

UNIVERSIDAD DE COSTA RICA

MODELOS LINEALES

MINOR HEADING

Pronóstico demográfico de Costa Rica

Autores

David Zumbado

Leonardo Blanco

Ignacio Barrantes

6 de octubre de 2022

Indice de contenidos

	5
1 Bitácora 1	6
1.1 Parte 1	6
1.1.1 Características generales de la tabla	6
1.1.2 Variables de estudio	6
1.1.3 Clasificación de las variables	8
1.2 Parte 2	9
1.2.1 Pregunta central de investigación	9
1.2.2 Objeto de la investigación	9
1.2.3 Conceptos delimitadores de la pregunta de investigación	9
1.2.4 Principios	10
2 Bitacora 2	11
2.1 Punto 1	11
2.2 Parte 2	14
2.2.1 Resumen de cinco números	14
2.2.2 Medidas de dispersión	14
2.2.3 Tablas de frecuencia	16
2.3 Parte 3: Propuesta de UVE	17
References	19

Listado de Figuras

2.1 Borrador de la UVE Heurística 18

Listado de Tablas

2.1	Primeras cinco filas y nueve columnas de la tabla de datos	12
2.2	Primeras cinco filas y segundas nueve columnas de la tabla de datos	13
2.3	Resumen de cinco números para algunas variables de la tabla de datos	14
2.4	Estadísticos de dispersión para algunas variables de la tabla de datos	14
2.5	Frecuencia de defunciones, población y tasa de mortalidad	16
2.6	Frecuencia de nacimientos y tasa de natalidad	17

1 Bitácora 1

En el proyecto se buscará realizar análisis demográfico, específicamente se centrará en el pronóstico de la cantidad de defunciones en Costa Rica.

1.1 Parte 1

1.1.1 Características generales de la tabla

La tabla de datos proviene del Instituto Nacional de Estadística y Censos (INEC) de Costa Rica y es de acceso público, descargable desde la página web del instituto, como puede consultarse en INEC (2021). Esta base presenta los principales indicadores demográficos anuales de Costa Rica durante el periodo 1950-2020. Incluye en total 18 variables, entre las cuales están el año, la población total al 30 de junio de cada año, desagregado también por sexo, así como la cantidad de defunciones.

1.1.2 Variables de estudio

Población de estudio: La población de estudio son aquellas personas que vivían en Costa Rica entre los años 1950-2020 y mueren en este periodo.

Muestra observada: La muestra observada, son todas aquellas personas que vivían en Costa Rica y al morir son registrados por el Instituto Nacional de Estadística y Censo.

Unidad estadística o individuos: La unidad estadística es el recuento anual de defunciones en Costa Rica.

Variables de estudio: Son un total de 18 variables, las cuales según los Indicadores demográficos 2019 y (2020) se tiene:

Año: Esta variable indica el año.

Total: Esta variable registra la población total.

Hombres: Esta variable registra la población total de hombres.

Mujeres: Esta variable se encarga de registrar el total de mujeres.

Nacimientos: Esta variable registra el total de nacimientos.

Defunciones: Esta variable registra el total de defunciones.

Defunciones infantiles: Esta variable registra las defunciones de infantes (niños y niñas).

Defunciones neonatales: Esta variable registra las defunciones de recién nacidos, hace referencia a la mortalidad de los nacidos antes de alcanzar los 28 días de edad.

Defunciones fetales: Esta variable registra las defunciones de fetos, se refiere a la mortalidad de un bebé antes o durante el parto.

Tasa de crecimiento: Esta variable registra la tasa de crecimiento de la población costarricense. Se refiere al crecimiento de la población entre dos fechas sin contemplar la migración.

Tasa de natalidad: Esta variable registra la tasa de nacimientos registrados en la población costarricense.

Tasa de mortalidad: Esta variable registra la tasa de muertes en su totalidad.

Tasa de mortalidad infantil Esta variable registra la tasa de muertes infantiles.

Tasa de mortalidad neonatal: Esta variable registra la tasa de muertes neonatales.

Tasa de mortalidad fetal: En esta variable se registra la tasa de muertes fetales.

Tasa global de fecundidad: Esta variable registra la tasa de fecundidad global (TGF), la cual indica cantidad de hijos e hijas que en promedio tendría cada mujer al final del periodo fértil, si durante su vida tiene sus hijos e hijas de acuerdo a las tasas de fecundidad por edad observadas en el país y año de interés y, además estas mujeres no están afectadas por la mortalidad desde el nacimiento hasta el final de periodo fértil.

