**Covid 19: Análisis Probit**

El siguiente análisis se realizó a partir de la base de datos diarios puesta a disposición del público por la Secretaría de Salud (SSA) (<https://coronavirus.gob.mx/datos/#DownZCSV>) del 14 de julio, 2020. Dado que no sabemos si las últimas personas en integrarse a la base podrían fallecer transcurrido cierto tiempo, la base se filtra para dejar sólo a pacientes que mostraron síntomas hace 15 días. Al filtrar por esto, el número de observaciones es de 296,024.

El objetivo de este análisis es encontrar la relación entre ciertas características de los pacientes con Covid-19 y la probabilidad de que fallezcan. Para este tipo de prácticas, es común utilizar el modelo probit para definir qué variables están relacionadas con esta probabilidad y la dirección/magnitud de esta relación.

Para el análisis se utilizaron las siguientes variables incluidas en la base de la SSA:

1. **Sexo**: Variable indicadora de sexo, 0 si es hombre, 1 si es mujer.
2. **Rangos de edad**: Variable indicadora por cada rango de edad, exceptuando el rango de edad <20 años. La variable tiene un valor de 0 si el paciente no tiene una edad incluida en el rango, 1 si sí.
3. **Comorbilidades**: Para las siguientes condiciones, un valor de 0 indica si el paciente no padece la condición y 1 si padece la condición.
   * Diabetes
   * EPOC
   * Asma
   * Inmunosupresión
   * Hipertensión
   * Cardiovascular
   * Obesidad
   * Renal crónica
   * Tabaquismo
   * Otra comorbilidad
4. **Entidad federativa:** Variable indicadora por entidad federativa, exceptuando la Ciudad de México. La variable tiene un valor de 1 si el paciente acudió a una unidad médica en la entidad federativa, 0 si no.
5. **Sector de unidad médica**: Variable indicadora por sector de unidad médica, donde la variable tiene valor de 1 si el paciente acudió a una unidad medica perteneciente a este sector, 0 si no. Se exceptúa el sector SSA (Secretaría de Salud).
6. **Interacciones**: Existen dos interacciones, sexo/comorbilidad y sexo/rango de edad. En este caso, la variable de interacción es 1 si es una mujer con comorbilidad en el caso de la interacción sexo/comorbilidad. En el segundo caso, la variable de interacción es 1 si es una mujer con una edad incluida el rango de edad.

Para medir cuál es el cambio en probabilidad que una variable tiene sobre la probabilidad de defunción de un paciente con Covid-19, se calculan los efectos marginales. Los efectos marginales miden el cambio de la probabilidad, en puntos porcentuales, de que un paciente con Covid-19 fallezca, relativo a ser un hombre de menos de 20 años sin comorbilidades en la Ciudad de México, que acudió a una unidad médica de la SSA.

Los siguientes cuadros presentan los efectos marginales para las variables utilizadas en el probit. Aunque están presentados en cuadros distintos, los resultados pertenecen a un mismo modelo. Los efectos marginales se calculan asumiendo que están en la media de las observaciones.

**Figura 1.**

A screenshot of a cell phone

Description automatically generated

Los efectos marginales para los rangos de edad se interpretan con respecto al rango de edad de menores de 20 años. Como se puede ver en la figura 1, para el rango de 20-30 años, existe una disminución de 2.75 puntos porcentuales asociada a la probabilidad de fallecer de Covid-19 al 1 por ciento de significancia, con respecto al rango de edad de menores de 20 años. Para el rango de 31-40 años, existe un aumento en 2.93 puntos porcentuales. Para los pacientes en los rangos de edad 41-50, 51-60, 60-70 y >70, existe un efecto marginal positivo y significativo en la probabilidad de que un paciente fallezca, relativo al rango de edad de menores de 20 años. El rango de edades de 41-50 está asociado a un incremento de 9.82 puntos porcentuales; el rango de 51-60 a un incremento de 15.85 puntos porcentuales; el rango de 61-70 a un incremento de 21.27 puntos porcentuales, mientras que el rango de edades de 70> está asociado a un incremento en 26.14 puntos porcentuales en la probabilidad de fallecer de Covid-19, relativo al grupo de menores de 20 años.

Para las mujeres, existe una menor probabilidad de fallecer de Covid-19 en cada rango de edad con respecto a los hombres. El efecto marginal es significativo para todos los rangos de edad excepto el de 20 a 30 años.

