

CONFERENCIAS STATA LATAM 2022

Herramientas y aplicaciones  
estadísticas para Ciencia de Datos

# Epidemiología

**Stata como apoyo para la vigilancia epidemiológica.**

Panorama epidemiológico y vigilancia epidemiológica.

Condiciones de riesgo.

Comparaciones de medias.

**La pandemia de Covid-19 en México.**

*Bogotá, Colombia. Septiembre 7, 2022*



## Herramientas y aplicaciones estadísticas para Ciencia de Datos

### Introducción.

[Panorama epidemiológico.](#)

[Vigilancia epidemiológica.](#)

[Datos y calidad.](#)

[Análisis exploratorio.](#)

[Magnitud.](#)

[Tiempo.](#)

[Lugar.](#)

[Variables personales.](#)

[Condiciones de riesgo.](#)

[Evidencia de causalidad.](#)

[Ejemplo: El lugar.](#)

[Otras condiciones.](#)

[Análisis de la evidencia.](#)

[Comparación de medias.](#)

[Estimación.](#)

[Una muestra.](#)

[Dos muestras.](#)

[Independientes.](#)

[Emparejadas.](#)

[Más de dos muestras.](#)

[Medidas de intervención.](#)

[Intervenciones inespecíficas.](#)

[Intervenciones específicas.](#)

[Conclusiones.](#)

[Preguntas.](#)

- 1. Introducción.**
- 2. El panorama epidemiológico.**
- 3. La vigilancia epidemiológica.**  
**Datos y calidad de la vigilancia.**  
**Análisis exploratorio.**  
**Magnitud.**  
**La variable tiempo.**  
**La variable lugar.**  
**Variables de las personas.**
- 4. Las condiciones de riesgo.**  
**Evidencia de causalidad.**  
**Ejemplo: El lugar.**  
**Otras condiciones.**
- 5. El análisis de la evidencia.**  
**Comparaciones de medias.**
- 6. Medidas de intervención.**
- 7. Conclusiones.**
- 8. Preguntas.**

## Herramientas y aplicaciones estadísticas para Ciencia de Datos

### Introducción.

Panorama epidemiológico.

Vigilancia epidemiológica.

Datos y calidad.

Análisis exploratorio.

Magnitud.

Tiempo.

Lugar.

Variables personales.

Condiciones de riesgo.

Evidencia de causalidad.

Ejemplo: El lugar.

Otras condiciones.

Análisis de la evidencia.

Comparación de medias.

Estimación.

Una muestra.

Dos muestras.

Independientes.

Emparejadas.

Más de dos muestras.

Medidas de intervención.

Intervenciones inespecíficas

Intervenciones específicas.

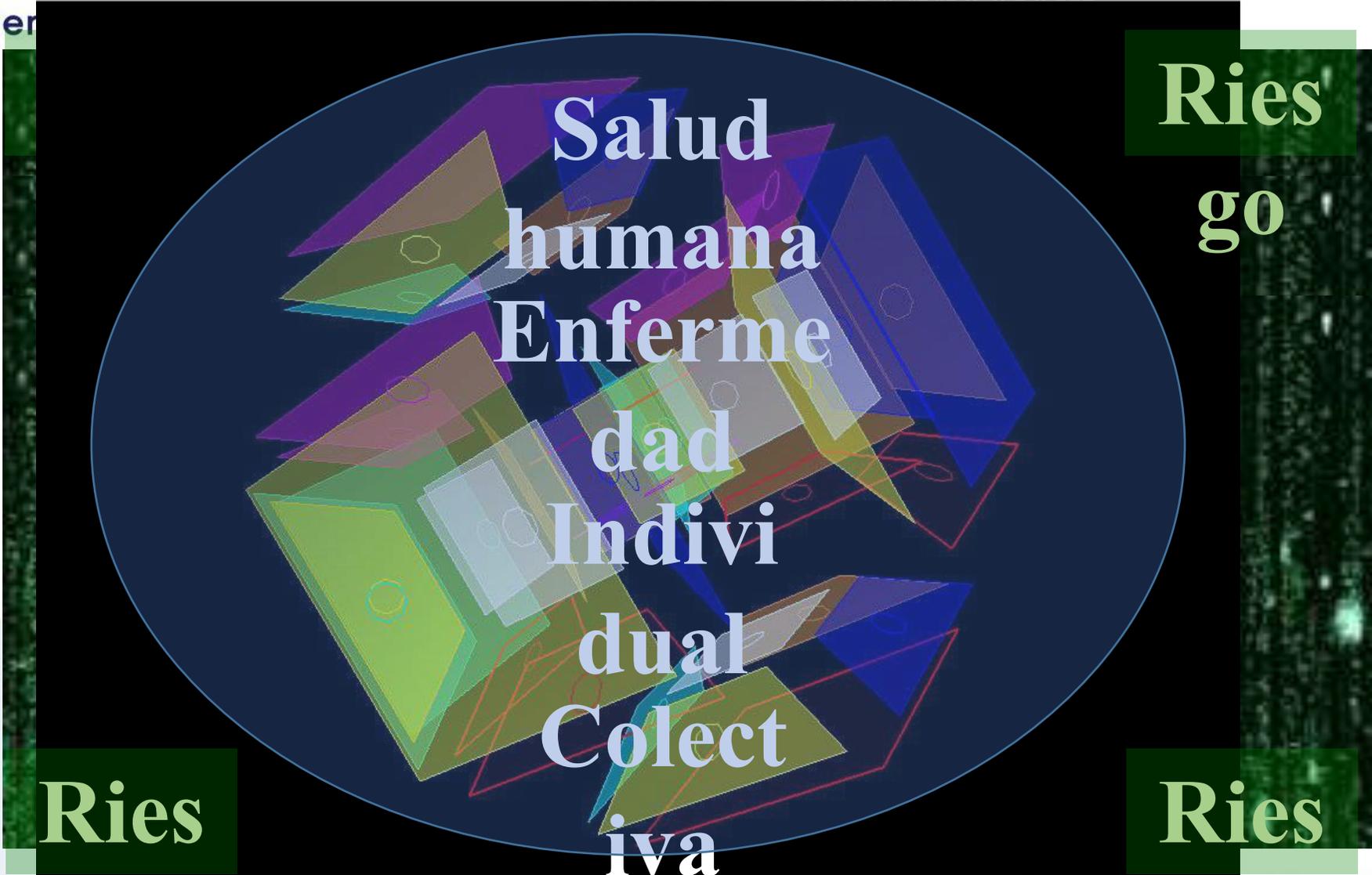
Conclusiones.

Preguntas.

- 1. Introducción.**
- 2. El panorama epidemiológico.**
- 3. La vigilancia epidemiológica.**  
*Datos y calidad de la vigilancia.*  
*Análisis exploratorio.*  
*Magnitud.*  
*La variable tiempo.*  
*La variable lugar.*  
*Variables de las personas.*
- 4. Las condiciones de riesgo.**  
*Evidencia de causalidad.*  
*Ejemplo: El lugar.*  
*Otras condiciones.*
- 5. El análisis de la evidencia.**  
*Comparaciones de medias.*
- 6. Medidas de intervención.**
- 7. Conclusiones.**
- 8. Preguntas.**

Introducción.

- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
      - Una muestra.
      - Dos muestras.
        - Independientes.
        - Emparejadas.
      - Más de dos muestras.
  - Medidas de intervención.
    - Intervenciones inespecíficas
    - Intervenciones específicas.
- Conclusiones.
- Preguntas.



Ries  
go

Ries  
go

Ries  
go

## Herramientas y aplicaciones estadísticas para Ciencia de Datos

Introducción.

Panorama epidemiológico.

Vigilancia epidemiológica.

Datos y calidad.

Análisis exploratorio.

Magnitud.

Tiempo.

Lugar.

Variables personales.

Condiciones de riesgo.

Evidencia de causalidad.

Ejemplo: El lugar.

Otras condiciones.

Análisis de la evidencia.

Comparación de medias.

Estimación.

Una muestra.

Dos muestras.

Independientes.

Emparejadas.

Más de dos muestras.

Medidas de intervención.

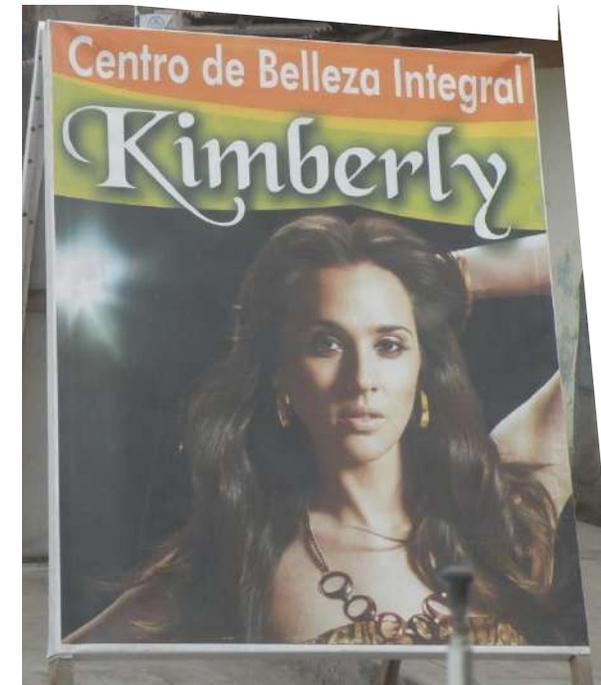
Intervenciones inespecíficas

Intervenciones específicas.

Conclusiones.

Preguntas.

## Panorama epidemiológico.



- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.**
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
    - Una muestra.
    - Dos muestras.
      - Independientes.
      - Emparejadas.
    - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

# Vigilancia epidemiológica.

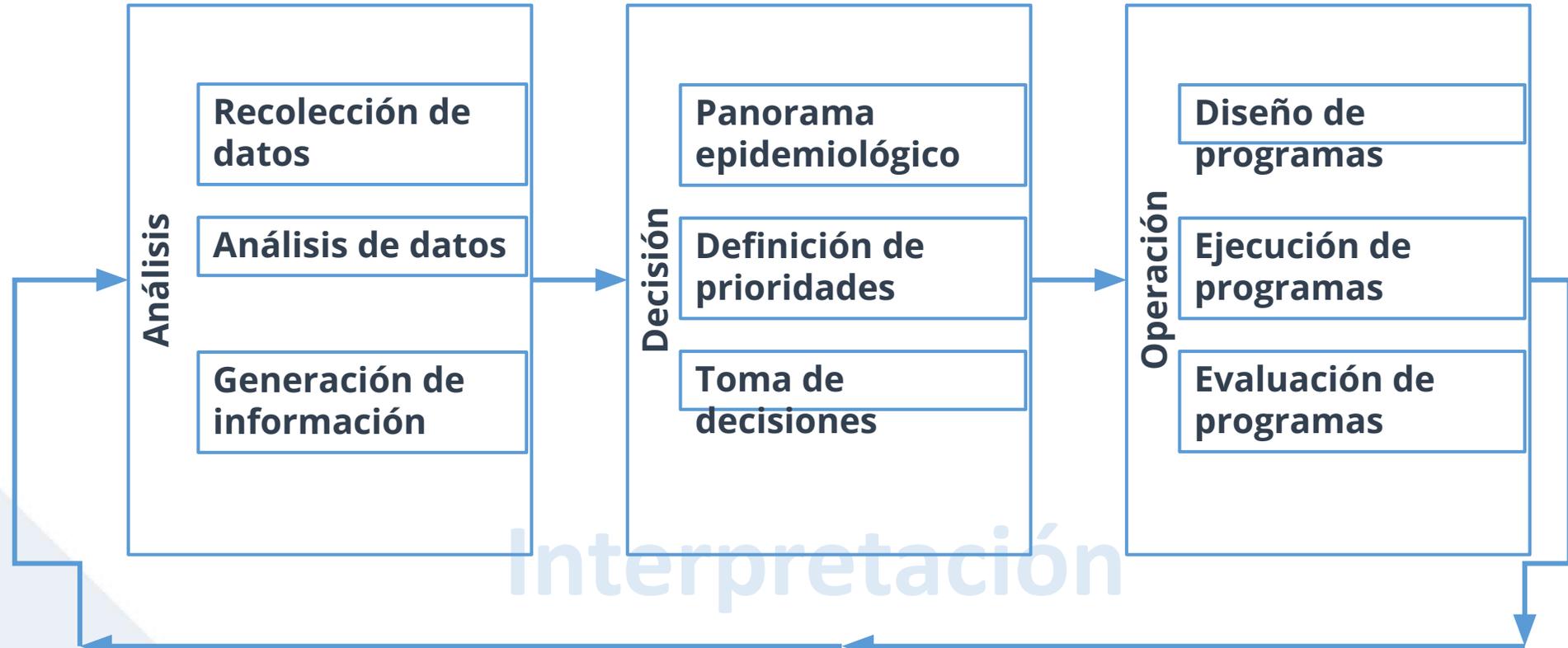
Insumos para el análisis.

Datos e información.

El archivo de datos.

# Sistema de vigilancia epidemiológica.

Conjunto de sistemas, básicos y específicos, de recolección y análisis de datos útiles para generar y divulgar información que sirva como base para tomar decisiones sobre planificación, ejecución y evaluación de intervenciones de prevención y control de problemas de salud, generalmente de Salud Pública.



**Panorama epidemiológico:** Información que describe la salud en una población humana y algunas de las condiciones, idealmente causales, que contribuyeron a determinarla.

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
    - Una muestra.
    - Dos muestras.
      - Independientes.
      - Emparejadas.
    - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

CONFERENCIAS STATA LATAM 2022

Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
- Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
  - Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
  - Una muestra.
  - Dos muestras.
    - Independientes.
    - Emparejadas.
  - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

El archivo de datos (importar).

Información recolectada de la realidad:

Registros  
Variables  
Campos

import delimited "C:\Covid-19 DGE Datos abiertos\  
Mexico 2022\2022\220813COVID19MEXICO.csv"

Archivo > Importar > Datos de texto (delimitados, \*, csv, ...)

Import delimited - Datos de texto Importación Delimitada

Archivo a importar: C:\Covid-19 DGE Datos abiertos - Parte 2, hasta dic 31\2022 Mexico\2022\220813COVID19MEXICO.csv

Delimitador: Automático

Usar la primera fila como nombre de variable: Automático

Mayúscula/minúscula: Inferior

Precisión de punto flotante: Usar estándar float

Codificación del texto: Automático UTF-8

Excluir comillas: Automático

Vincular comillas: Suelta

Reglas de análisis numérico

Utilice el análisis basado en la configuración regional para los números

Local: español (México)

Cambiar separador decimal: [ ]

Cambiar separador de agrupación: [ ]

Vista previa:

#	fecha_actualizacion	id_registro	origen	sector	entidad_um	sexo	entidad_nac	entidad_res	municipio_res	tipo_paciente	fecha_ingreso	fecha_sintomas
2	2022-08-13	z53cb3	2	12	09	2	09	09	002	1	2021-01-16	2021-01-13
3	2022-08-13	z2a3cf	1	6	24	2	24	24	028	1	2020-02-20	2020-02-18
4	2022-08-13	z3d1e2	2	12	09	1	09	09	005	1	2020-04-22	2020-04-20
5	2022-08-13	z17f15	1	12	07	1	07	07	101	1	2020-03-26	2020-03-23
6	2022-08-13	zr49f	2	12	19	1	19	19	089	1	2020-05-20	2020-05-19
7	2022-08-13	z4b9a5	1	12	29	2	29	29	013	2	2020-05-20	2020-05-15
8	2022-08-13	z29dac	2	12	09	1	09	09	010	1	2020-09-12	2020-09-10
9	2022-08-13	z2159	2	12	19	1	19	19	026	1	2020-07-10	2020-07-06
10	2022-08-13	z21c19	1	12	19	2	19	19	026	1	2020-11-01	2020-11-01
11	2022-08-13	z4bf07	2	12	31	1	31	31	041	1	2020-06-13	2020-06-10
12	2022-08-13	z45dcb	2	12	09	2	09	09	003	1	2020-08-18	2020-08-16
13	2022-08-13	z477ab	1	12	12	2	12	12	059	2	2020-06-02	2020-05-27
14	2022-08-13	z5293	1	12	29	1	29	29	005	1	2020-07-09	2020-07-01
15	2022-08-13	z1d9a5	1	12	22	2	09	22	014	1	2020-12-09	2020-12-03

Variable Tipo

fecha_actualizacion	str
id_registro	str
origen	numeric
sector	numeric
entidad_um	numeric
sexo	numeric
entidad_nac	numeric
entidad_res	numeric
municipio_res	numeric
tipo_paciente	numeric
fecha_ingreso	str

Nota: la importación de cadenas de caracteres como datos numéricos puede resultar en la pérdida de datos.

OK Cancelar Enviar



# Herramientas y aplicaciones estadísticas para Ciencia de Datos

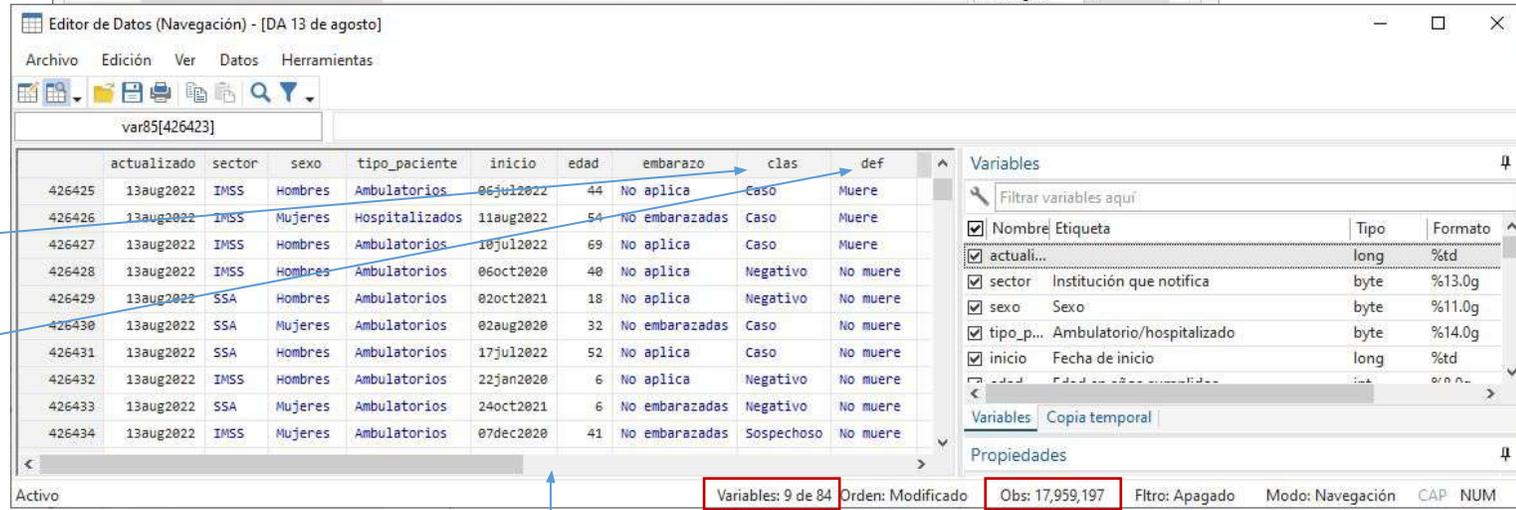
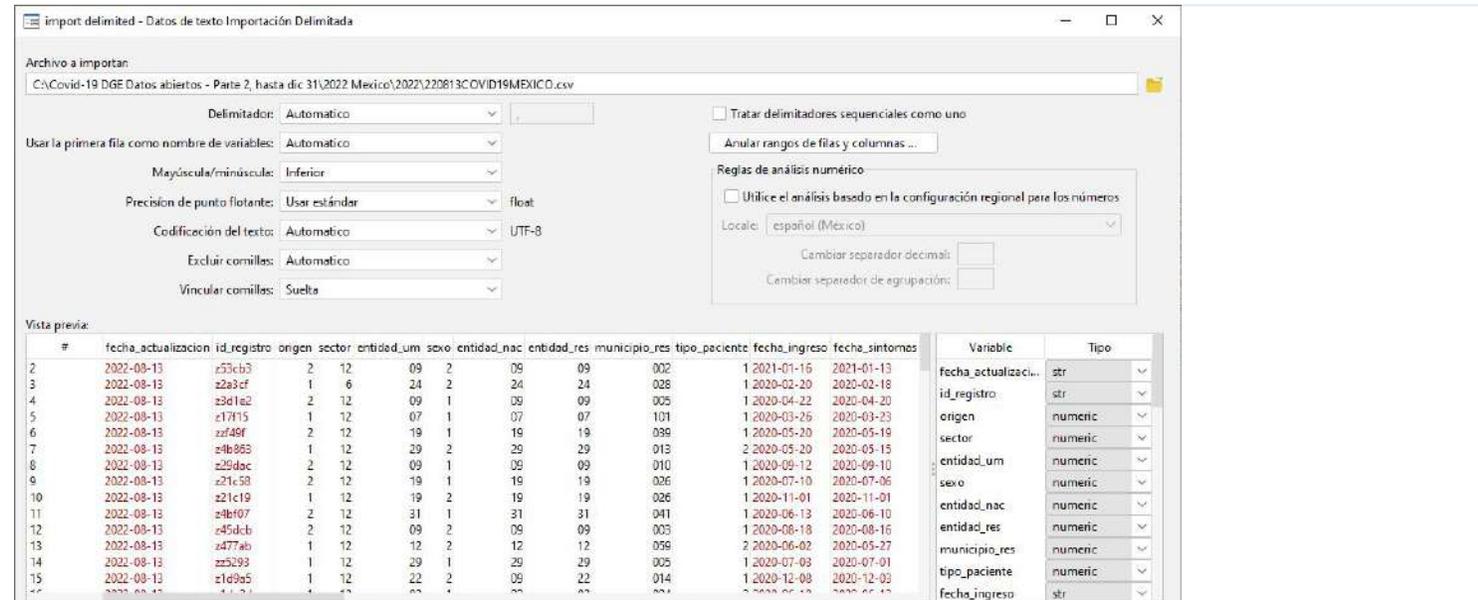
- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

El archivo de datos (importado).

Información recolectada de la realidad:

Registros (revisión, duplicados)  
 Variables (etiquetas, revisión)  
 Campos (etiquetas, códigos, revisión)

Datos > Crear o cambiar datos > Otras comandos para transformar variables > Recodificar variable categórica >



\*\*\* Recodificar la variable clasificacion\_final

```
recode clasificacion_final (1 = 1 "Caso") (2 = 1 "Caso") (3 = 1 "Caso") (4 = 4 "No válido") (5 = 4 "No válido") (6 = 0 "Sospechoso") (7 = 3 "Negativo") (8 = 1 "Caso") (9 = 0 "Sospechoso"), generate(clas)
```

label variable clas "Clasificación"

```
*recode muerte (nonmissing = 1 "Muerte"), generate(def)
```

```
brow actualizado sector sexo tipo_paciente inicio edad embarazo clas def
```



- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
    - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
      - Una muestra.
      - Dos muestras.
        - Independientes.
        - Emparejadas.
      - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas.
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

# Vigilancia epidemiológica.

Estudios descriptivos.

Análisis exploratorio (descripción).

Magnitud.

# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

## Daños a la salud (magnitud).

Número de casos notificados y de muertes registradas para cada periodo anual.

Datos para estimar la morbilidad y de la mortalidad para cada periodo.

use DA 13 de agosto.dta

table ( añoini ) ( evol ) if evol<3, missing 1

table ( añodef ) ( evol ) if evol==2, missing 2

Estadísticas > Sumarios, tablas, y tests estadísticos > Tablas de frecuencias, sumario estadístico y resultados de comandos

Cuadro\* núm. 1a. Covid-19. Casos, por fecha de inicio del cuadro clínico, y defunciones, por fecha de ocurrencia, por años, notificados hasta el sábado 13 de agosto, último día de la semana epidemiológica número 32 de 2022, en la República Mexicana.

Periodo de notificación.	<span style="border: 1px solid red; border-radius: 50%; padding: 2px;">1</span> Número de casos <sup>2</sup> .	<span style="border: 1px solid red; border-radius: 50%; padding: 2px;">2</span> Número de defunciones <sup>3</sup> .
2020	1,563,135	149,455
2021	2,526,649	154,570
2022	2,835,884	24,699
2020-2022 <sup>1</sup>	6,925,668	328,724

\* Elaborado a partir de los datos abiertos publicados por la Dirección General de Epidemiología el 13 de agosto de 2022.

Fuente: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>

<sup>1</sup> Total de casos y defunciones desde el inicio de la pandemia hasta la fecha de la publicación.

<sup>2</sup> Total de casos notificados que iniciaron un cuadro clínico de Covid-19 en cada periodo.

<sup>3</sup> Defunciones por Covid-19 que ocurrieron durante cada periodo, sin considerar las fechas en que iniciaron el cuadro clínico.



- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
    - Una muestra.
    - Dos muestras.
      - Independientes.
      - Emparejadas.
    - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas.
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

# Vigilancia epidemiológica.

Panorama epidemiológico.

Análisis exploratorio.

La variable tiempo.

# Herramientas y aplicaciones estadísticas para Ciencia de Datos

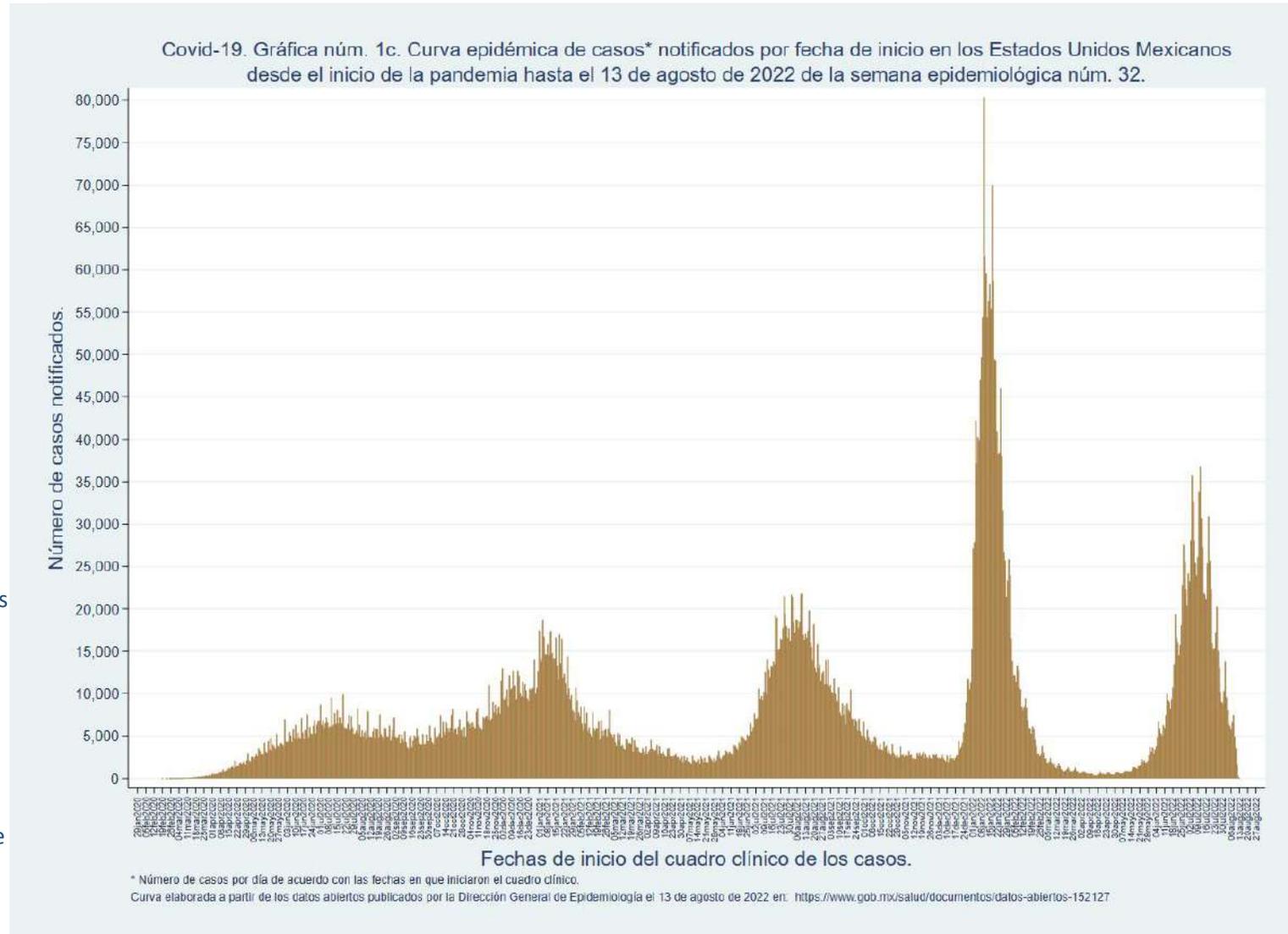
- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
    - Una muestra.
    - Dos muestras.
      - Independientes.
      - Emparejadas.
    - Más de dos muestras.
  - Medidas de intervención.
    - Intervenciones inespecíficas
    - Intervenciones específicas.
- Conclusiones.
- Preguntas.

## Curva epidémica (casos).

### Número de casos notificados por día para los años 2020, 2021 y 2022.

twoway (histogram inicio if evol<3, discrete frequency fcolor(red%40) lcolor(brown) lwidth(vthin)), ytitle(Número de casos notificados.) ytitle(, size(large) color(navy8)) ylabel(#20, labcolor(dknavy) angle(horizontal) format(%9.0fc)) ymtick(, labsz(sm)) xtitle(Fechas de inicio del cuadro clínico de los casos.) xtitle(, size(large) color(navy8)) xlabel(#150, labsz(vsm)) labcolor(dknavy) angle(vertical) title(Covid-19. Gráfica núm. 1c. Curva epidémica de casos\* notificados por fecha de inicio en los Estados Unidos Mexicanos, size(large) color(navy8)) subtitle(desde el inicio de la pandemia hasta el 13 de agosto de 2022 de la semana epidemiológica núm. 32., size(large) color(navy8)) caption(Curva elaborada a partir de los datos abiertos publicados por la Dirección General de Epidemiología el 13 de agosto de 2022 en: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>, size(sm) color(navy8)) note(\* Número de casos por día de acuerdo con las fechas en que iniciaron el cuadro clínico., size(sm) color(dknavy)) legend(order(1 "Casos" 2 "Defunciones") color(navy8)) scale(.45)

Gráficos > Gráficos de doble entrada (dispersión, línea, etc.) \ Gráficos \ Crear \ Gráficos básicos \ Barra



# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
  - Una muestra.
  - Dos muestras.
    - Independientes.
    - Emparejadas.
  - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

## Curva epidémica (muertes).

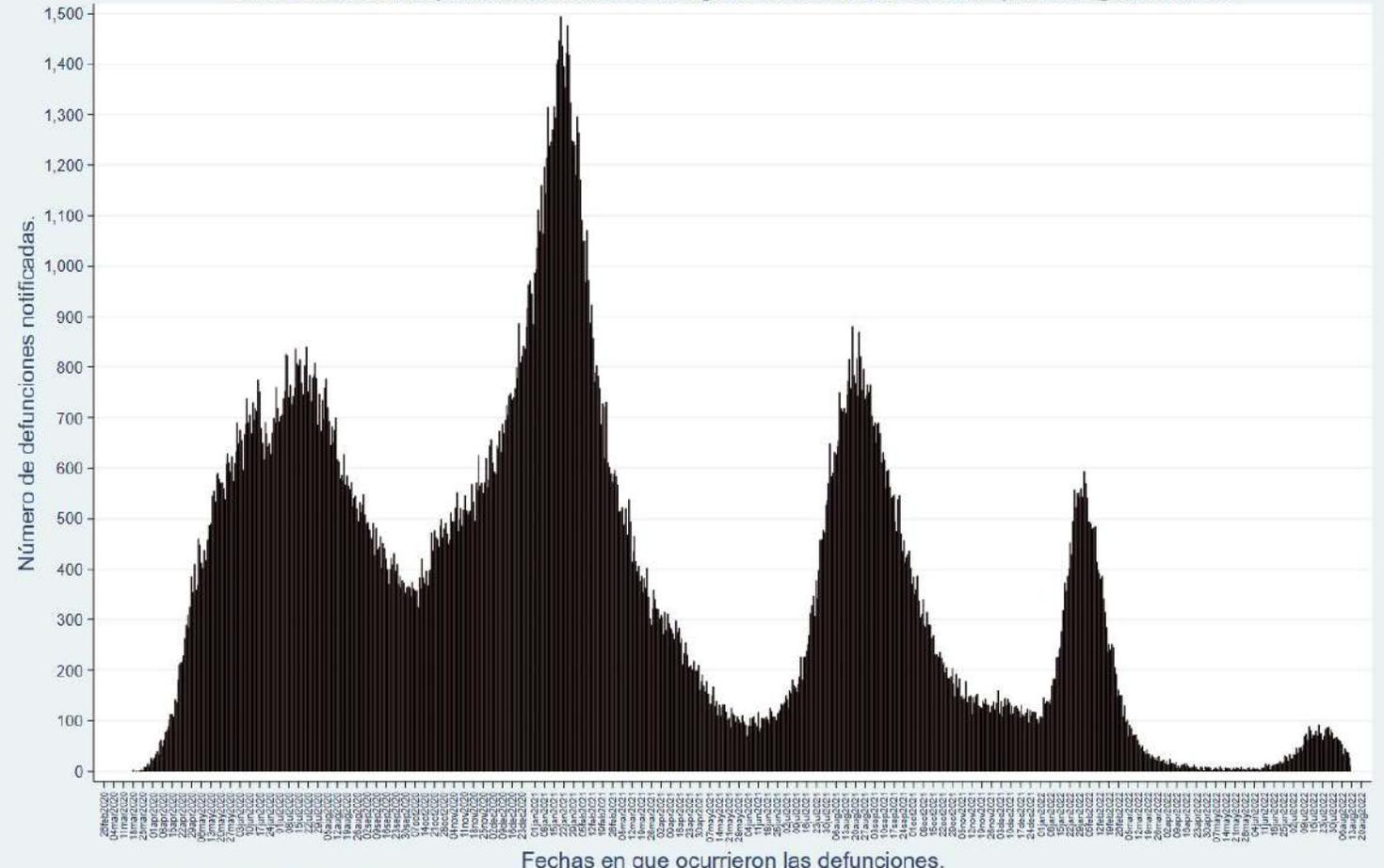
### Número de muertes registradas por día para los años 2020, 2021 y 2022.

```

twoway (histogram muerte if evol==2 &
    errorindef!=1, discrete frequency
    fcolor(maroon%60) lcolor(black) lwidth(vthin)),
    ytitle(Número de defunciones notificadas.) ytitle(
    size(large) color(navy8)) ylabel(#20, labcolor(dknavy),
    angle(horizontal) format(%9.0fc)) ymtick(
    labszsize(small)) xtitle(Fechas en que ocurrieron las
    defunciones.) xtitle(, size(large) color(navy8))
    xlabel(#150, labszsize(vsmall) labcolor(dknavy)
    angle(vertical)) title(Covid-19. Gráfica núm. 1d.
    Curva epidémica de defunciones* por fecha en que
    se registró la muerte en la República Mexicana,
    size(large) color(navy8)) subtitle(desde el inicio de la
    pandemia hasta el 13 de agosto de 2022 de la
    semana epidemiológica núm. 32., size(large)
    color(navy8)) caption(Curva elaborada a partir de los
    datos abiertos publicados por la Dirección General
    de Epidemiología el 13 de agosto de 2022 en:
    https://www.gob.mx/salud/documentos/datos-abie
    rtos-152127, size(small) color(navy8)) note(*
    Número de defunciones por día de acuerdo con las
    fechas en que ocurrieron las muertes., size(small)
    color(dknavy)) legend(order(1 "Casos" 2
    "Defunciones") color(navy8)) scale(.45)
    
```

Gráficos > Gráficos de doble entrada (dispersión, línea, etc.) \ Gráficos \ Crear \ Gráficos básicos \ Barra

Covid-19. Gráfica núm. 1d. Curva epidémica de defunciones\* por fecha en que se registró la muerte en la República Mexicana desde el inicio de la pandemia hasta el 13 de agosto de 2022 de la semana epidemiológica núm. 32.