Tasa bruta de reproducción: Esta variable hace referencia a el cantidad de hijas que en promedio tendría cada mujer al final del periodo fértil, si durante su vida tiene sus hijos e hijas de acuerdo a las tasas de fecundidad por edad observadas en el país y año de interés y, además estas mujeres no están afectadas por la mortalidad desde el nacimiento hasta el final de periodo fértil.

Tasa neta de reproducción: es el número de hijas que en promedio tendría cada mujer al final del periodo fértil, si durante su vida tiene sus hijos e hijas de acuerdo a las tasas de fecundidad por edad observadas en el país y año de interés y, además estas mujeres sí están afectadas por la mortalidad por edad observada en el país y año de interés desde el nacimiento hasta el final de periodo fértil.

1.1.3 Clasificación de las variables

Todas las variables de la tabla utilizada son numéricas donde cinco identifican cantidades absolutas registradas con respecto a la cantidad de nacimientos, defunciones, etc, y nueve variables son tasas de variabilidad con respecto al año anterior, o como un porcentaje de la población.

1.2 Parte 2

1.2.1 Pregunta central de investigación

La pregunta formulada es: ¿Cómo se puede realizar un pronóstico de la serie de defunciones totales anuales de Costa Rica?

1.2.2 Objeto de la investigación

La cantidad de defunciones totales anuales ocurridas en Costa Rica y registradas por el Instituto Nacional de Estadística y Censos del año 1950 hasta el año 2020.

1.2.3 Conceptos delimitadores de la pregunta de investigación

- Defunciones totales anuales: Según Macció, Centro Latinoamericano de Demografía, et al. (1985), la mortalidad o defunción se emplea para expresar la acción de la muerte sobre la población. Esta se mide en valores absolutos y sobre el año calendario.
- Pronóstico: Según Brownlee (2020) la realización de predicciones sobre el futuro se denomina extrapolación en el tratamiento estadístico clásico de los datos de las series temporales. Los campos más modernos que se centran en el tema, lo denominan pronóstico de series temporales. El pronóstico consiste en tomar modelos ajustados a datos históricos y utilizarlos para predecir observaciones futuras.
- Serie: De acuerdo a Macció, Centro Latinoamericano de Demografía, et al. (1985) una series cronológica o crónica se le llama a “la que forman los valores sucesivos que una cierta variable ha tomado en el transcurso del tiempo”.

1.2.4 Principios

Rees (2020) afirma que una de las mayores contribuciones de la demografía ha sido el desarrollo de modelos para la proyección de futuras poblaciones, lo cuál está estrechamente relacionado con la proyección de defunciones. De acuerdo al autor, dos de los modelos más importantes y usados en Demografía cuando a proyección poblacional se refiere, están el de Cohorte-Componente (*The Cohort-Component Model for Projecting the Population*) y el de Proyecciones Probabilísticas (*Probabilistic Projections*).

El modelo de Cohorte-Componente utiliza datos de nacimientos, muertes y migración (a estos factores se les conoce como componentes de cambio) a nivel de cohortes. De acuerdo al Diccionario Demográfico Multilingüe, un cohorte es “conjunto de individuos que han vivido un cierto acontecimiento durante un mismo período” (1985); y en el marco del modelo puede referirse a grupos de edad. Usualmente, los cohortes están desagregados por sexo (Wilson y Rees 2021).

Por otro lado, el modelo de Proyecciones Probabilísticas, el cual también es establecido en Rees (2020) trata de estimar la población futura a través de las distribuciones de error de ciertos componentes como la tasa de fertilidad. Con estas se pueden proyectar año a año y bajo la combinación de cientos de estos componentes se puede construir escenarios posibles. Cada una de estas distribuciones se puede derivar utilizando datos históricos y a través del uso de modelos de series de tiempo auto-regresivas, o de censos o opiniones experta. Bajo este enfoque se construyen varios escenarios alrededor de una mediana y que se dispersan producto de las distribuciones de incertidumbre pero que al final logran construir un análogo del intervalo de confianza.

