Universidad Internacional de La Rioja

Facultad de Ingeniería y Tecnología

Máster Universitario en Análisis y Visualización de Datos Masivos /

Visual Analitics & Big Data

ANALISIS MUESTRAL A LA PREDICCION POBLACIONAL

Actividad de estudio presentado por:	Juan David Escobar Escobar
Tipo de trabajo:	Actividad 3
Modalidad:	Individual
Director/a:	PhD. Yamila García Martínez
Fecha:	Marzo 2022

Índice de contenidos

Índice de imágenes

1. Contrastes de hipótesis

Son un procedimiento estadístico para decidir si una afirmación sobre una población objeto de estudio es cierta o falsa a partir de los datos. Los contrastes de hipótesis son una herramienta de amplio uso en la ciencia, la filosofía, la antropología, entre otras áreas, y es usada para contrastar y decidir sobre posibles teorías científicas asociadas a un fenómeno de una o mas hipótesis.

El primer paso para contrastar una hipótesis es definir la hipótesis de afirmación o hipótesis nula H0 y la hipótesis alternativa de negación H1, esta emplea los operadores lógicos de *<, > o <>*. La siguiente imagen muestra algunos ejemplos de H1 para proporciones medias y desviaciones estándar:

Proporciones:	$H_1: p > 0.5$	$H_1: p < 0.5$	$H_1: p \neq 0.5$
Medias:	$H_1: \mu > 98.6$	$H_1: \mu < 98.6$	$H_1: \mu \neq 98.6$
Desviaciones estándar:	$H_1: \sigma > 15$	$H_1: \sigma < 15$	$H_1: \sigma \neq 15$

Ilustración 1. Ejemplos de Hipotesis Alternativa

Fuente: (Triola, 2009)

Para llevar a cabo la validación de veracidad de una hipótesis nula establecida en un problema, es esencial calcular el Estadístico de prueba *(z, t o X^2)*.

• **Media**	Estadístico de prueba para proporciones	$z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}}$
• Proporción		
• Varianza	Estadístico de prueba para medias	$z = \dfrac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$ o $t = \dfrac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n}}}$
• Diferencia de medias		
• Diferencia de proporciones	Estadístico de prueba para desviaciones estándar	$\chi^2 = \dfrac{(n-1)s^2}{\sigma^2}$

Ilustración 2. Estadístico de prueba

Fuente: (Triola, 2009)

Otro concepto relevante para tener en cuenta para el contraste de hipótesis es el de Región critica, nivel de significancia, valor critico y valor P.

La región critica es el conjunto de todos los valores del estadístico de prueba que hacen que rechacemos H0. El nivel de significancia alfa *(0.05, 0.01 o 0,10)*, es la probabilidad de que el estadístico de prueba caiga en la región critica, cuando H0 = verdadero (probabilidad de cometer el error de rechazar H0 cuando es verdadera).

El valor critico separa la región critica (donde se rechaza H0), este valor depende de la naturaleza de H0, la distribución muestral y el nivel de significancia alfa, ejemplo alfa = 0.05, z=1.645. Colas de una distribución (Dos colas, cola izquierda, cola derecha), son la distribución de las regiones

extremas limitadas por los valores críticos. En la distribución bilateral, el valor de alfa de significancia esta dividido equitativamente entre las 2 colas de la región critica.

A	$H_0: \theta = \theta_0$ $H_1: \theta \neq \theta_0$	Dos Colas Bilateral	
B	$H_0: \theta \geq \theta_0$ $H_1: \theta < \theta_0$	Cola Unilateral (Izquierda)	
C	$H_0: \theta \leq \theta_0$ $H_1: \theta > \theta_0$	Cola Unilateral (Derecha)	
D	$H_0: \theta_1 \leq \theta \leq \theta_2$ $H_1: \theta < \theta_1$ o $\theta > \theta_2$	Dos Colas Bilateral	

Ilustración 3. Tipos de contraste de Hipotesis

Fuente: (Triola, 2009)

El ultimo paso para contrastar una hipótesis es rechazar o aceptar la hipótesis nula, para ello debemos contar con un Criterio de decisión, el cual se puede llevar a cabo mediante el método tradicional (prueba de hipótesis), el método de valor P o intervalos de confianza, a continuación, se detallan las condiciones de rechazo o aceptación para cada método:

Método tradicional:	*Rechace* H_0 si el estadístico de prueba cae dentro de la región crítica. *No rechace* H_0 si el estadístico de prueba no cae dentro de la región crítica.
Método del valor *P*:	*Rechace* H_0 si el valor de $P \leq \alpha$ (donde α es el nivel de significancia, tal como 0.05). *No rechace* H_0 si el valor $P > \alpha$.
Otra opción:	En vez de usar un nivel de significancia como $\alpha = 0.05$, simplemente identifique el valor P y deje la decisión al lector.
Intervalos de confianza:	Como un estimado del intervalo de confianza de un parámetro de población contiene los valores posibles de tal parámetro, rechace la aseveración de que el parámetro de población tiene un valor que no está incluido en el intervalo de confianza.

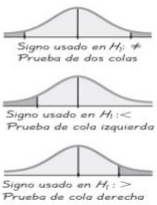


Ilustración 4. Metodos de contraste de hipotesis

Fuente: (Triola, 2009)

2. Algoritmo

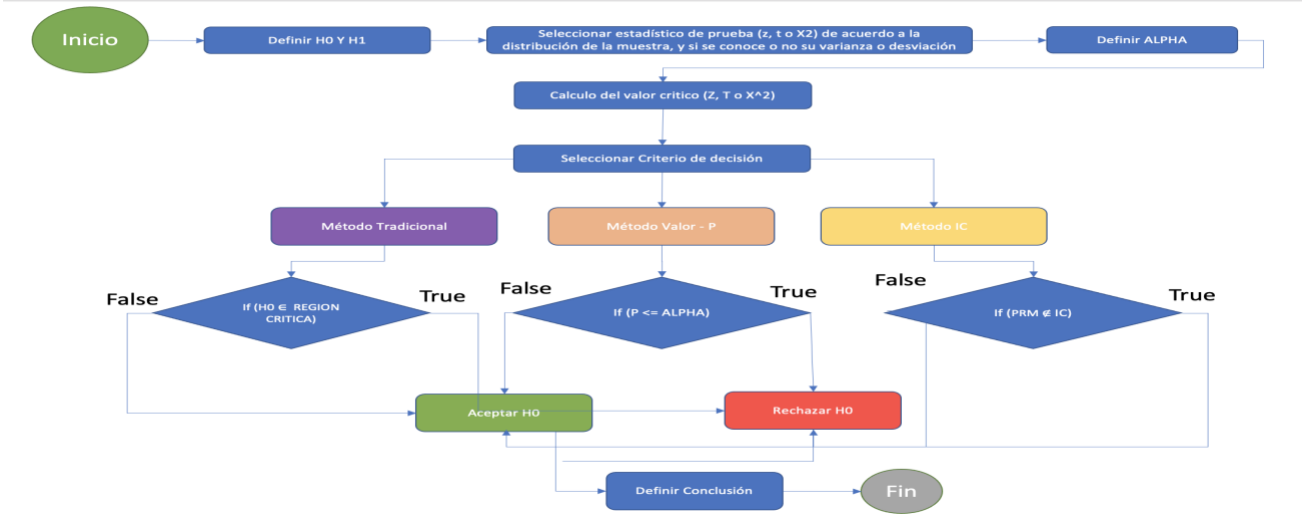


Ilustración 5. Algoritmo para el contraste de hipótesis

3. Escenarios contraste de hipótesis (1 y 2 poblaciones)

Contraste de Hipótesis para 1 población:

Población 1: Conjunto de datos asociado a la asignación de dosis de vacuna contra COVID – 19, desde el mes de febrero de 2021 hasta el mes de febrero de 20222 en los territorios o departamentos de Colombia. (Datos tomados de https://www.datos.gov.co/Salud-y-Protecci-n-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr/data).

Contraste de hipótesis: Para el año 2021 se tuvo un total de 520.620 casos positivos de COVID-19 en la ciudad de Medellín, Colombia, se toma una muestra aleatoria simple de 35 personas, de las cuales se contagiaron 18 Hombres y 17 Mujeres. Con un nivel de confianza de 0,05 ¿Se puede afirmar que los contagios se presentan mas en los hombres de Medellín?

