

## Tarea 1: Estadística Descriptiva

Métodos Estadísticos - 0603

Lic. Ciencia de Datos

IIMAS - UNAM

22 de marzo del 2021

### 1. Censo de Población y Vivienda 2020

1.1. Generar una base de datos a nivel municipio extrayendo directamente los sub-totales para cada municipio y usa esta base para responder las siguientes preguntas.

- ¿De qué tamaño es la base de datos? (renglones  $\times$  columnas en MB)  
22,3 MB
- No filtres y agregues sobre las localidades de cada municipio. ¿Por qué esto sería una pésima idea?

Es una mala idea agregar por cada localidad porque el INEGI, para ciertas localidades con muy baja densidad poblacional, pone valores nulos a campos que darían indicación muy particular sobre un factor. Por ejemplo, la localidad Granja Adelita del municipio de Aguascalientes, Aguascalientes, cuenta con una población de 5 personas, y para las columnas que indican el número de población femenina y masculina, el INEGI pone un valor de \*. Dicho esto, si sumáramos por cada localidad, en el total para cada municipio perderíamos información sobre las localidades.

1.2. Genera el porcentaje de población económicamente activa <sup>1</sup> para cada una de las 32 entidades federativas así como para las 16 alcaldías de la CDMX.

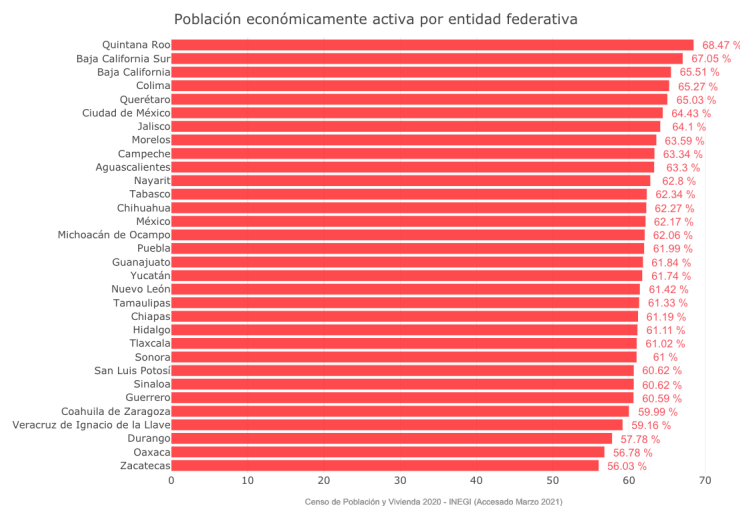


Figura 1: Población económicamente activa para las 32 entidades federativas

De la figura 1 podemos observar que todas las entidades se concentran en valores muy similares (cerca de los sesenta), concretamente, el promedio de población económicamente activa en el país es del 62,06 % con una desviación estándar de 2,6607. Con respecto al análisis por alcaldía de la figura 2, podemos observar que de igual manera, el porcentaje de población económicamente activa ronda los sesenta, concretamente, las alcaldías tienen una media de

<sup>1</sup>Población de 12 años en adelante económicamente activa

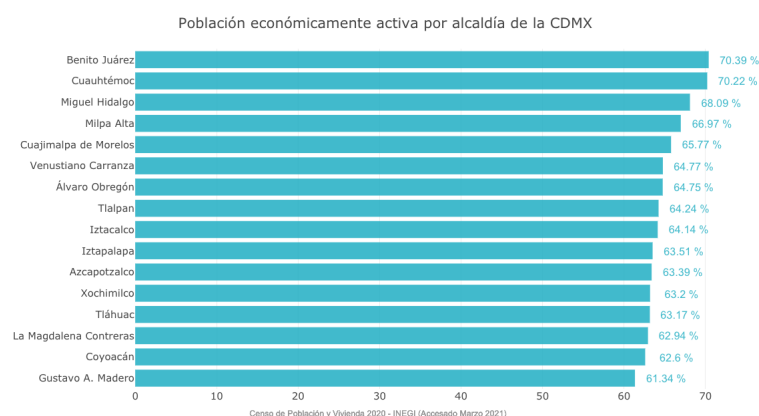


Figura 2: Población económicamente activa para las 16 alcaldías de la Ciudad de México

64,96 % con una desviación estándar de 2,6656. Con estas medidas de dispersión, podemos concluir que la Ciudad de México tiene, en promedio, una mayor población económicamente activa con respecto al promedio nacional.

