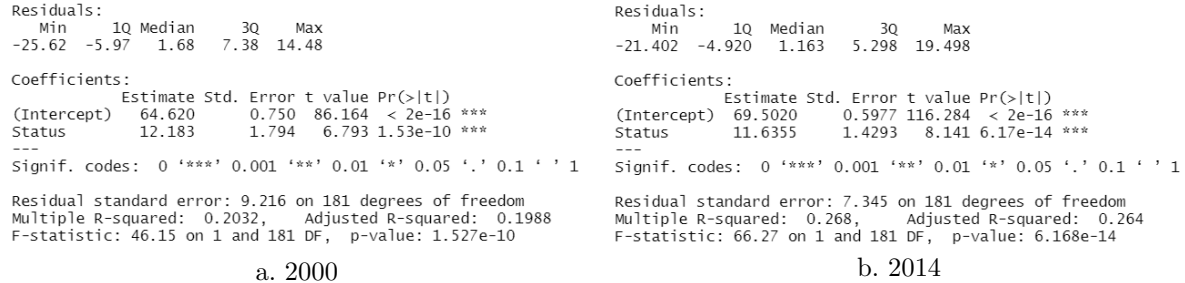


Figura 2: Regresión lineal simple con factor *Status*

Podemos ver en la figura 2 que el factor predictor *Status* es estadísticamente significativos para ambos con un nivel de significancia de 0, este resultado era esperado ya que varios estudios como el de [eur, 2014] revelan que hay diferencias en la *esperanza de vida* entre los países desarrollados y en vía de desarrollo. Dicho resultado nos motivó a separar ambos escenarios con el propósito de comparar los factores elegidos en tanto en tiempo como en *Status*.

Por lo que finalmente contamos con los datos de:
Año 2000:

- Países en desarrollo
- Países desarrollados

Año 2014:

- Países en desarrollo
- Países desarrollados

6.2. Regresión Lineal Múltiple

En esta sección se presentan los resultados de las regresión lineal múltiple realizadas con mínimos cuadrados ordinarios para cada set de datos.