\* Número de defunciones por día de acuerdo con las fechas en que ocurrieron las muertes.  
 Curva elaborada a partir de los datos abiertos publicados por la Dirección General de Epidemiología el 13 de agosto de 2022 en: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>



# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
  - Una muestra.
  - Dos muestras.
    - Independientes.
    - Emparejadas.
  - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

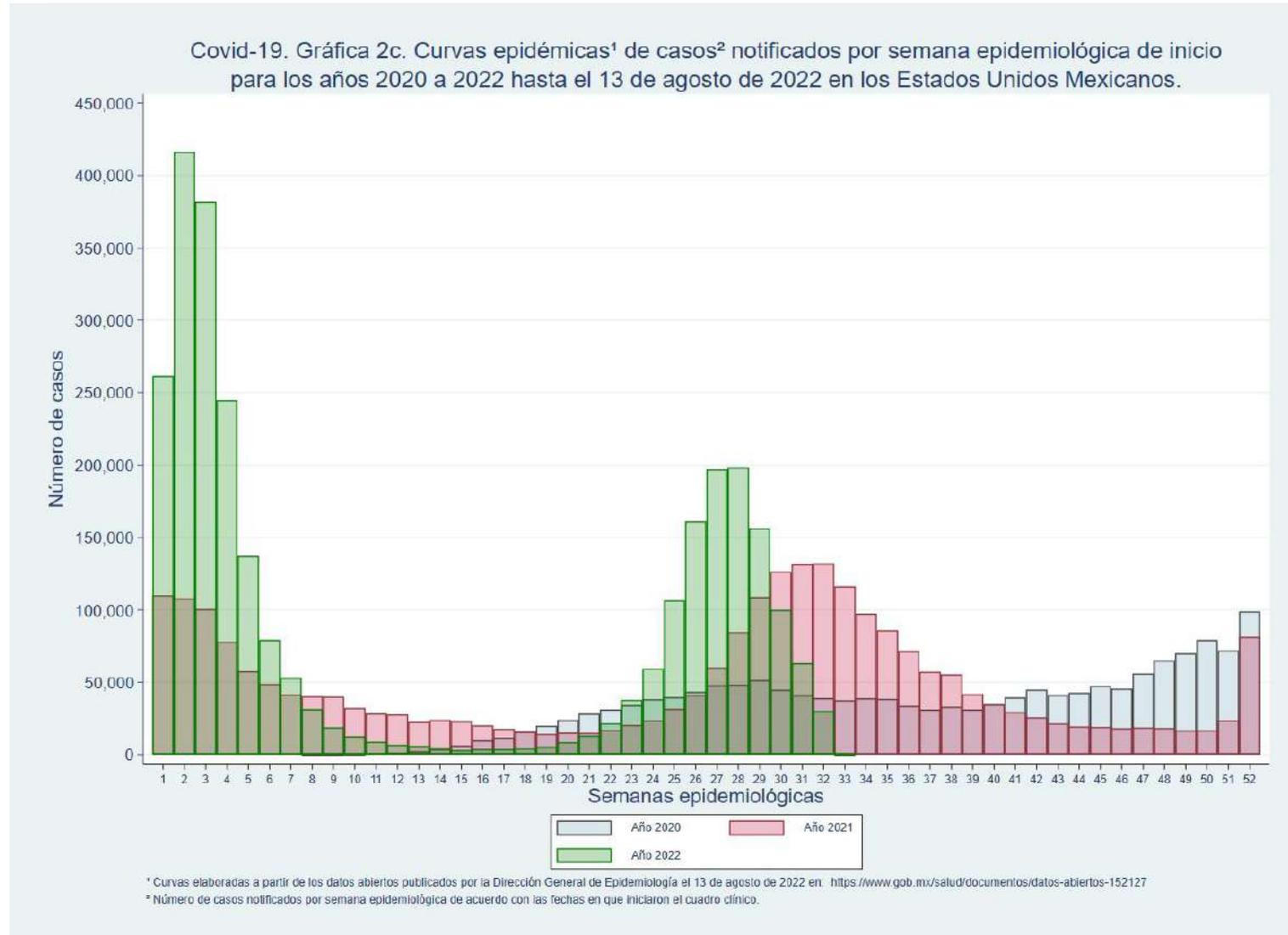
## Curva epidémica (casos).

Número de casos notificados por semana para los años 2020, 2021 y 2022.

```

twoway (histogram semini if añoini==2020 & evol<3,
discrete frequency fcolor(bluishgray%99) lcolor(gs4)
lwidth(medium)) (histogram semini if añoini==2021 &
evol<3, discrete frequency fcolor(cranberry%25)
lcolor(maroon)) (histogram semini if añoini==2022 &
evol<3, discrete frequency fcolor(green%25)
lcolor(green)), ytitle(Número de casos) ytitle(, size(large)
color(navy8)) yscale(lcolor(navy8)) ylabel(#10,
labsize(medium) labcolor(navy8) angle(horizontal)
format(%9.0gc) tcolor(navy8) xtitle(Semanas
epidemiológicas) xtitle(, size(large) color(navy8))
xlabel(#53, labsize(small) labcolor(navy8)) title(Covid-19.
Gráfica 2c. Curvas epidémicas1 de casos2 notificados por
semana epidemiológica de inicio, size(vlarge)
color(navy8)) subtitle(para los años 2020 a 2022 hasta el
13 de agosto de 2022 en los Estados Unidos Mexicanos.,
size(vlarge) color(navy8)) caption(2 Número de casos
notificados por semana epidemiológica de acuerdo con
las fechas en que iniciaron el cuadro clínico., size(small)
color(navy8)) note(1 Curvas elaboradas a partir de los
datos abiertos publicados por la Dirección General de
Epidemiología el 13 de agosto de 2022 en:
https://www.gob.mx/salud/documentos/datos-abiertos-
152127, size(small) color(navy8)) legend(order(1 "Año
2020" 2 "Año 2021" 3 "Año 2022") size(small)
color(navy8)) scale(0.45)
    
```

Gráficos > Gráficos de doble entrada (dispersión, línea, etc.) \ Gráficos \ Crear \ Gráficos básicos \ Barra



# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
  - Una muestra.
  - Dos muestras.
    - Independientes.
    - Emparejadas.
  - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

## Curva epidémica (muertes).

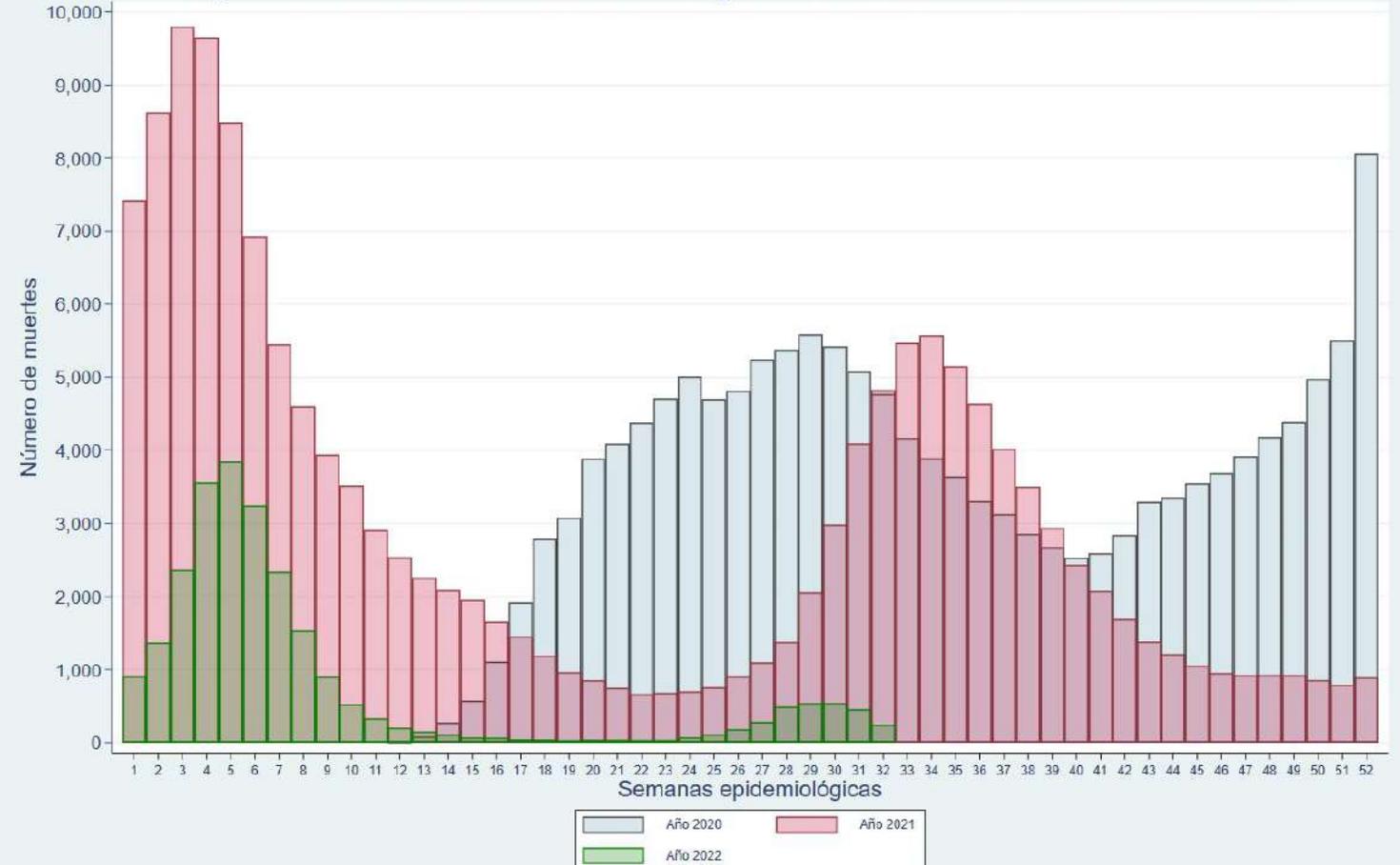
Número de muertes registradas por semana para los años 2020, 2021 y 2022.

```

twoway (histogram semdef if añoodef==2020 & evol==2,
discrete frequency fcolor(bluishgray%99) lcolor(gs4)
lwidth(medium)) (histogram semdef if añoodef==2021 &
evol==2, discrete frequency fcolor(cranberry%25)
lcolor(maroon)) (histogram semdef if añoodef==2022 &
evol==2, discrete frequency fcolor(green%25)
lcolor(green) lwidth(medium)), ytitle(Número de
muertes) ytitle(, size(large) color(navy8))
yscale(lcolor(navy8)) ylabel(#10, labsize(medium)
labcolor(navy8) angle(horizontal) format(%9.0gc)
tlcolor(navy8)) xtitle(Semanas epidemiológicas) xtitle(,
size(large) color(navy8)) xlabel(#53, labsize(small)
labcolor(navy8)) title(Covid-19. Gráfica 2d. Curvas
epidémicas1 de muertes2 registradas por semana
epidemiológica de ocurrencia, size(vlarge) color(navy8))
subtitle( para los años 2020 a 2022 hasta el 13 de agosto
de 2022 en los Estados Unidos Mexicanos., size(vlarge)
color(navy8)) caption(2 Número de muertes registradas
por semana epidemiológica de acuerdo con las fechas en
que ocurrieron las defunciones., size(small) color(navy8))
note(1 Curvas elaboradas a partir de los datos abiertos
publicados por la Dirección General de Epidemiología el
13 de agosto de 2022 en:
https://www.gob.mx/salud/documentos/datos-abiertos-152127,
size(small) color(navy8)) legend(order(1 "Año
2020" 2 "Año 2021" 3 "Año 2022")) size(small)
color(navy8)) scale(0.45)
    
```

Gráficos > Gráficos de doble entrada (dispersión, línea, etc.) \ Gráficos \ Crear \ Gráficos básicos \ Barra

Covid-19. Gráfica 2d. Curvas epidémicas<sup>1</sup> de muertes<sup>2</sup> registradas por semana epidemiológica de ocurrencia para los años 2020 a 2022 hasta el 13 de agosto de 2022 en los Estados Unidos Mexicanos.



<sup>1</sup> Curvas elaboradas a partir de los datos abiertos publicados por la Dirección General de Epidemiología el 13 de agosto de 2022 en: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>  
<sup>2</sup> Número de muertes registradas por semana epidemiológica de acuerdo con las fechas en que ocurrieron las defunciones.



# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

## Curva epidémica (hospitalización).

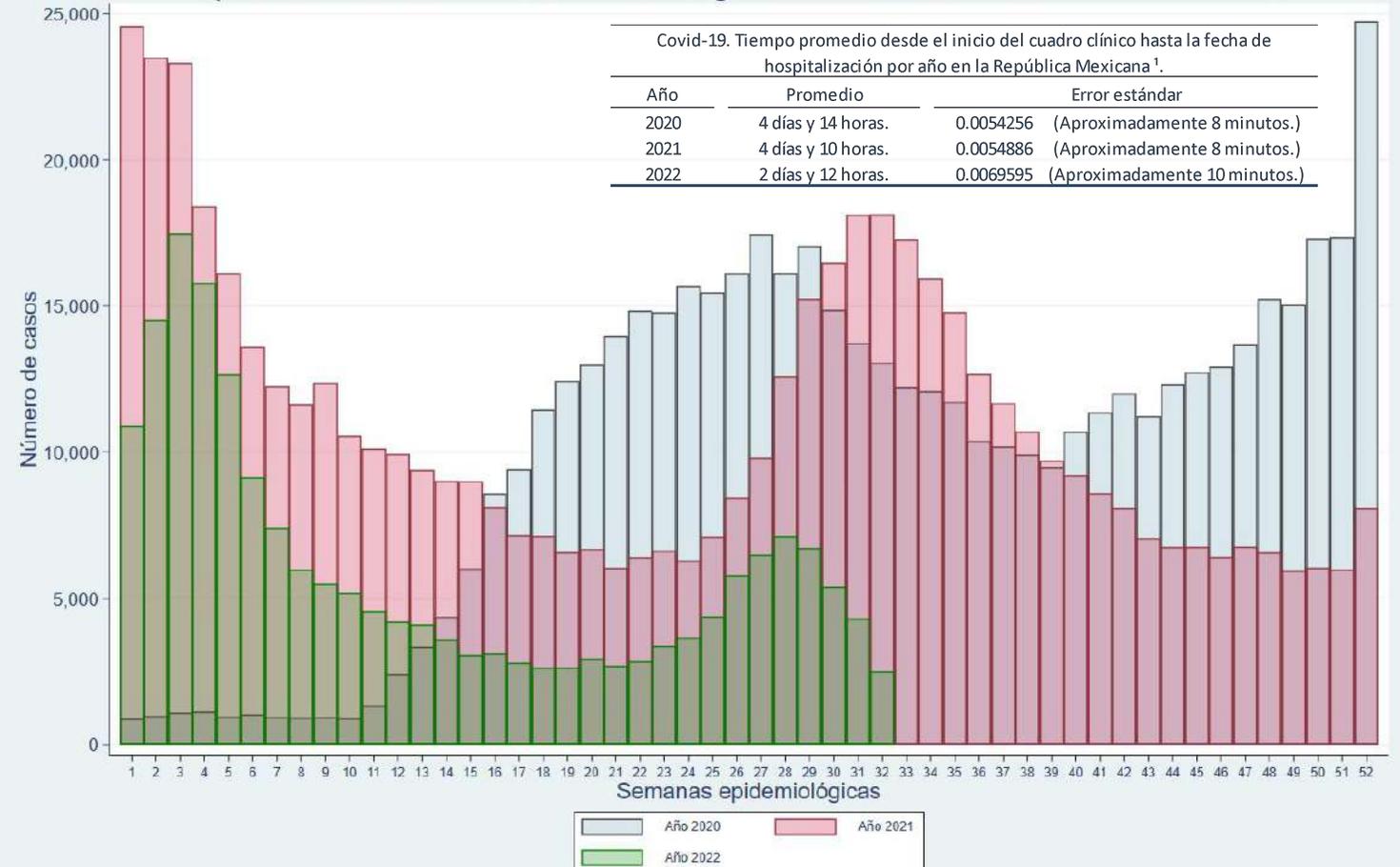
### Número de casos hospitalizados por semana para los años 2020, 2021 y 2022.

```

twoway (histogram semini if añoini==2020 & hosp==1,
discrete frequency fcolor(bluishgray%99) lcolor(gs4)
lwidth(medium)) (histogram semini if añoini==2021 &
hosp==1, discrete frequency fcolor(cranberry%25)
lcolor(maroon)) (histogram semini if añoini==2022 &
hosp==1, discrete frequency fcolor(green%25)
lcolor(green)), ytitle(Número de casos) ytitle(, size(large)
color(navy8)) yscale(lcolor(navy8)) ylabel(#10,
labsize(medium) labcolor(navy8) angle(horizontal)
format(%9.0gc) tcolor(navy8) xtitle(Semanas
epidemiológicas) xtitle(, size(large) color(navy8))
xlabel(#53, labsize(small) labcolor(navy8)) title(Covid-19.
Gráfica 2h. Curvas epidémicas1 de hospitalizaciones2
registradas por semana epidemiológica de inicio,
size(vlarge) color(navy8)) subtitle(para los años 2020 a
2022 hasta el 13 de agosto de 2022 en los Estados Unidos
Mexicanos., size(vlarge) color(navy8)) caption(2 Número de
personas hospitalizadas por semana epidemiológica de
acuerdo con las fechas en que iniciaron un cuadro clínico
sospechoso de Covid-19 independientemente de su
diagnóstico final., size(small) color(navy8)) note(1 Curvas
elaboradas a partir de los datos abiertos publicados por la
Dirección General de Epidemiología el 13 de agosto de
2022 en:
https://www.gob.mx/salud/documentos/datos-abiertos-152127,
size(small) color(navy8)) legend(order(1 "Año 2020"
2 "Año 2021" 3 "Año 2022")) size(small) color(navy8))
scale(0.45)
    
```

Gráficos > Gráficos de doble entrada (dispersión, línea, etc.) \ Gráficos \ Crear \ Gráficos básicos \ Barra

Covid-19. Gráfica 2h. Curvas epidémicas<sup>1</sup> de hospitalizaciones<sup>2</sup> registradas por semana epidemiológica de inicio para los años 2020 a 2022 hasta el 13 de agosto de 2022 en los Estados Unidos Mexicanos.



<sup>1</sup> Curvas elaboradas a partir de los datos abiertos publicados por la Dirección General de Epidemiología el 13 de agosto de 2022 en: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>  
<sup>2</sup> Número de personas hospitalizadas por semana epidemiológica de acuerdo con las fechas en que iniciaron un cuadro clínico sospechoso de Covid-19 independientemente de su diagnóstico final.



CONFERENCIAS STATA LATAM 2022

Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
- Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
    - Una muestra.
    - Dos muestras.
      - Independientes.
      - Emparejadas.
    - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

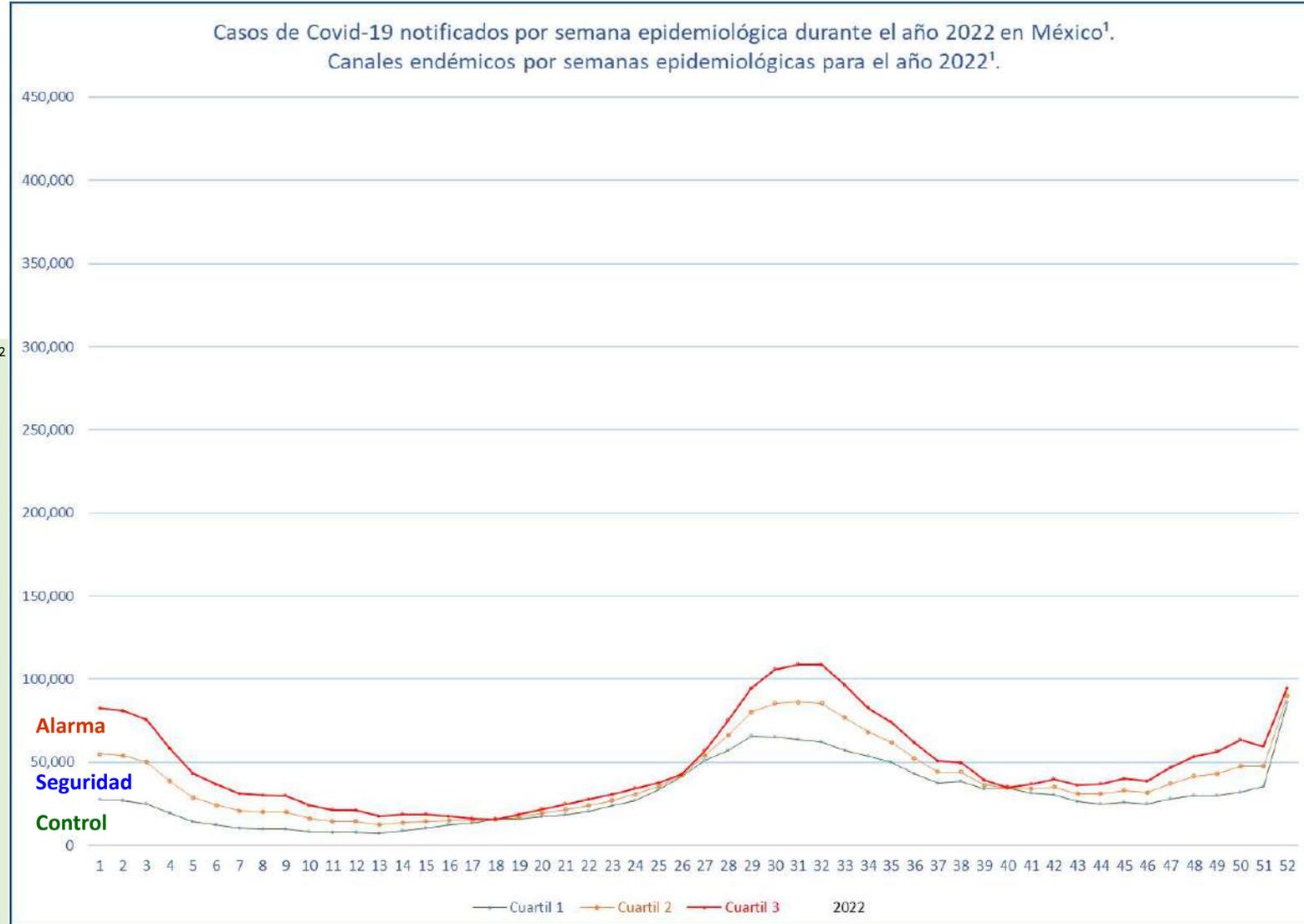
Canales endémicos.

Límite de la zona de alarma. Cuartil 3.

tab semini añoini if clas==1

SE	2020	2021	2022	Q1	Q2	Q3	2022
1	0	110,067		27,516.8	55,033.5		
2	0	108,121		27,030.3	54,060.5		
3	0	100,464		25,116.0	50,232.0		
4	0	77,680		19,420.0	38,840.0		
5	0	57,622		14,405.5	28,811.0		
6	0	48,810		12,202.5	24,405.0		
7	0	41,544		10,386.0	20,772.0		
8	5	40,222		10,059.3	20,113.5		
9	60	39,987		10,041.8	20,023.5		
10	222	32,192		8,214.5	16,207.0		
11	790	28,376		7,686.5	14,583.0		
12	1,329	27,749		7,934.0	14,539.0		
13	2,249	22,661		7,352.0	12,455.0		
14	3,732	24,024		8,805.0	13,878.0		
15	6,264	22,988		10,445.0	14,626.0		
16	9,818	20,201		12,413.8	15,009.5		
17	11,828	17,830		13,328.5	14,829.0		
18	15,881	15,929		15,893.0	15,905.0		
19	19,955	14,304		15,716.8	17,129.5		
20	23,981	15,230		17,417.8	19,605.5		
21	28,225	15,014		18,316.8	21,619.5		
22	31,185	17,089		20,613.0	24,137.0		
23	34,064	20,432		23,840.0	27,248.0		
24	38,240	23,406		27,114.5	30,823.0		
25	39,673	31,614		33,628.8	35,643.5		
26	43,387	41,017		41,609.5	42,202.0		
27	47,975	60,155		51,020.0	54,065.0		
28	48,296	84,638		57,381.5	66,467.0		
29	51,461	108,960		65,835.8	80,210.5		
30	44,763	126,217		65,126.5	85,490.0		
31	40,936	131,421		63,557.3	86,178.5		
32	39,013	131,857		62,224.0	85,435.0		
33	37,492	116,300		57,194.0	76,896.0		

Excel: =CUARTIL.INC(\$scasos por semana en 2020;\$scasos por semana en 2021,3)



Nota: Los valores para los cuartiles se calcularon con Excel porque permite realizar los cálculos con base en la amplitud de la serie de datos.



CONFERENCIAS STATA LATAM 2022

Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

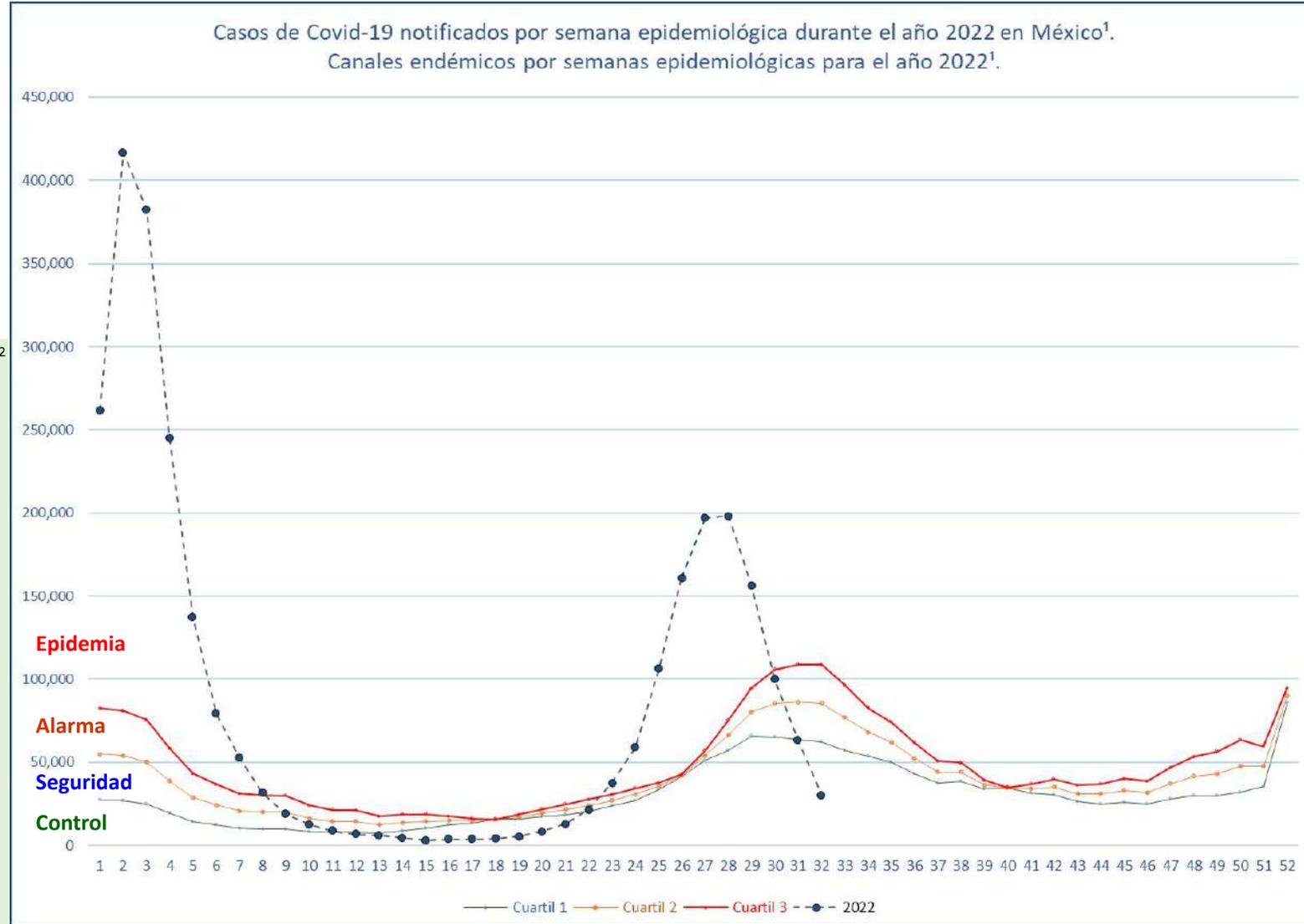
Canales endémicos.