- 1.3. Genera el porcentaje de población de 5 años en adelante que hable alguna lengua indígena así como el porcentaje de población de 15 años en adelante que se analfabeta para cada uno de los municipios del país. ¿Existe alguna relación entre ambas variables?

En el país existen 2469 municipios por lo que un diagrama de dispersión no reflejará de manera adecuada la posible correlación entre el porcentaje de personas que hablan alguna lengua indígena y el porcentaje de población analfabeta. Sin embargo, en la figura 3 se muestra el panorama general del país.

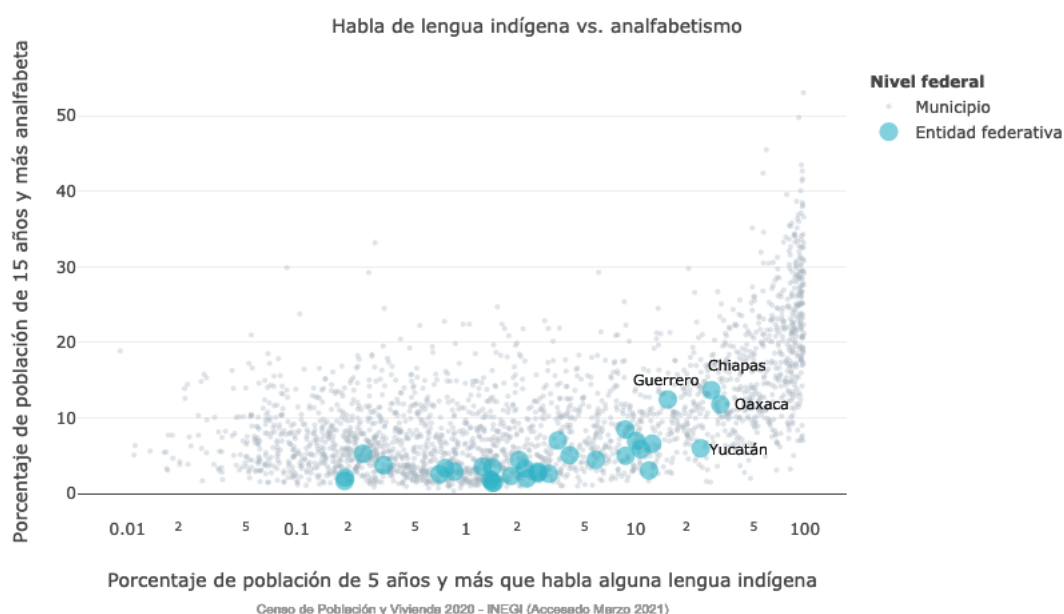


Figura 3: Porcentaje de personas que hablan lengua indígena vs. porcentaje de población analfabeta con escala logarítmica

De la figura 3 podemos observar una ligera tendencia de correlación positiva, sin embargo, para poder dar una respuesta más concluyente, se tomó como base la siguiente tabla que muestra el coeficiente de Pearson para las entidades federativas que mostraran una correlación positiva o negativa:

Entidad Federativa	Coefficiente de Pearson	Cantidad de Municipios
Aguascalientes	-0.2921	11
Baja California	0.9965	6
Chihuahua	0.9419	67
Nayarit	0.9611	20
Nuevo León	-0.4889	51
Tamaulipas	-0.3484	43