2 Bitacora 2

2.1 Punto 1

En la Tabla 2.1 y Tabla 2.2 se presentan las primeras cinco líneas de la tabla de datos para las columnas 1 a 9 y 10 a 18 respectivamente. En conformidad con Wickham y Grolemund (2016), una tabla en formato *tidy* cumple que:

- Cada variable posee su propia columna
- Cada observación posee su propia fila
- Cada valor posee su propia celda

Se puede apreciar en las dos tablas mencionadas que la tabla de datos cumple con el formato *tidy*, pues las variables están dispuestas a lo largo de las columnas y las observaciones a lo largo de las filas.

Tabla 2.1: Primeras cinco filas y nueve columnas de la tabla de datos

Año	Población total	Población de hombres	Población de mujeres	Nacimientos	Defunciones	Defunciones infantiles	Defunciones neonatales	Defunciones fetales
1 950	868 934	438 185	430 749	37 248	9 769	3 358	974	1 103
1 951	897 630	452 702	444 928	39 239	9 631	3 467	953	1 062
1 952	929 173	468 613	460 560	42 461	9 902	3 811	1 083	1 012
1 953	962 485	485 464	477 021	42 817	10 312	4 017	1 161	1 021
1 954	997 535	503 219	494 316	48 157	9 713	3 856	1 210	1 135
1 955	1 035 424	522 378	513 046	48 903	9 998	4 065	1 233	1 095

Nota
El conteo de población se realiza al 30 de junio de cada año.

Tabla 2.2: Primeras cinco filas y segundas nueve columnas de la tabla de datos

Tasa de crecimiento	Tasa de natalidad	Tasa de mortalidad	Tasa de mortalidad infantil	Tasa de mortalidad neonatal	Tasa de mortalidad fetal	Tasa global de fecundidad	Tasa bruta de reproducción	Tasa neta de reproducción
31.62381	42.86632	11.242511	90.15249	26.14906	29.61233	-	-	-
32.98464	43.71400	10.729365	88.35597	24.28706	27.06491	-	-	-
35.04084	45.69763	10.656788	89.75295	25.50576	23.83364	-	-	-
33.77195	44.48589	10.713933	93.81788	27.11540	23.84567	-	-	-
38.53900	48.27600	9.737002	80.07143	25.12615	23.56874	-	-	-
37.57398	47.22993	9.655948	83.12373	25.21318	22.39126	6.96	-	-

Nota
El conteo de población se realiza al 30 de junio de cada año.

Tabla 2.3: Resumen de cinco números para algunas variables de la tabla de datos

	Población total	Población de hombres	Población de mujeres	Nacimientos	Defunciones	Defunciones infantiles	Defunciones neonatales	Defunciones fetales
Mínimo	868 934	438 185	430 749.0	37 248	8 596.0	462.0	344.0	329
Primer cuartil	1 638 283	827 590	810 692.5	59 589	9 897.5	765.0	537.0	526
Mediana	2 646 142	1 339 573	1 306 569.0	70 004	11 376.0	1 345.0	807.0	691
Tercer cuartil	4 054 418	2 053 052	2 001 366.5	75 794	15 704.5	3 543.5	1 173.5	959
Máximo	5 111 238	2 575 550	2 535 688.0	84 337	26 209.0	4 889.0	1 688.0	1 261

Tabla 2.4: Estadísticos de dispersión para algunas variables de la tabla de datos

	Población total	Población de hombres	Población de mujeres	Nacimientos	Defunciones	Defunciones infantiles	Defunciones neonatales	Defunciones fetales
Desviación estándar	1 340 677	677 255.2	663 434.1	11 274.6	4 428.495	1 396.199	366.604	241.5763
Rango intercuartílico	2 416 136	1 225 461.5	1 190 674.0	16 205.0	5 807.000	2 778.500	636.500	433.0000

2.2 Parte 2

2.2.1 Resumen de cinco números

En la tabla [Tabla 2.3](#) se detalla el resumen de cinco números para algunas variables relevantes de la tabla de datos.

2.2.2 Medidas de dispersión

En la tabla [Tabla 2.4](#) se detallan los estadísticos de desviación estándar y rango intercuartílico para algunas variables relevante de la tabla de datos.

Se observa que la variable población de hombres presenta mayor dispersión respecto población de mujeres, esta mayor variabilidad en la población de hombres se puede deber a diversos factores, entre ellos : una diferencia en el número de defunciones, nacimientos, o número de migrantes en la población masculina en relación con la de población femenina.

Una observación importante, es que los nacimientos presentan una mayor dispersión respecto a el número de defunciones. Es decir, se observa que el número defunciones anuales presenta menor desviación respecto a la media (desviación estándar) y una menor diferencia entre tercer cuartil y primer cuartil de defunciones (IQR).