Vigilancia semanal durante el año en curso, en este caso 2022.

```
tab semini añoini if clas==1
```

SE	2020	2021	2022	Q1	Q2	Q3	2022
1	0	110,067		27,516.8	55,033.5	82,550.3	
2	0	108,121		27,030.3	54,060.5	81,090.8	
3	0	100,464		25,116.0	50,232.0	75,348.0	
4	0	77,680		19,420.0	38,840.0	58,260.0	
5	0	57,622		14,405.5	28,811.0	43,216.5	
6	0	48,810		12,202.5	24,405.0	36,607.5	
7	0	41,544		10,386.0	20,772.0	31,158.0	
8	5	40,222		10,059.3	20,113.5	30,167.8	
9	60	39,987		10,041.8	20,023.5	30,005.3	
10	222	32,192		8,214.5	16,207.0	24,199.5	
11	790	28,376		7,686.5	14,583.0	21,479.5	
12	1,329	27,749		7,934.0	14,539.0	21,144.0	
13	2,249	22,661		7,352.0	12,455.0	17,558.0	
14	3,732	24,024		8,805.0	13,878.0	18,951.0	
15	6,264	22,988		10,445.0	14,626.0	18,807.0	
16	9,818	20,201		12,413.8	15,009.5	17,605.3	
17	11,828	17,830		13,328.5	14,829.0	16,329.5	
18	15,881	15,929		15,893.0	15,905.0	15,917.0	
19	19,955	14,304		15,716.8	17,129.5	18,542.3	
20	23,981	15,230		17,417.8	19,605.5	21,793.3	
21	28,225	15,014		18,316.8	21,619.5	24,922.3	
22	31,185	17,089		20,613.0	24,137.0	27,661.0	
23	34,064	20,432		23,840.0	27,248.0	30,656.0	
24	38,240	23,406		27,114.5	30,823.0	34,531.5	
25	39,673	31,614		33,628.8	35,643.5	37,658.3	
26	43,387	41,017		41,609.5	42,202.0	42,794.5	
27	47,975	60,155		51,020.0	54,065.0	57,110.0	
28	48,296	84,638		57,381.5	66,467.0	75,552.5	
29	51,461	108,960		65,835.8	80,210.5	94,585.3	
30	44,763	126,217		65,126.5	85,490.0	105,853.5	
31	40,936	131,421		63,557.3	86,178.5	108,799.8	
32	39,013	131,857		62,224.0	85,435.0	108,646.0	
33	37,492	116,300		57,194.0	76,896.0	96,598.0	

Excel: Copiar los datos semanales que produce la orden de Stata



Nota: Los valores para los cuartiles se calcularon con Excel porque permite realizar los cálculos con base en la amplitud de la serie de datos.



# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
    - Una muestra.
    - Dos muestras.
      - Independientes.
      - Emparejadas.
    - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

## Curva epidémica (casos).

Casos notificados por día durante las últimas cuatro semanas.

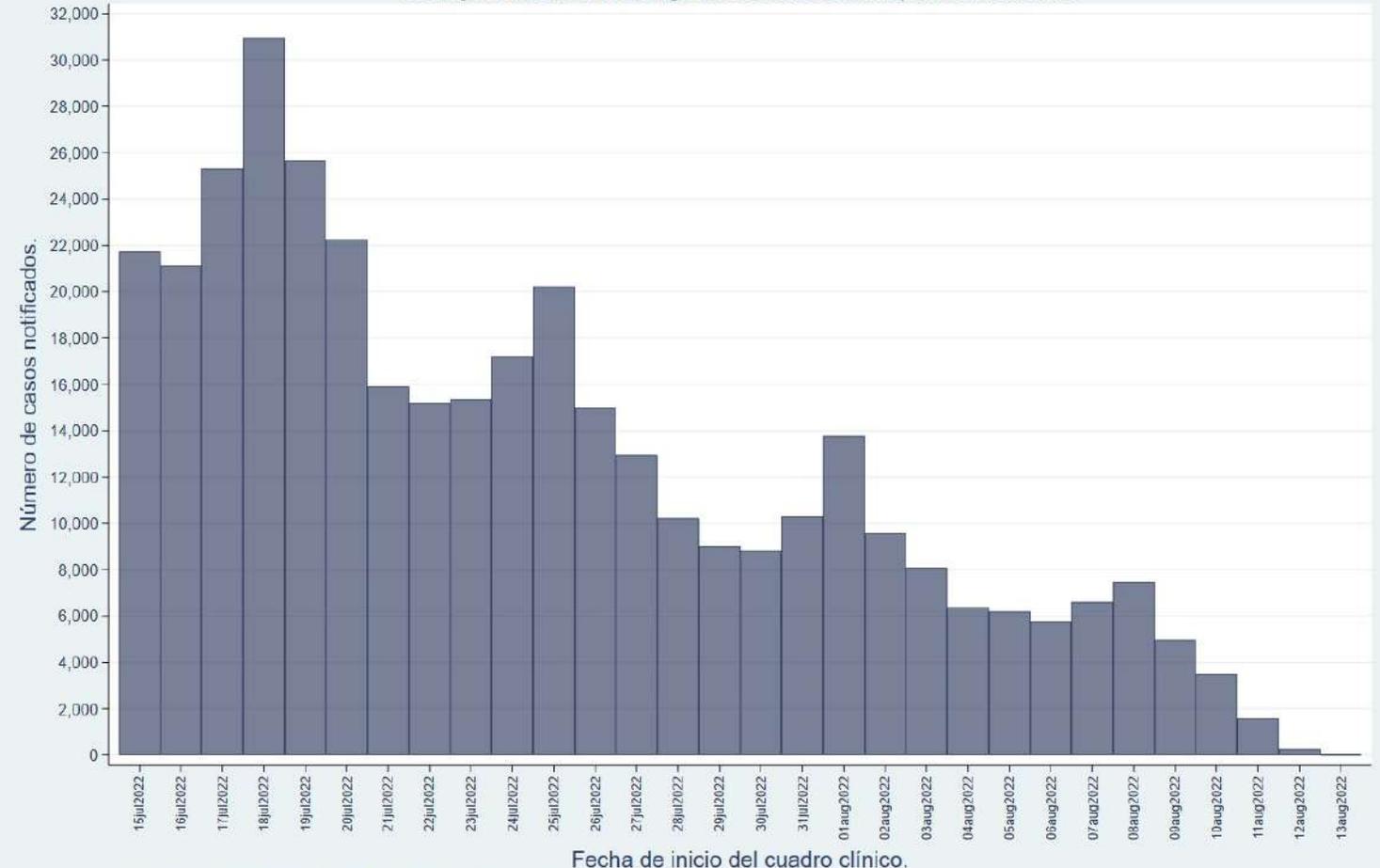
use DA 13 de agosto.dta

```

twoway (histogram inicio if evol < 3 & diapandemia
>= 921, discrete frequency fcolor(dknavy%60)
lcolor(dknavy) lwidth(vthin)), ytitle(Número de casos
notificados.) ytitle(, size(large) color(navy8))
ylabel(#20, labcolor(dknavy) angle(horizontal)
format(%9.0fc)) ymtick(, labszsize(small)) xtitle(Fecha
de inicio del cuadro clínico.) xtitle(, size(large)
color(navy8)) xlabel(#30, labszsize(small)
labcolor(dknavy) angle(vertical)) title(Covid-19.
Gráfica núm. 3c. Curva epidémica de casos*
notificados por fecha de inicio del cuadro clínico
desde el, size(large) color(navy8)) subtitle(15 de julio
hasta al 13 de agosto de 2022 en la República
Mexicana., size(large) color(navy8)) caption(Curva
elaborada a partir de los datos abiertos publicados
por la Dirección General de Epidemiología el 13 de
agosto de 2022 en:
https://www.gob.mx/salud/documentos/datos-abie
rtos-152127, size(small) color(navy8)) note(*
Número de casos por día de acuerdo con las fechas
en que iniciaron el cuadro clínico., size(small)
color(dknavy)) scale(.45)
    
```

Gráficos > Gráficos de doble entrada (dispersión, línea, etc.) \ Gráficos \ Crear \ Gráficos avanzados \ Histograma

Covid-19. Gráfica núm. 3c. Curva epidémica de casos\* notificados por fecha de inicio del cuadro clínico desde el 15 de julio hasta al 13 de agosto de 2022 en la República Mexicana.



\* Número de casos por día de acuerdo con las fechas en que iniciaron el cuadro clínico.  
 Curva elaborada a partir de los datos abiertos publicados por la Dirección General de Epidemiología el 13 de agosto de 2022 en: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>



# Herramientas y aplicaciones estadísticas para Ciencia de Datos

Introducción.  
 Panorama epidemiológico.  
 Vigilancia epidemiológica.  
 Datos y calidad.  
 Análisis exploratorio.  
 Magnitud.  
 Tiempo.  
 Lugar.  
 Variables personales.  
 Condiciones de riesgo.  
 Evidencia de causalidad.  
 Ejemplo: El lugar.  
 Otras condiciones.  
 Análisis de la evidencia.  
 Comparación de medias.  
 Estimación.  
 Una muestra.  
 Dos muestras.  
 Independientes.  
 Emparejadas.  
 Más de dos muestras.  
 Medidas de intervención.  
 Intervenciones inespecíficas  
 Intervenciones específicas.  
 Conclusiones.  
 Preguntas.

## Curva epidémica (muertes).

Muertes registradas por día durante las últimas cuatro semanas.

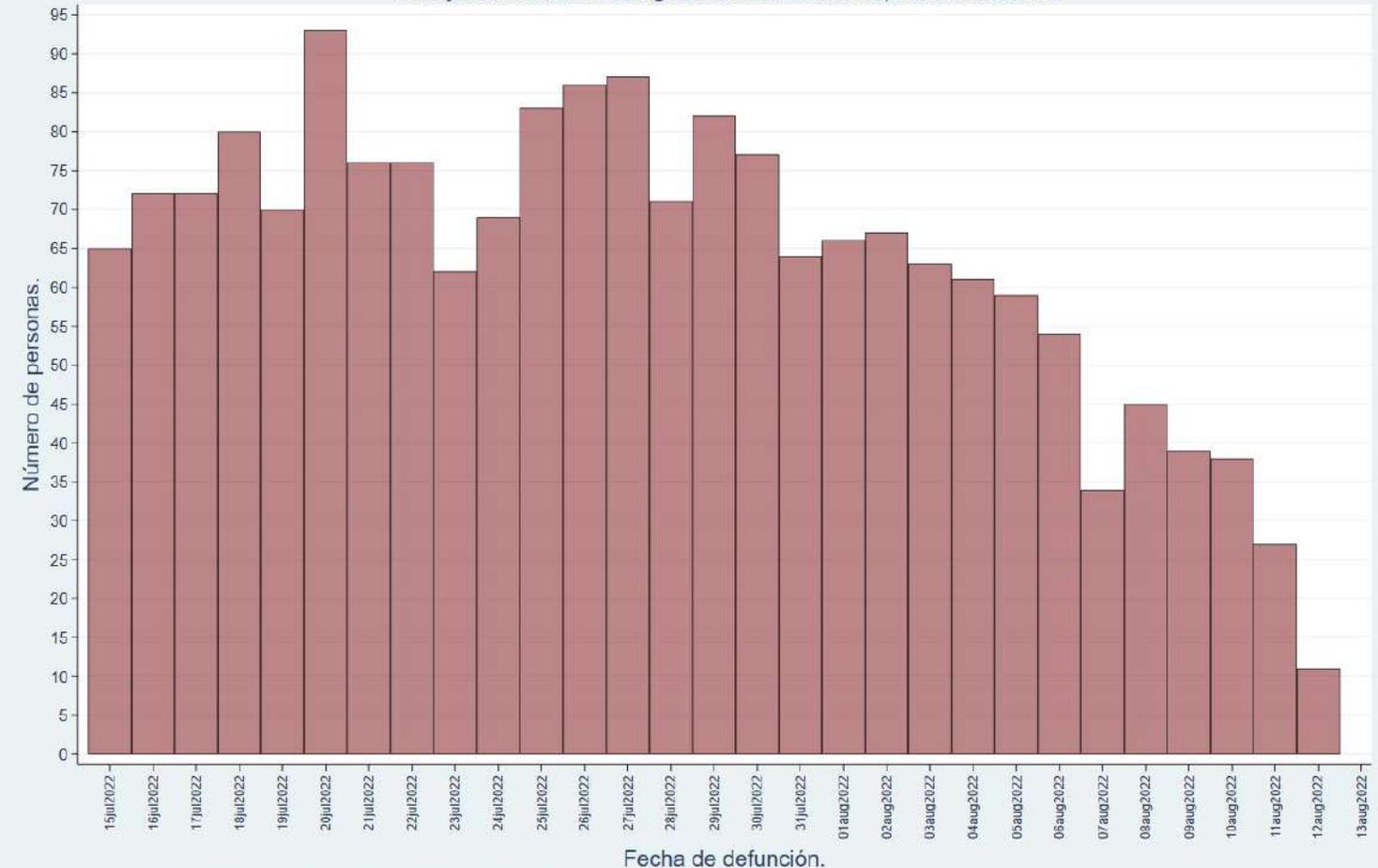
use DA 13 de agosto.dta

```
twoway (histogram muerte if evol == 2 &
defpandemia >= 921 & erroringdef!=1, discrete
frequency fcolor(maroon%60) lcolor(black)
lwidth(vthin)), ytitle(Número de personas.) ytitle(
size(large) color(navy8)) ylabel(#20, labcolor(dknavy)
angle(horizontal) ) ymtick(, labsize(small))
xtitle(Fecha de defunción.) xtitle(, size(large)
color(navy8)) xlabel(#30, labsize(small)
labcolor(dknavy) angle(vertical)) title(Covid-19.
Gráfica núm. 3d. Curva epidémica de muertes*
notificadas por fecha de defunción desde el
size(large) color(navy8)) subtitle(15 de julio hasta al
size(large) color(navy8)) caption(Curva elaborada a
partir de los datos abiertos publicados por la
Dirección General de Epidemiología el 13 de agosto
de 2022 en:
```

```
https://www.gob.mx/salud/documentos/datos-abie
rtos-152127, size(small) color(navy8)) note(*
Número de defunciones por día de acuerdo con las
fechas en que ocurrieron las muertes., size(small)
color(dknavy)) scale(.45)
```

Gráficos > Gráficos de doble entrada (dispersión, línea, etc.) \ Gráficos \ Crear \ Gráficos avanzados \ Histograma

Covid-19. Gráfica núm. 3d. Curva epidémica de muertes\* notificadas por fecha de defunción desde el 15 de julio hasta al 13 de agosto de 2022 en la República Mexicana.



\* Número de defunciones por día de acuerdo con las fechas en que ocurrieron las muertes.

Curva elaborada a partir de los datos abiertos publicados por la Dirección General de Epidemiología el 13 de agosto de 2022 en: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
    - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
      - Una muestra.
      - Dos muestras.
        - Independientes.
        - Emparejadas.
      - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas.
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

# Vigilancia epidemiológica.

Panorama epidemiológico.

Análisis exploratorio.

La variable lugar.

CONFERENCIAS STATA LATAM 2022

Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

Distribución espacial (casos y muertes).

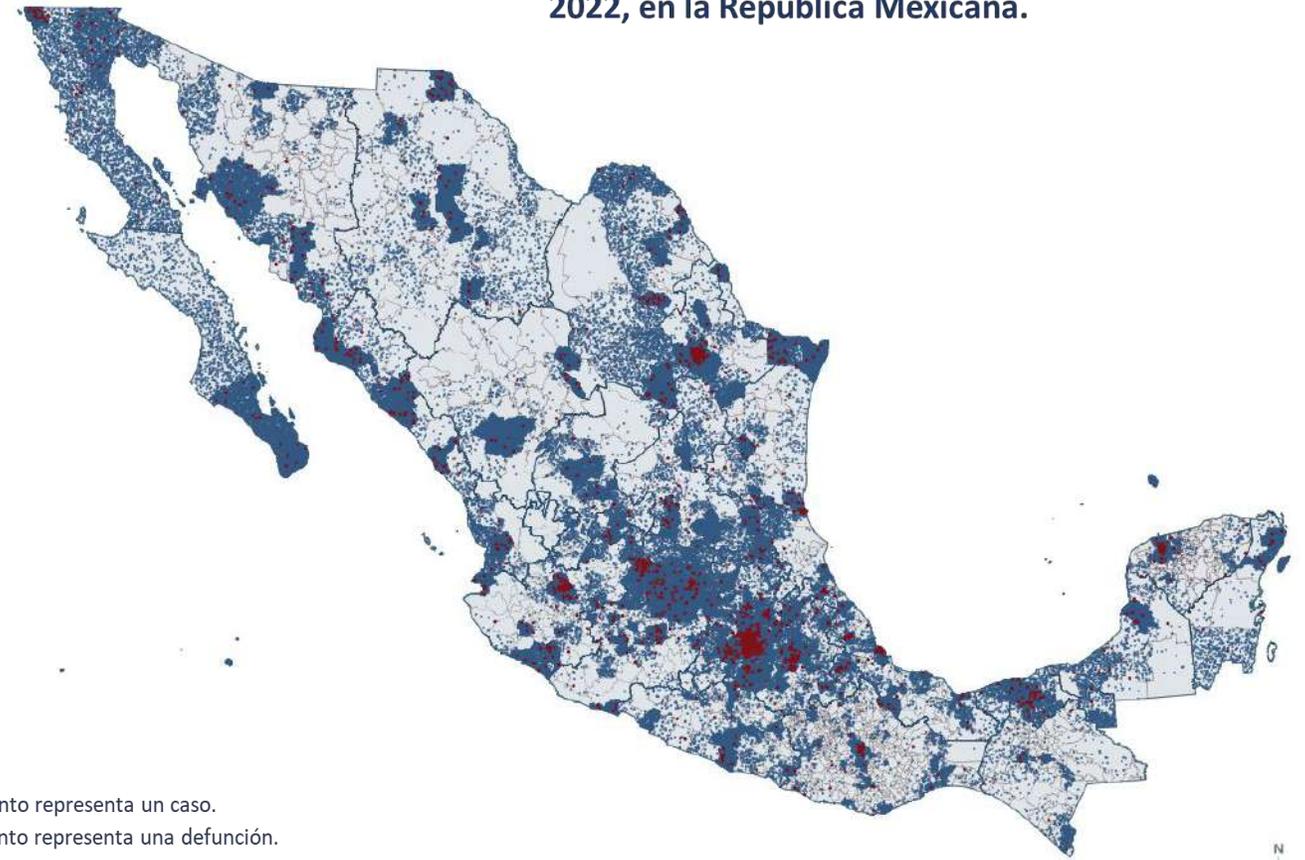
Casos y defunciones por municipio desde el 15 de julio hasta el 13 de agosto de 2022 en México.

use DA 13 de agosto.dta

table (objectid) (evol) if evol < 3 & diapandemia >= 921

Código geográfico de municipio	Evolución		Total
	Caso vive	Caso muere	
1	6,281	11	6,292
2	3,940	4	3,944
3	2,163	2	2,165
4	4,837	12	4,849
5	3,438	11	3,449
6	4,108	13	4,121
7	7,341	11	7,352
8	10,312	19	10,331
9	904	3	907
10	2,985	4	2,989
11	3,466	9	3,475
12	5,643	11	5,654
13	3,244	5	3,249
14	3,124	1	3,125
15	10,649	19	10,668
16	1,683	1	1,684
17	1		1
18	9		9
19	4		4
20	175		175
21	1,974	2	1,976
23	14	1	15
24	38		38
25	8		8

Covid-19. Mapa núm. 3. Casos notificados y defunciones registradas durante las últimas 4 semanas, desde el 15 de julio hasta el 13 de agosto de 2022, en la República Mexicana.



- Un punto representa un caso.
- Un punto representa una defunción.

Elaborado a partir de los datos abiertos publicados por la Dirección General de Epidemiología el 13 de agosto de 2022.

<https://www.gob.mx/salud/documentos/datos-abiertos-152127>

Nota: Los puntos están distribuidos aleatoriamente dentro del polígono que representa el área de cada municipio.

Nota: Para generar el mapa se utilizó un Sistema de Información Geográfica, en este caso QGIS versión 3.26.2 "Buenos Aires".



**CONFERENCIAS STATA LATAM 2022**

**Herramientas y aplicaciones estadísticas para Ciencia de Datos**

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
  - Una muestra.
  - Dos muestras.
    - Independientes.
    - Emparejadas.
  - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

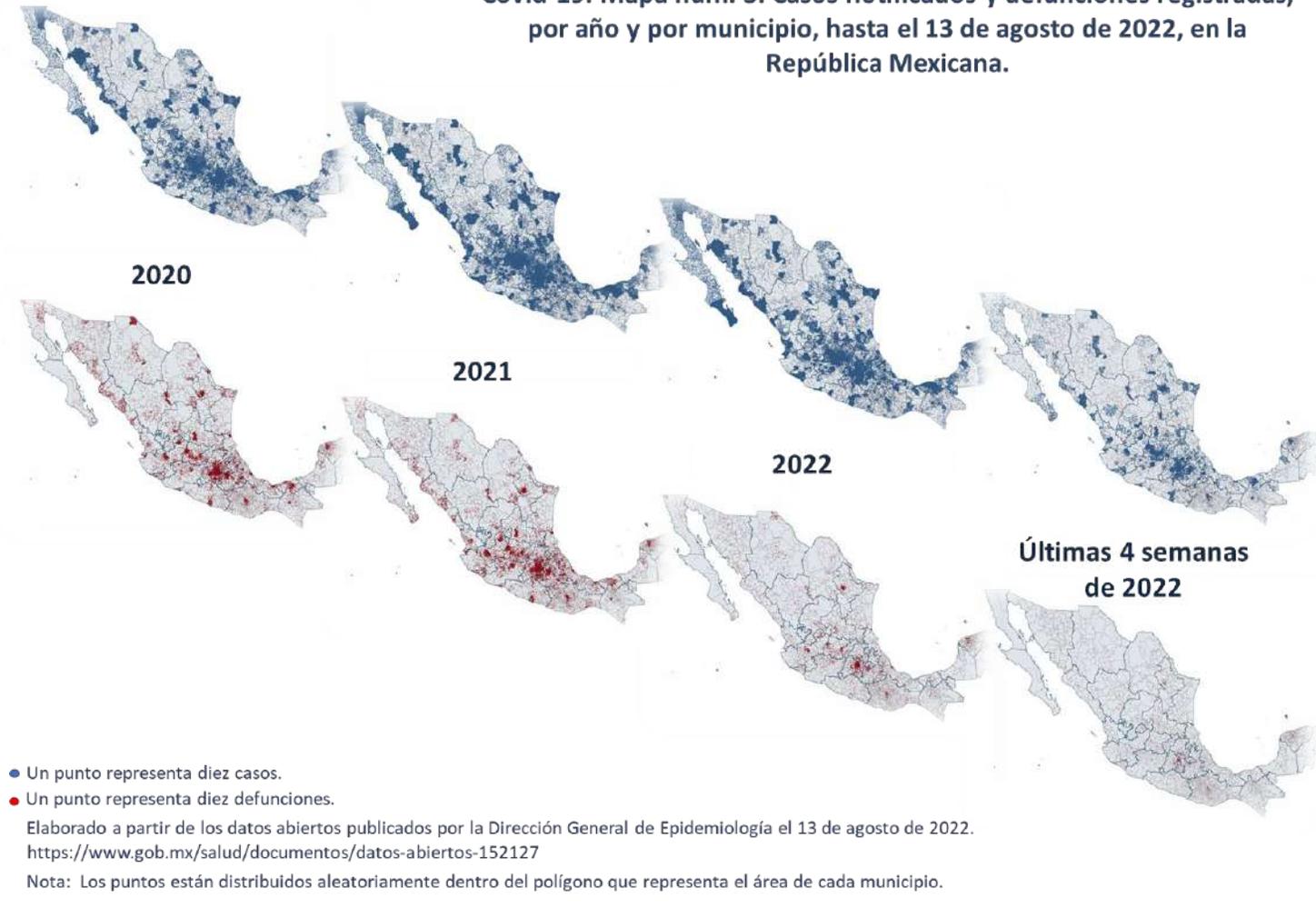
**Distribución temporal y espacial (casos y muertes).**

**Casos y defunciones por municipio para diferentes periodos de evolución de la pandemia de Covid-19 en México.**

```
table (objectid) (evol) (añoini) if evol > 0 & evol < 3 & objectid<50
```

añoini = 2020			
	Caso vive	Evolución Caso muere	Total
Código geográfico de municipio			
1	34,607	1,227	35,834
2	19,793	834	20,627
3	12,154	509	12,663
añoini = 2021			
	Caso vive	Evolución Caso muere	Total
Código geográfico de municipio			
1	67,211	1,034	68,245
2	32,884	647	33,531
3	16,580	326	16,906
añoini = 2022			
	Caso vive	Evolución Caso muere	Total
Código geográfico de municipio			
1	59,200	146	59,346
2	36,414	76	36,490
3	18,170	63	18,233

**Covid-19. Mapa núm. 3. Casos notificados y defunciones registradas, por año y por municipio, hasta el 13 de agosto de 2022, en la República Mexicana.**



- Un punto representa diez casos.
- Un punto representa diez defunciones.

Elaborado a partir de los datos abiertos publicados por la Dirección General de Epidemiología el 13 de agosto de 2022.  
<https://www.gob.mx/salud/documentos/datos-abiertos-152127>

Nota: Los puntos están distribuidos aleatoriamente dentro del polígono que representa el área de cada municipio.

Nota: Para generar el mapa se utilizó un Sistema de Información Geográfica, en este caso QGIS versión 3.26.2 "Buenos Aires".



CONFERENCIAS STATA LATAM 2022

Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
      - Una muestra.
      - Dos muestras.
        - Independientes.
        - Emparejadas.
      - Más de dos muestras.
  - Medidas de intervención.
    - Intervenciones inespecíficas
    - Intervenciones específicas.
  - Conclusiones.
  - Preguntas.

Casos y defunciones por entidad federativa.

Población afectada por el daño a la salud ocasionado por una agente patógeno, que ocasiona enfermedad en esa población durante un periodo.

use DE 13 de agosto.dta  
table (entidad\_res) (evol) if clas == 1

Estado de residencia	Evolución		Total
	Caso vive	Caso muere	
Aguascalientes	78,488	3,536	82,024
Baja California	145,420	12,381	157,801
Baja California Sur	119,786	2,725	122,511
Campeche	40,196	2,274	42,470
Coahuila	168,470	8,915	177,385
Colima	64,404	2,232	66,636
Chiapas	41,310	2,428	43,738
Chihuahua	149,610	10,084	159,694
Ciudad de México	1,661,833	43,390	1,705,223
Durango	74,619	3,650	78,278
Guerrero	3,540,685	114,476	3,655,161
Hidalgo	3,082,841	114,648	3,197,489
Jalisco	8,348,151	277,705	8,625,856
Estado de México	16,992,418	686,678	17,679,096
Michoacán	4,748,846	107,803	4,856,649
Morelos	1,971,520	85,351	2,056,871
Nayarit	1,235,456	71,136	1,306,592
Nuevo León	5,784,442	390,447	6,174,889
Oaxaca	4,132,148	144,546	4,276,694
Puebla	6,583,278	203,911	6,787,189
Querétaro	2,368,467	174,032	2,542,499
Quintana Roo	1,857,985	111,112	1,969,097
San Luis Potosí	2,822,255	229,859	3,052,114
Sinaloa	3,026,943	171,086	3,198,029
Sonora	2,944,840	190,815	3,135,655
Tabasco	2,402,598	212,421	2,615,019
Tamaulipas	3,527,735	172,679	3,700,414
Tlaxcala	1,342,977	55,310	1,398,287
Veracruz	8,062,579	219,179	8,281,758
Yucatán	2,320,898	135,413	2,456,311
Zacatecas	1,622,138	77,535	1,700,673
<b>Nacional</b>	<b>126,014,024</b>	<b>6,925,668</b>	<b>132,939,692</b>

Estadísticas > Sumarios, tablas y tests estadísticos > Tablas de frecuencias, sumario estadístico y resultados de comandos \ Principal \ if/in

Covid-19. Cuadro núm 2\*. Morbilidad. Incidencia acumulada de casos por Entidad Federativa desde el inicio de la pandemia al 13 de agosto de 2022, de acuerdo con los datos publicados el por la Secretaría de Salud, en los Estados Unidos Mexicanos.

Entidad Federativa	Población en riesgo y casos notificados por estado.		Incidencia acumulada <sup>1</sup> con intervalos de confianza de 95%		
	Población en riesgo <sup>4</sup>	Casos notificados	Incidencia acumulada <sup>1</sup>	Límite inferior <sup>2</sup>	Límite superior <sup>3</sup>
Aguascalientes	1,425,607	82,024			
Baja California	3,769,020	157,801			
Baja California Sur	798,447	122,511			
Campeche	928,363	42,470			
Coahuila	3,146,771	177,385			
Colima	731,391	66,636			
Chiapas	5,543,828	43,738			
Chihuahua	3,741,869	159,694			
Ciudad de México	9,209,944	1,705,223			
Durango	1,832,650	78,278			
Guanajuato	6,166,934	343,766			
Guerrero	3,540,685	114,476			
Hidalgo	3,082,841	114,648			
Jalisco	8,348,151	277,705			
Estado de México	16,992,418	686,678			
Michoacán	4,748,846	107,803			
Morelos	1,971,520	85,351			
Nayarit	1,235,456	71,136			
Nuevo León	5,784,442	390,447			
Oaxaca	4,132,148	144,546			
Puebla	6,583,278	203,911			
Querétaro	2,368,467	174,032			
Quintana Roo	1,857,985	111,112			
San Luis Potosí	2,822,255	229,859			
Sinaloa	3,026,943	171,086			
Sonora	2,944,840	190,815			
Tabasco	2,402,598	212,421			
Tamaulipas	3,527,735	172,679			
Tlaxcala	1,342,977	55,310			
Veracruz	8,062,579	219,179			
Yucatán	2,320,898	135,413			
Zacatecas	1,622,138	77,535			
<b>Nacional</b>	<b>126,014,024</b>	<b>6,925,668</b>			

\* Elaborado a partir de los datos abiertos de la Dirección General de Epidemiología (DGE/SSA) publicados el 13 de agosto de 2022 en: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>  
<sup>1</sup> Casos residentes de los estados por cada mil habitantes. <sup>2</sup> Límite inferior de la incidencia acumulada calculado con 95% de confianza. <sup>3</sup> Límite superior de la incidencia acumulada calculado con 95% de confianza.  
<sup>4</sup> Fuente: INEGI. Censo de Población y Vivienda 2020. Consultada en el sitio: <https://www.inegi.org.mx/sistemas/01ap/Proyectos/bd/censos/cv2020/pt.asp>



# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
- Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
    - Una muestra.
    - Dos muestras.
      - Independientes.
      - Emparejadas.
    - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

## Morbilidad por entidad federativa.

Magnitud relativa del daño a la salud ocasionado por una agente patógeno, que ocasiona enfermedad en una población durante un periodo.

display "Aguascalientes"  
cii proportions 1425607 82024

```

Aguascalientes
.
. cii proportions 1425607 82024
    
```

Variable	Obs	Proportion	Std. err.	Binomial exact [95% conf. interval]	
	1,425,607	.0575362	.000195	.0571545	.0579197

```
. display %10.1f r(proportion) * 1000
57.5
```

```
. display %10.1f r(lb) * 1000
57.2
```

```
. display %10.1f r(ub) * 1000
57.9
```

## Órdenes de ejecución inmediata.

**Covid-19. Cuadro núm 2\*. Morbilidad. Incidencia acumulada de casos por Entidad Federativa desde el inicio de la pandemia al 13 de agosto de 2022, de acuerdo con los datos publicados el por la Secretaría de Salud, en los Estados Unidos Mexicanos.**

Entidad Federativa	Población en riesgo y casos notificados por estado.		Incidencia acumulada <sup>1</sup> con intervalos de confianza de 95%		
	Población en riesgo <sup>4</sup>	Casos notificados	Incidencia acumulada <sup>1</sup>	Límite inferior <sup>2</sup>	Límite superior <sup>3</sup>
Aguas calientes	1,425,607	82,024	57.5	57.2	57.9
Baja California	3,769,020	157,801	41.9	41.7	42.1
Baja California Sur	798,447	122,511	153.4	152.6	154.2
Campeche	928,363	42,470	45.7	45.3	46.2
Coahuila	3,146,771	177,385	56.4	56.1	56.6
Colima	731,391	66,636	91.1	90.5	91.8
Chiapas	5,543,828	43,738	7.9	7.8	8.0
Chihuahua	3,741,869	159,694	42.7	42.5	42.9
Ciudad de México	9,209,944	1,705,223	185.2	184.9	185.4
Durango	1,832,650	78,278	42.7	42.4	43.0
Guanajuato	6,166,934	343,766	55.7	55.6	55.9
Guerrero	3,540,685	114,476	32.3	32.1	32.5
Hidalgo	3,082,841	114,648	37.2	37.0	37.4
Jalisco	8,348,151	277,705	33.3	33.1	33.4
Estado de México	16,992,418	686,678	40.4	40.3	40.5
Michoacán	4,748,846	107,803	22.7	22.6	22.8
Morelos	1,971,520	85,351	43.3	43.0	43.6
Nayarit	1,235,456	71,136	57.6	57.2	58.0
Nuevo León	5,784,442	390,447	67.5	67.3	67.7
Oaxaca	4,132,148	144,546	35.0	34.8	35.2
Puebla	6,583,278	203,911	31.0	30.8	31.1
Querétaro	2,368,467	174,032	73.5	73.1	73.8
Quintana Roo	1,857,985	111,112	59.8	59.5	60.1
San Luis Potosí	2,822,255	229,859	81.4	81.1	81.8
Sinaloa	3,026,943	171,086	56.5	56.3	56.8
Sonora	2,944,840	190,815	64.8	64.5	65.1
Tabasco	2,402,598	212,421	88.4	88.1	88.8
Tamaulipas	3,527,735	172,679	48.9	48.7	49.2
Tlaxcala	1,342,977	55,310	41.2	40.8	41.5
Veracruz	8,062,579	219,179	27.2	27.1	27.3
Yucatán	2,320,898	135,413	58.3	58.0	58.6
Zacatecas	1,622,138	77,535	47.8	47.5	48.1
<b>Nacional</b>	<b>126,014,024</b>	<b>6,925,668</b>	<b>55.0</b>	<b>54.9</b>	<b>55.0</b>