Cuadro 1: Coeficientes de Pearson para las entidades federativas con  $\rho \leq 0$  y  $\rho \geq 0,75$

A continuación se muestran las gráficas de dispersión para las entidades del cuadro 1

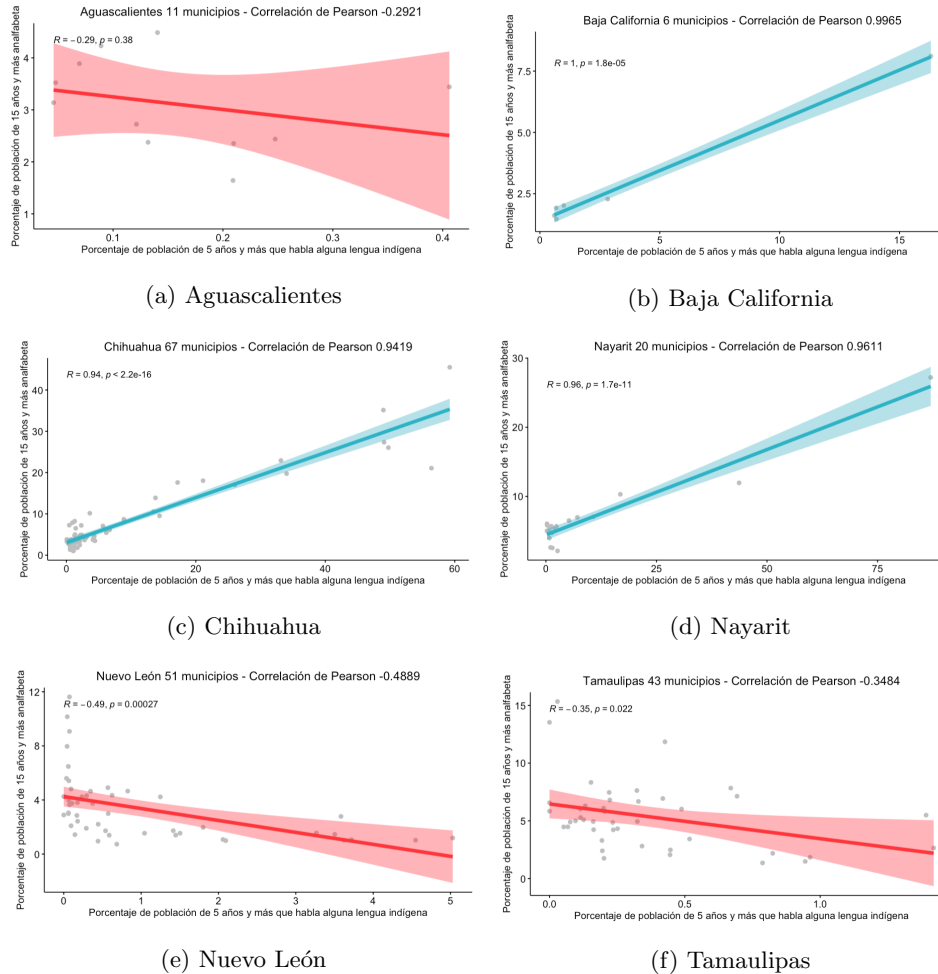


Figura 4: Porcentaje de personas que hablan lengua indígena vs. porcentaje de población analfabeta de las entidades federativas del cuadro 1

Del análisis de la figura 5, podemos observar que sí existe una correlación positiva entre el porcentaje de personas que hablan alguna lengua indígena y el analfabetismo en los estados de Baja California, Chihuahua y Nayarit. Para las demás entidades federativas, no se puede afirmar la existencia de esta correlación. Como punto final, en la siguiente figura se muestra el diagrama de dispersión de las entidades federativas que tienen los mayores índices de habla indígena y analfabetismo, sin embargo, con base en el coeficiente de Pearson, no se puede asegurar la existencia de esta correlación de manera certera:

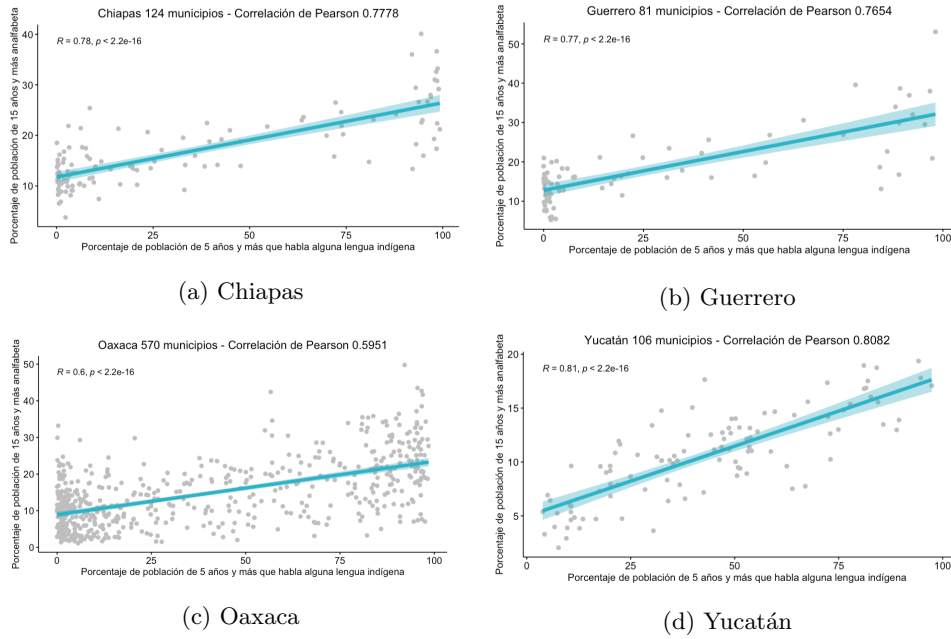


Figura 5: Porcentaje de personas que hablan lengua indígena vs. porcentaje de población analfabeta de las entidades federativas del cuadro 1

## 2. Pandemia de COVID-19 en la CDMX <sup>2</sup>

### 2.1. Hospitalizaciones

2.1.1. A partir de la fecha de ingreso (variable *fecha\_ingreso* en formato *as.Date*) primero junio de 2020 (un lunes, esto es importante) calcula el porcentaje de hospitalización por día:

$$\text{porcentaje hospitalizaciones por día} = \frac{\text{hospitalizados por día (de los infectados)}}{\text{infectados de COVID-19} \times \text{día}}$$

Genera una gráfica de serie de tiempo con esta información (en el eje *x* la *fecha\_ingreso* y en el eje *y* los %)

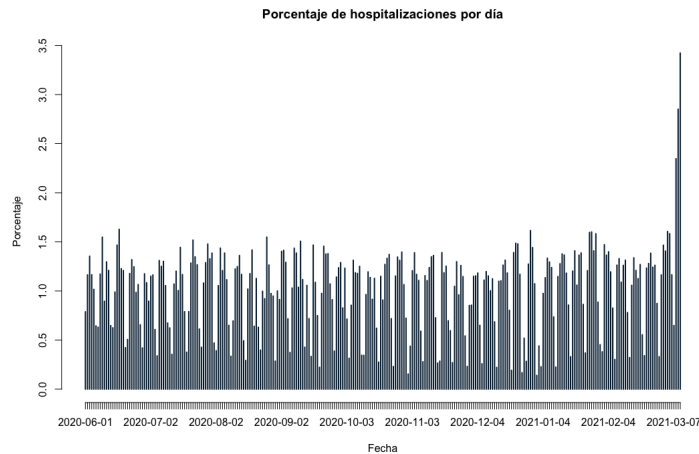


Figura 6: Porcentaje de hospitalizaciones por día desde el lunes, 1 de junio de 2020 hasta el miércoles 10 de marzo de 2021

<sup>2</sup>Esta sección se trabajó gracias a los datos obtenidos del portal de datos abiertos de la CDMX [2].