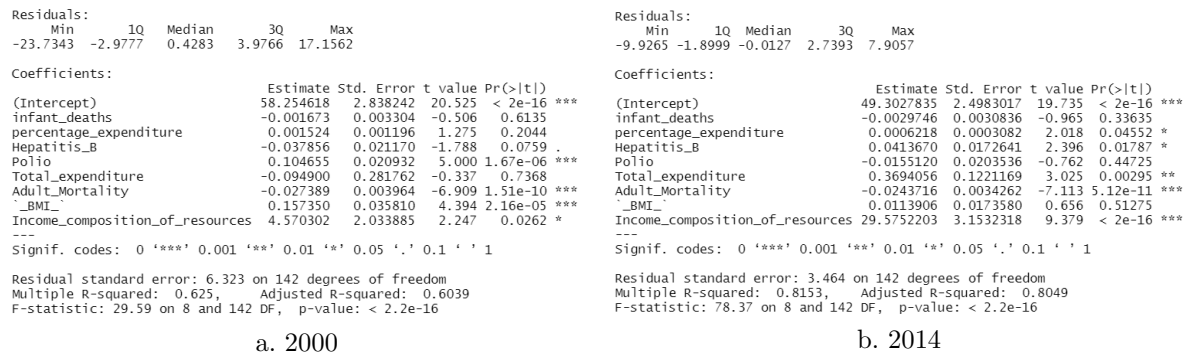


Figura 3: Regresión lineal múltiple países en desarrollo

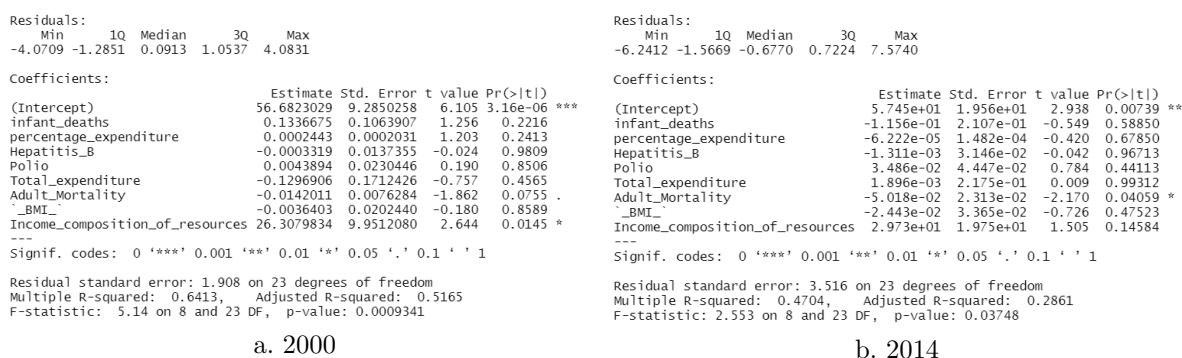


Figura 4: Regresión lineal múltiple países desarrollados

A continuación, se presentan los análisis correspondientes para comprobar si se cumplen los supuestos que debe cumplir la regresión lineal.

6.3. Prueba de Homocedasticidad

Realizamos la prueba diagnóstica de Homocedasticidad por medio del test de Breusch-Pagan.

Datos	BP	df	p-value
2000 Países en desarrollo	25.808	8	0.001133
2014 Países en desarrollo	15.844	8	0.04467
2000 Países desarrollados	10.701	8	0.2192
2014 Países desarrollados	5.3936	8	0.7148

Cuadro 3: Test de Breusch-Pagan

Como se puede observar en el cuadro 3 hay evidencias de falta de homocedasticidad en ambas regresiones de los países en desarrollo.

6.4. Mínimos Cuadrados Generalizados

Debido a los resultados obtenidos en la sección 6.3, donde ambas regresiones de países en desarrollo no cumplen con el supuesto de homocedasticidad, se realizaron nuevamente la regresiones ahora con mínimos cuadrados generalizados.

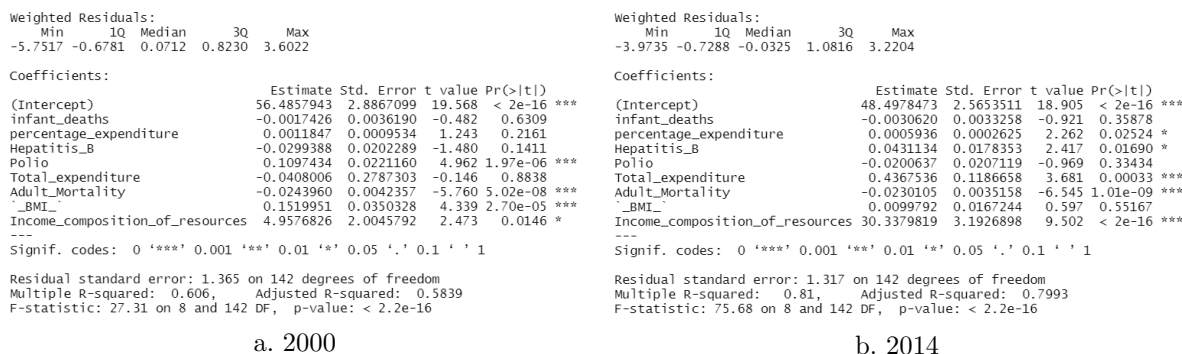


Figura 5: Regresión por Mínimos Cuadrados Generalizados

6.4.1. Prueba de Homocedasticidad

En esta sección se repite la prueba de Breusch-Pagan para las regresiones que presentaban p-valores menores a 0,05 en la sección 6.3.

Datos	BP	df	p-value
2000 Países en desarrollo	2.3556	8	0.9681
2014 Países en desarrollo	6.471	8	0.5946

Cuadro 4: Test de Breusch-Pagan para MCG

En el cuadro 4 se puede evidenciar el cambio de los p-valores de ambas regresiones, estos resultados muestran que no hay evidencia de falta de homocedasticidad para ninguna de las regresiones.

Ahora, como ya se cumplen los supuestos de homocedasticidad, a partir de las siguientes secciones, se tomarán en cuenta estas regresiones en lugar de las anteriores.

6.5. No autocorrelación

Procedemos a realizar el test de Durbin-Watson para comprobar que no haya auto-correlación.