Contraste de Hipótesis para 2 poblaciones:

- **Población 1**: Conjunto de datos asociado a los casos de VIH confirmados entre 2008 -2021 en la ciudad Bogotá, Colombia (Datos tomados de datosabiertos.bogota.gov.co).
- **Población 2**: Conjunto de datos asociado a los casos de VIH confirmados entre 2008 -2021 en la ciudad Medellín, Colombia (Datos tomados de medata.gov.co).
- **Contraste de hipótesis:** ¿Existe diferencia entre la cantidad de personas confirmadas con VIH entre la ciudad de Medellín y Bogotá?

4. Contraste de Hipótesis 1 población

Planteamiento estudio: Según el articulo publicado por (consultorsalud, 2021) Gracias a las investigaciones recientes se sabe que uno de los principales factores de riesgo es la edad avanzada, sin embargo, los hombres mayores que fallecen a causa del coronavirus doblan la cifra de inmunidad que los hombres tras ser vacunadas contra la gripe. De otro lado, los hombres infectados con VIH tienen tendencia hacia una carga viral más alta que las mujeres infectadas con el virus.

Contraste de hipótesis: Para el año 2021 se tuvo un total de 520.620 casos positivos de COVID-19 en la ciudad de Medellín, Colombia, se toma una muestra aleatoria simple de 35 personas, de las cuales se contagiaron 18 Hombres y 17 Mujeres. Con un nivel de confianza de 0,05 ¿Se puede afirmar que los contagios se presentan mas en los hombres de Medellín?

Tipo de estudio Básico descriptivo, inferencial (contraste de hipótesis).

4.1. Población

La población esta constituida por 520.620personas de la ciudad de Medellín contagiadas en el año 2021, individuos de diferentes edades, pertenencias étnicas y sexo. En la Ilustración 1 se describe el detalle demográfico de la población. La edad media de las personas es de 40 años, de las cuales 54% son hombres y 46% mujeres. Los datos se agrupan en 3 grupos de edades, de jóvenes (0.89,37.7] con un 52% de participación, adultos (37.7,74.3] con un 43% de participación y adulto mayor (74.3,111] con un 39% de participación.

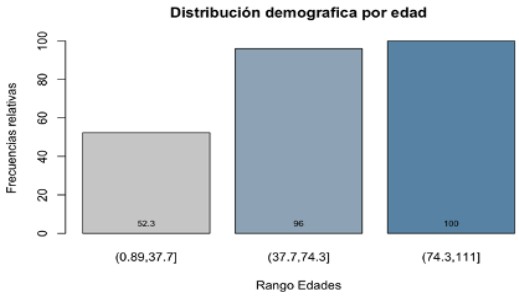


Ilustración 6. Distribución demográfica por Edad

La **Ilustración 6** representa dos series de tiempo para los hombres y mujeres infectados por covid-19 en el año 2021, ambas muestran una tendencia similar, con un pico máximo de 743 casos en hombres y 828 mujeres en el mes de junio de 2021.

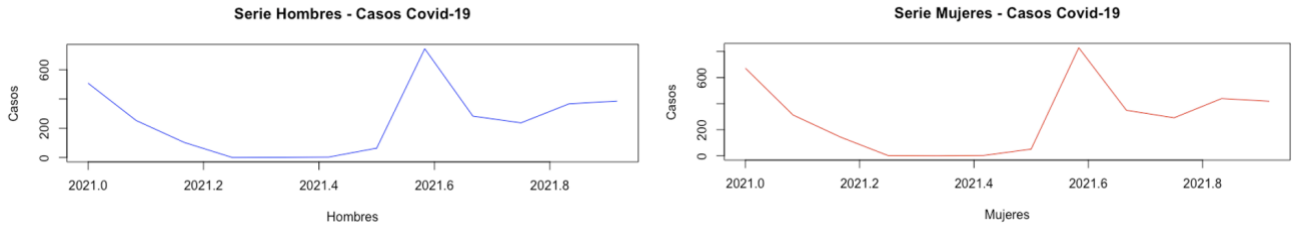


Ilustración 7. Series de tiempo casos positivos por Covid-19 hombres y mujeres [2021]

Contraste de Hipótesis

- **Z** = 0.1690309; no se encuentra en el IC (0.06270678 -0.06270678),
- **p-valor** = 0.8657724; p-valor > 0.05.