Análogamente, se observa que para los tipos de defunción: infantil, neonatales y fetales. Se evidencia una mayor dispersión para las defunciones infantiles, seguidas de las neonatales y por último, con menor dispersión las defunciones fetales.

Finalmente, se observa una diferencia significativa entre la medidas de dispersión desviación estándar y rango intercuartílico (IQR), esto se debe a la sensibilidad de la desviación estándar a valores extremos, es decir es posible que existan valores extremos (muy alto o muy bajos respecto a la media) y por esta razón ambas medidas difieran considerablemente, sin embargo para este caso en particular el orden en el grado de dispersión (observe que la dispersión disminuye al avanzar en la tabla de izquierda a derecha) se mantiene para ambos medidas.

Tabla 2.5: Frecuencia de defunciones, población y tasa de mortalidad

Defunciones	Frec. de defunciones	Población total	Frec. de poblacion	Tasa de mortalidad	Frec. de tasa
(8.58e+03,1.04e+04]	23	(8.65e+05,1.29e+06]	11	(3.72,4.41]	39
(1.04e+04,1.21e+04]	18	(1.29e+06,1.72e+06]	9	(4.41,5.1]	8
(1.21e+04,1.39e+04]	4	(1.72e+06,2.14e+06]	8	(5.1,5.78]	1
(1.39e+04,1.56e+04]	8	(2.14e+06,2.57e+06]	6	(5.78,6.46]	3
(1.56e+04,1.74e+04]	5	(2.57e+06,2.99e+06]	6	(6.46,7.14]	3
(1.74e+04,1.92e+04]	4	(2.99e+06,3.41e+06]	5	(7.14,7.83]	3
(1.92e+04,2.09e+04]	3	(3.41e+06,3.84e+06]	5	(7.83,8.51]	6
(2.09e+04,2.27e+04]	2	(3.84e+06,4.26e+06]	6	(8.51,9.19]	1
(2.27e+04,2.44e+04]	3	(4.26e+06,4.69e+06]	7	(9.19,9.88]	3
(2.44e+04,2.62e+04]	1	(4.69e+06,5.12e+06]	8	(10.6,11.3]	4

2.2.3 Tablas de frecuencia

De la Tabla 2.5 se observa que históricamente las defunciones totales de la mayoría de los años cae dentro del primero y segundo intervalo, con una distribución más uniforme dentro de los intervalos más altos. Por su parte la tasa de mortalidad bruta, la cual se aprecia en las columnas cinco y seis, se muestra que en la mayoría de años la tasa se ha mantenido entre 3.72 y 4.41. Esto quiere decir que la cantidad de defunciones relativo a la población ha sido más o menos constante en la mayoría de años. Por su parte, la población total se ha distribuido más uniformemente en todos los intervalos por lo que se concluye que el crecimiento población se ha movido más rápido que las defunciones.

En la tabla Tabla 2.6 se observa que los nacimientos en nuestro país a lo largo de los años tiene una distribución más uniforme para los últimos intervalos de la tabla.

Por otro lado, para el caso de la tasa de nacimientos se observa un comportamiento menos uniforme a diferencia de los observado para la tasa de mortalidad. En este caso, se observa que la frecuencia oscila entre valores menores a 6 pero con ciertos repuntes altos de frecuencia para ciertos intervalos por lo que la tasa de nacimientos no se mantiene en cierto intervalo específico.

Tabla 2.6: Frecuencia de nacimientos y tasa de natalidad

Nacimientos	Frec. de nacimientos	Tasa de nacimientos	Frec. de tasa
(3.72e+04,4.2e+04]	2	(11.3,15.4]	8
(4.2e+04,4.67e+04]	2	(15.4,19.6]	12
(4.67e+04,5.14e+04]	2	(19.6,23.7]	6
(5.14e+04,5.61e+04]	4	(23.7,27.8]	5
(5.61e+04,6.08e+04]	10	(27.8,31.9]	19
(6.08e+04,6.55e+04]	7	(31.9,36.1]	3
(6.55e+04,7.02e+04]	9	(36.1,40.2]	1
(7.02e+04,7.49e+04]	15	(40.2,44.3]	3
(7.49e+04,7.96e+04]	9	(44.3,48.4]	11
(7.96e+04,8.44e+04]	11	(48.4,52.6]	3