\* Elaborado a partir de los datos abiertos de la Dirección General de Epidemiología (DGE/SSA) publicados el 13 de agosto de 2022 en: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>  
<sup>1</sup> Casos residentes de los estados por cada mil habitantes. <sup>2</sup> Límite inferior de la incidencia acumulada calculado con 95% de confianza. <sup>3</sup> Límite superior de la incidencia acumulada calculado con 95% de confianza.  
<sup>4</sup> Fuente: INEGI. Censo de Población y Vivienda 2020. Consultada en el sitio: <https://www.inegi.org.mx/sistemas/01ap/Proyectos/bd/censos/cv2020/pt.asp>



# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
  - Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
  - Una muestra.
  - Dos muestras.
    - Independientes.
    - Emparejadas.
  - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

## Morbilidad por entidad federativa.

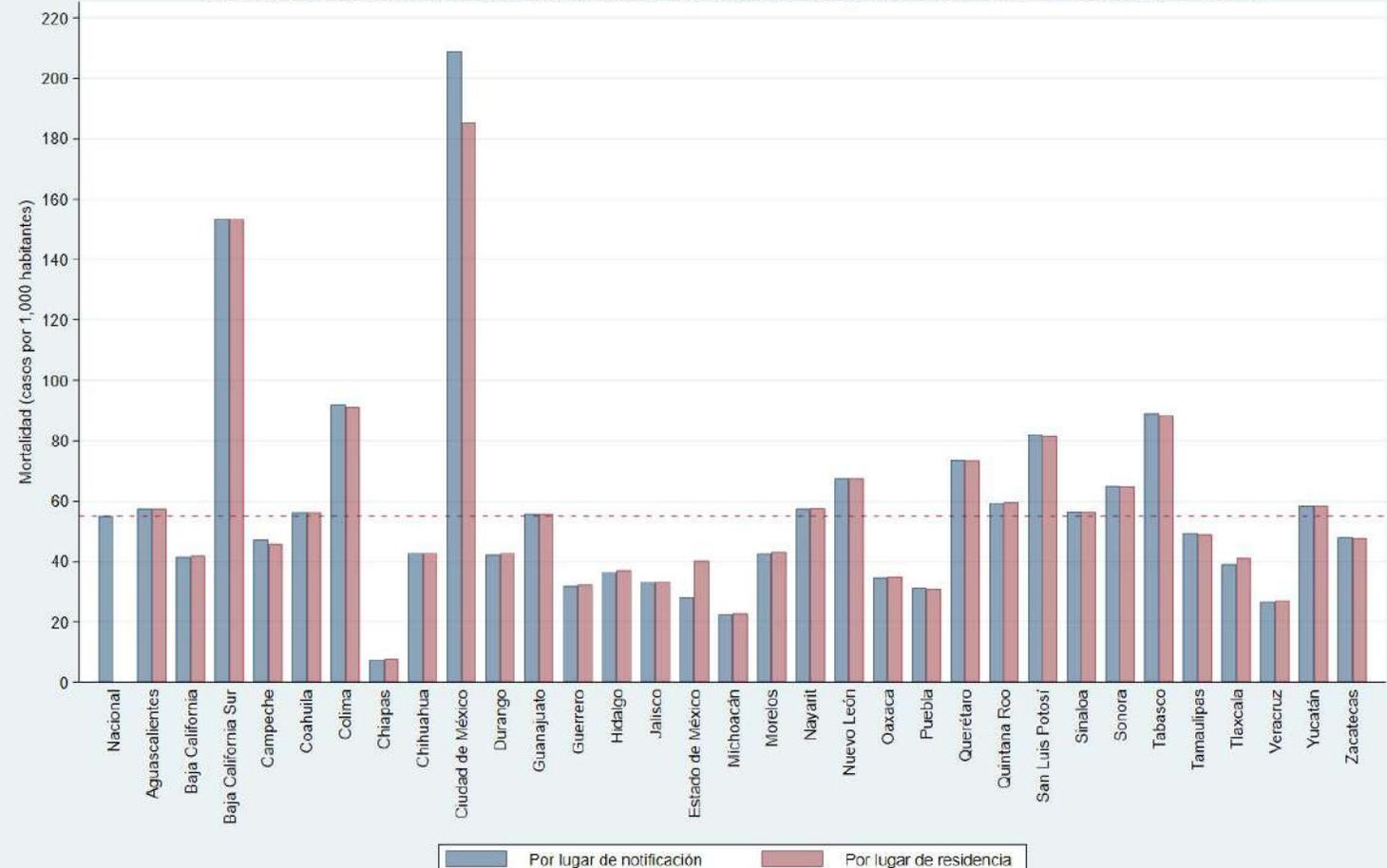
Magnitud relativa del daño a la salud ocasionado por una agente patógeno, que ocasiona enfermedad en una población durante un periodo.

use letalidad

```
graph bar (sum) morbocu (sum) morbres, over(edo, label(angle(vertical))) bar(1, fcolor(%50) lwidth(vthin)) bar(2, fcolor(%50) lwidth(vthin)) ytitle(Mortalidad (casos por 1,000 habitantes)) yline(54.96, lwidth(thin) lpattern(vshortdash)) ylabel(0(20)220, angle(horizontal)) title(Covid-19. Gráfica núm 3a. Morbilidad específica* por Entidad Federativa de ocurrencia y de residencia de los) subtitle(casos notificados en la República Mexicana desde el inicio de la pandemia hasta el 13 de agosto de 2022., size(large) color(dknavy)) caption(Mortalidad estimada a partir de los datos abiertos publicados por la Dirección General de Epidemiología el 13 de agosto de 2022 en el sitio: https://www.gob.mx/salud/documentos/datos-abiertos-152127, size(small)) note(* Número de casos por cada mil habitantes., size(small)) legend(order(1 "Por lugar de ocurrencia" 2 "Por lugar de residencia")) scale(.45)
```

Gráficos > Gráficos de barra \Principal \ Categoría

Covid-19. Gráfica núm 3a. Morbilidad específica\* por Entidad Federativa de notificación y de residencia de los casos notificados en la República Mexicana desde el inicio de la pandemia hasta el 13 de agosto de 2022.



\* Número de casos por cada mil habitantes.  
Mortalidad estimada a partir de los datos abiertos publicados por la Dirección General de Epidemiología el 13 de agosto de 2022 en el sitio: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>



# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
  - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
  - Una muestra.
  - Dos muestras.
    - Independientes.
    - Emparejadas.
  - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

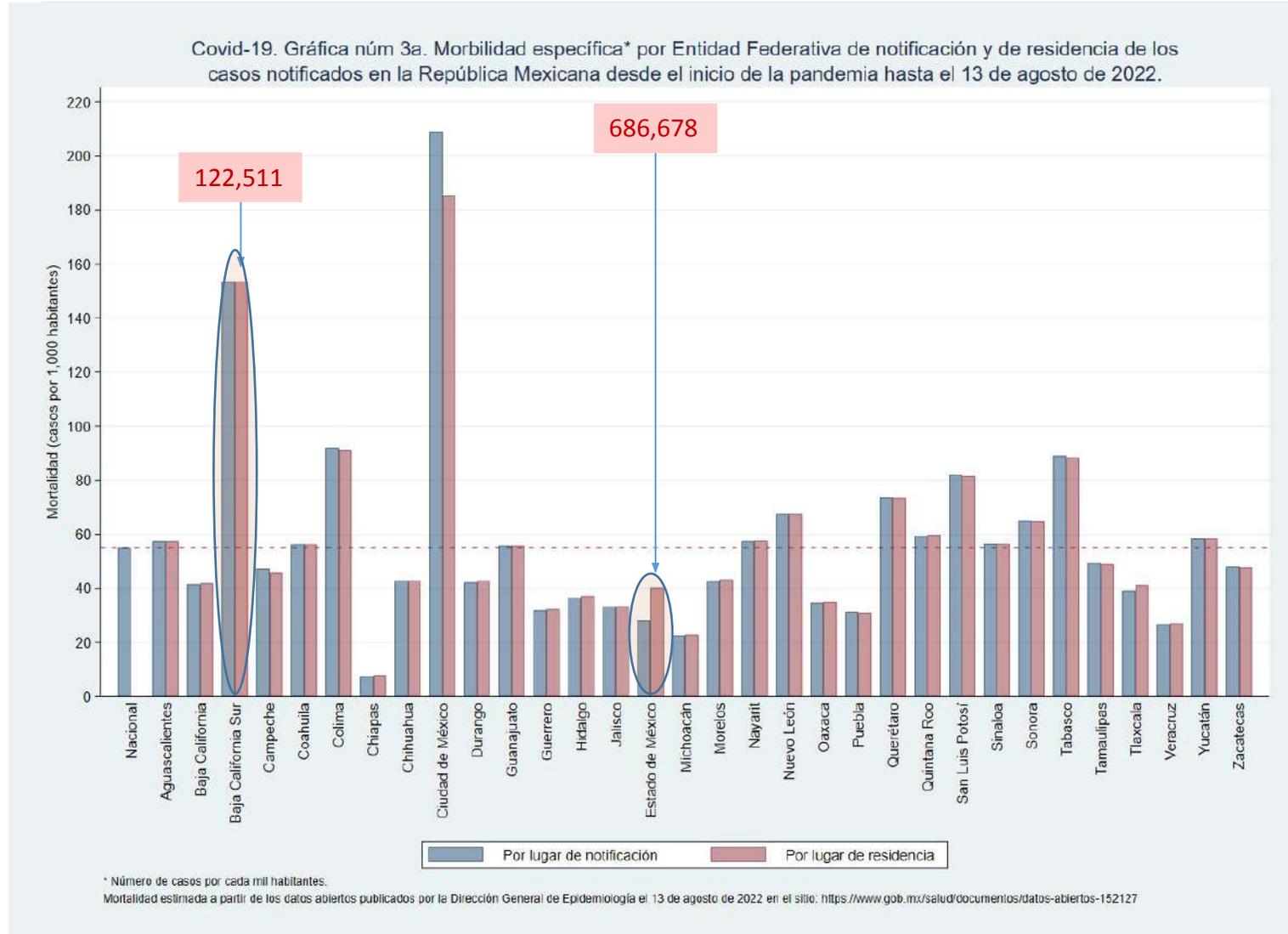
## Morbilidad por entidad federativa.

Magnitud relativa del daño a la salud ocasionado por una agente patógeno, que ocasiona enfermedad en una población durante un periodo.

**Baja California Sur**  
*Casos = 122,511*

**Estado de México**  
*Casos = 686,678*

Gráficos > Gráficos de barra \Principal \ Categoría



Introducción.  
Panorama epidemiológico.  
Vigilancia epidemiológica.  
Datos y calidad.  
Análisis exploratorio.  
Magnitud.  
Tiempo.  
Lugar.  
Variables personales.  
Condiciones de riesgo.  
Evidencia de causalidad.  
Ejemplo: El lugar.  
Otras condiciones.  
Análisis de la evidencia.  
Comparación de medias.  
Estimación.  
Una muestra.  
Dos muestras.  
Independientes.  
Emparejadas.  
Más de dos muestras.  
Medidas de intervención.  
Intervenciones inespecíficas  
Intervenciones específicas.  
Conclusiones.  
Preguntas.

# Vigilancia epidemiológica.

Panorama epidemiológico.

Análisis exploratorio.

Las variables de las personas.

# Herramientas y aplicaciones estadísticas para Ciencia de Datos

## Gráficas de caja y línea.

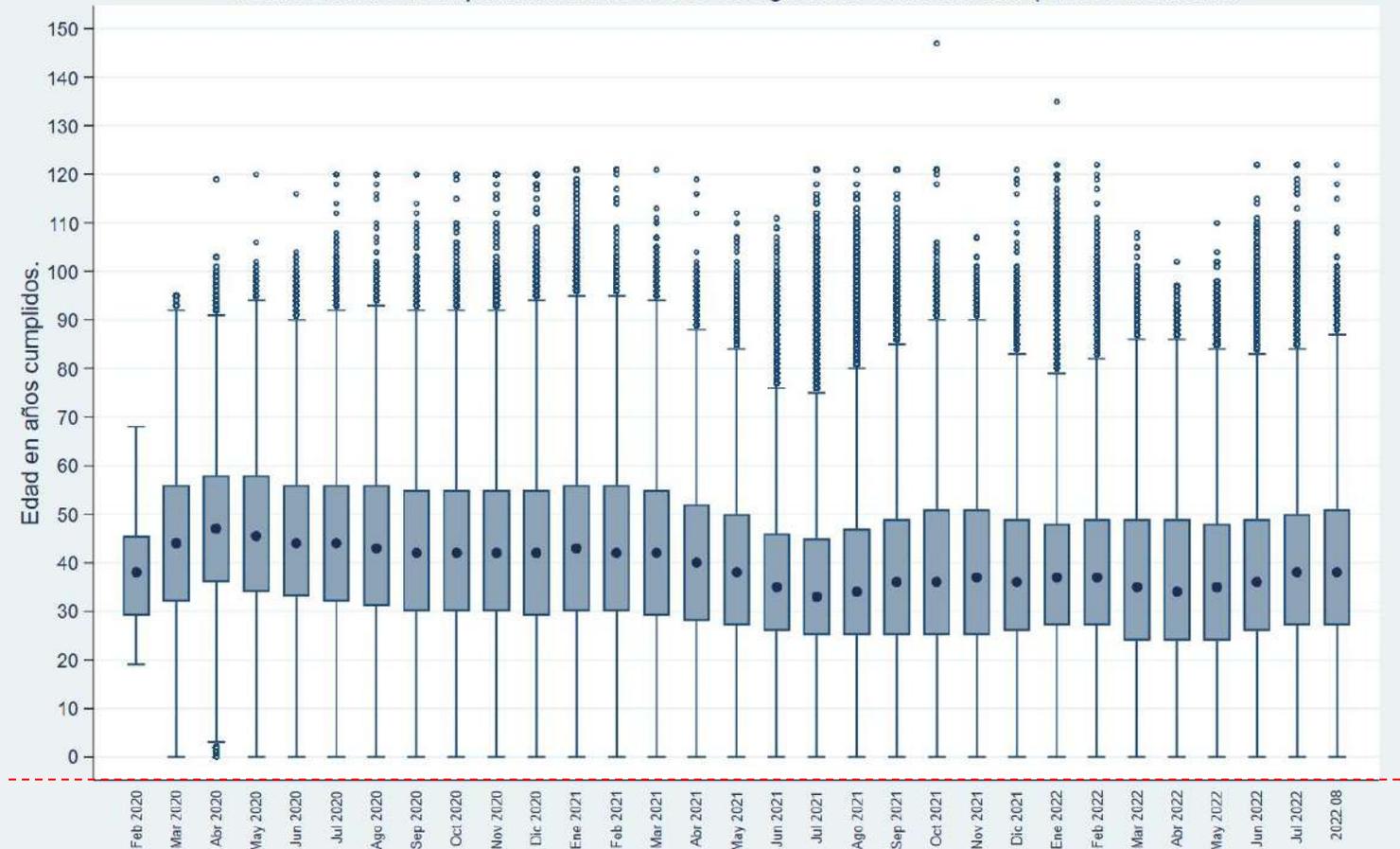
Variación de la edad de los casos por mes.

```

use DA 13 de agosto
graph box edad if evol <= 2, over(añosmesesini,
gap(50) label(labcolor("dnavy") angle(vertical)
labsize(vsmall)) axis(lcolor(dnavy)))
medtype(marker) medmarker(mcolor(dnavy)
msize(small)) marker(1, msize(vsmall)
msymbol(circle_hollow)) ytitle(Edad en años
cumplidos.) ytitle(, size(medsmall) color(dnavy))
ylabel(#25, labels labsize(small) labcolor(dnavy)
angle(horizontal) format(%9.0fc)) ymtick(,
labcolor(dnavy) angle(horizontal) format(%9.0fc)
valuelabel) title(Covid-19. Gráfica núm. 4c. Gráficas
de caja y línea de la edad* de los casos por mes de
inicio del cuadro clínico, size(medium) color(dnavy))
subtitle(desde el inicio de la pandemia hasta el 13 de
agosto de 2022 en la República Mexicana.,
size(medium) color(dnavy)) caption(Gráfica
elaborada a partir de los datos abiertos publicados
por la Dirección General de Epidemiología el 13 de
agosto de 2022 en:
https://www.gob.mx/salud/documentos/datos-abier-
tos-152127, size(vsmall) color(dnavy))
note(Medianas cuartiles y valores extremos de edad
de los casos para cada mes desde el inicio de la
pandemia., size(vsmall) color(dnavy)) scale(0.65)
    
```

Gráficos > Gráficos de caja \Principal \ Categoría

Covid-19. Gráfica núm. 4c. Gráficas de caja y línea de la edad\* de los casos por mes de inicio del cuadro clínico desde el inicio de la pandemia hasta el 13 de agosto de 2022 en la República Mexicana.



Medianas cuartiles y valores extremos de edad de los casos para cada mes desde el inicio de la pandemia.

Gráfica elaborada a partir de los datos abiertos publicados por la Dirección General de Epidemiología el 13 de agosto de 2022 en: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>

# Herramientas y aplicaciones estadísticas para Ciencia de Datos

## Gráficas de caja y línea.

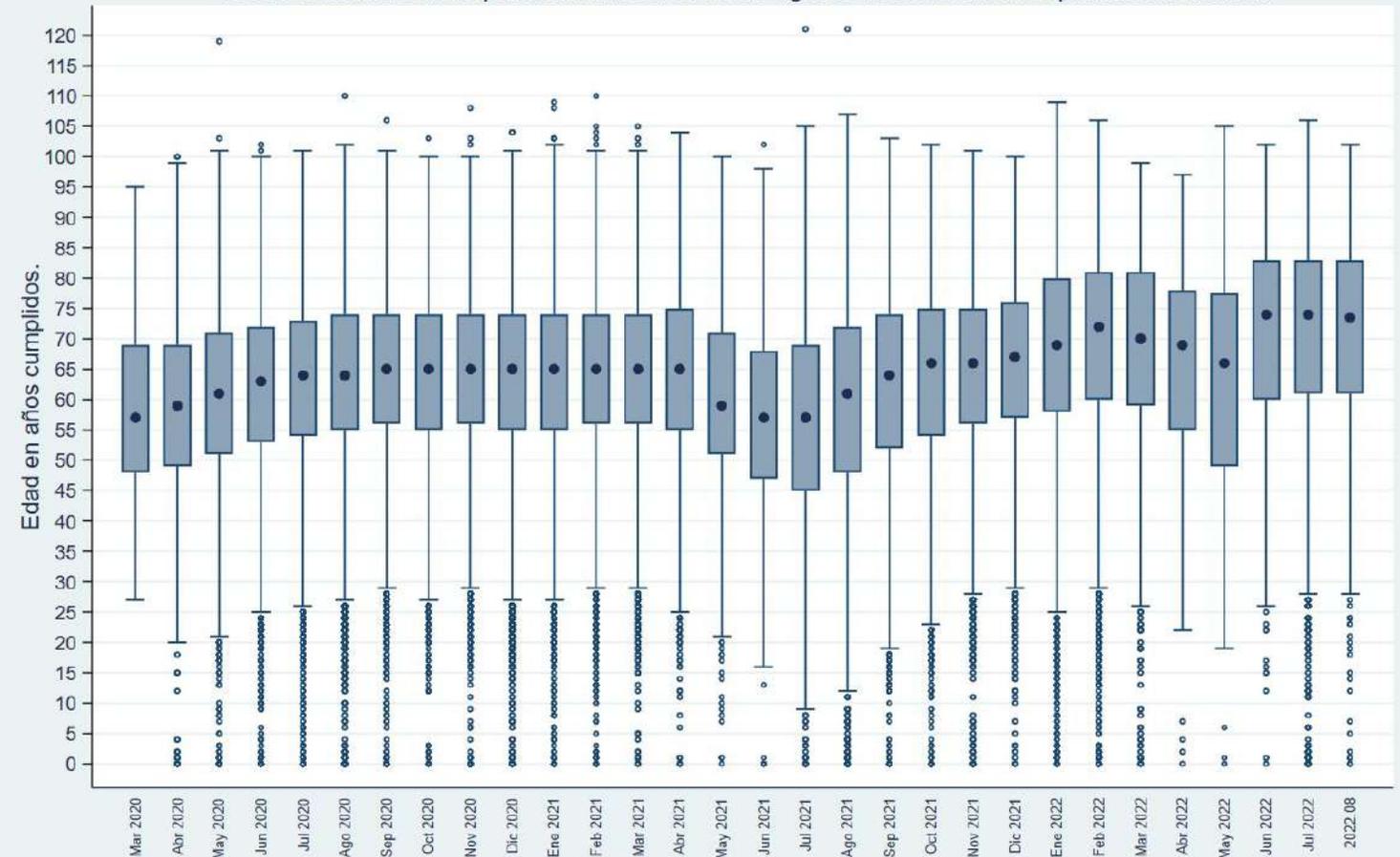
### Variación de la edad de las defunciones por mes.

```

use DA 13 de agosto
graph box edad if evol == 2 & erroringdef!=1,
over(añosmesesdef, gap(50) label(labcolor("dknavy")
angle(vertical) labsize(vsmall)) axis(lcolor(dknavy)))
medtype(marker) medmarker(mcolor(dknavy)
msize(small)) marker(1, msize(vsmall)
msymbol(circle_hollow)) ytitle(Edad en años
cumplidos.) ytitle(, size(medsmall) color(dknavy))
ylabel(#25, labels labsize(small) labcolor(dknavy)
angle(horizontal) format(%9.0fc)) ymtick(,
labcolor(dknavy) angle(horizontal) format(%9.0fc)
valuelabel) title(Covid-19. Gráfica núm. 4d. Gráficas
de caja y línea de la edad* de las defunciones por
mes en que ocurrió la muerte, size(medium)
color(dknavy)) subtitle(desde el el inicio de la
pandemia hasta el 13 de agosto de 2022 en la
República Mexicana., size(medium) color(dknavy))
caption(Gráfica elaborada a partir de los datos
abiertos publicados por la Dirección General de
Epidemiología el 13 de agosto de 2022 en:
https://www.gob.mx/salud/documentos/datos-abier-
tos-152127, size(vsmall) color(dknavy))
note(Medianas cuartiles y valores extremos de edad
de los casos que fallecieron para cada mes desde el
inicio de la pandemia., size(vsmall) color(dknavy))
scale(0.65)
    
```

Gráficos > Gráficos de caja \Principal \ Categoría

Covid-19. Gráfica núm. 4d. Gráficas de caja y línea de la edad\* de las defunciones por mes en que ocurrió la muerte desde el el inicio de la pandemia hasta el 13 de agosto de 2022 en la República Mexicana.



Medianas cuartiles y valores extremos de edad de los casos que fallecieron para cada mes desde el inicio de la pandemia.

Gráfica elaborada a partir de los datos abiertos publicados por la Dirección General de Epidemiología el 13 de agosto de 2022 en: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

CONFERENCIAS STATA LATAM 2022

Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
- Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
  - Una muestra.
  - Dos muestras.
    - Independientes.
    - Emparejadas.
  - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

Calidad de la información.

Covid-19 y embarazo en edades seleccionadas.

use DA 13 de agosto

tab edad evol if evol<3 & embarazo==1

Edad en años cumplidos	Evolución		Total
	Caso vive	Caso_muer	
0	19	0	19
1	7	0	7
2	2	0	2
3	1	0	1
5	1	0	1
6	1	0	1
7	2	0	2
8	2	0	2
9	4	0	4
10	7	0	7
11	3	0	3
12	13	0	13
13	31	0	31
14	88	0	88
15	269	1	270
16	575	3	578
17	899	4	903
18	1,248	1	1,249
19	1,767	6	1,773
20	2,331	12	2,343

Estadísticas > Sumarios, tablas y tests estadísticos > Tablas de frecuencias, sumario estadístico y resultados de comandos \ Principal \ if/in

Cuadro\* núm. 30. Covid-19. Número de casos en mujeres embarazadas menores de 20 años y mayores de 59 años desde el inicio de la pandemia hasta el sábado 13 de agosto, último día de la semana epidemiológica número 32 de 2022, en la República Mexicana.

Edad en años cumplidos	Número de casos en embarazadas	Edad en años cumplidos	Número de casos en embarazadas	Edad en años cumplidos	Número de casos en embarazadas	Edad en años cumplidos	Número de casos en embarazadas
Menores de un año	19	60 años	7	80 años	2	100 años	2
1 año	7	61 años	4	81 años	2	101 años	0
2 años	2	62 años	9	82 años	2	102 años	0
3 años	1	63 años	5	83 años	2	103 años	0
4 años	0	64 años	2	84 años	0	104 años	0
5 años	1	65 años	3	85 años	0	105 años	0
6 años	1	66 años	3	86 años	1	106 años	0
7 años	2	67 años	4	87 años	0	107 años	0
8 años	2	68 años	2	88 años	0	108 años	0
9 años	4	69 años	2	89 años	0	109 años	1
10 años	7	70 años	1	90 años	1	110 años	0
11 años	3	71 años	2	91 años	0		
12 años	13	72 años	2	92 años	0		
13 años	31	73 años	4	93 años	1		
14 años	88	74 años	2	94 años	0		
15 años	270	75 años	2	95 años	1		
16 años	578	76 años	1	96 años	0		
17 años	903	77 años	0	97 años	1		
18 años	1,249	78 años	3	98 años	0		
19 años	1,773	79 años	1	99 años	0		
<b>Total</b>	<b>4,954</b>	<b>Total</b>	<b>59</b>	<b>Total</b>	<b>13</b>	<b>Total</b>	<b>3</b>

\* Elaborado a partir de los datos abiertos de la Dirección General de Epidemiología (DGE/SSA) publicados el 13 de agosto de 2022 en: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>

El archivo de datos.

Información recolectada de la realidad:

- Registros (revisión, duplicados)
- Variables (etiquetas, revisión)
- Campos (etiquetas, códigos, **revisión**)



**CONFERENCIAS STATA LATAM 2022**  
**Herramientas y aplicaciones estadísticas para Ciencia de Datos**

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
- Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
  - Una muestra.
  - Dos muestras.
    - Independientes.
    - Emparejadas.
  - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

**Calidad de la información.**  
**Covid-19 y embarazo en edades seleccionadas.**

```
clear all
import delimited "C:\Users\mario MG\Desktop\Datos mex\COVID19MEXICO2021.csv "
keep fecha_actualizacion sexo fecha_sintomas fecha_def edad embarazo clasificacion_final
generate int def1 = 1 if fecha_def != "9999-99-99"
replace def = 0 if fecha_def == "9999-99-99"
recode def1 (1 = 1 "Muere") (0 = 2 "No muere"),
generate(def)
drop def1
generate int emb1 = 1 if embarazo==1
replace emb1 = 0 if embarazo > 1 & sexo == 2
recode emb1 (1 = 1 "Embarazada") (0 = 2 "No"),
generate(embarazada)
drop emb1
recode sexo (1 = 1 "Hombre") (2 = 2 "Mujer"), generate(Sexo)
recode clasificacion_final (1 = 1 "Caso") (2 = 1 "Caso") (3 = 1 "Caso") (4 = 4 "No válido") (5 = 4 "No válido") (6 = 0 "Sospechoso") (7 = 3 "Negativo"), generate(clas)
label variable def "Defunciones"
label variable embarazada "Embarazo"
label variable Sexo "Sexo"
label variable clas "Clasificación"
save "C:\Users\mario MG\Desktop\Datos mex\Covid-19 mex 2021.dta"
table (edad) (clas) (embarazada) if edad<10 & clas<2
table (edad) (clas) (embarazada) if edad>60 & clas<2
table (def) (clas) if clas<2
```

Estadísticas > Sumarios, tablas y tests estadísticos > Tablas de frecuencias, sumario estadístico y resultados de comandos \ Principal \ if/in

Embarazo = Embarazada			
	Clasificación		
	Sospechoso	Caso	Total
EDAD			
0	1	9	10
1	1	4	5
3	1	1	2
6	1	1	2
8		1	1
Total	4	16	20

Embarazo = Embarazada			
	Clasificación		
	Sospechoso	Caso	Total
EDAD			
61		1	1
62	1	4	5
63		1	1
64	2		2
65		1	1
66		1	1
67		1	1
68	1	1	2
69		1	1
71		1	1
72		1	1
74		1	1
75		2	2
79		1	1
80	1	1	2
81		1	1
82		1	1
83		2	2
93	1		1
97		1	1
Total	6	23	29

**Covid-19 en México.**  
**Número de casos en mujeres embarazadas menores de 10 años y mayores de 59 años durante el año 2021 en México, de acuerdo con los datos actualizados, revisados y publicados el 9 de agosto de 2022<sup>1</sup> por la Secretaría de Salud de México.**

**El archivo de datos.**  
 Información recolectada de la realidad:

Registros (revisión, duplicados)  
 Variables (etiquetas, revisión)  
 Campos (etiquetas, códigos, **revisión**)

<sup>1</sup> Elaborado a partir de los datos abiertos actualizados el 9 de agosto de 2022, publicados por la DGE en: <https://www.gob.mx/salud/documentos/datos-abiertos-bases-historicas-direccion-general-de-epidemiologia>  
 Cierre Datos Abiertos Históricos 2021



CONFERENCIAS STATA LATAM 2022

Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
- Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
  - Una muestra.
  - Dos muestras.
    - Independientes.
    - Emparejadas.
  - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

Pirámide de población (casos).

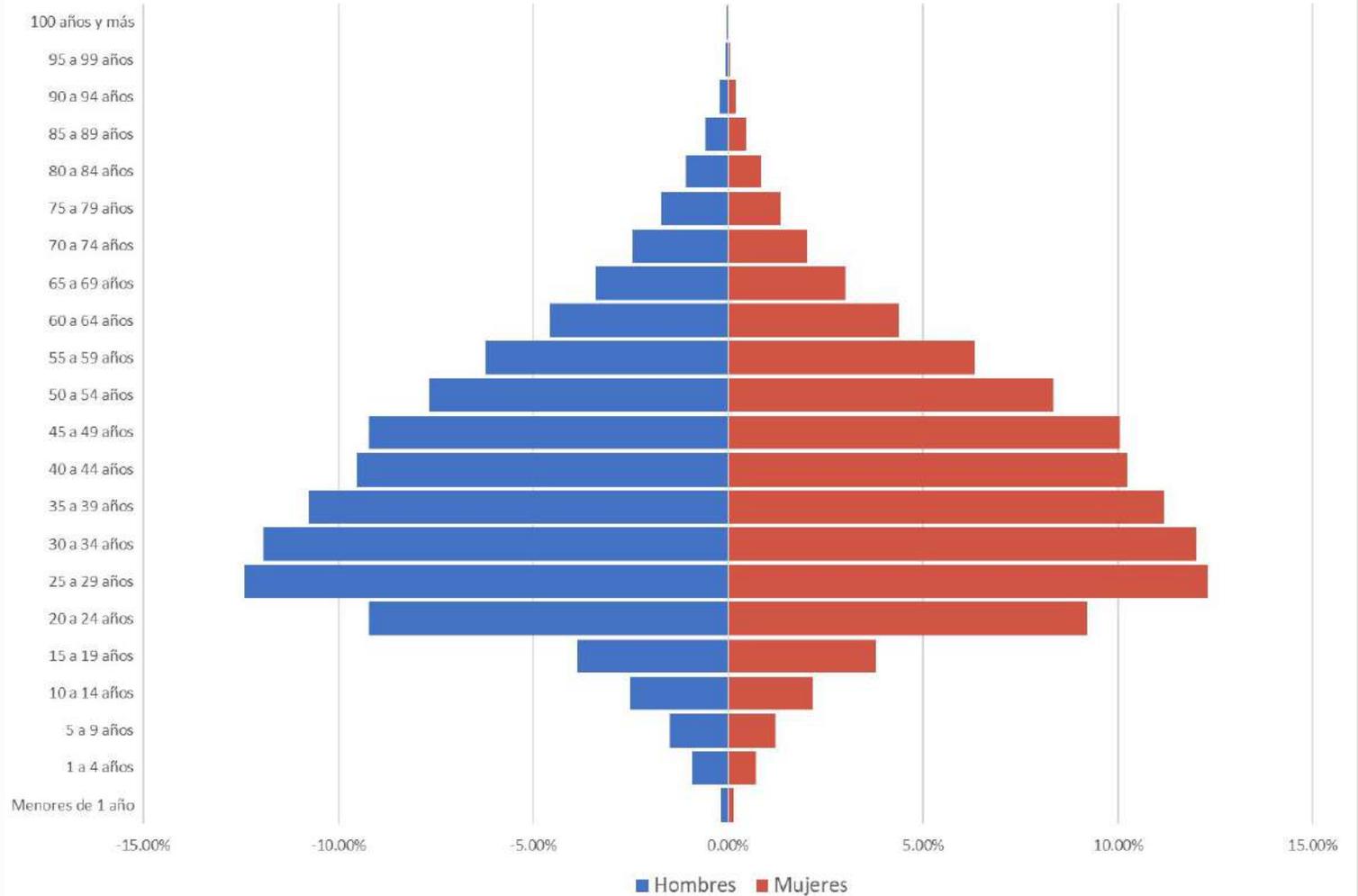
Casos por grupos de edad y sexo.

use DA 13 de agosto  
table (gedadm) (sexo) (evol) if evol<3

Stata				Excel	
Grupos de edad	Total	Hombres	Mujeres	Porc. de hombres	Porc. de mujeres
Menores de 1 año	10,651	5,753	4,898	-0.18%	0.13%
1 a 4 años	56,178	30,123	26,055	-0.93%	0.71%
5 a 9 años	92,960	48,451	44,509	-1.49%	1.21%
10 a 14 años	161,609	82,022	79,587	-2.53%	2.16%
15 a 19 años	264,843	125,681	139,162	-3.87%	3.78%
20 a 24 años	637,401	299,037	338,364	-9.21%	9.20%
25 a 29 años	855,061	402,556	452,505	-12.40%	12.30%
30 a 34 años	828,772	387,238	441,534	-11.92%	12.00%
35 a 39 años	761,139	349,469	411,670	-10.76%	11.19%
40 a 44 años	686,207	309,660	376,547	-9.53%	10.24%
45 a 49 años	668,451	299,097	369,354	-9.21%	10.04%
50 a 54 años	555,546	248,968	306,578	-7.67%	8.34%
55 a 59 años	434,639	202,089	232,550	-6.22%	6.32%
60 a 64 años	309,365	148,450	160,915	-4.57%	4.38%
65 a 69 años	220,652	109,721	110,931	-3.38%	3.02%
70 a 74 años	154,267	79,818	74,449	-2.46%	2.02%
75 a 79 años	105,633	55,630	50,003	-1.71%	1.36%
80 a 84 años	65,922	34,967	30,955	-1.08%	0.84%
85 a 89 años	36,476	19,060	17,416	-0.59%	0.47%
90 a 94 años	14,313	7,087	7,226	-0.22%	0.20%
95 a 99 años	3,969	1,937	2,032	-0.06%	0.06%
100 años y más	1,614	903	711	-0.03%	0.02%
Total	6,925,668	3,247,717	3,677,951	-100.00%	100.00%

Excel: Insertar gráfico > Barras agrupadas

Pirámide de población de casos<sup>1</sup> de Covid-19 por grupos de edad y sexo desde el inicio de la pandemia hasta el 13 de agosto de 2022 en México<sup>2</sup>.