- 2.1.2. Utiliza la función *frollmean* de la librería *data.table* para obtener una media móvil con una ventana de 7 días. Genera una gráfica de serie de tiempo con esta información.

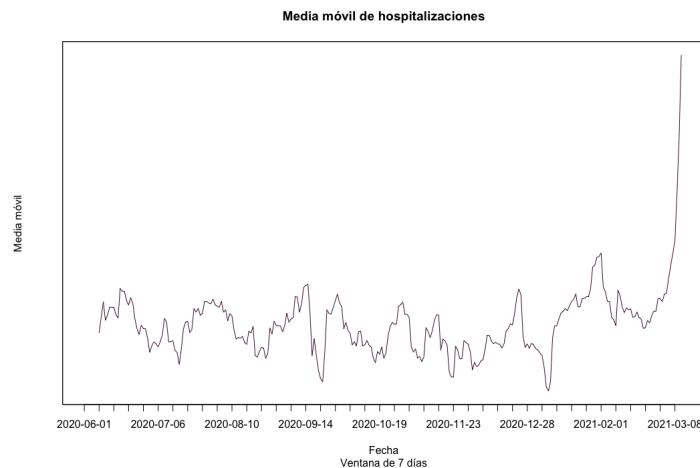


Figura 7: Media móvil con ventana de 7 días de las hospitalizaciones desde el lunes, 1 de junio de 2020 hasta el miércoles 10 de marzo de 2021

- 2.1.3. Comenta y compara las dos gráficas anteriores (es fácil interpretar la primer gráfica, la segunda, etc.)

Sin considerar los últimos días—dado que se estima que los registros completos de nuevos casos, fallecimientos y hospitalizaciones tardan entre 14 y 21 días en estar correctamente registrados en la base de datos—se puede observar que en la primera gráfica tenemos que los hospitales han estado constantemente recibiendo entre el 1% y 2% de contagiados por COVID-19 diariamente. Esto quiere decir que si reducimos al máximo la cantidad de contagios, los hospitales reducirán en automático la saturación. De la misma manera, si aumenta el número de contagios, los hospitales comenzarán a recibir una cantidad proporcional de pacientes y se saturará el sistema de salud. Esta gráfica es una muestra clara del porqué se deben restringir las actividades no esenciales para que los hospitales puedan seguir operando y atender a todos los pacientes que requieran la atención. Adicionalmente, es interesante este resultado, dado que no se muestra que haya diferencias estacionales en el porcentaje de hospitalizados por día. Es decir, independientemente del número de contagios que haya diariamente, podemos decir con certeza que habrá entre el 1% y 2% de hospitalizados. Finalmente, sería interesante revisar esta gráfica una vez que las variantes recién descubiertas del virus comiencen a ser la cepa predominante en México y revisar si aumenta o se reduce el número de hospitalizaciones.

En contraste, la segunda gráfica sí es sensible a las fechas en las que uno esté realizando las observaciones, dado que los valores de esta gráfica están directamente relacionados con el número de contagios. En particular, podemos observar que el promedio móvil incrementa en las fechas en las que hay puentes vacacionales o días de asueto, la reducción de las medidas sanitarias y la temporada de fiestas decembrinas. En general, en el periodo bajo estudio no es tan sencillo de observar con esta gráfica—debido a que el promedio es poco estable—pero hay dos grandes tendencias. Una pequeña reducción en el número de casos, seguido de un aumento durante las últimas semanas del 2020 y una reducción a partir de enero de 2021. Este resultado nos puede decir que el número de hospitalizaciones incrementa después de unas semanas de que ocurren los contactos de contagio. Por ejemplo, una persona que estuvo en contacto con un enfermo y se contagió durante la noche del 24 de diciembre, probablemente habrá ingresado la primera semana de enero—coincidente con el tiempo de convalecencia de la enfermedad.<sup>3</sup>[3]

<sup>3</sup>Es una conclusión parecida a la que llegó Julio Frenk en su artículo “Covid-19: Balance catastrófico” publicado en la revista Nexos en la edición de octubre de 2020.