Datos	Autocorrelation	D-W statistic	p-value
2000 Países en desarrollo	-0.06498883	2.122429	0.51
2014 Países en desarrollo	-0.1170381	2.231331	0.208
2000 Países desarrollados	0.1493885	1.660739	0.19
2014 Países desarrollador	-0.1257035	2.220173	0.632

Cuadro 5: Test de Durbin-Watson

En el cuadro 5 se evidencia que no hay evidencia de auto-correlación para ninguna de las regresiones.

6.6. Prueba de Normalidad

A continuación se presentaran os resultados de Test de Jarque Bera para comprobar la normalidad de residuales.

Datos	x-squared	p-value
2000 Países en desarrollo	14.234	0.0008114
2014 Países en desarrollo	8.1589	0.01692
2000 Países desarrollados	0.1096	0.9467
2014 Países desarrollador	7.6445	0.02188

Cuadro 6: Jarque Bera test

Para tres de nuestras regresiones se puede encontrar evidencia en contra de la normalidad de los residuales, como se observa en el cuadro 6. Sin embargo, al tratarse de muestras grandes, por los teoremas centrales del límite, esto no debe suponer problemas.

6.7. Intervalos de confianza

A continuación se presentan los intervalos de confianza de las cuatro regresiones.

Factores	2.5 %	97.5 %
Intercepto	50.7793143926	62.192274139
infant deaths	-0.0088966319	0.005411511
percentage expenditure	-0.0006999858	0.003069427
Hepatitis B	-0.0699274940	0.010049880
Polio	0.066024288	0.153462570
Total expenditure	-0.5917977581	0.510196473
Adult Mortality	-0.0327691196	-0.016022843
BMI	0.0827419577	0.221248324
Income composition of sources	0.9950082705	8.920356838

Cuadro 7: Intervalos de confianza 2000 países en desarrollo

Factores	2.5 %	97.5 %
Intercepto	4.342663e+01	53.569061501
infant deaths	-9.636444e-03	0.003512435
percentage expenditure	7.475246e-05	0.001112393
Hepatitis B	7.856429e-03	0.078370385
Polio	-6.100713e-02	0.020879702
Total expenditure	2.021737e-01	0.671333493
Adult Mortality	-2.996067e-02	-0.016060370
BMI	-2.308184e-02	0.043040310
Income composition of sources	2.402664e+01	36.649326117

Cuadro 8: Intervalos de confianza 2014 países en desarrollo

Factor	2.5 %	97.5 %
Intercepto	16.995599745	97.9025529942
infant deaths	-0.551461411	0.3202463274
percentage expenditure	-0.000368802	0.0002443602
Hepatitis B	-0.066396574	0.0637750085
Polio	-0.057138873	0.1268579101
Total expenditure	-0.448116738	0.4519088614
Adult Mortality	-0.098018704	-0.0023416103
BMI	-0.094038482	0.0451853563
Income composition of resources	-11.124232258	70.5833902040

Cuadro 9: Intervalos de confianza 2000 países desarrollados

Factor	2.5 %	97.5 %
Intercepto	37.4747635805	7.588984e+01
infant deaths	-0.0864184790	3.537534e-01
percentage expenditure	-0.0001758203	6.643895e-04
Hepatitis B	-0.0287459921	2.808221e-02
Polio	-0.0432819581	5.206076e-02
Total expenditure	-0.4839329196	2.245517e-01
Adult Mortality	-0.0299817214	1.579453e-03
BMI	-0.0455181775	3.823750e-02
Income composition of resources	5.7223411169	4.689363e+01

Cuadro 10: Intervalos de confianza 2014 países desarrollados

6.8. Análisis de las estimaciones

En la figura 4, podemos observar que ambas observaciones no se cuentan con los mismos factores estadísticamente significativos, ya que mientras en el 2000 el *índice de desarrollo humano* es estadísticamente significativo al 1 % en el 2014 este factor no lo es, si no que lo es la **mortalidad en adultos**.