<sup>1</sup> Porcentaje de casos por sexo y por grupos de edad.

<sup>2</sup> Elaborado a partir de los datos abiertos publicados por la DGE el 13 de agosto de 2022 en: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>



CONFERENCIAS STATA LATAM 2022

Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
- Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
    - Una muestra.
    - Dos muestras.
      - Independientes.
      - Emparejadas.
    - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

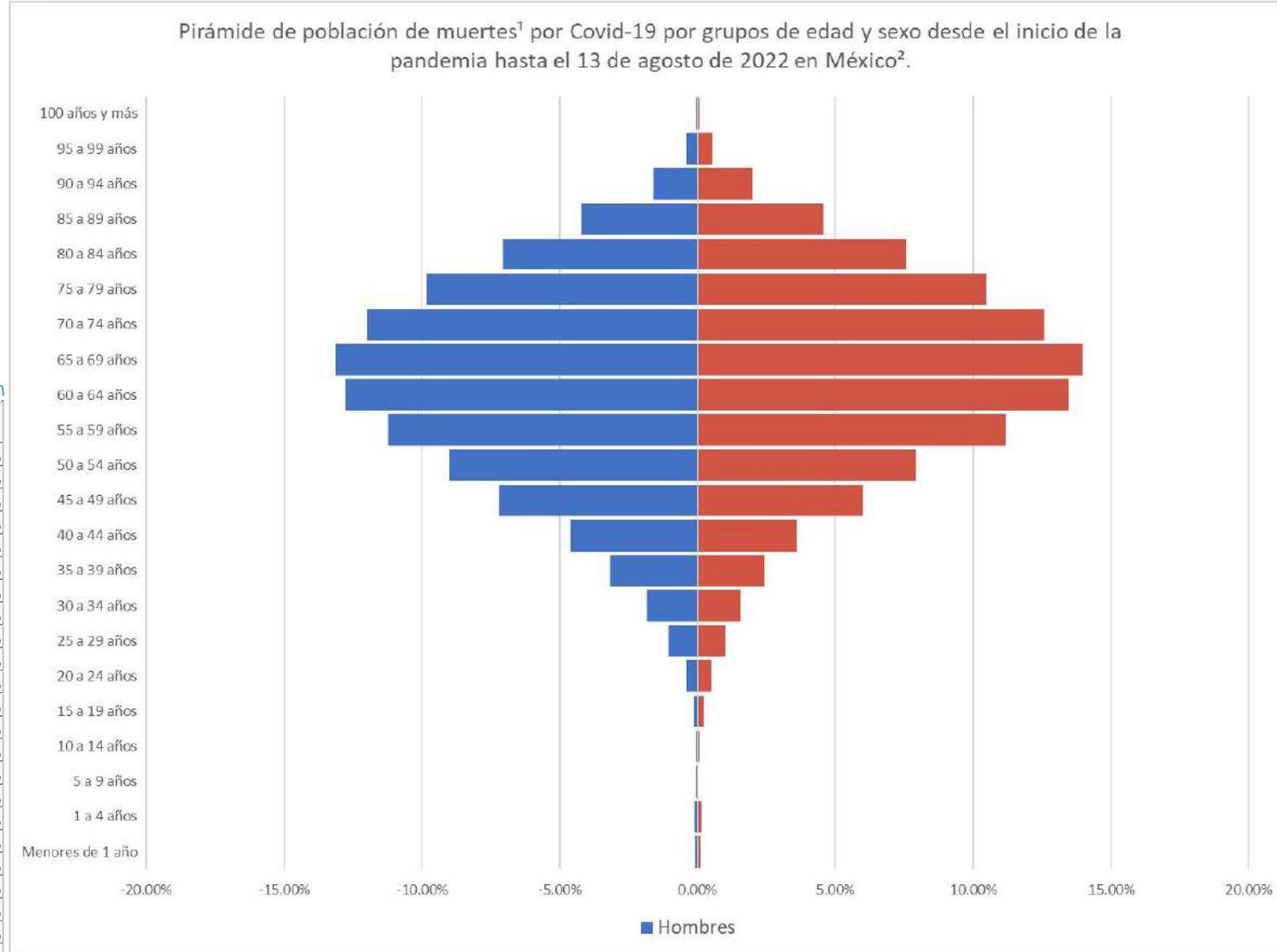
Pirámide de población (muertes).

Muertes por grupos de edad y sexo.

use DA 13 de agosto  
table (gedadm) (sexo) (evol) if evol<3

Grupos de edad	Stata			Excel	
	Total	Hombres	Mujeres	Porc. de hombres	Porc. de mujeres
Menores de 1 año	309	170	139	-0.08%	0.11%
1 a 4 años	376	197	179	-0.10%	0.14%
5 a 9 años	152	90	62	-0.04%	0.05%
10 a 14 años	232	134	98	-0.07%	0.08%
15 a 19 años	553	264	289	-0.13%	0.23%
20 a 24 años	1,460	852	608	-0.42%	0.48%
25 a 29 años	3,399	2,115	1,284	-1.05%	1.02%
30 a 34 años	5,713	3,740	1,973	-1.85%	1.56%
35 a 39 años	9,495	6,445	3,050	-3.19%	2.41%
40 a 44 años	13,850	9,310	4,540	-4.60%	3.59%
45 a 49 años	22,121	14,536	7,585	-7.19%	6.00%
50 a 54 años	28,277	18,255	10,022	-9.02%	7.93%
55 a 59 años	36,849	22,707	14,142	-11.23%	11.18%
60 a 64 años	42,823	25,816	17,007	-12.76%	13.45%
65 a 69 años	44,266	26,572	17,694	-13.14%	13.99%
70 a 74 años	40,177	24,260	15,917	-11.99%	12.59%
75 a 79 años	33,138	19,870	13,268	-9.82%	10.49%
80 a 84 años	23,846	14,274	9,572	-7.06%	7.57%
85 a 89 años	14,251	8,490	5,761	-4.20%	4.56%
90 a 94 años	5,699	3,209	2,490	-1.59%	1.97%
95 a 99 años	1,512	852	660	-0.42%	0.52%
100 años y más	226	120	106	-0.06%	0.08%
Total	328,724	202,278	126,446	-100.00%	100.00%

Excel: Insertar gráfico > Barras agrupadas



<sup>1</sup> Porcentaje de casos por sexo y por grupos de edad.

<sup>2</sup> Elaborado a partir de los datos abiertos publicados por la DGE el 13 de agosto de 2022 en: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>



CONFERENCIAS STATA LATAM 2022

Herramientas y aplicaciones estadísticas para Ciencia de Datos

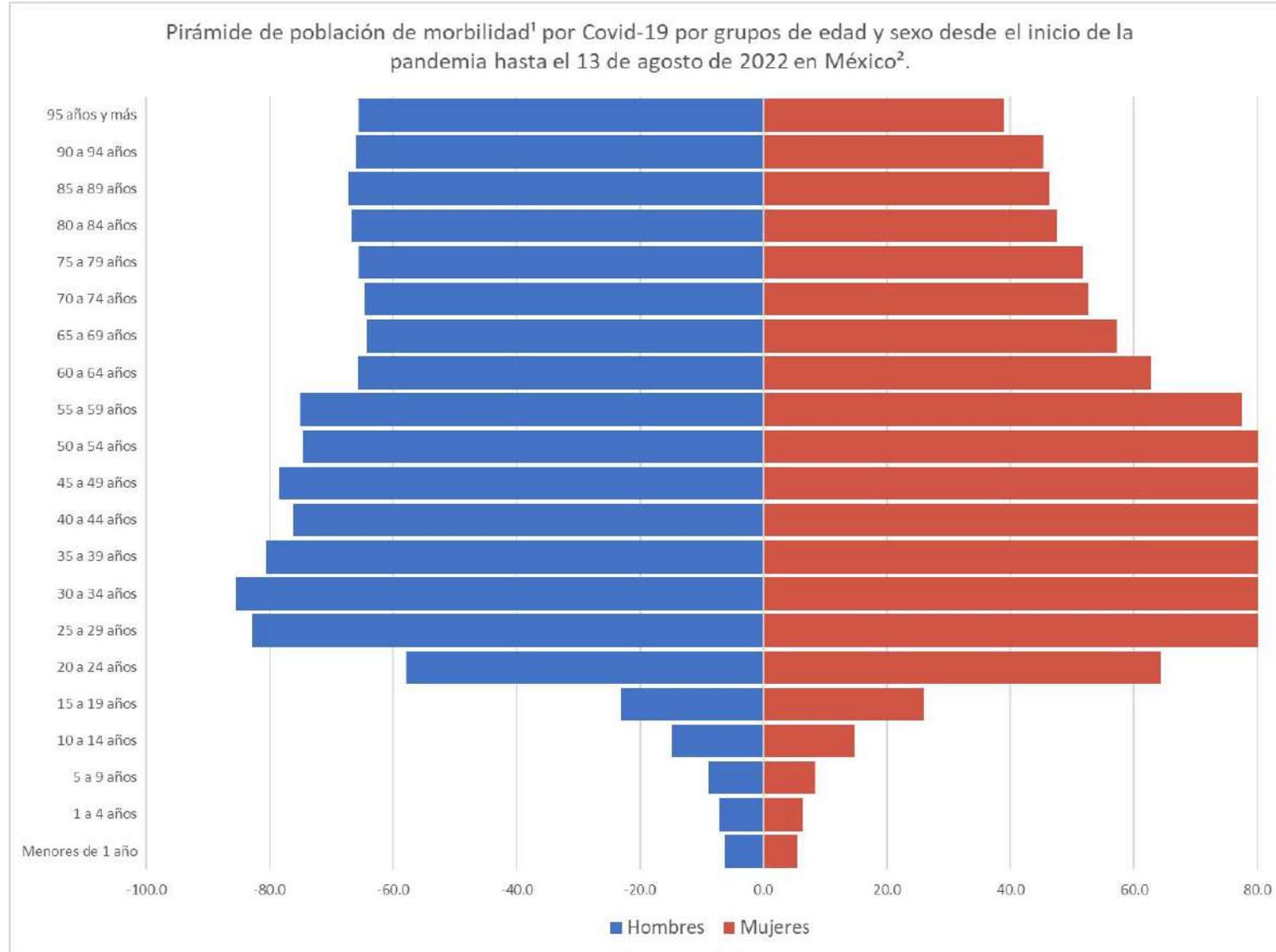
Pirámide de población (morbilidad).

Casos por grupos de edad y sexo por cada mil habitantes.

Población por grupos de edad en el 2020 en la República Mexi			
Grupos de edad	Total	Hombres	Mujeres
Menores de 1 año	1,812,977	916,140	896,837
De 1 a 4 años	8,234,388	4,161,342	4,073,046
De 5 a 9 años	10,764,379	5,453,091	5,311,288
De 10 a 14 años	10,943,540	5,554,260	5,389,280
De 15 a 19 años	10,806,690	5,462,150	5,344,540
De 20 a 24 años	10,422,095	5,165,884	5,256,211
De 25 a 29 años	9,993,001	4,861,404	5,131,597
De 30 a 34 años	9,420,827	4,527,726	4,893,101
De 35 a 39 años	9,020,276	4,331,530	4,688,746
De 40 a 44 años	8,503,586	4,062,304	4,441,282
De 45 a 49 años	7,942,413	3,812,344	4,130,069
De 50 a 54 años	7,037,532	3,332,163	3,705,369
De 55 a 59 años	5,695,958	2,692,976	3,002,982
De 60 a 64 años	4,821,062	2,257,862	2,563,200
De 65 a 69 años	3,645,077	1,706,850	1,938,227
De 70 a 74 años	2,647,340	1,233,492	1,413,848
De 75 a 79 años	1,814,582	847,898	966,684
De 80 a 84 años	1,175,364	523,812	651,552
De 85 a 89 años	659,245	283,351	375,894
De 90 a 94 años	266,806	107,358	159,448
De 95 a 99 años	95,205	36,615	58,590
De 100 años y más	18,295	6,644	11,651
No especificado	273,386	136,194	137,192
Total	126,014,024	61,473,390	64,540,634

Fuente: INEGI. Censo de Población y Vivienda 2020. Tabulados del Cuestionario Básico. <https://www.inegi.org.mx/sistemas/Olap/Proyectos/bd/censos/cpv2020/pt.asp>

Excel: Insertar gráfico > Barras agrupadas



<sup>1</sup> Casos notificados por cada mil habitantes.

<sup>2</sup> Elaborado a partir de los datos abiertos publicados por la DGE el 13 de agosto de 2022 en: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.



CONFERENCIAS STATA LATAM 2022

Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
- Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
  - Una muestra.
  - Dos muestras.
    - Independientes.
    - Emparejadas.
  - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

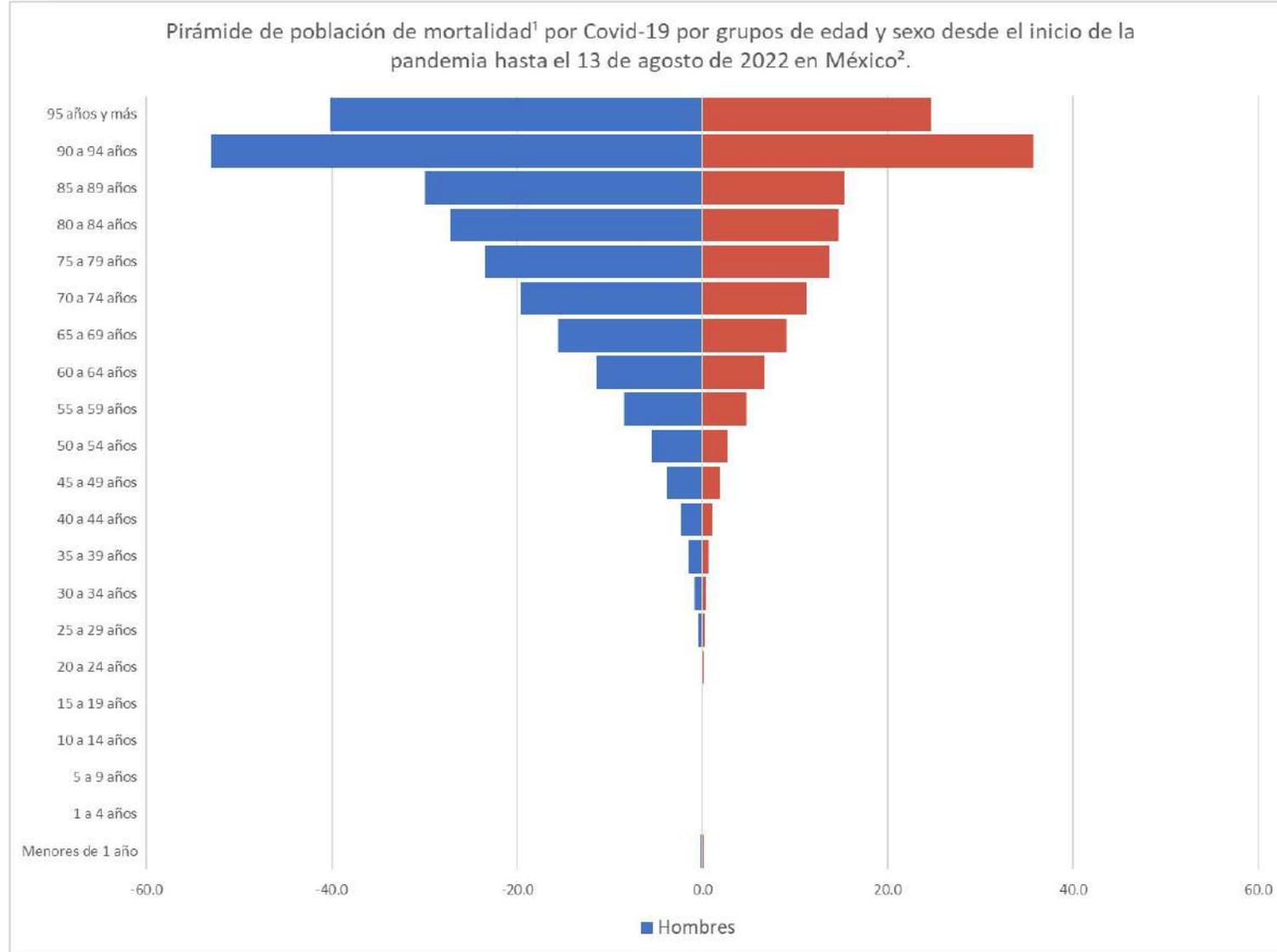
Pirámide de población (mortalidad).

Muertes por grupos de edad y sexo por cada mil habitantes.

Población por grupos de edad en el 2020 en la República Mexi			
Grupos de edad	Total	Hombres	Mujeres
Menores de 1 año	1,812,977	916,140	896,837
De 1 a 4 años	8,234,388	4,161,342	4,073,046
De 5 a 9 años	10,764,379	5,453,091	5,311,288
De 10 a 14 años	10,943,540	5,554,260	5,389,280
De 15 a 19 años	10,806,690	5,462,150	5,344,540
De 20 a 24 años	10,422,095	5,165,884	5,256,211
De 25 a 29 años	9,993,001	4,861,404	5,131,597
De 30 a 34 años	9,420,827	4,527,726	4,893,101
De 35 a 39 años	9,020,276	4,331,530	4,688,746
De 40 a 44 años	8,503,586	4,062,304	4,441,282
De 45 a 49 años	7,942,413	3,812,344	4,130,069
De 50 a 54 años	7,037,532	3,332,163	3,705,369
De 55 a 59 años	5,695,958	2,692,976	3,002,982
De 60 a 64 años	4,821,062	2,257,862	2,563,200
De 65 a 69 años	3,645,077	1,706,850	1,938,227
De 70 a 74 años	2,647,340	1,233,492	1,413,848
De 75 a 79 años	1,814,582	847,898	966,684
De 80 a 84 años	1,175,364	523,812	651,552
De 85 a 89 años	659,245	283,351	375,894
De 90 a 94 años	266,806	107,358	159,448
De 95 a 99 años	95,205	36,615	58,590
De 100 años y más	18,295	6,644	11,651
No especificado	273,386	136,194	137,192
Total	126,014,024	61,473,390	64,540,634

Fuente: INEGI. Censo de Población y Vivienda 2020. Tabulados del Cuestionario Básico. <https://www.inegi.org.mx/sistemas/Olap/Proyectos/bd/censos/cpv2020/pt.asp>

Excel: Insertar gráfico > Barras agrupadas



<sup>1</sup> Porcentaje de casos por sexo y por grupos de edad.  
<sup>2</sup> Elaborado a partir de los datos abiertos publicados por la DGE el 13 de agosto de 2022 en: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>



CONFERENCIAS STATA LATAM 2022

Herramientas y aplicaciones estadísticas para Ciencia de Datos

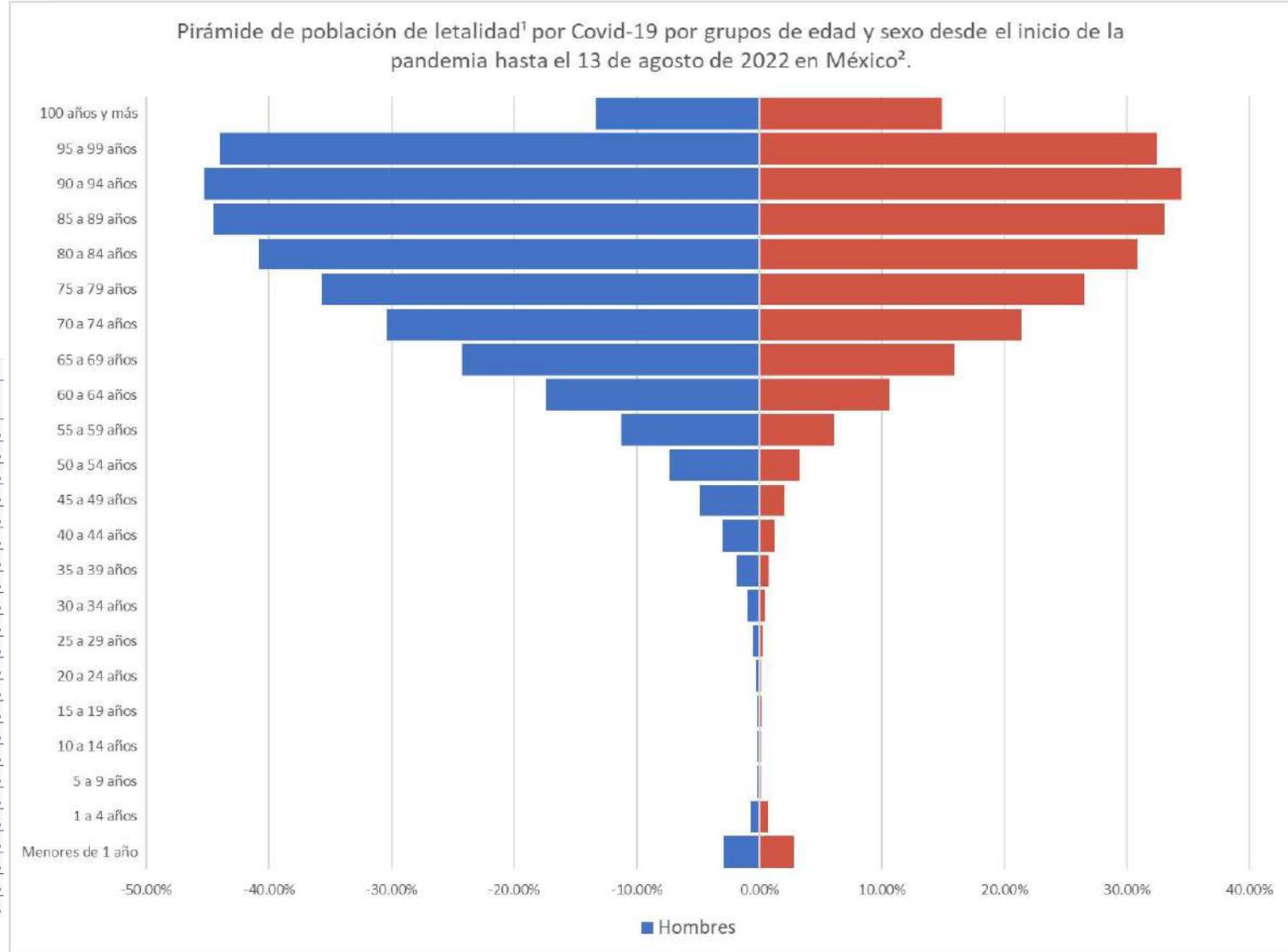
- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
    - Una muestra.
    - Dos muestras.
      - Independientes.
      - Emparejadas.
    - Más de dos muestras.
  - Medidas de intervención.
    - Intervenciones inespecíficas
    - Intervenciones específicas.
- Conclusiones.
- Preguntas.

Pirámide de población (letalidad).

Porcentaje de casos que mueren por sexo y grupos de edad.

Grupos de edad	Muertes		Casos		Letalidad	
	Hombres	Mujeres	Hombres	Mujeres	Porc. de hombres	Porc. de mujeres
Menores de 1 año	170	139	5,753	4,898	-2.95%	2.84%
1 a 4 años	197	179	30,123	26,055	-0.65%	0.69%
5 a 9 años	90	62	48,451	44,509	-0.19%	0.14%
10 a 14 años	134	98	82,022	79,587	-0.16%	0.12%
15 a 19 años	264	289	125,681	139,162	-0.21%	0.21%
20 a 24 años	852	608	299,037	338,364	-0.28%	0.18%
25 a 29 años	2,115	1,284	402,556	452,505	-0.53%	0.28%
30 a 34 años	3,740	1,973	387,238	441,534	-0.97%	0.45%
35 a 39 años	6,445	3,050	349,469	411,670	-1.84%	0.74%
40 a 44 años	9,310	4,540	309,660	376,547	-3.01%	1.21%
45 a 49 años	14,536	7,585	299,097	369,354	-4.86%	2.05%
50 a 54 años	18,255	10,022	248,968	306,578	-7.33%	3.27%
55 a 59 años	22,707	14,142	202,089	232,550	-11.24%	6.08%
60 a 64 años	25,816	17,007	148,450	160,915	-17.39%	10.57%
65 a 69 años	26,572	17,694	109,721	110,931	-24.22%	15.95%
70 a 74 años	24,260	15,917	79,818	74,449	-30.39%	21.38%
75 a 79 años	19,870	13,268	55,630	50,003	-35.72%	26.53%
80 a 84 años	14,274	9,572	34,967	30,955	-40.82%	30.92%
85 a 89 años	8,490	5,761	19,060	17,416	-44.54%	33.08%
90 a 94 años	3,209	2,490	7,087	7,226	-45.28%	34.46%
95 a 99 años	852	660	1,937	2,032	-43.99%	32.48%
100 años y más	120	106	903	711	-13.29%	14.91%
Total	202,278	126,446	3,247,717	3,677,951	-6.23%	3.44%

Excel: Insertar gráfico > Barras agrupadas



<sup>1</sup> Porcentaje de casos por sexo y por grupos de edad.

<sup>2</sup> Elaborado a partir de los datos abiertos publicados por la DGE el 13 de agosto de 2022 en: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>



- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
      - Una muestra.
      - Dos muestras.
        - Independientes.
        - Emparejadas.
      - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

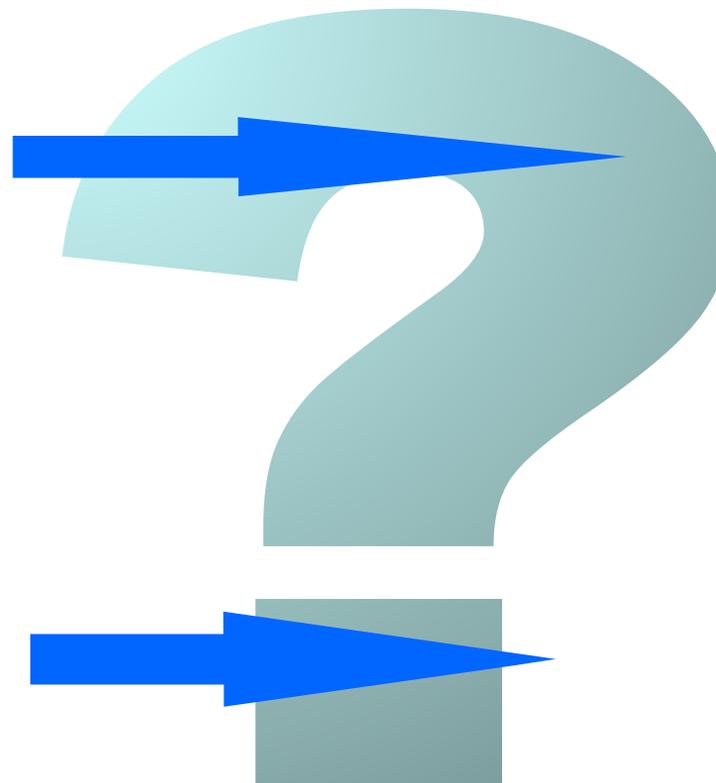
# Vigilancia epidemiológica.

Condiciones de riesgo.  
Evidencia de causalidad.

Múltiples dimensiones.

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
    - Una muestra.
    - Dos muestras.
      - Independientes.
      - Emparejadas.
    - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

## Asociación y causalidad.



*Un científico se parece más a un jugador de apuestas que a un profeta*

## ¿Evidencia de causalidad?

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
    - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
    - Una muestra.
    - Dos muestras.
      - Independientes.
      - Emparejadas.
    - Más de dos muestras.
  - Medidas de intervención.
    - Intervenciones inespecíficas.
    - Intervenciones específicas.
- Conclusiones.
- Preguntas.

- **Conocimiento imperfecto.**
- **Koch-Henle (1887); Bradford Hill (1965)**
- **Criterios de causalidad.**
  - **Fuerza de asociación, Consistencia, Especificidad.**
  - **Temporalidad, Gradiente biológico, Plausibilidad biológica.**
  - **Coherencia, Evidencia experimental y Analogía.**
- **No existen criterios definitivos para establecer causalidad.**
- **Evidencia, inferencia, ciencia y política.**

**No todo lo que puede ser contado cuenta... y no todo lo que cuenta puede ser contado.**

*Albert Einstein.*



- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
    - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
      - Una muestra.
      - Dos muestras.
        - Independientes.
        - Emparejadas.
      - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas.
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

# Vigilancia epidemiológica.

Condiciones de riesgo.  
Evidencia de causalidad.

Ejemplo: El lugar.

# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

## Morbilidad por entidad federativa.

Magnitud relativa del daño a la salud ocasionado por una agente patógeno, que ocasiona enfermedad en una población durante un periodo.