- 2.1.4. Observando con mucha atención la gráfica con los promedios móviles, quizá tengas que modificar tu escala (hacer un acercamiento sobre el eje  $y$  para ver claramente lo que sucede). Obtén cada una de las fechas en las que se observan crestas (pequeñas montañas). Investiga en internet y responde: ¿qué sucedió en esas fechas? Si fueron días festivos, ¿cuáles? ¿Periodos vacacionales? ¿Cambios de color en el semáforo?

Si se observan las fechas en las que hubo un aumento significativo del número de casos, obtenemos las siguientes: 31 de julio de 2020, 5 de octubre de 2020, 1 de diciembre de 2020 y 21 de enero de 2021. ¿Cómo se pueden analizar las observaciones?

Para el primer aumento de casos, consideramos que este proviene del cambio en la metodología del manejo del semáforo epidemiológico y el fin de la jornada de “Sana distancia” o fase 3 de las fases epidemiológicas. Adicionalmente, estos casos están relacionados con el paulatino regreso a los centros de trabajo y el aumento parcial de la movilidad de las personas—incluyendo aquellas que salieron por motivos vacacionales debido al término del ciclo escolar, a pesar de que no se considere como causa principal.

En octubre, el aumento de casos coincide con las fechas de celebración del 16 de septiembre—el inicio de la Independencia en México—coincidente con los 15 a 21 días de que se efectuaron las celebraciones. Por lo tanto, se puede considerar que aumentó la movilidad de las personas y el contacto entre familias o conocidos en lugares cerrados y con poca protección o cuidado personal. De igual manera, se puede observar un aumento significativo durante los primeros días de diciembre, coincidiendo con el día de asueto por el 20 de noviembre de 2021—inicio de la Revolución mexicana—y el inicio de actividades decembrinas, especialmente compras y reuniones de los centros de trabajo o posadas. En este particular, consideramos que fue tardío el anuncio del semáforo rojo en la CDMX, el cual fue indicado hasta el 18 de diciembre. Peor aún, este inició el lunes 21 de diciembre, permitiendo que las personas continuaran con sus actividades durante todo el fin de semana.

El 21 de enero, el pico más alto en el número de hospitalizaciones se explica por las fiestas decembrinas y de Año Nuevo. Reuniones, fiestas y alta movilidad tienen una alta correlación con el aumento en el número de casos y, como se ha visto con la gráfica anterior, con el número de hospitalizaciones. Este periodo de tiempo fue el que más complicaciones generó para el sistema de salud y donde más fallecimientos se presentaron—desafortunadamente. Es más que claro que las reuniones, contacto en lugares cerrados y alta movilidad son los factores principales de propagación de la enfermedad, algo que no se ha aplicado del todo bien durante la implementación de los semáforos epidemiológicos en rojo, gracias a pocas campañas de concientización y el discurso de que la falsa dicotomía: salvar la economía o salvar vidas.

## 2.2. Comorbilidades en los infectados por COVID-19

- 2.2.1. Calcula la probabilidad condicional de fallecer por COVID-19 dado cada una de las comorbilidades en la base de datos  $P(F|C_j)$ ,  $j = 1, 2, \dots$ . Obviamente solo considerando los enfermos de COVID-19. Haz una gráfica de barras y comenta al respecto.