2.3 Parte 3: Propuesta de UVE

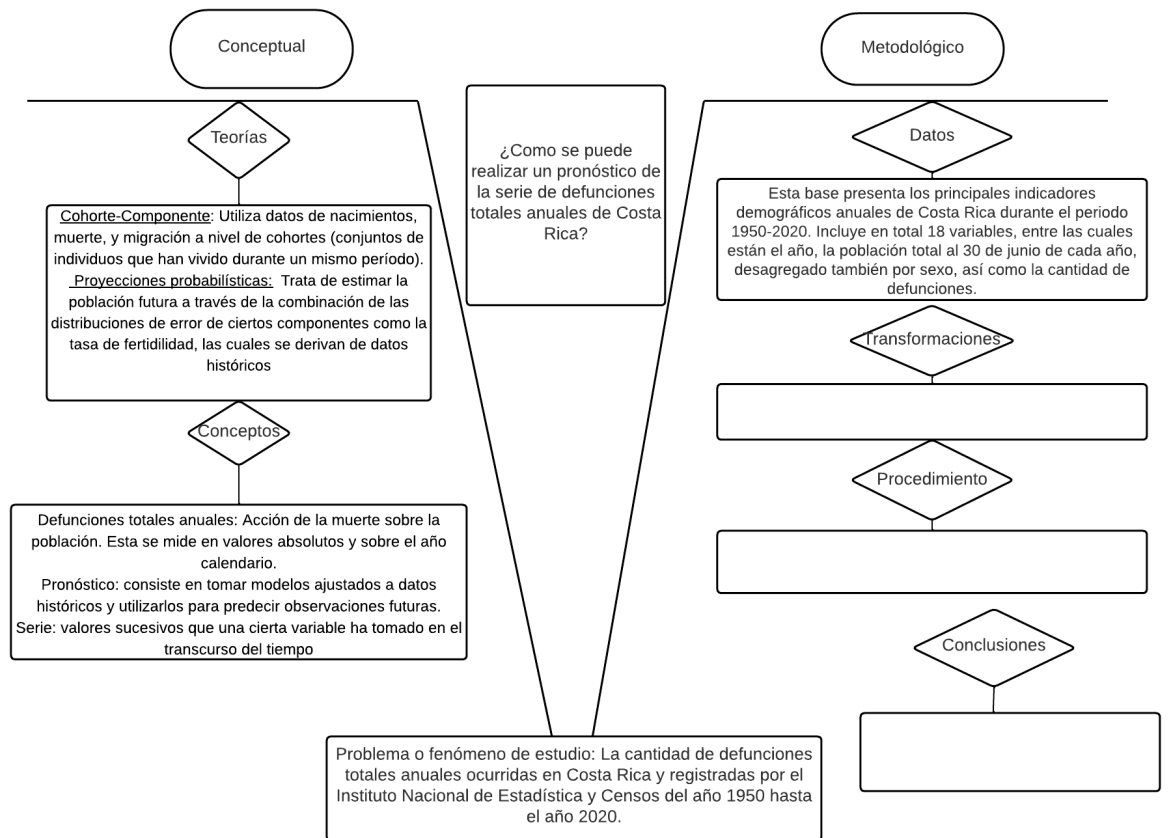


Figura 2.1: Borrador de la UVE Heurística

References

- Brownlee, Jason. 2020. *Introduction to time series forecasting with Python*. eBook.
- Indicadores demográficos 2019. 2020.
- INEC. 2021. Estadísticas demográficas. 1950-2020. Principales Indicadores Demográficos. <https://www.inec.go.cr/documento/estadisticas-demograficas-1950-2020-principales-indicadores-demograficos%7D>.
- Macció, Guillermo A, Santiago Centro Latinoamericano de Demografia, et al. 1985. Diccionario demográfico multilingüe.
- Rees, Philip. 2020. Demography. En *International Encyclopedia of Human Geography (Second Edition)*, editado por Audrey Kobayashi, Second Edition, 239-56. Oxford: Elsevier. [https://doi.org/https://doi.org/10.1016/B978-0-08-102295-5.10252-5](https://doi.org/10.1016/B978-0-08-102295-5.10252-5).
- Wickham, Hadley, y Garrett Grolemund. 2016. *R for data science: import, tidy, transform, visualize, and model data*. " O'Reilly Media, Inc."
- Wilson, Tom, y Philip Rees. 2021. A brief guide to producing a national population projection. *Australian Population Studies* 5 (1): 77-100.