Conclusiones: No existe base estadística para confirmar que los contagios por covid-19 se presentaron en mayor proporción en los hombres de Medellín en el año 2021.

El numero de casos de contagios por Covid-19 se ha estabilizado en los últimos años, gracias a las vacunas, pero hay puntos en el tiempo en los cuales se dan picos de contagios por la aparición de una nueva variante, para el año actual muchos países a disminuido el régimen de protocolos por riesgo de

contagio, aunque se nota una disminución los últimos tres años, es importante que las personas continúen cuidándose para evitar mas riesgo de muertes y otras pandemias.

5. Contraste de Hipótesis 2 poblaciones

Planteamiento estudio: Las infecciones de transmisión sexual siguen siendo un problema de salud publica en Colombia, con tendencia a creciente en los casos reportados (Pinzón, 2020). Mas de 1600 casos de VIH se detectaron el año pasado 2021 en Medellín (Rosales, 2021).

Desde el 1 de enero hasta el 20 de junio de 2020 se confirmaron 2323 casos de VIH/Sida en la ciudad de Bogotá, para el mismo periodo del año anterior 2021 se habían notificado 1526 casos, presentando un aumento del 26% en el numero de casos de la ciudad capital (saludcapital, 2021).

Contraste de hipótesis: ¿Existe diferencia entre la cantidad de personas confirmadas con VIH entre las principales ciudades de Colombia Medellín y Bogotá entre los años 2008 y 2020?

Tipo de estudio

Básico descriptivo, inferencial (contraste de hipótesis).

5.1. Población

La población esta constituida por 3.876 personas de la ciudad de Medellín y 12.894 en la ciudad de Bogotá contagiadas por VIH entre los años 2008 y 2021, individuos de diferentes localidades. La media de contagios por localidad en Medellín entre 2008 y 2021 es de 4 personas, mientras que en Bogotá es de 106.

Ilustración 8. Ranking casos confirmados por VIH en las localidades de Medellín [2008 - 2021]

En la **Ilustración 8** se describe el detalle de casos confirmados por VIH en las 13 localidades o barrios mas afectados en la ciudad de Medellín entre el año 2008 y 2021.

Ilustración 9. Ranking casos confirmados por VIH en las localidades de Bogota [2008 - 2021]

En la **Ilustración 9** se describe el detalle de casos confirmados por VIH en las 15 localidades o barrios mas afectados en la ciudad de Bogotá entre el año 2008 y 2021.

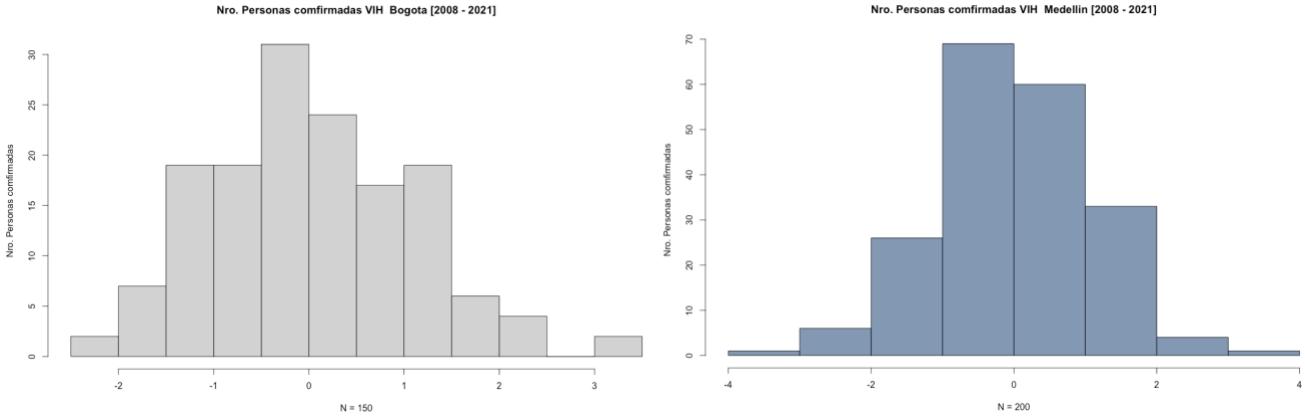


Ilustración 10. Distribución M.A.S casos VIH Medellín y Bogotá [2008 - 2020]

En la **Ilustración 10** se detalla la distribución para cada una de las muestras aleatorias para la población de personas confirmadas de la ciudad de Medellín y Bogotá entre los años 2008 y 2021.