Con los resultados obtenidos, es posible observar que la diferencia entre las tres comorbilidades con mayor probabilidad de estar presentes en casos de fallecidos por COVID-19 respecto a las demás es del orden de 0,3, un elevado factor de diferencia. Adicionalmente, coincide que estas tres son una de las causas principales de complicaciones, enfermedades y muertes en México. Por lo tanto, si uno padece hipertensión, diabetes u obesidad, las probabilidades de fallecer—si se llegara a enfermar de COVID-19 son altas. Debería ser prioridad del gobierno vacunar a personas con estas condiciones de inmediato para reducir el número de fallecidos. Sin embargo, sería complejo realizar un ordenado proceso de vacunación con bajo esta premisa.

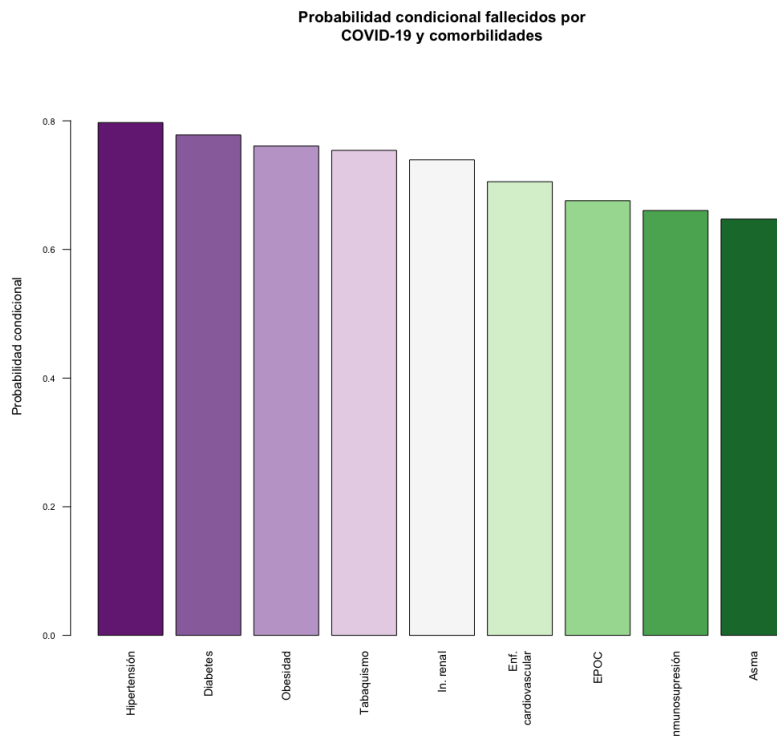


Figura 8: Probabilidad condicional de fallecer por COVID-19 dada cada comorbilidad

2.2.2. Piensa muy bien la siguiente pregunta, ¿qué comorbilidad es asociada a un mayor número de fallecimientos por COVID-19 en la CDMX? Justifica tu respuesta de la siguiente manera: si en la CDMX hubiera 100,000 infectados por COVID-19, obtén el número promedio de fallecimientos asociados a cada comorbilidad.

Considerando las 3 grandes comorbilidades descritas anteriormente, de las cuales tenemos un porcentaje de fallecimiento del: 0,729 % para hipertensión, 0,608 % para diabetes y 0,385 % para las personas que padecen obesidad, si se considerara el número de fallecidos a partir de 100,000 contagios por COVID-19 en la CDMX, obtendríamos que el siguiente número de personas habría fallecido, segregado por la comorbilidad que hubieran tenido, sería el descrito a continuación:

- Hipertensión: 729
- Diabetes: 608
- Obesidad: 385
- Tabaquismo: 174
- Insuficiencia renal: 107
- Enfermedad cardiovascular: 85
- EPOC: 79
- Inmunosupresión: 41
- Asma: 25

Con lo anterior, la comorbilidad que más se presenta en los fallecidos por COVID-19 en la CDMX es la hipertensión, seguida por la diabetes y la obesidad. En general, esto no solo es una alerta para enfrentar los casos de COVID-19, es un recordatorio de las condiciones de salud en las que se encuentra la población mexicana y de la CDMX. Son alarmantes los altos niveles de padecimientos por hipertensión, diabetes y obesidad. Es decir, son urgentes campañas de salud enfocadas en reducir estas enfermedades y que se

promueva el mejoramiento de la alimentación de los ciudadanos mexicanos.