### Baja California Sur

Casos = 122,511

Población = 798,447

$$\frac{122,511}{798,447} = 0.153436$$

$$0.057536 \times 1,000 = 153.4$$

### Estado de México

Casos = 686,678

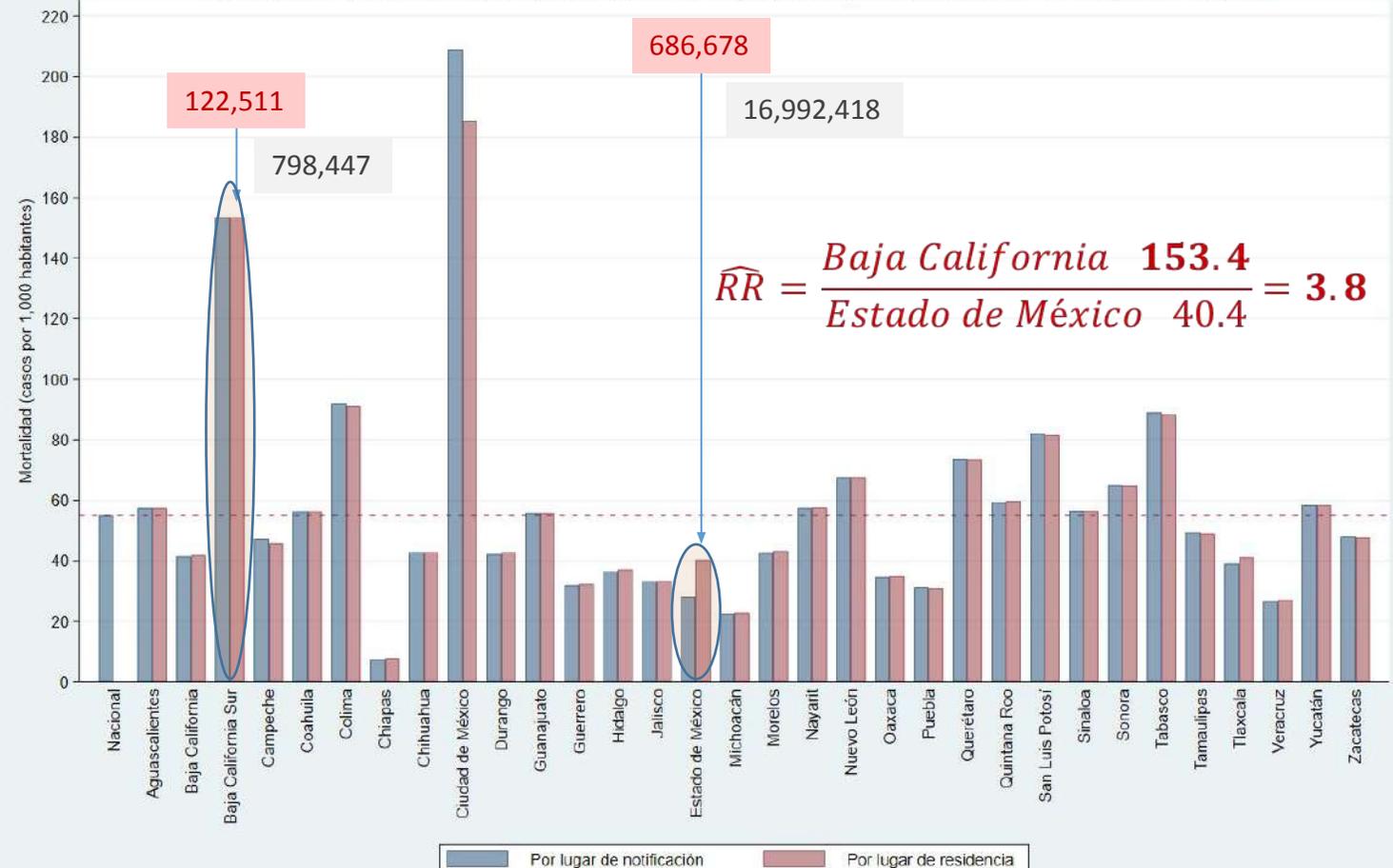
Población = 16,992,418

$$\frac{686,678}{16,992,418} = 0.04041$$

$$0.04041 \times 1,000 = 40.4$$

Gráficos > Gráficos de barra \Principal \ Categoría

Covid-19. Gráfica núm 3a. Morbilidad específica\* por Entidad Federativa de notificación y de residencia de los casos notificados en la República Mexicana desde el inicio de la pandemia hasta el 13 de agosto de 2022.



$$\widehat{RR} = \frac{\text{Baja California } 153.4}{\text{Estado de México } 40.4} = 3.8$$

\* Número de casos por cada mil habitantes. Mortalidad estimada a partir de los datos abiertos publicados por la Dirección General de Epidemiología el 13 de agosto de 2022 en el sitio: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>

# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

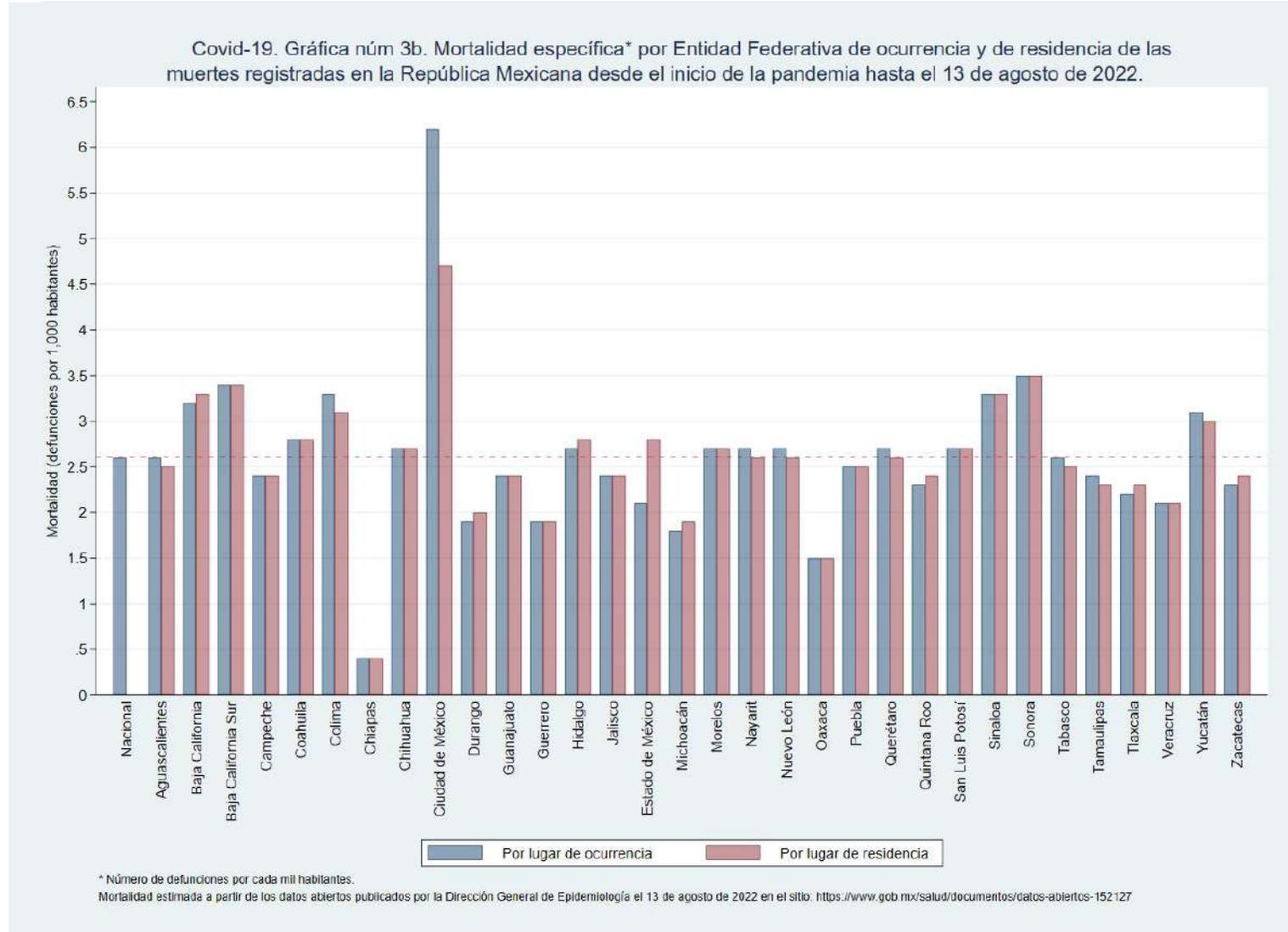
## Mortalidad por entidad federativa.

Magnitud relativa del daño a la salud ocasionado por un agente patógeno que ocasiona la muerte en una población durante un periodo.

use letalidad

```
graph bar (sum) mortocu (sum) mortres, over(edo,
label(angle(vertical))) bar(1, fcolor(%50)
lwidth(vthin)) bar(2, fcolor(%50) lwidth(vthin))
ytile(Mortalidad (defunciones por 1,000
habitantes)) yline(2.6086, lwidth(thin)
lpattern(vshortdash)) ylabel(0(0.5)6.5,
angle(horizontal)) title(Covid-19. Gráfica núm 3b.
Mortalidad específica* por Entidad Federativa de
ocurrencia y de residencia de las) subtitle(muertes
registradas en la República Mexicana desde el inicio
de la pandemia hasta el 13 de agosto de 2022.,
size(large) color(dknavy)) caption(Mortalidad
estimada a partir de los datos abiertos publicados
por la Dirección General de Epidemiología el 13 de
agosto de 2022 en el sitio:
https://www.gob.mx/salud/documentos/datos-abie
rtos-152127, size(small)) note(* Número de
defunciones por cada mil habitantes., size(small))
legend(order(1 "Por lugar de ocurrencia" 2 "Por
lugar de residencia")) scale(.45)
```

Gráficos > Gráficos de barra \Principal \ Categoría



# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

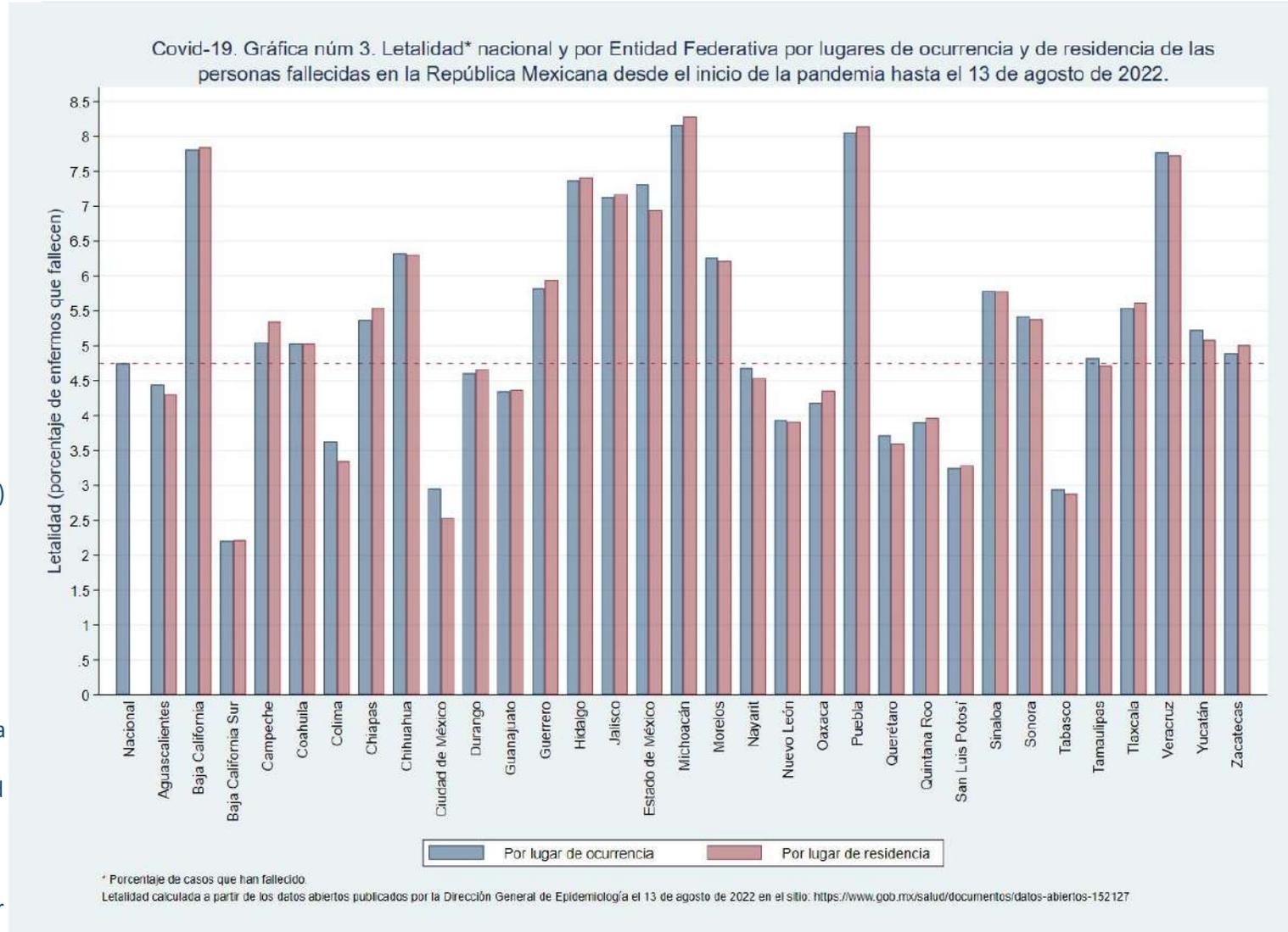
## Letalidad por entidad federativa.

Magnitud relativa del daño a la salud ocasionado por una enfermedad que ocasiona la muerte a los enfermos, en un lugar durante un periodo.

use letalidad

```
graph bar (sum) letocu (sum) letres, over(edo,
label(angle(vertical))) bar(1, fcolor(%50) lwidth(thin))
bar(2, fcolor(%50) lwidth(thin)) ytitle(Letalidad
(porcentaje de enfermos que fallecen)) ytitle(,
size(medlarge) color(dknavy)) yline(4.746459114,
lwidth(thin) lpattern(vshortdash)) ylabel(0(0.5)8.5,
angle(horizontal)) ymtick(, ticks) title(Covid-19.
Gráfica núm 3c. Letalidad* nacional y por Entidad
Federativa de ocurrencia y de residencia de las)
subtitle(personas fallecidas en la República Mexicana
desde el inicio de la pandemia hasta el 13 de agosto
de 2022., size(large) color(dknavy)) caption(Letalidad
calculada a partir de los datos abiertos publicados
por la Dirección General de Epidemiología el 13 de
agosto de 2022 en el sitio:
https://www.gob.mx/salud/documentos/datos-abier-
tos-152127, size(small)) note(* Porcentaje de casos
que han fallecido., size(small)) legend(order(1 "Por
lugar de ocurrencia" 2 "Por lugar de residencia"))
scale(.45)
```

Gráficos > Gráficos de barra \ Principal \ Categoría



- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
      - Una muestra.
      - Dos muestras.
        - Independientes.
        - Emparejadas.
      - Más de dos muestras.
  - Medidas de intervención.
    - Intervenciones inespecíficas
    - Intervenciones específicas.
- Conclusiones.
- Preguntas.

# Vigilancia epidemiológica.

Condiciones de riesgo.  
Evidencia de causalidad.

Otras condiciones.

CONFERENCIAS STATA LATAM 2022

**Herramientas y aplicaciones estadísticas para Ciencia de Datos**

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

**Estimación de riesgos  
Letalidad por Covid-19 en México y diabetes mellitus.**  
(riesgo de morir por enfermarse).

Porcentajes de enfermos de Covid-19 con y sin diabetes mellitus que mueren.

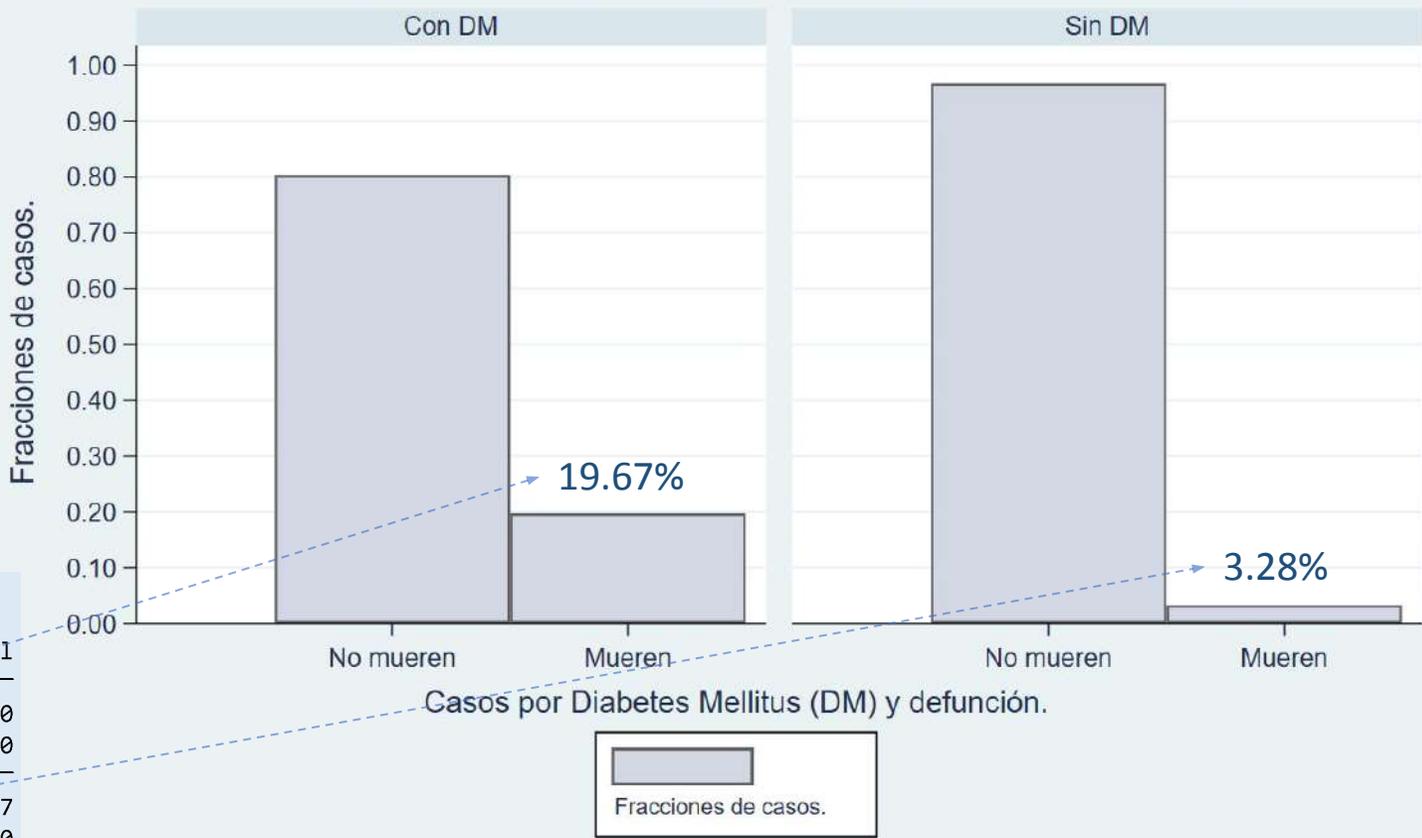
use 01BD Riesgos 13 de agosto 2022.dta

01BD Riesgos 13 de agosto 2022.dta

tab2 diabetes def if rdiab<2 & rdef<2 & rcase==1, row

Diabetes Mellitus (DM)	Defunción		Total
	Mueren	No mueren	
Con DM	120,655 19.67	492,855 80.33	613,510 100.00
Sin DM	206,378 3.28	6,085,559 96.72	6,291,937 100.00
Total	327,033 4.74	6,578,414 95.26	6,905,447 100.00

Covid-19 y Diabetes mellitus (DM). Fracciones de casos que sobreviven y que mueren desde el inicio de la pandemia hasta el 13 de agosto de 2022 en México.



\* Análisis elaborado a partir de los datos abiertos publicados por la Dirección General de Epidemiología el 13 de agosto de 2022 en: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>

Conclusiones: La letalidad (riesgo de morir) en un paciente con Covid-19 que además tiene diabetes mellitus es de 19.67%, la letalidad de los casos sin diabetes mellitus es de 3.28%.



CONFERENCIAS STATA LATAM 2022

**Herramientas y aplicaciones estadísticas para Ciencia de Datos**

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

**Estimación de riesgos  
Riesgo relativo de letalidad por Covid-19 en México y diabetes mellitus.**

Letalidad por Covid-19 en casos con y sin diabetes mellitus.

use 01BD Riesgos 13 de agosto 2022.dta

table (diabetes) (def) if rdiab <2 & rdef <2 & rcase ==1

**Letalidad por Covid-19 y diabetes mellitus en México<sup>1</sup>.**

Riesgo relativo (razón de riesgos) de letalidad por Covid-19 en casos con diabetes mellitus hasta el 13 de agosto de 2022 en México.

	Diabetes mellitus [DM]		
	Exposed	Unexposed	Total
Cases	120655	206378	327033
Noncases	492855	6085559	6578414
<b>Total</b>	<b>613510</b>	<b>6291937</b>	<b>6905447</b>
<b>Risk</b>	<b>.1966635</b>	<b>.0328004</b>	<b>.0473587</b>
	<b>Point estimate</b>	<b>[95% conf. interval]</b>	
<b>Risk difference</b>	<b>.1638631</b>	<b>.1628588</b>	<b>.1648674</b>
<b>Risk ratio</b>	<b>5.995766</b>	<b>5.956315</b>	<b>6.035478</b>
Attr. frac. ex.	.8332156	.832111	.834313
Attr. frac. pop	.3074052		

chi2(1) =332695.71 Pr>chi2 = 0.0000

<sup>1</sup>Elaborado a partir de los datos abiertos publicados por la Dirección General de Epidemiología el 13 de agosto de 2022. Publicados en el sitio: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>

Conclusiones: El riesgo relativo de morir de los pacientes con Covid-19 que además tienen diabetes en comparación de aquellos sin diabetes mellitus es de 5.99 (un exceso de 499%).



**CONFERENCIAS STATA LATAM 2022**  
**Herramientas y aplicaciones estadísticas para Ciencia de Datos**

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
- Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
    - Una muestra.
    - Dos muestras.
      - Independientes.
      - Emparejadas.
    - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

**Estimación de riesgos**  
**Fracción atribuible de letalidad por Covid-19 en México y diabetes mellitus.**

**Letalidad ¿evitable? por Covid-19 y diabetes mellitus.**

use 01BD Riesgos 13 de agosto 2022.dta

cs rdef rdiab if rdiab <2 & rdef <2 & rcase ==1

$120,655 \times 0.8332156 = 100,532$

**Letalidad por Covid-19 y diabetes mellitus en México<sup>1</sup>.**

Riesgo relativo (razón de riesgos) de letalidad por Covid-19 en casos con diabetes mellitus hasta el 13 de agosto de 2022 en México.

	Diabetes mellitus [DM]		
	Exposed	Unexposed	Total
Cases	120655	206378	327033
Noncases	492855	6085559	6578414
<b>Total</b>	<b>613510</b>	<b>6291937</b>	<b>6905447</b>
Risk	<b>.1966635</b>	<b>.0328004</b>	.0473587
	<b>Point estimate</b>		<b>[95% conf. interval]</b>
Risk difference	.1638631	.1628588	.1648674
Risk ratio	5.995766	5.956315	6.035478
<b>Attr. frac. ex.</b>	<b>.8332156</b>	<b>.832111</b>	<b>.834313</b>
Attr. frac. pop	.3074052		

chi2(1) =332695.71 Pr>chi2 = 0.0000

<sup>1</sup>Elaborado a partir de los datos abiertos publicados por la Dirección General de Epidemiología el 13 de agosto de 2022. Publicados en el sitio: <https://www.gob.mx/salud/documentos/datos-abiertos-152127>

**Conclusiones:**  
 La fracción atribuible de letalidad en pacientes con Covid-19 que además tiene diabetes mellitus es de 83.3%.  
 Si se evitara la diabetes en estos casos se evitarían en promedio 100,532 defunciones.



Introducción.  
Panorama epidemiológico.  
Vigilancia epidemiológica.  
Datos y calidad.  
Análisis exploratorio.  
Magnitud.  
Tiempo.  
Lugar.  
Variables personales.  
Condiciones de riesgo.  
Evidencia de causalidad.  
Ejemplo: El lugar.  
Otras condiciones.  
Análisis de la evidencia.  
Comparación de medias.  
Estimación.  
Una muestra.  
Dos muestras.  
Independientes.  
Emparejadas.  
Más de dos muestras.  
Medidas de intervención.  
Intervenciones inespecíficas  
Intervenciones específicas.  
Conclusiones.  
Preguntas.

# Vigilancia epidemiológica.

Análisis de la evidencia.

Inferencia estadística.

Estimación de un parámetro.

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
      - Una muestra.
      - Dos muestras.
        - Independientes.
        - Emparejadas.
      - Más de dos muestras.
  - Medidas de intervención.
    - Intervenciones inespecíficas.
    - Intervenciones específicas.
- Conclusiones.
- Preguntas.

## Comparación de medias (Prueba t de student).

## t de Student.

### Origen del término.

Prueba desarrollada en 1908 por William Sealy Gosset.

Gosset, quien trabajaba en la fábrica de cerveza Guinness, que no permitía que sus empleados publicaran artículos científicos por razones de seguridad industrial, publicaba el resultado de sus desarrollos utilizando el seudónimo de "Student".

### Descripción.

La distribución t de Student estima el valor de una media muestral pequeña que se extrae de una población con una distribución normal, cuya desviación estándar desconocemos.

La distribución t de Student se asemeja a una normal, salvo que esta última tiene las colas más anchas que la t de Student.

No depende de la media y varianza como sucede en la distribución normal, sino de los grados de libertad de la distribución t de Student

### Objetivo.

Estimar una media.

Comparar una media.

Comparar las medias de dos poblaciones.

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
      - Una muestra.
      - Dos muestras.
        - Independientes.
        - Emparejadas.
      - Más de dos muestras.
  - Medidas de intervención.
    - Intervenciones inespecíficas.
    - Intervenciones específicas.
- Conclusiones.
- Preguntas.

**Comparación de medias**  
(Pruebas de t de Student).

Estimación de una media.

Comparación de una media.

Comparación de dos medias independientes.

Comparación de dos medias emparejadas.

## Supuestos.

### Escala cuantitativa continua.

La variable está medida en una **escala cuantitativa continua**.  
Estrictamente no es una prueba recomendada cuando la escala de medición es cuantitativa discreta.

### Normalidad.

Los datos siguen una **distribución de probabilidad normal**.  
En cada grupo los datos siguen una distribución normal respecto a su media.  
No debe haber valores atípicos significativos.

### Homocedasticidad.

Las varianzas de los dos grupos de datos son iguales.

### Independencia.

Los grupos de comparación son independientes.  
No hay relación entre los elementos en un grupo en comparación con los de otro grupo.

### Selección aleatoria.

Cada grupo es una muestra aleatoria de su población de origen.  
Cada elemento tiene una probabilidad conocida de ser seleccionado como parte de la muestra.

# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
    - Una muestra.
    - Dos muestras.
      - Independientes.
      - Emparejadas.
      - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

**Inferencia estadística**  
(estimación de parámetros a partir de estadísticos)

Estimación de una media.

Ejemplo: Estimación, con 95% de confianza, de la media de edad de los casos de Covid-19 (N=6,925,68) en México a partir de una muestra de 500 casos (n=500).

use DA 13 de agosto casos.dta

sum edad 1

sample 500, count 2

mean edad 3

Estadísticas > Muestreo repetido > Exter muestra aleatoria  
Estadísticas > Sumarios, tablas y tests estadísticos > Sumario y estadísticas descriptivas > Medias

```
. summ edad
```

Variable	Obs	Mean	Std. dev.	Min	Max
<span style="border: 1px solid red; border-radius: 50%; padding: 2px;">1</span> edad	6,925,668	39.93159	16.8804	0	147

```
2 . sample 500, count  
(6,925,168 observations deleted)
```

```
. mean edad
```

Mean estimation		Number of obs = 500	
	Mean	Std. err.	[95% conf. interval]
<span style="border: 1px solid red; border-radius: 50%; padding: 2px;">3</span> edad	39.264	.7471929	37.79597 40.73203

$$IC_{0.95} = [37.8, 40.7] \quad ?$$

Conclusión: Con una probabilidad de 95% ( $IC_{95}$ ) el valor promedio (media aritmética) de la edad de los casos de Covid-19 en México (parámetro) coincide con alguno de los valores contenidos en el intervalo desde 37.8 años (límite inferior) hasta 40.7 años (límite superior).

Introducción.  
Panorama epidemiológico.  
Vigilancia epidemiológica.  
Datos y calidad.  
Análisis exploratorio.  
Magnitud.  
Tiempo.  
Lugar.  
Variables personales.  
Condiciones de riesgo.  
Evidencia de causalidad.  
Ejemplo: El lugar.  
Otras condiciones.  
Análisis de la evidencia.  
Comparación de medias.  
Estimación.  
Una muestra.  
Dos muestras.  
Independientes.  
Emparejadas.  
Más de dos muestras.  
Medidas de intervención.  
Intervenciones inespecíficas  
Intervenciones específicas.  
Conclusiones.  
Preguntas.

# Vigilancia epidemiológica.

Análisis de la evidencia.

Comparaciones de medias.

El caso de una muestra.

# CONFERENCIAS STATA LATAM 2022

## Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
  - Una muestra.
  - Dos muestras.
    - Independientes.
    - Emparejadas.
  - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

### Inferencia estadística (Comparación de una media).

#### Una muestra.

Ejemplo: Se requiere probar, con 95% de confianza, si la media aritmética de edad de los casos de Covid-19 en México es diferente de 41 años. Se utilizará una muestra de 500 casos (n=500).

use DA 13 de agosto Casos Muestra de 500.dta

ttest edad == 41

1

```
. ttest edad == 41
One-sample t test
-----
Variable | Obs      Mean   Std. err.   Std. dev.   [95% conf. interval]
-----+-----
edad     | 500     39.264   .7471929    16.70774    37.79597   40.73203

      mean = mean(edad)                                t = -2.3234
H0: mean = 41                                         Degrees of freedom = 499

      Ha: mean < 41                                Ha: mean = 41                                Ha: mean > 41
Pr(T < t) = 0.0103                                Pr(|T| > |t|) = 0.0206                                Pr(T > t) = 0.9897
```

$$H_1: \mu \neq 41$$

$$H_0: \mu = 41 \quad \text{Regla: se rechazará } H_0 \text{ si } p \leq 0.05$$

$$IC_{0.95} = [37.8, 40.7] \quad \text{---} \quad 40.7 < 41$$

Conclusión: Con una probabilidad mayor de 95% ( $IC_{95}$ ) la evidencia muestral contradice a la hipótesis nula ( $H_0$ ).

Decisión: Se rechaza la hipótesis nula ( $H_0$ ).

En castellano: Con 95% de confianza se puede concluir que la media de edad de los casos de Covid-19 en México es diferente de 41 años.



- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
    - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
      - Una muestra.
      - Dos muestras.
        - Independientes.
        - Emparejadas.
        - Más de dos muestras.
  - Medidas de intervención.
    - Intervenciones inespecíficas.
    - Intervenciones específicas.
- Conclusiones.
- Preguntas.

# Vigilancia epidemiológica.

Análisis de la evidencia.  
Comparaciones de medias.

El caso de dos muestras.

Introducción.  
Panorama epidemiológico.  
Vigilancia epidemiológica.  
Datos y calidad.  
Análisis exploratorio.  
Magnitud.  
Tiempo.  
Lugar.  
Variables personales.  
Condiciones de riesgo.  
Evidencia de causalidad.  
Ejemplo: El lugar.  
Otras condiciones.  
Análisis de la evidencia.  
Comparación de medias.  
Estimación.  
Una muestra.  
Dos muestras.  
Independientes.  
Emparejadas.  
Más de dos muestras.  
Medidas de intervención.  
Intervenciones inespecíficas  
Intervenciones específicas.  
Conclusiones.  
Preguntas.

# Vigilancia epidemiológica.

Análisis de la evidencia.

Comparaciones de medias.

Dos muestras independientes.

# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

## Comparación de medias.

Dos muestras independientes.  
Prueba t de Student para muestras independientes.

El estudio de bajo peso al nacer.

## Descripción.

En 1986, en Springfield, Massachusetts, USA, el Baystate Medical Center recopiló datos sobre el peso al nacer de 189 niños con el objetivo de utilizar estos datos para crear algunos modelos para determinar qué factores intervienen en la determinación de si un niño nace con un peso saludable o no.

Se quiere probar que el hábito de fumar durante el embarazo afecta el peso del recién nacido ocasionando menor peso al nacer .

use bpn (lbw).dta

Archivo > abrir

Editor de Datos (Navegación) - [bpn (lbw)]

Archivo Edición Ver Datos Herramientas

	bpn	edad	raza	fuma	fuma2	popre	hta	iu	png
1	0. No	32	1. Blanca	0. No fuma	No fuma	0	0. No	0. No	3860
2	1. Si	19	1. Blanca	1. Si fuma	Fuma	2	0. No	1. Si	1885
3	0. No	19	1. Blanca	1. Si fuma	Fuma	0	1. Si	0. No	3629
4	0. No	22	1. Blanca	1. Si fuma	Fuma	0	0. No	0. No	3132
5	1. Si	22	1. Blanca	1. Si fuma	Fuma	1	0. No	1. Si	2187
6	0. No	22	1. Blanca	0. No fuma	No fuma	0	0. No	0. No	2637
7	1. Si	19	1. Blanca	0. No fuma	No fuma	0	0. No	0. No	2082
8	0. No	18	1. Blanca	1. Si fuma	Fuma	0	0. No	1. Si	2600
9	0. No	18	1. Blanca	1. Si fuma	Fuma	0	0. No	0. No	3856
10	0. No	21	1. Blanca	0. No fuma	No fuma	0	0. No	0. No	3062
11	0. No	25	1. Blanca	0. No fuma	No fuma	0	0. No	0. No	4153
12	0. No	33	1. Blanca	1. Si fuma	Fuma	0	0. No	0. No	3033
13	1. Si	21	1. Blanca	1. Si fuma	Fuma	0	1. Si	0. No	2495
14	0. No	29	1. Blanca	1. Si fuma	Fuma	0	0. No	0. No	2663
15	0. No	18	1. Blanca	1. Si fuma	Fuma	0	0. No	1. Si	3076
16	0. No	22	1. Blanca	0. No fuma	No fuma	0	0. No	0. No	3027
17	0. No	28	1. Blanca	0. No fuma	No fuma	0	0. No	0. No	2877
18	0. No	27	1. Blanca	1. Si fuma	Fuma	0	0. No	0. No	2922
19	1. Si	31	1. Blanca	1. Si fuma	Fuma	1	0. No	0. No	2353

Variables

Filtrar variables aquí

Variable	Etiqueta	Tipo	Formato
<input checked="" type="checkbox"/> raza	Raza	byte	%9.0g
<input checked="" type="checkbox"/> fuma	La madre fumó durante el embarazo	byte	%10.0g
<input checked="" type="checkbox"/> fuma2	La madre fumó durante el embarazo	float	%9.0g
<input checked="" type="checkbox"/> pppre	Antecedente de partos prematuros (número)	byte	%8.0g
<input checked="" type="checkbox"/> hta	Antecedente materno de hipertensión arterial	byte	%8.0g
<input checked="" type="checkbox"/> iu	Irritabilidad uterina	byte	%8.0g
<input type="checkbox"/> nvp	Número de visitas prenatales en el 1er trimestre	int	%8.0g

Propiedades

Variables

Nombre	fuma2
Etiqueta	La madre fumó durante el embarazo
Tipo	float
Formato	%9.0g
Etiqueta de valor	Fuma2
Notas	

Datos

Activo Variables: 9 de 12 Orden: Dataset Obs: 189 Filtro: Apagado Modo: Navegación CAP: NUM

$$H_1: bpn(fuma) < bpn(no fuma)$$

$$H_0: bpn(fuma) \geq bpn(no fuma)$$

Regla: se rechazará  $H_0$  si  $p \leq 0.05$

Conclusión.