Contraste de Hipótesis Bilateral

- H0 = μ1 == μ2
- H1 = μ1 <> μ2
- IC = (1.96, 1.96)
- Z = 0.16; no se encuentra en el IC,
- p-valor = 0.94; p-valor > 0.05.

Conclusiones: no existe base estadística para encontrar diferencias entre la cantidad de personas confirmadas por VIH entre entre las principales ciudades de Colombia Medellín y Bogotá entre los años 2008 y 2020.

La mayoría de los casos confirmados por VIH se da en los barrios o localidades de mas bajo estrato social, o bajos recursos. Estos casos se dan entre personas homosexuales, trabajadoras sociales, indigentes y personas de bajos recursos.

El gobierno colombiano debe poner foco de concientización y cuidado de la salud, e inversión en educación en las localidades mas afectadas por la situación.

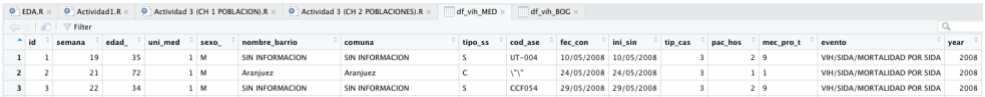
A pesar de la gravedad de la enfermedad, existen avances en términos de tratamientos, lo cual ha ayudado a disminuir la taza de mortalidad por VIH, en los últimos años.

6. Anexos

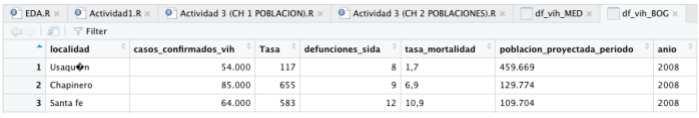
Se adjuntan diferentes artefactos en repositorio personal de GitHub utilizados para llevar a cabo el informe estadístico entre los cuales están:

- **Base de datos:** Casos_positivos_de_COVID-19_en_Colombia.csv, osb_vihsida.csv y sivigila_vih.csv.
- **Informe estadístico**: ESCOBAR_ESCOBAR_JUAN_DAVID.pdf
- **Scripts R-Studio**: Actividad 3 (CH 2 POBLACIONES).R, Actividad 3 (CH 1 POBLACION).R
- **Herramientas**: R-Studio (https://www.rstudio.com/)
- **Librerias**: readr, dplyr, RSQLite, gsubfn, proto, sqldf, lattice, survival, Formula, ggplot2, Hmisc, ggpubr, nortest, car, feather, reshape2, tseries, xts y astsa.

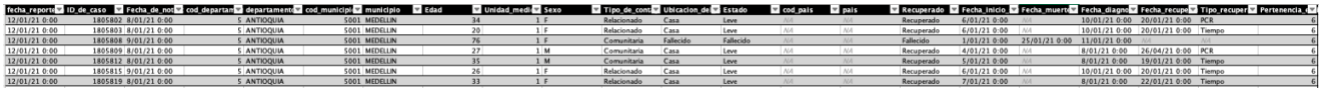
7. Base de datos



	id	semana	edad_	uni_med	sexo_	nombre_barrio	comuna	tipo_ss	cod_ase	fec_con	ini_sin	tip_cas	pac_hos	mec_pro_t	evento	year
1	1	19	35	1	M	SIN INFORMACION	SIN INFORMACION	S	UT-004	10/05/2008	10/05/2008	3	2	9	VIH/SIDA/MORTALIDAD POR SIDA	2008
2	2	21	72	1	M	Aranjuez	Aranjuez	C	\"\"	24/05/2008	24/05/2008	3	1	1	VIH/SIDA/MORTALIDAD POR SIDA	2008
3	3	22	34	1	M	SIN INFORMACION	SIN INFORMACION	S	CCF054	29/05/2008	29/05/2008	3	2	9	VIH/SIDA/MORTALIDAD POR SIDA	2008



	localidad	casos_confirmados_vih	Tasa	defunciones_sida	tasa_mortalidad	poblacion_proyectada_periodo	anio
1	Usaqu\u00e9n	54.000	117	8	1,7	459.669	2008
2	Chapinero	85.000	655	9	6,9	129.774	2008
3	Santa fe	64.000	583	12	10,9	109.704	2008