Aún así, es importante hacer comparaciones con otros países—por ejemplo Pakistán—quienes tienen una mayor prevalencia de diabetes y mayores niveles de pobreza. Sin embargo, el nivel de muerte que ha alcanzado este país asiático es mucho menor al de México, a pesar de contar con más de 200,000 de personas en su territorio. Es decir, no se enfrentó de manera oportuna ni agresiva la situación. Por lo tanto, se volvió más compleja la respuesta, debido a la prevalencia de comorbilidades y, aunado a la laxa respuesta gubernamental y la reestructuración del sistema de salud, la pandemia nos ha golpeado muy fuerte—incluso a comparación de otros países.

### 2.3. Conclusión y reflexión a partir de los resultados

Gracias a la exploración de los datos de comorbilidad y porcentaje de hospitalizaciones encontrado con ayuda de la base de datos bajo análisis, se puede concluir que existe una relación directa entre el número de hospitalizados y el aumento en el número de contagios, además de poder decir que no existe una relación entre el aumento de contagios y el porcentaje de población contagia que debe hospitalizarse. Es decir, la probabilidad de que un contagiado deba hospitalizarse es independiente a la cantidad de contagios globales por fecha. Por lo cual, si se mantiene bajo el número de contagios, se mantendrá bajo el número de hospitalizaciones.

También, es interesante analizar que el impacto que ha tenido la pandemia en México tiene su origen en el rezago que ha habido desde hace muchos años en el sistema de salud. Este agravante, se ha acrecentado con la reestructuración desafortunada que está en proceso en las dependencias responsables de salvaguardar la salud de los mexicanos. Este desmantelamiento—originado en el cada vez menor presupuesto otorgado a estas dependencias desde hace cinco años—ha alcanzado su cima con la desaparición del Seguro Popular y la creación del INSABI, un órgano que no tiene reglamentación clara de operación ni tiene experiencia en su dirección. Como consecuencia, la atención del COVID en México ha estado completamente segregada en la capacidad estatal de las distintas dependencias locales, lo cual no es el caso bajo estudio pero sí tiene inflexión, gracias al papel que ha jugado la CDMX en la pandemia. [?]

Finalmente, no se debe perder de vista que las comorbilidades—características de la población mexicana—que han incrementado el número de padecimientos, costos de atención y mortalidad de aquellos que se han contagiado de COVID-19, es un malestar social que no desaparecerá una vez que se tenga contagiada la pandemia. Las comorbilidades son en sí una epidemia y emergencia en el ámbito de la salud y es urgente su atención. Tenemos que aprender de todo lo acontecido durante este año, y contando, de las omisiones, errores y desmantelamiento institucional que han provocado tantas pérdidas humanas y han afectado de manera incalculable la vida de los mexicanos.

## Referencias

- [1] INEGI, “Censo de Población y Vivienda 2020”, revisado el 22 de marzo del 2021 en: [https://www.inegi.org.mx/programas/ccpv/2020/#Datos\\_abiertos](https://www.inegi.org.mx/programas/ccpv/2020/#Datos_abiertos)
- [2] Datos Abiertos Gobierno de la CDMX. (2021, marzo 10) Covid-19 SINAVE Ciudad de México. Secretaría de Salud Gobierno de la CDMX. Obtenida el viernes, 19 de marzo de 2021 de [https://archivo.datos.cdmx.gob.mx/sisver\\_public.csv](https://archivo.datos.cdmx.gob.mx/sisver_public.csv)
- [3] Frenk, Julio, Gómez Dantés, Octavio. (2020, octubre). “Covid-19: Balance Catastrófico”. *Nexos*, número 514, 32-36.