Decisión.



```
. ttest edad, by(def)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
No muere	474	38.05485	.742706	16.16986	36.59544	39.51426
Muere	26	61.30769	1.879255	9.582356	57.4373	65.17809
Combined	500	39.264	.7471929	16.70774	37.79597	40.73203
diff		-23.25284	3.203501		-29.54688	-16.9588

```

diff = mean(No muere) - mean(Muere)          t = -7.2586
H0: diff = 0                                Degrees of freedom = 498

Ha: diff < 0                                Ha: diff != 0                                Ha: diff > 0
Pr(T < t) = 0.0000                          Pr(|T| > |t|) = 0.0000                          Pr(T > t) = 1.0000

```

La edad de los casos que sobreviven es, en promedio, 23 años menor que la quienes fallecen.

$H_1: edad(sobreviven) < edad(mueren)$   
 $H_0: edad(sobreviven) \geq edad(mueren)$

**Regla: se rechazará  $H_0$  si  $p \leq 0.05$**

Conclusión: Con una probabilidad mayor de 95% ( $IC_{95}$ ) la evidencia muestral contradice a la hipótesis nula ( $H_0$ ).

Decisión: Se rechaza la hipótesis nula ( $H_0$ ).

# CONFERENCIAS STATA LATAM 2022

## Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas.
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

**Comparación de medias.**  
 Dos muestras independientes.  
 Prueba t de student para muestras independientes.

La pandemia de Covid-19 en México.  
 Se tiene una muestra de 500 casos de Covid-19 seleccionada al azar del total de los casos registrados.

Se quiere probar que la edad de los casos que sobreviven es menor que la de los casos que fallecen.  
 use DA 13 de agosto Casos Muestra de 500

ttest edad, by(def)

En castellano: En el caso de México, la edad de los casos de Covid-19 que han fallecido es, en promedio, 23 años y 3 meses menor que la edad de los casos que han muerto.



# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

## Comparación de medias.

Dos muestras independientes.  
Prueba t de Student para muestras independientes.

## El estudio de bajo peso al nacer.

### Descripción.

En 1986, en Springfield, Massachusetts, USA, el Baystate Medical Center recopiló datos sobre el peso al nacer de 189 niños con el objetivo de utilizar estos datos para crear algunos modelos para determinar qué factores intervienen en la determinación de si un niño nace con un peso saludable o no.

Este conjunto de datos se encuentra disponible en varios formatos y algunas de las variables que incluye son el peso del producto al nacer, la edad, la raza y el peso de la madre y si la madre fumó durante el embarazo.

use `bpn (lbw).dta`

Archivo > abrir

Editor de Datos (Navegación) - [bpn (lbw)]

Archivo Edición Ver Datos Herramientas

	bpn	edad	raza	fuma	fuma2	pppre	hta	iu	png
1	0. No	32	1. Blanca	0. No fuma	No fuma	0	0. No	0. No	3860
2	1. Si	19	1. Blanca	1. Si fuma	Fuma	2	0. No	1. Si	1885
3	0. No	19	1. Blanca	1. Si fuma	Fuma	0	1. Si	0. No	3629
4	0. No	22	1. Blanca	1. Si fuma	Fuma	0	0. No	0. No	3132
5	1. Si	22	1. Blanca	1. Si fuma	Fuma	1	0. No	1. Si	2187
6	0. No	22	1. Blanca	0. No fuma	No fuma	0	0. No	0. No	2637
7	1. Si	19	1. Blanca	0. No fuma	No fuma	0	0. No	0. No	2082
8	0. No	18	1. Blanca	1. Si fuma	Fuma	0	0. No	1. Si	2500
9	0. No	18	1. Blanca	1. Si fuma	Fuma	0	0. No	0. No	3856
10	0. No	21	1. Blanca	0. No fuma	No fuma	0	0. No	0. No	3062
11	0. No	25	1. Blanca	0. No fuma	No fuma	0	0. No	0. No	4153
12	0. No	33	1. Blanca	1. Si fuma	Fuma	0	0. No	0. No	3033
13	1. Si	21	1. Blanca	1. Si fuma	Fuma	0	1. Si	0. No	2495
14	0. No	29	1. Blanca	1. Si fuma	Fuma	0	0. No	0. No	2663
15	0. No	18	1. Blanca	1. Si fuma	Fuma	0	0. No	1. Si	3076
16	0. No	22	1. Blanca	0. No fuma	No fuma	0	0. No	0. No	3027
17	0. No	28	1. Blanca	0. No fuma	No fuma	0	0. No	0. No	2877
18	0. No	27	1. Blanca	1. Si fuma	Fuma	0	0. No	0. No	2922
19	1. Si	31	1. Blanca	1. Si fuma	Fuma	1	0. No	0. No	2353

Variables

Nombre	Etiqueta	Tipo	Formato
<input checked="" type="checkbox"/> fuma	La madre fumó durante el embarazo	byte	%10.0g
<input checked="" type="checkbox"/> fuma2	La madre fumó durante el embarazo	float	%9.0g
<input checked="" type="checkbox"/> pppre	Antecedente de partos prematuros (número)	byte	%8.0g
<input checked="" type="checkbox"/> hta	Antecedente materno de hipertensión arterial	byte	%8.0g
<input checked="" type="checkbox"/> iu	Irritabilidad uterina	byte	%8.0g
<input type="checkbox"/> nvp	Número de visitas prenatales en el 1er trimestre	int	%8.0g
<input checked="" type="checkbox"/> png	Peso al nacer en gramos	int	%8.0g

Propiedades

Variables

Nombre	png
Etiqueta	Peso al nacer en gramos
Tipo	int
Formato	%8.0g
Etiqueta de valor	
Notas	

Datos

Variables: 9 de 12 Orden: Dataset Obs: 189 Filtro: Apagado Modo: Navegación CAP: NUM

Algunas características importantes de este conjunto de datos son que ninguna de las variables predictoras está altamente correlacionada, no faltan valores en los datos y las distribuciones de las variables pueden considerarse normales.

Esos datos se han utilizado como ejemplo para diferentes análisis en varios textos, por ejemplo *“Applied Logistic Regression”* de David W. Hosmer, Stanley Lemeshow y Rodney X. Sturdivant o *“An Introduction to Stata for Health Researchers”* de Svend Juul y Morten Frydenberg.

Introducción.  
Panorama epidemiológico.  
Vigilancia epidemiológica.  
Datos y calidad.  
Análisis exploratorio.  
Magnitud.  
Tiempo.  
Lugar.  
Variables personales.  
Condiciones de riesgo.  
Evidencia de causalidad.  
Ejemplo: El lugar.  
Otras condiciones.  
Análisis de la evidencia.  
Comparación de medias.  
Estimación.  
Una muestra.  
Dos muestras.  
Independientes.  
Emparejadas.  
Más de dos muestras.  
Medidas de intervención.  
Intervenciones inespecíficas  
Intervenciones específicas.  
Conclusiones.  
Preguntas.

# Vigilancia epidemiológica.

Análisis de la evidencia.

Comparaciones de medias.

Dos muestras emparejadas.

### Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

#### Comparación de medias.

Dos muestras emparejados.

Prueba t de student para muestras emparejadas.

#### El estudio Sepsis.

Descripción.

Se administró ibuprofeno a un grupo de pacientes y placebo a otro grupo, se registró inicialmente la temperatura basal, la temperatura a las 2 horas y a las 4 horas, después cada 4 horas hasta las 44 horas, y finalmente a las 72, 96 y 120 horas.

Se quiere probar que la administración de ibuprofeno reduce la temperatura corporal en pacientes sépticos febriles después de dos horas de ser administrado.

Archivo > abrir

	trat	raza	apache	evol	followup	temp0c	temp1c	diftempc
1	Placebo	Blanca	27	Muerto	50	35.22222	34.30889	.8333321
2	Ibuprofeno	Negra	14	Vivo	720	38.66666	37.22222	1.444443
3	Placebo	Negra	33	Muerto	33	38.33333	37.16667	1.166664
4	Ibuprofeno	Blanca	3	Vivo	720	38.33333	37.88889	.4444456
5	Placebo	Blanca	5	Vivo	720	38.55556	38.55556	0
6	Ibuprofeno	Blanca	13	Vivo	720	38.16666	38.33333	-.1666679
7	Ibuprofeno	Blanca	34	Muerto	45	38.38889	37.61111	.7777786
8	Placebo	Blanca	11	Vivo	720	38.27778	37.94445	.3333321
9	Ibuprofeno	Blanca	25	Vivo	720	38.22223	37.61111	.6111145
10	Placebo	Blanca	20	Vivo	720	37.22222	36.66667	.5555534
11	Ibuprofeno	Blanca	21	Vivo	720	38.11111	38.55556	-.4444456
12	Placebo	Blanca	14	Vivo	720	38.61111	38.72222	-.1111107
13	Ibuprofeno	Blanca	19	Muerto	94	38.55556	37.66667	.8888893
14	Placebo	Blanca	23	Vivo	720	37.88889	37.55555	.3333321
15	Ibuprofeno	Blanca	22	Muerto	243	38.44444	38.16666	.2777786
16	Placebo	Blanca	16	Muerto	250	38.11111	38	.1111107
17	Placebo	Negra	17	Vivo	720	38.11111	37.72222	.3888893
18	Placebo	Blanca	14	Vivo	720	38.66666	38.66666	0
19	Ibuprofeno	Blanca	21	Vivo	720	37	37.11111	-.1111145

$$H_1: temp1c < temp0c$$

$$H_0: temp1c \geq temp0c$$

**Regla: se rechazará  $H_0$  si  $p \leq 0.05$**

Conclusión.

Decisión.

## CONFERENCIAS STATA LATAM 2022

# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

## Comparación de medias.

Dos muestras emparejadas.

Prueba t de student para muestras emparejadas.

## El estudio Sepsis.

Descripción.

Se administró ibuprofeno a un grupo de pacientes y placebo a otro grupo, se registró inicialmente la temperatura basal, la temperatura a las 2 horas y a las 4 horas, después cada 4 horas hasta las 44 horas, y finalmente a las 72, 96 y 120 horas.

Se quiere probar que la administración de ibuprofeno reduce la temperatura corporal en pacientes sépticos febriles después de dos horas de ser administrado.

```
ttest temp1c = temp0c if trat == 1
```

```
. ttest temp1c = temp0c if trat == 1
```

Paired t test

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
temp1c	208	37.5117	.0714197	1.030029	37.3709	37.6525
temp0c	208	38.0031	.0829791	1.196742	37.83951	38.16669
diff	208	-.4913999	.0478614	.6902662	-.5857581	-.3970417

mean(diff) = mean(temp1c - temp0c)

t = -10.2672

H0: mean(diff) = 0

Degrees of freedom = 207

Ha: mean(diff) < 0

Ha: mean(diff) != 0

Ha: mean(diff) > 0

Pr(T < t) = 0.0000

Pr(|T| > |t|) = 0.0000

Pr(T > t) = 1.0000

**$H_1: temp1c < temp0c$**

**$H_0: temp1c \geq temp0c$**

**Regla: se rechazará  $H_0$  si  $p \leq 0.05$**

Conclusión: Con una probabilidad mayor de 95% ( $IC_{95}$ ) la evidencia muestral contradice a la hipótesis nula ( $H_0$ ).

Decisión: Se rechaza la hipótesis nula ( $H_0$ ).

La media de temperatura disminuye después de dos horas de administrar ibuprofeno.

En castellano: La administración de ibuprofeno en pacientes sépticos febriles reduce la temperatura corporal 0.49° C, en promedio, después de dos horas de su administración.

Introducción.  
Panorama epidemiológico.  
Vigilancia epidemiológica.  
Datos y calidad.  
Análisis exploratorio.  
Magnitud.  
Tiempo.  
Lugar.  
Variables personales.  
Condiciones de riesgo.  
Evidencia de causalidad.  
Ejemplo: El lugar.  
Otras condiciones.  
Análisis de la evidencia.  
Comparación de medias.  
Estimación.  
Una muestra.  
Dos muestras.  
Independientes.  
Emparejadas.  
Más de dos muestras.  
Medidas de intervención.  
Intervenciones inespecíficas  
Intervenciones específicas.  
Conclusiones.  
Preguntas.

# Vigilancia epidemiológica.

Análisis de la evidencia.

Comparaciones de medias.

Más de dos muestras.

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
    - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
      - Una muestra.
      - Dos muestras.
        - Independientes.
        - Emparejadas.
      - Más de dos muestras.
  - Medidas de intervención.
    - Intervenciones inespecíficas.
    - Intervenciones específicas.
- Conclusiones.
- Preguntas.

## Comparación de más de dos medias (Análisis de varianza - ANOVA).

# Análisis de varianza.

## Origen del término.

ANOVA es la contracción del nombre en inglés, **AN**alisis **Of** **VA**riance.

Fue desarrollada por el biólogo Ronald Aylmer Fisher en 1919, cuando trabajaba en una estación agrícola experimental donde se requería grandes datos de cultivos de casi un siglo originando la genética de poblaciones, el diseño experimental y el concepto de hipótesis nula y fue un fuerte opositor de la estadística bayesiana y defendió que una correlación positiva entre dos factores no implica necesariamente una asociación causal entre ellos.

## Descripción.

Método para comparar las medias de más de dos poblaciones utilizado en general, pero no exclusivamente, para el análisis de diseños experimentales.

## Objetivo.

Analizar simultáneamente las medias de más de dos poblaciones.

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
      - Una muestra.
      - Dos muestras.
        - Independientes.
        - Emparejadas.
      - Más de dos muestras.
    - Medidas de intervención.
      - Intervenciones inespecíficas.
      - Intervenciones específicas.
  - Conclusiones.
  - Preguntas.

## Comparación de más de dos medias (Análisis de varianza - ANOVA).

## Supuestos.

### Escala cuantitativa continua.

La variable está medida en una **escala cuantitativa continua**.  
Estrictamente no es una prueba recomendada cuando la escala de medición es cuantitativa discreta.

### Normalidad.

Los datos siguen una distribución de probabilidad normal.  
En cada grupo los datos siguen una **distribución normal** respecto a su media.  
No debe haber valores atípicos significativos.

### Homocedasticidad.

Las varianzas de las poblaciones de procedencia de los grupos son iguales.  
Esta característica se conoce como **homocedasticidad**.

### Independencia.

Los grupos de comparación son independientes.  
No hay relación entre los elementos en un grupo en comparación con los de otro grupo.

### Selección aleatoria.

Cada grupo es una muestra aleatoria de su población de origen.  
Cada elemento tiene una probabilidad conocida de ser seleccionado como parte de la muestra.

# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

## Comparación de más de dos medias.

Tres muestras independientes.

Prueba análisis de varianza de una vía.

## El estudio Bajo Peso al nacer (bpn).

Se tiene una muestra de 189 recién nacidos de mujeres con diferentes características.

Se quiere probar el peso al nacer es diferente para los recién nacidos de madres de diferente raza.

use bpn (lbw).dta

Archivo > abrir

	id	bpn	edad	raza	pppre	fuma	hta	png	
1	207	0. No	32	1. Blanca	0	0. No fuma	0. No	3860	
2	23	1. Sí	19	1. Blanca	2	1. Sí fuma	0. No	1885	
3	187	0. No	19	1. Blanca	0	1. Sí fuma	1. Sí	3629	
4	148	0. No	22	1. Blanca	0	1. Sí fuma	0. No	3132	
5	42	1. Sí	22	1. Blanca	1	1. Sí fuma	0. No	2187	
6	92	0. No	22	1. Blanca	0	0. No fuma	0. No	2637	
7	33	1. Sí	19	1. Blanca	0	0. No fuma	0. No	2082	
8	89	0. No	18	1. Blanca	0	1. Sí fuma	0. No	2608	
9	205	0. No	18	1. Blanca	0	1. Sí fuma	0. No	3856	
10	131	0. No	21	1. Blanca	0	0. No fuma	0. No	3062	
11	221	0. No	25	1. Blanca	0	0. No fuma	0. No	4153	
12	127	0. No	33	1. Blanca	0	1. Sí fuma	0. No	3033	
13	84	1. Sí	21	1. Blanca	0	1. Sí fuma	1. Sí	2495	
14	94	0. No	29	1. Blanca	0	1. Sí fuma	0. No	2663	
15	133	0. No	18	1. Blanca	0	1. Sí fuma	0. No	3076	
16	204	0. No	22	1. Blanca	0	0. No fuma	0. No	3827	
17	112	0. No	28	1. Blanca	0	0. No fuma	0. No	2877	
18	125	0. No	27	1. Blanca	0	1. Sí fuma	0. No	2922	
19	56	1. Sí	31	1. Blanca	1	1. Sí fuma	0. No	2353	

$H_1$ : Al menos en una de las parejas posibles el peso al nacer del producto es diferente

$H_0$ :  $png(Blanca) = png(Negra) = png(Otra)$

Regla: se rechazará  $H_0$  si  $p \leq 0.05$

Conclusión.

Decisión.

CONFERENCIAS STATA LATAM 2022

Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

Comparación de más de dos medias.

Tres muestras independientes.

Prueba análisis de varianza de una vía.

El estudio Bajo Peso al nacer (bpn).

Se tiene una muestra de mujeres que tuvieron un parto.

Se quiere probar el peso al nacer es diferente para los recién nacidos de madres de diferente raza.

Las medias y el análisis de varianza.  
use bpn (lbw).dta

oneway png raza, tabulate

```
. oneway png raza, tabulate
```

Raza	Summary of Peso al nacer en gramos		
	Mean	Std. dev.	Freq.
1. Blanca	3103.0104	727.87244	96
2. Negra	2719.6923	638.68388	26
3. Otra	2804.0149	721.30115	67
Total	2944.2857	729.01602	189

Analysis of variance					
Source	SS	df	MS	F	Prob > F
Between groups	5048361.06	2	2524180.53	4.95	0.0081
Within groups	94866937.5	186	510037.298		
Total	99915298.6	188	531464.354		

Bartlett's equal-variances test: chi2(2) = 0.6560 Prob>chi2 = 0.720

$H_1$ : al menos en una de las parejas posibles el peso al nacer del producto es diferente

$H_0$ :  $png(Blanca) = png(Negra) = png(Otra)$

Regla: rechaza  $H_0$  si  $p \leq 0.05$

Conclusión: Con una probabilidad de 95% ( $IC_{95}$ ) la evidencia muestral contradice a la hipótesis nula ( $H_0$ ).

Decisión: Se rechaza la hipótesis nula ( $H_0$ ).

En castellano: Al menos en una de las parejas posibles de los pesos de los recién nacidos de madres de diferente raza es diferente.

¿Cuál pareja o cuales parejas serán diferentes?



CONFERENCIAS STATA LATAM 2022

Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

**Comparación de más de dos medias.**

Tres muestras independientes.

Prueba análisis de varianza de una vía.

Pruebas post-hoc.

(Stata ofrece tres opciones: Bonferroni, Scheffe y Sidak)

**El estudio Bajo Peso al nacer (bpn).**

Se tiene una muestra de mujeres que tuvieron un parto.

Se quiere probar el peso al nacer es diferente para los recién nacidos de madres de use bpn (lbw).dta diferente raza.

oneway png raza, bonferroni tabulate noanova

Raza	Summary of Peso al nacer en gramos		
	Mean	Std. dev.	Freq.
1. Blanca	3103.0104	727.87244	96
2. Negra	2719.6923	638.68388	26
3. Otra	2804.0149	721.30115	67
Total	2944.2857	729.01602	189

Comparison of Peso al nacer en gramos by Raza (Bonferroni)

Row Mean- Col Mean	1. Blanc	2. Negra
2. Negra	-383.318 <b>0.048</b>	
3. Otra	-298.995 <b>0.028</b>	84.3226 1.000

$H_1$ : al menos en una de las parejas posibles el peso al nacer del producto es diferente

$H_0$ :  $png(Blanca) = png(Negra) = png(Otra)$

**Regla: rechaza  $H_0$  si  $p \leq 0.05$**

Conclusión: Con una probabilidad de 95% ( $IC_{95}$ ) la evidencia muestral contradice a la hipótesis nula ( $H_0$ ).

Decisión: Se rechaza la hipótesis nula ( $H_0$ ).

Nota aclaratoria: No es plausible que la raza, en este caso el color de la piel, por si misma ocasione el bajo peso al nacer sino que otras condiciones asociadas a esta variable son las que probablemente sean las responsables de este efecto.

En castellano: El peso promedio de los recién nacidos de madres de raza blanca es diferente los pesos de los productos de las mujeres de las otras razas. Los de las mujeres de raza negra y de otra raza no lo son.



CONFERENCIAS STATA LATAM 2022

Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

Comparación de más de dos medias.

Tres muestras independientes.

Prueba análisis de varianza de una vía.

La edad de los casos de Covid-19 en tres estados de México.

Se quiere probar que las edades de los casos de Covid-19 en tres estados, Coahuila, Michoacán y Puebla, son diferentes.

use DA 13 de agosto Casos Muestra de 500

Archivo > abrir

Editor de Datos (Navegación) - [sepsis]

Archivo Edición Ver Datos Herramientas

	trat	raza	apache	evol	followup	temp0c	temp1c	diftempc
16	Placebo	Blanca	16	Muerto	250	38.11111	38	.1111107
17	Placebo	Negra	17	Vivo	720	38.11111	37.72222	-.3888893
18	Placebo	Blanca	14	Vivo	720	38.66666	38.66666	0
19	Ibuprofeno	Blanca	21	Vivo	720	37	37.11111	-.1111145
20	Ibuprofeno	Otra	20	Vivo	720	39.22222	37.72222	1.5
21	Placebo	Blanca	24	Muerto	215	39.27777	39.38889	-.1111145
22	Ibuprofeno	Blanca	19	Muerto	654	38.5	37.61111	.8888893
23	Placebo	Blanca	14	Vivo	720	38.44444	38.66666	-.2222214
24	Ibuprofeno	Blanca	26	Muerto	582	38.27778	37.88889	.3888931
25	Placebo	Blanca	22	Vivo	720	35.66666	36.77777	-1.1111111
26	Ibuprofeno	Negra	23	Muerto	295	38.66666	38.16666	.5
27	Placebo	Blanca	18	Muerto	266	38.27778	37.61111	.6666679
28	Placebo	Negra	16	Vivo	720	39.27777	37.38889	1.888885
29	Ibuprofeno	Blanca	6	Vivo	720	37.27778	37.16667	.1111107
30	Placebo	Blanca	16	Vivo	720	38.77778	39.44444	-.6666641
31	Ibuprofeno	Negra	25	Vivo	720	35.27778	36.27778	-1
32	Ibuprofeno	Blanca	12	Vivo	720	38.5	38.27778	.2222214
33	Placebo	Blanca	11	Vivo	720	39	38.5	.5
34	Placebo	Blanca	17	Vivo	720	37.77778	37.77778	0

Variables

Filtrar variables aquí

Variable	Etiqueta	Tipo	Formato
<input checked="" type="checkbox"/> Nombre	Etiqueta		
<input checked="" type="checkbox"/> raza	Raza	float	%9.0g
<input checked="" type="checkbox"/> apache	Baseline APACHE Score	float	%9.0g
<input checked="" type="checkbox"/> evol	Evolución mortal a los 30 días	float	%9.0g
<input checked="" type="checkbox"/> followup	Follow-up (hours)	float	%9.0g
<input checked="" type="checkbox"/> temp0c	Temperatura basal	float	%9.0g
<input checked="" type="checkbox"/> temp1c	Temperatura después de 2 horas	float	%9.0g
<input checked="" type="checkbox"/> diftempc	Diferencia de temperatura a las 2 horas en grados centigrados	float	%9.0g

Propiedades

Variables

Nombre	diftempc
Etiqueta	Diferencia de temperatura a las 2 horas en grados cen
Tipo	float
Formato	%9.0g
Etiqueta de valor	
Notas	

Datos

Activo Variables: 8 de 13 Orden: Dataset Obs: 455 Filtro: Apagado Modo: Navegación CAP NUM

$H_1$ : Al menos en una de las parejas posibles la edad de los casos es diferente

$H_0$ : edad(Coahuila) = edad(Michoacán) = edad(Puebla)

Regla: se rechazará  $H_0$  si  $p \leq 0.05$

Conclusión.

Decisión.

CONFERENCIAS STATA LATAM 2022

Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

Comparación de más de dos medias.

Tres muestras independientes.

Prueba análisis de varianza de una vía.

La edad de los casos de Covid-19 en tres estados de México.

Se quiere probar que las edades de los casos de Covid-19 en tres estados, Coahuila, Michoacán y Puebla, son diferentes.

use DA 13 de agosto Casos Muestra de 500

```
oneway edad entidad_res if entidad_res == 5 | entidad_res == 16 | entidad_res == 21, tabulate
```

```
. oneway edad entidad_res if entidad_res == 5 | entidad_res == 16 | entidad_res == 21, tabulate
```

Estado de residencia	Summary of Edad en años cumplidos		
	Mean	Std. dev.	Freq.
Coahuila	31.636364	15.324965	11
Michoacán	51.166667	16.819001	12
Puebla	41.785714	16.56771	14
Total	41.810811	17.652255	37

Analysis of variance					
Source	SS	df	MS	F	Prob > F
Between groups	2189.10641	2	1094.55321	4.12	0.0250
Within groups	9028.56926	34	265.546155		
Total	11217.6757	36	311.602102		

Bartlett's equal-variances test: chi2(2) = 0.0982 Prob>chi2 = 0.952

$H_1$ : al menos en una de las parejas posibles la edad de los casos es diferente

$H_0$ : edad(Coahuila) = edad(Michoacán) = edad(Puebla)

Regla: se rechazará  $H_0$  si  $p \leq 0.05$

Conclusión: Con una probabilidad de 95% ( $IC_{95}$ ) la evidencia muestral contradice a la hipótesis nula ( $H_0$ ).

Decisión: Se rechaza la hipótesis nula ( $H_0$ ).

En castellano: Al menos en una de las parejas posibles de los estados de Coahuila, Michoacán y Puebla la edad promedio de los casos de Covid-19 es diferente.

¿Cuál pareja o cuales parejas serán diferentes?



CONFERENCIAS STATA LATAM 2022

Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

**Comparación de más de dos medias.**

Tres muestras independientes.  
 Prueba análisis de varianza de una vía.  
 Pruebas post-hoc.  
 (Stata ofrece tres opciones: Bonferroni, Scheffe y Sidak)

La edad de los casos de Covid-19 en tres estados de México.

Se quiere probar que las edades de los casos de Covid-19 en tres estados, Coahuila, Michoacán y Puebla, son diferentes.

use DA 13 de agosto Casos Muestra de 500