Analizando la figura 5, regresión para los países en desarrollo estimado mediante mínimos cuadrados generalizado, obtuvimos algunos factores predictivos que son estadísticamente significativos para ambas observaciones, como es el caso del *índice de desarrollo humano*, *la mortalidad en adultos*. Además contamos con otros factores que cambian con el escenario, en el 2000 también encontramos que el nivel de *inmunización del polio* es estadísticamente significativo, junto con el *BMI*. Por otro lado en el 2014 contamos con que el *gasto del gobierno en salud como porcentaje del PIB*, *el hepatitis B* y *gasto general del gobierno en salud* son estadísticamente significativos.

Finalmente, comparando los países en desarrollo y desarrollados en ambos escenarios, vemos que no contamos con los mismos factores predictivos estadísticamente significativos en ambos periodos. Por

los resultados que los países en desarrollo cuentan con más factores estadísticamente significativos que los países en desarrollo, lo que puede significar que los predictores elegidos afectan más a los países en desarrollo, lo cual teóricamente es coherente para los casos de las enfermedades como la Hepatitis B y el polio, lo cuales presentan un mayor impacto en países en vía de desarrollo de África y Asia.

7. Conclusiones

Las regresiones lineales planteadas pueden ser consideradas confiables ya que cumplen con las condiciones necesarias de una regresión lineal múltiple.

Adicionalmente nos sorprendió evidenciar que muchos de los predictores elegidos bajo fundamento teórico no fueron estadísticamente significativos en los países desarrollados, sin embargo, notamos que factores no tan usuales incluidos como los son *los índices de inmunización e índice de desarrollo humano* sí terminaron afectando más que todo a los países que se encuentran en desarrollo. Lo que teóricamente era esperable, ya que las enfermedades elegidas, como fue mencionado anteriormente, afecta más que todo a países en vía de desarrollo en los continentes de África y Asia.

Referencias

- [eur, 2014] 2014 (Mar). *Economic growth and life expectancy – do wealthier countries...*
- [Andreev *et al.*, 2006] Andreev, Kirill F, Vaupel, James W, *et al.* 2006. Forecasts of cohort mortality after age 50. *Max Planck Institute for Demographic Research Working Paper*, **12**, 2006.
- [Balan, 2011] Balan, Christiana. 2011. Statistical Analysis Of The Determinants Of Life Expectancy In Romania. *Romanian Journal of Regional Science*, **5**(2), 25–38.
- [Briceño *et al.*, n.d.] Briceño, Yerrina, Olaya, Cherly, Jacinto, Michelle, Arana, Alejandra, Rodríguez, Andrea, *et al.* Factores que afectan la esperanza de vida al nacer.
- [Hazan, 2011] Hazan, Moshe. 2011. Life expectancy and schooling: new insights from cross-country data. *Journal of Population Economics*, **25**(10), 1–12.
- [Kontis *et al.*, 2017] Kontis, Vasilis, Bennett, James E, Mathers, Colin D, Li, Guangquan, Foreman, Kyle, & Ezzati, Majid. 2017. Future life expectancy in 35 industrialised countries: projections with a Bayesian model ensemble. *The Lancet*, **389**(10076), 1323–1335.
- [Max Roser & Ritchie, 2013] Max Roser, Esteban Ortiz-Ospina, & Ritchie, Hannah. 2013. Esperanza de vida.
- [Monsef & Mehrjardi, 2015] Monsef, Abdalali, & Mehrjardi, Abolfazl Shahmohammadi. 2015. Determinants of Life Expectancy: A Panel Data Approach. *Asian Economic and Financial Review*, **5**(11), 1251–1257.
- [Pascariu *et al.*, 2018] Pascariu, Marius D., Canudas-Romo, Vladimir, & Vaupel, James W. 2018. The double-gap life expectancy forecasting model. *Insurance: Mathematics and Economics*, **78**, 339–350. Longevity risk and capital markets: The 2015–16 update.
- [Pathirathne & Sooriyarachchi, 2019] Pathirathne, Linali, & Sooriyarachchi, MR. 2019. Factors Affecting Life Expectancy: A Global Perspective.
- [Sormin *et al.*, 2019] Sormin, Mhd. Khafiroh Zamzamy, Sihombing, Poltak, Amalia, A., Wanto, Anjar, Hartama, Dedy, & Chan, Defri Muhammad. 2019. Predictions of World Population Life Expectancy Using Cyclical Order Weight / Bias. *Journal of Physics: Conference Series*, **1255**(1), 012017.