fecha_reporte	ID_de_caso	Fecha_de_not	cod_departamento	departamento	cod_municipio	municipio	Edad	Unidad_med	Sexo	Tipo_de_cont	Ubicacion_de	Estado	cod_pais	pais	Recuperado	Fecha_inicio	Fecha_muerte	Fecha_diagno	Fecha_recupe	Tipo_recuper	Pertenencia_
12/01/21 0:00	1805802	8/01/21 0:00	5	ANTIOQUIA	5001	MEDELLIN	34	1	F	Relacionado	Casa	Leve	NA	NA	Recuperado	6/01/21 0:00	NA	10/01/21 0:00	20/01/21 0:00	PCR	6
12/01/21 0:00	1805803	8/01/21 0:00	5	ANTIOQUIA	5001	MEDELLIN	20	1	F	Relacionado	Casa	Leve	NA	NA	Recuperado	6/01/21 0:00	NA	10/01/21 0:00	20/01/21 0:00	Tiempo	6
12/01/21 0:00	1805808	9/01/21 0:00	5	ANTIOQUIA	5001	MEDELLIN	76	1	F	Comunitaria	Fallecido	Fallecido	NA	NA	Fallecido	1/01/21 0:00	21/01/21 0:00	11/01/21 0:00	NA	NA	6
12/01/21 0:00	1805809	8/01/21 0:00	5	ANTIOQUIA	5001	MEDELLIN	27	1	M	Comunitaria	Casa	Leve	NA	NA	Recuperado	4/01/21 0:00	NA	8/01/21 0:00	26/04/21 0:00	PCR	6
12/01/21 0:00	1805812	8/01/21 0:00	5	ANTIOQUIA	5001	MEDELLIN	35	1	M	Comunitaria	Casa	Leve	NA	NA	Recuperado	5/01/21 0:00	NA	8/01/21 0:00	19/01/21 0:00	Tiempo	6
12/01/21 0:00	1805815	9/01/21 0:00	5	ANTIOQUIA	5001	MEDELLIN	26	1	F	Relacionado	Casa	Leve	NA	NA	Recuperado	6/01/21 0:00	NA	10/01/21 0:00	20/01/21 0:00	Tiempo	6
12/01/21 0:00	1805819	8/01/21 0:00	5	ANTIOQUIA	5001	MEDELLIN	33	1	F	Relacionado	Casa	Leve	NA	NA	Recuperado	7/01/21 0:00	NA	8/01/21 0:00	22/01/21 0:00	Tiempo	6

8. Bibliografía

Triola, M. F. (2009). ESTADISTÍCA (Vol. DÉCIMA EDICIÓN). (PEARSON, Ed.) Ciudad de Mexico, Mexico: Greg Tobin.

consultorsalud. (22 de Enero de 2021). ¿Por qué el Covid-19 afecta más a los hombres que a las mujeres? Obtenido de consultorsalud.com: https://consultorsalud.com/por-que-el-covid-19-afecta-mas-a-hombres/

Pinzón, M. (10 de 11 de 2020). clinicamedellin. (C. Medellin, Productor) Obtenido de voz del especialista: https://www.clinicamedellin.com/contacto-vital/voz-del-especialista/siempre-hay-tiempo-para-hablar-del-sida-o-el-vih/

Rosales, J. S. (1 de 12 de 2021). bluradio. (bluradio, Productor) Obtenido de Más de 1.600 casos de VIH se han detectado en lo corrido de este año en Medellín:

https://www.bluradio.com/blu360/antioquia/mas-de-1-600-casos-de-vih-se-han-detectado-en-lo-corrido-de-este-ano-en-medellin
Camacho, D. D. (30 de 9 de 2021). TRansformación de datos no normales [Video].
Prof, D. J. (s.f.). rua.ua. Obtenido de Contrastes para los parámetros de dos poblaciones Normales: https://rua.ua.es/dspace/bitstream/10045/17038/1/T6-7_Contr_2_Pob_Normales.pdf
saludcapital. (17 de 8 de 2021). saludcapital. Obtenido de Datos de Salud: https://saludata.saludcapital.gov.co/osb/index.php/datos-de-salud/enfermedades-trasmisibles/covid19/