```
oneway edad entidad_res if entidad_res == 5 |
entidad_res == 16 | entidad_res == 21, bonferroni
tabulate noanova
```

Estado de residencia	Summary of Edad en años cumplidos		
	Mean	Std. dev.	Freq.
Coahuila	31.636364	15.324965	11
Michoacán	51.166667	16.819001	12
Puebla	41.785714	16.56771	14
Total	41.810811	17.652255	37

Comparison of Edad en años cumplidos by Estado de residencia (Bonferroni)

Row Mean- Col Mean	Coahuila	Michoacán
Michoacán	19.5303 0.021	
Puebla	10.1494 0.394	-9.38095 0.458

$H_1$ : al menos en una de las parejas posibles la edad de los casos es diferente

$H_0$ : edad(Coahuila) = edad(Michoacán) = edad(Puebla)

**Regla: se rechazará  $H_0$  si  $p \leq 0.05$**

Conclusión: Con una probabilidad de 95% ( $IC_{95}$ ) la evidencia muestral contradice a la hipótesis nula ( $H_0$ ).

Decisión: Se rechaza la hipótesis nula ( $H_0$ ).

En castellano: La edad promedio de los casos de Covid-19 de los estados de Coahuila y Michoacán es diferente. Las edades de Coahuila y Puebla y de Michoacán y Puebla no lo son.



- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
    - Una muestra.
    - Dos muestras.
      - Independientes.
      - Emparejadas.
    - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

# Vigilancia epidemiológica.

Medidas de intervención.

Protección inespecífica.

Limitación del contacto.

## Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
  - Una muestra.
  - Dos muestras.
    - Independientes.
    - Emparejadas.
  - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

Existe evidencia abrumadora de que la inhalación del coronavirus 2 del síndrome respiratorio agudo severo (SARS-CoV-2) representa una ruta de transmisión importante para la enfermedad por coronavirus 2019 (COVID-19). Existe una necesidad urgente de armonizar los debates sobre los modos de transmisión del virus en todas las disciplinas para garantizar las estrategias de control más eficaces y proporcionar una orientación clara y coherente al público. Para ello, debemos aclarar la terminología para distinguir entre aerosoles y gotitas utilizando un umbral de tamaño de 100  $\mu\text{m}$ , no los 5  $\mu\text{m}$  históricos ( 1 ). Este tamaño separa de manera más efectiva su comportamiento aerodinámico, la capacidad de ser inhalado y la eficacia de las intervenciones.

Los virus en gotitas (mayores de 100  $\mu\text{m}$ ) normalmente caen al suelo en segundos a 2 m de la fuente y pueden ser rociados como pequeñas balas de cañón sobre individuos cercanos. Debido a su rango de recorrido limitado, el distanciamiento físico reduce la exposición a estas gotas. Los virus en aerosoles (menos de 100  $\mu\text{m}$ ) pueden permanecer suspendidos en el aire durante muchos segundos a horas, como el humo, y ser inhalados. Están muy concentrados cerca de una persona infectada, por lo que pueden infectar a las personas más fácilmente en las proximidades. Pero los aerosoles que contienen virus infecciosos ( 2 ) también pueden viajar más de 2 m y acumularse en el aire interior mal ventilado, lo que lleva a eventos de superpropagación ( 3 ).

Las personas con COVID-19, muchas de las cuales no presentan síntomas, liberan miles de aerosoles cargados de virus y muchas menos gotitas al respirar y hablar ( 4 - 6 ). Por lo tanto, es mucho más probable que uno inhale aerosoles que una gota ( 7 ), por lo que el equilibrio de la atención debe centrarse en la protección contra la transmisión aérea. Además de los mandatos existentes sobre el uso de mascarillas, el distanciamiento social y los esfuerzos de higiene, instamos a los funcionarios de salud pública a que agreguen una guía clara sobre la importancia de trasladar las actividades al aire libre, mejorar el aire interior mediante ventilación y filtración y mejorar la protección para los trabajadores de alto riesgo. ( 8 ).

### Transmisión aérea de SARS-CoV-2

Kimberly A. Prather <sup>1,\*</sup>, Linsey C. Marr <sup>2,\*</sup>, Robert T. Schooley <sup>3</sup>, Melissa A. McDiarmid

+ Ver todos los autores y afiliaciones

Science 16 de octubre de 2020:  
Vol. 370, Número 6514, págs. 303-304  
DOI: 10.1126 / science.abf0521

16 de octubre de 2020

Vol 370, número 6514



<sup>1</sup> Fuente:

[https://science.sciencemag.org/content/370/6514/303.2?utm\\_campaign=toc\\_sci-mag\\_2020-10-15&et rid=40167825&et\\_cid=3521427](https://science.sciencemag.org/content/370/6514/303.2?utm_campaign=toc_sci-mag_2020-10-15&et rid=40167825&et_cid=3521427)

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

**nature**

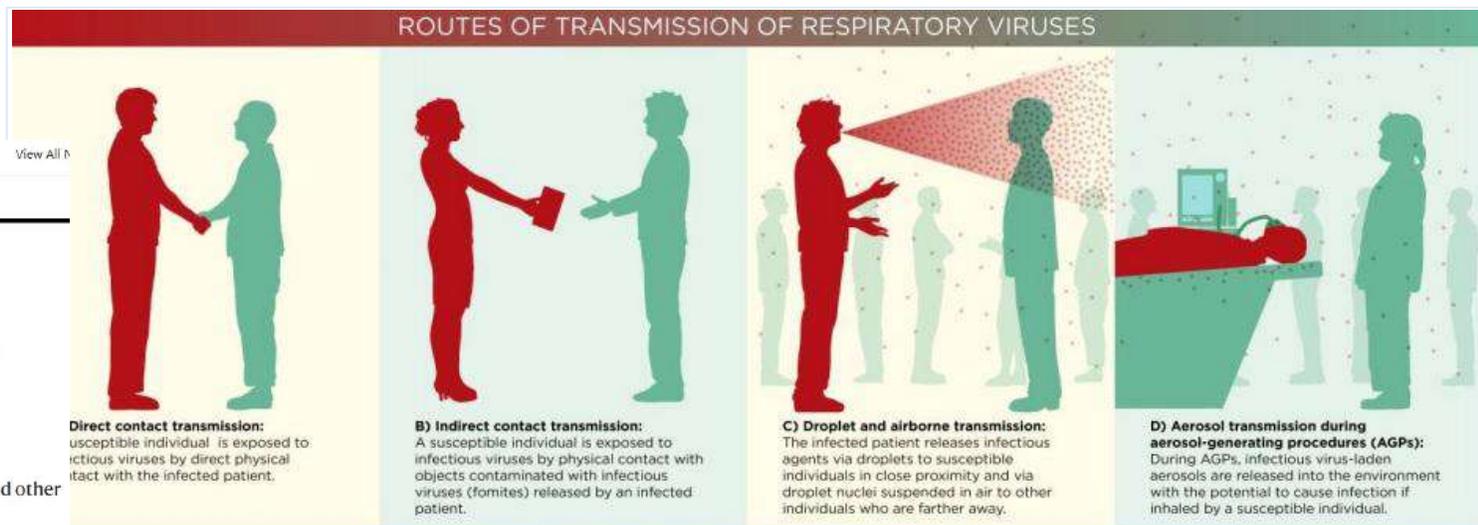
Explore Content | Journal Information | Publish With Us | Subscribe

nature > news feature > article

NEWS FEATURE · 29 JANUARY 2021

## COVID-19 rarely spreads through surfaces. So why are we still deep cleaning?

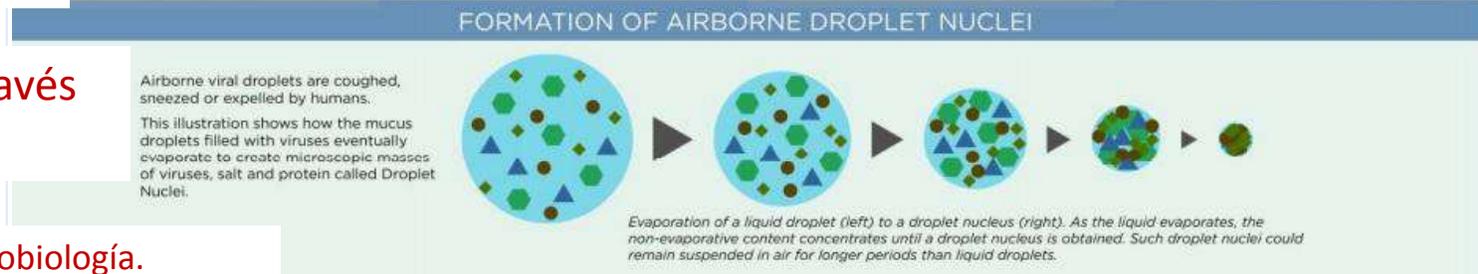
The coronavirus behind the pandemic can linger on doorknobs and other surfaces, but these aren't a major source of infection.



## COVID-19 rara vez se propaga a través de superficies<sup>1</sup>.

Emanuel Goldman. Profesor de microbiología. Facultad de Medicina de Rutgers New Jersey en Newark

Puede permanecer en las manijas de las puertas y otras superficies, pero estas no son una fuente importante de infección.



## Distanciamiento<sup>2</sup>.

<sup>2</sup> Fuente: <https://www.intramed.net/contenido.asp?contenido=96298>



<sup>1</sup> Fuente: [https://www.nature.com/articles/d41586-021-00251-4?fbclid=IwAR23\\_J60hJsKLTATMpTQaKajQud1oDhvvgVOEGX4zOxxvj7eOjUJoyzLFs](https://www.nature.com/articles/d41586-021-00251-4?fbclid=IwAR23_J60hJsKLTATMpTQaKajQud1oDhvvgVOEGX4zOxxvj7eOjUJoyzLFs)



# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.



## Riesgo de transmisión del SARS-CoV-2 de personas asintomáticas en diferentes entornos y para diferentes tiempos de ocupación, ventilación y niveles de hacinamiento (ignorando la variación en la susceptibilidad y las tasas de diseminación viral)

Tipo y nivel de actividad grupal	Ocupación baja			Ocupación alta		
	Al aire libre y bien ventilado	Interior y bien ventilado	Mal ventilado	Al aire libre y bien ventilado	Interior y bien ventilado	Mal ventilado
<b>Utilizando cubrebocas, contacto durante poco tiempo</b>						
Silencioso	Green	Green	Green	Green	Green	Yellow
Hablando	Green	Green	Green	Green	Green	Yellow
Gritando o cantando	Green	Green	Yellow	Yellow	Yellow	Red
<b>Utilizando cubrebocas, contacto durante tiempo prolongado</b>						
Silencioso	Green	Green	Yellow	Green	Yellow	Red
Hablando	Green	Green	Yellow	Yellow	Yellow	Red
Gritando o cantando	Green	Yellow	Red	Yellow	Red	Red
<b>Sin utilizar cubrebocas, contacto durante poco tiempo</b>						
Silencioso	Green	Green	Yellow	Yellow	Yellow	Red
Hablando	Green	Yellow	Yellow	Yellow	Red	Red
Gritando o cantando	Yellow	Yellow	Red	Red	Red	Red
<b>Sin utilizar cubrebocas, contacto durante tiempo prolongado</b>						
Silencioso	Green	Yellow	Red	Yellow	Red	Red
Hablando	Yellow	Yellow	Red	Red	Red	Red
Gritando o cantando	Yellow	Red	Red	Red	Red	Red

Riesgo de transmisión  
 Bajo ■ Moderado ■ Alto ■

\* Caso límite que depende en gran medida de las definiciones cuantitativas de distanciamiento, número de individuos y tiempo de exposición.

Nicholas R Jones et al. BMJ 2020;370:bmj.m3223

©2020 by British Medical Journal Publishing Group <https://www.bmj.com/content/370/bmj.m3223>



<sup>1</sup> Fuente: <https://www.bmj.com/content/370/bmj.m3223>



- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
      - Una muestra.
      - Dos muestras.
        - Independientes.
        - Emparejadas.
      - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

# Vigilancia epidemiológica.

Medidas de intervención.

Protección inespecífica.

Acciones contra el agente.

## Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
    - Una muestra.
    - Dos muestras.
      - Independientes.
      - Emparejadas.
    - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

### Agua y jabón... un invisible acto de magia.

- Cuando la piel se contamina con estos coronavirus, los virus no pueden penetrar la piel, pero se quedan esperando la oportunidad para encontrar alguna "puerta de entrada", como las mucosas de la boca y de la nariz o las conjuntivas de los ojos.
- Este coronavirus está protegido por una doble envoltura de lípidos (grasas).
- Al lavarse las manos o la piel con agua y jabón uno de los extremos de las moléculas de jabón se une a las grasas, en este caso a la envoltura que protege al virus y el otro extremo se une a las moléculas de agua.
- Las fuerzas de atracción de estas uniones son tan fuertes que desprenden la grasa de la envoltura, separándola en pedazos cada vez más pequeños, dejando el interior del virus sin protección, lo que termina destruyéndolo.
- Los pedazos de grasa que formaban la envoltura y los restos del virus son arrastrados por el agua que en su recorrido los elimina de la superficie donde se encontraban.

Para que esta **magia** surta efecto es necesario **mantener el acto** durante **20 segundos**.



### La magia del jabón... y del agua.



- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
      - Una muestra.
      - Dos muestras.
        - Independientes.
        - Emparejadas.
      - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

# Vigilancia epidemiológica.

Medidas de intervención.

Protección específica.

Tratamiento y vacunación.

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

## El remdesivir y el interferón fracasan en el megastudio de la OMS sobre los tratamientos del COVID-19

Por Kai Kupferschmidt | 16 de octubre de 2020 3:45 a. M.

Ciencia • COVID-19 informes s es apoyado por el Centro Pulitzer y la Fundación Heising-Simons

SUSCRÍBETE A NUESTRO BOLETÍN DIARIO

(Obtenga más contenido excelente como este directamente para usted)

Dirección de correo electrónico \*

Uno de los ensayos más grandes del mundo de las terapias COVID-19 publicó ayer sus tan esperados resultados provisionales, y son una decepción. Ninguno de los cuatro tratamientos del ensayo Solidarity, en el que se inscribieron más de 11.000 pacientes en 400 hospitales de todo el mundo, aumentó la supervivencia, ni siquiera el tan publicitado medicamento antiviral remdesivir. Científicos de la Organización Mundial de la Salud (OMS) publicaron los datos como una preimpresión en medRxiv anoche antes de su publicación prevista en el *New England Journal of Medicine*.



Sin embar  
aclarar los  
contra el C  
mostrado  
dice Jerem  
funcione, p  
dice el cier

Acceso /

Temas de salud ▼ Países ▼ Centro de prensa ▼ Emergencias ▼

## La OMS interrumpe los grupos de tratamiento de la COVID-19 con hidroxiquina y con la combinación lopinavir/ritonavir

4 de julio de 2020 | Comunicado de prensa

La OMS ha aceptado hoy la recomendación del Grupo directivo internacional del ensayo Solidarity de interrumpir los grupos de tratamiento de la COVID-19 con hidroxiquina y lopinavir/ritonavir. La OMS puso en marcha este ensayo con objeto de encontrar un tratamiento eficaz para los pacientes hospitalizados con COVID-19.

El Grupo directivo internacional formuló esta recomendación tras comparar los resultados provisionales obtenidos en este ensayo entre el grupo de tratamiento con hidroxiquina y el grupo de tratamiento con lopinavir/ritonavir con el tratamiento de referencia, así como los datos procedentes de todos los ensayos presentados en la Cumbre de investigación e innovación sobre la COVID-19 organizada por la OMS los días 1 y 2 de julio.

De acuerdo con estos resultados provisionales, ni la hidroxiquina y la combinación lopinavir/ritonavir reducen la mortalidad en los enfermos de COVID-19 hospitalizados o dan lugar a una disminución muy leve. Por consiguiente, los investigadores del ensayo Solidarity interrumpirán de inmediato estos tratamientos.

Aunque estos resultados provisionales no proporcionan pruebas sólidas de que alguno de estos fármacos cause un aumento de la mortalidad, al examinar los datos analíticos y clínicos obtenidos en el estudio Discovery, vinculado al ensayo Solidarity, se observaron indicios de posibles efectos adversos que también se notificarán en una publicación con revisión científica externa.

Esta decisión se aplica a la continuación del ensayo Solidarity en pacientes hospitalizados pero no afecta a la posible evaluación de los efectos de la hidroxiquina y de la combinación lopinavir/ritonavir en pacientes no hospitalizados como profilaxis previa o posterior a la exposición a la COVID-19. Se están ultimando los resultados provisionales del ensayo Solidarity con el fin de someterlos a una revisión científica externa y publicarlos.

# Tratamiento antiviral específico.

SECCIONES

SIPSE.COM  
Información en todo momento



## Por información falsa, SocArXiv retira estudio de la CDMX sobre ivermectina

El documento tenía ya 11 mil descargas antes de que fuera retirado de la plataforma.

POR EL UNIVERSAL | SÁBADO, 5 DE FEBRERO, 2022



Hoteles en Zacatlán Desde \$452 Reservo ahora

Hoteles en Tecoluitla Desde \$552 Reservo ahora

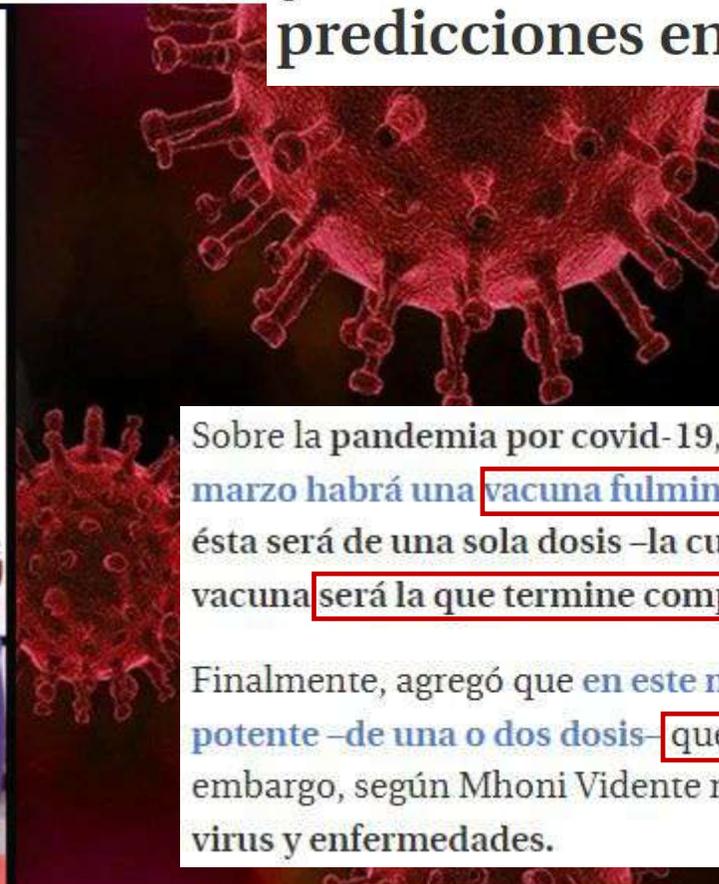
Hoteles en Acapulco Desde \$439 Reservo ahora

## ¿La ciencia no es una opción?

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
  - Una muestra.
  - Dos muestras.
    - Independientes.
    - Emparejadas.
  - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas.
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

¿En marzo 2021 se acabará la pandemia por covid-19? ¿No habrá más contagios?  
¿Habrá nuevos medicamentos? Mhoni Vidente predice lo que sucederá en México y el mundo con el coronavirus.

### ¿Fin del covid-19? Mhoni Vidente lanza predicciones en México para marzo 2021



Sobre la pandemia por covid-19, Mhoni Vidente reveló que **en el mes de marzo habrá una vacuna fulminante contra el coronavirus.** Aseguró que ésta será de una sola dosis –la cual surgirá en EU o Israel–. Explicó que esta vacuna **será la que termine completamente con los contagios.**

Finalmente, agregó que **en este mes saldrá a la luz un medicamento potente –de una o dos dosis– que podría acabar con el covid-19.** Sin embargo, según Mhoni Vidente no debemos confiarnos, pues prevé nuevos virus y enfermedades.

Mhoni Vidente predice para marzo 2021: el fin del covid-19, nuevos medicamentos y guerras. (Especial)

<sup>1</sup> Fuente: <https://www.milenio.com/virales/mhoni-vidente-predice-covid19-mexico-predicciones-marzo-2021>

## OMS<sup>1</sup>. Vacunas: Lista para uso de emergencia (1 de 4).

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

Status of COVID-19 Vaccines within WHO EUL/PQ evaluation process

	Manufacturer / WHO EUL holder	Name of Vaccine	NRA of Record	Platform	EOI accepted	Pre-submission meeting held	Dossier accepted for review*	Status of assessment**	Decision date***
1.	 BioNTech Manufacturing GmbH	BNT162b2/COMIRNATY Tozinameran (INN)	EMA	Nucleoside modified mRNA	✓	✓	✓	Finalized:	31/12/2020
USFDA			Additional sites: – Baxter Oncology GmbH Germany (DP) 30/06/2021 – Novartis Switzerland 08/07/2021 – Mibe (Dermapharm) Germany (DP) 16/07/2021 – Delpharm, Saint-Remy FRANCE (DP) 17/09/2021 – Sanofi-Aventis Deutschland GmbH Germany (DP) 18/06/2021 – Siegfried Hameln GmbH, Germany (DP) 11/11/2021 – Patheon Italia S.p.A, Italy (DP) 07/12/2021 – Catalent Agnani 21/01/2022 – Exela Pharma Sciences, LLC, NC 16/03/2022  Diluent suppliers: – Pfizer Perth, Australia 20/09/2021 – Fresenius Kabi, USA 20/09/2021 – Pfizer Manufacturing Belgium 30/11/2021 – Kwang Myung Pharm Co., Ltd. 14/01/2022  Shelf life extension: 12 months at -70 to -90°C 18/05/2022 Booster dose approved for adults 18 years of age and older 17/12/2022 Age extension to children 5 – 11 years of age 12/02/2022						
2.	 AstraZeneca, AB	AZD1222 Vaxzevria	EMA	Recombinant ChAdOx1 adenoviral vector encoding the Spike protein antigen of the SARS-CoV-2.	✓	✓	✓	Finalized:	16 April 2021
MFDS KOREA			Core data finalized					16 April 2021	
Japan MHLW/PMDA			Finalized: Additional sites: – SK-Catalent 16/04/2021 – Wuxi (DS) 30/04/2021 – Chemo Spain 04/06/2021 – Amylin Ohio US (DP) 23/07/2021 – Wuxi Biologics, Germany (DP) 06/03/2022					15 Feb 2021	
Australia TGA			Finalized					09 July 2021	
			Additional sites: Nipro Pharma Corporation Ise, Japan 11 October 2021					09 July 2021	
5.					✓	✓	Finalized Additional site: Siam Bioscience Co., Ltd Thailand	11 October 2021	

<sup>1</sup> Fuente: [https://extranet.who.int/pqweb/sites/default/files/documents/Status\\_COVID\\_VAX\\_07July2022.pdf](https://extranet.who.int/pqweb/sites/default/files/documents/Status_COVID_VAX_07July2022.pdf)  
 Nota: Información actualizada hasta el 2 de septiembre de 2022.

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
- Magnitud.
- Tiempo.
- Lugar.
- Variables personales.
- Condiciones de riesgo.
- Evidencia de causalidad.
- Ejemplo: El lugar.
- Otras condiciones.
- Análisis de la evidencia.
- Comparación de medias.
- Estimación.
- Una muestra.
- Dos muestras.
- Independientes.
- Emparejadas.
- Más de dos muestras.
- Medidas de intervención.
- Intervenciones inespecíficas
- Intervenciones específicas.
- Conclusiones.
- Preguntas.

## OMS<sup>1</sup>. Vacunas: Lista para uso de emergencia (2 de 4).

Vaccines								Guidance Document 07 July 2022	
	Manufacturer / WHO EUL holder	Name of Vaccine	NRA of Record	Platform	EOI accepted	Pre-submission meeting held	Dossier accepted for review*	Status of assessment**	Decision date***
14.	NOVAVAK	NVX-CoV2373/Novavaxoid	EMA	Recombinant nanoparticle prefusion spike protein formulated with Matrix-M™ adjuvant	✓	✓	Rolling data started 19 August 2021	Finalized	20 December 2021
15.	康希诺生物 CanSinoBio	Ad5-nCoV	NMPA	Recombinant Novel Coronavirus Vaccine (Adenovirus Type 5 Vector)	✓	✓	Rolling data started 09 August 2021	Finalized	19 May 2022
16.	RUSSIAN DIRECT INVESTMENT FUND	Sputnik V	Russian NRA	Human Adenovirus Vector-based Covid-19 vaccine	Additional information submitted	Several meetings have been and continue to be held.	"Rolling" submission incomplete.	Process restarted, awaiting completion of rolling submission and CAPAs to last inspection	Anticipated date will be set once all data is submitted and follow-up of inspection observations completed.
16.	RUSSIAN DIRECT INVESTMENT FUND	Sputnik V	Russian NRA	Human Adenovirus Vector-based Covid-19 vaccine	Additional information submitted	Several meetings have been and continue to be held.	"Rolling" submission incomplete.		To be confirmed
21.	Shifa Pharmad - Barkat	Coviran® vaccine	Iran Food Drug Administration (IFDA)	Inactivated, produced in Vero cells	✓	Presubmission meeting held on 26 January 2022	Rolling data starting in July	Ongoing	To be confirmed
22.	CIGB	Abdaia	CECMED	Protein subunit	✓	✓	Rolling data starting June	Ongoing	To be confirmed
23.	SK Bioscience	Novavaxoid prefilled syringe	MFDS (Korea)	Recombinant nanoparticle prefusion spike protein formulated with Matrix-M™ adjuvant	✓	✓	Rolling data pending	Ongoing	To be confirmed
24.	Biological E	Corbevax	DCGI India	RBD antigen of SARS CoV-2 (Covid-19)	✓	✓	Rolling data started 10 <sup>th</sup> of June	Ongoing	To be confirmed
25.	SK Bioscience	GBPS10	MFDS (Korea)	Recombinant protein subunit	EOI under review				
26.	WestVac Biopharma	Recombinant COVID-19 Vaccine	NMPA China	Recombinant SARS-CoV-2 S-RBD protein	EOI under review				
27.	Nanogen	Nanocovax	Drug Administration of Vietnam	Recombinant Spike protein	EOI under review				
28.	Cinnagen	SpikoGen	Iran Food Drug Administration (IFDA)	Recombinant Protein	EOI under review				
29.	R-PHARM	Vaccine R-COV1	Russian NRA	Recombinant ChAdOx1 adenoviral vector encoding the Spike protein antigen of the SARS-CoV-2.	EOI under review				
30.	Arcturus Therapeutics	ARCT-154	Drug Administration of Vietnam	RNA Vaccine	EOI under review				
31.	Bio-Manguinhos/Fiocruz	AZD1222	ANVISA	Recombinant ChAdOx1 adenoviral vector encoding the Spike protein antigen of the SARS-CoV-2.	EOI under review				

<sup>1</sup> Fuente: [https://extranet.who.int/pqweb/sites/default/files/documents/Status\\_COVID\\_VAX\\_07July2022.pdf](https://extranet.who.int/pqweb/sites/default/files/documents/Status_COVID_VAX_07July2022.pdf)  
 Nota: Información actualizada hasta el 2 de septiembre de 2022.

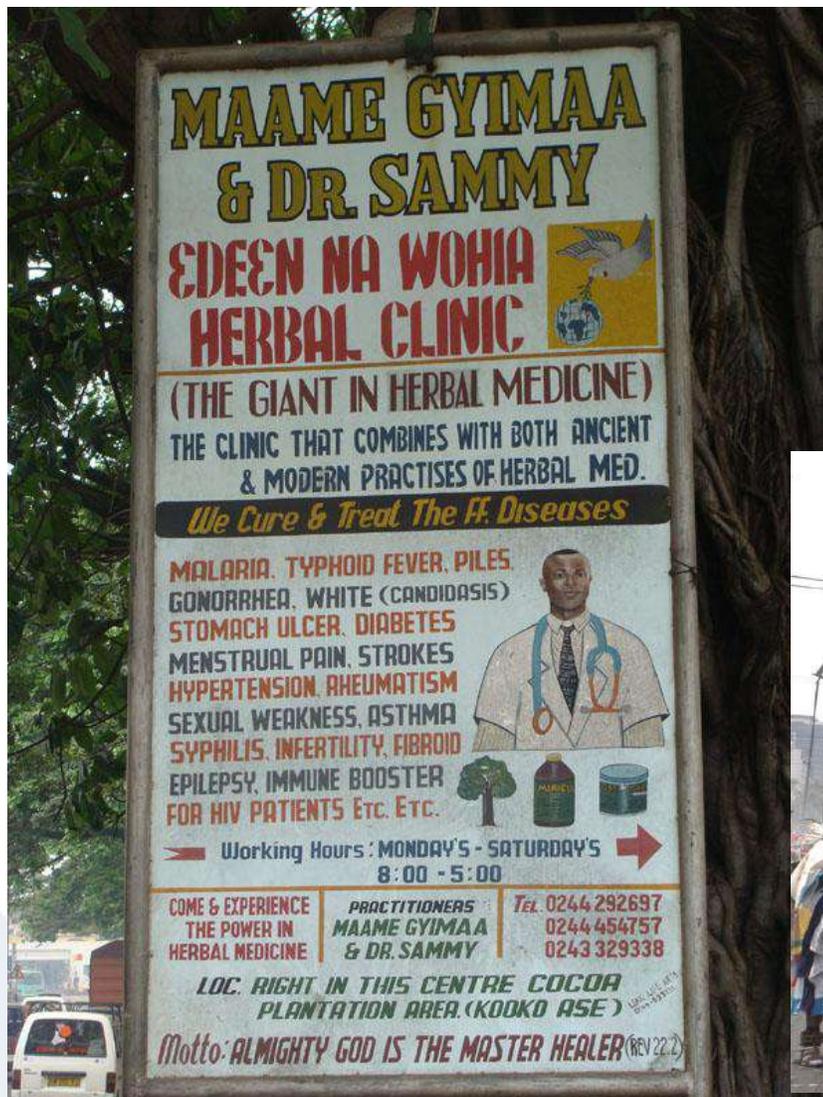
# Herramientas y aplicaciones estadísticas para Ciencia de Datos

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
- Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
      - Una muestra.
      - Dos muestras.
        - Independientes.
        - Emparejadas.
      - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.



- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
    - Una muestra.
    - Dos muestras.
      - Independientes.
      - Emparejadas.
    - Más de dos muestras.
  - Medidas de intervención.
    - Intervenciones inespecíficas
    - Intervenciones específicas.
- Conclusiones.
- Preguntas.

## ¿La ciencia no es una opción?



SACERDOTISA ETÍOPE. Ouma Acha fuma, bebe café y mastica qat para expulsar a los malos espíritus





- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
  - Estimación.
    - Una muestra.
    - Dos muestras.
      - Independientes.
      - Emparejadas.
    - Más de dos muestras.
- Medidas de intervención.
  - Intervenciones inespecíficas
  - Intervenciones específicas.
- Conclusiones.
- Preguntas.

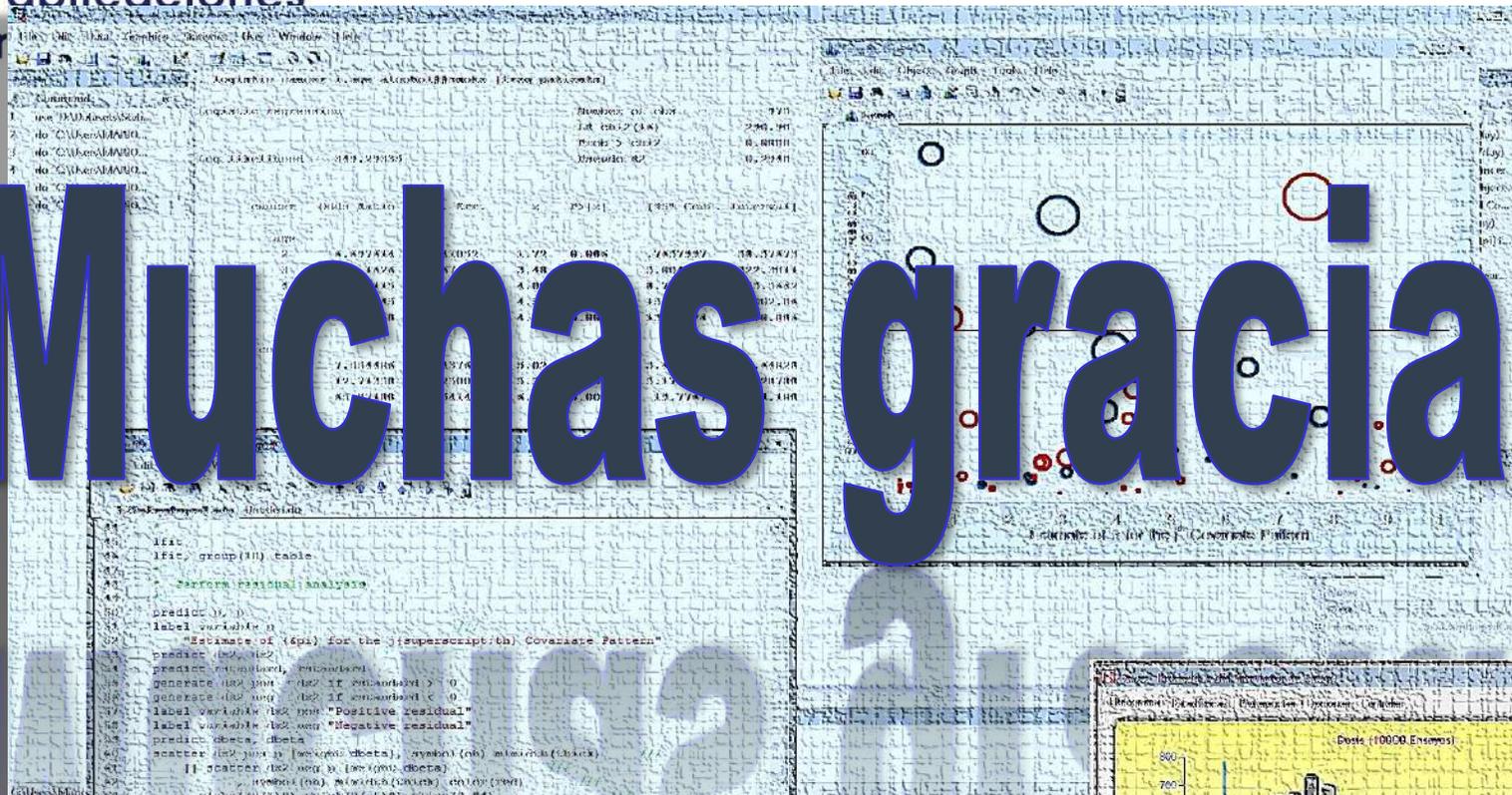
**“No hay mejor vigilancia  
epidemiológica que la que  
incluye la participación activa de  
una población adecuadamente  
informada.”**

***Doc.  
Amo de casa.***

- Introducción.
- Panorama epidemiológico.
- Vigilancia epidemiológica.
- Datos y calidad.
- Análisis exploratorio.
  - Magnitud.
  - Tiempo.
  - Lugar.
  - Variables personales.
- Condiciones de riesgo.
  - Evidencia de causalidad.
    - Ejemplo: El lugar.
  - Otras condiciones.
- Análisis de la evidencia.
  - Comparación de medias.
    - Estimación.
      - Una muestra.
      - Dos muestras.
        - Independientes.
        - Emparejadas.
      - Más de dos muestras.
  - Medidas de intervención.
    - Intervenciones inespecíficas
    - Intervenciones específicas.
- Conclusiones.
- Preguntas.

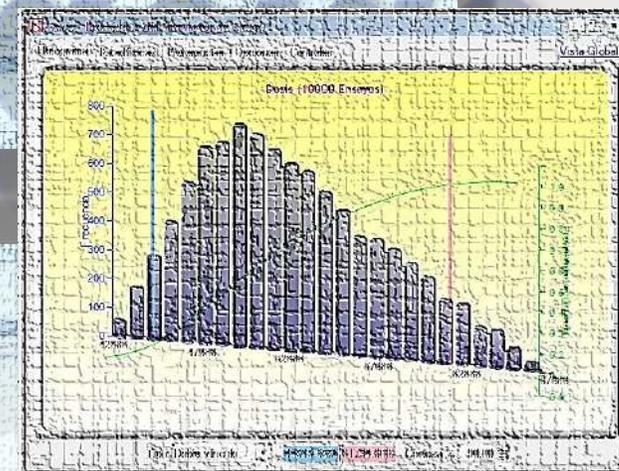
# Preguntas.

# Muchas gracias



```
*** Estimation results ***
*** Dependent Variable: y ***
*** Number of observations = 100 ***
*** Number of variables = 4 ***
*** F(3,96) = 100.0000 ***
*** Prob > F = 0.0000 ***
*** R-squared = 0.7500 ***
*** Adjusted R-squared = 0.7400 ***
*** Total variance = 100.0000 ***
*** Explained variance = 75.0000 ***
*** Residual variance = 25.0000 ***
*** Total sum of squares = 100.0000 ***
*** Explained sum of squares = 75.0000 ***
*** Residual sum of squares = 25.0000 ***
*** Root MSE = 0.5000 ***

*** Coefficients ***
*** Variable | Coefficient | Std. Err. | z | P | [95% Conf. Interval] ***
***-----+-----|-----|-----|---|---|-----|
*** _cons | 4.000000 | 0.000000 | . | . | 4.000000 - 4.000000 ***
*** beta1 | 4.000000 | 0.000000 | . | . | 4.000000 - 4.000000 ***
*** beta2 | 4.000000 | 0.000000 | . | . | 4.000000 - 4.000000 ***
*** beta3 | 4.000000 | 0.000000 | . | . | 4.000000 - 4.000000 ***
```



# Conferencias Stata LATAM 2022

## Herramientas y aplicaciones estadísticas para Ciencia de Datos

Organiza:



Conozca más sobre STATA  
escaneando el código QR.