La figura 2 presenta los efectos marginales de las variables comorbilidades por sexo. Los resultados indican que ser mujer con una comorbilidad está asociado a una probabilidad más alta de fallecer por Covid-19 con respecto a ser hombre con una comorbilidad para diabetes e hipertensión. Para el resto de las comorbilidades, no hay suficiente evidencia estadística como para no rechazar que la diferencia entre hombres y mujeres sea de 0. Una vez controlando por rangos de edad, los efectos marginales de las comorbilidades indican en su mayoría una mayor probabilidad de fallecer por Covid-19. La condición renal crónica aumenta la probabilidad de fallecer en 6, mientras que la diabetes e inmunosupresión aumentan esta probabilidad en 4 puntos porcentuales. Tener obesidad aumenta la probabilidad en 3.36 puntos porcentuales. Padecer EPOC e hipertensión aumentan la probabilidad en 1.7 y 0.9 puntos porcentuales. Los pacientes que padecen asma tienen una menor probabilidad de fallecer por Covid-19 en 1.3 puntos porcentuales. Para el tabaquismo y condición cardiovascular, no hay evidencia suficiente para no rechazar que el efecto sea 0.

**Figura 2.**

A screenshot of a cell phone

Description automatically generated

Por sector, los efectos marginales se interpretan con respecto a las unidades médicas de la Secretaria de Salud (SSA) y se muestran en la figura 3.

Las unidades médicas del IMSS están asociadas a un incremento en 8.5 puntos porcentuales; las unidades del ISSTE a un incremento en 5.1 puntos porcentuales. El sector privado y el sector municipal están asociados a una disminución de 4.2 y 8.75 puntos porcentuales en la probabilidad de que un paciente con Covid-19 fallezca, con respecto a las unidades médicas de la SSA. Para las unidades médicas de la Secretaría de Defensa Nacional (SEDENA), Secretaría de Marina (SEMAR) y el sector universitario, no se cuenta con suficiente evidencia estadística como para no rechazar que el efecto marginal sea 0.

**Figura 3.**

A screenshot of a cell phone

Description automatically generated

La figura 4 muestra los efectos marginales por entidad federativa de la unidad médica. Cada efecto marginal indica el cambio en la probabilidad de que un paciente con Covid-19 fallezca en esa entidad, con respecto a la Ciudad de México. Como podemos ver, las entidades del noreste/bajío están asociadas con menores probabilidades de que un paciente con Covid-19 fallezca, mientras que el centro sureste está asociado a probabilidades más altas.

**Figura 4. Efectos marginales por estado, base Ciudad de México**

A screenshot of a cell phone

Description automatically generated

Esto podría ser por que la pandemia empezó antes en algunas entidades, sin embargo, también podría estar asociado a características de la población en la entidad, de las unidades médicas del estado, y del manejo de la pandemia a nivel estatal.

Resultado de la regresión sobre variables estatales

# Tabla de resultados de la regresión

* **Variable dependiente: Coeficiente probit de los estados**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Modelo inicial |  | Modelo final |  |
| Constante | 8.5675 | \* | 10.2863 | \*\* |
|  | (-4.9138) |  | (-3.866) |  |
| ln(Días desde 10 muertes) | -7.5892 |  | -9.5134 | \*\* |
|  | (-4.823) |  | (-3.8257) |  |
| ln(Días desde 10 muertes)^2 | 0.2097 |  | 0.2599 | \*\* |
|  | (-0.1266) |  | (-0.0995) |  |
| ln(Camas) | 0.0777 |  |  |  |
|  | (-0.1577) |  |  |  |
| ln(Densidad) | 0.0223 |  |  |  |
|  | (-0.0274) |  |  |  |
| ln(Edad) | 0.5111 |  |  |  |
|  | (-0.5772) |  |  |  |
| ln(Escolaridad) | 0.3411 |  |  |  |
|  | (-0.3921) |  |  |  |
| ln(PIB per cápita) | -0.9604 | \* | -0.506 | \*\*\* |
|  | (-0.55) |  | (-0.1593) |  |
| Pob. urbana | -0.0575 |  |  |  |
|  | (-0.0875) |  |  |  |
| Pobreza | 0.0303 |  |  |  |
|  | (-0.1219) |  |  |  |
| Sin salud | -0.0629 |  |  |  |
|  | (-0.3354) |  |  |  |
| Informalidad | -0.0639 |  |  |  |
|  | (-0.128) |  |  |  |
| Turismo | 0.1366 |  |  |  |
|  | (-0.1369) |  |  |  |

Errores estándar en paréntesis.

\* p<0.1, \*\* p<0.05, \*\*\*p<0.01

En el modelo inicial se utilizaron las siguientes variables:

* Días desde las 10 muertes, representan el número de días que pasaron desde que el estado acumuló 10 muertes por COVID hasta el 12 de julio de 2020. Se toma el logaritmo natural de esta variable, y también se incluye el cuadrado de ese logaritmo. Este dato se obtuvo de la base de datos diarios de COVID-19 de la SSA.
* Camas, es el número de camas hospitalarias por cada 100 mil habitantes en cada estado. El dato se obtuvo del Catálogo CLUES de la SSA.
* Densidad, representa la densidad de población de cada estado de acuerdo con la Encuesta Intercensal 2015.
* Edad, representa la edad mediana de la población de cada estado de acuerdo con la Encuesta Intercensal 2015.
* Escolaridad, representa el grado de escolaridad promedio de la población de cada estado de acuerdo con la Encuesta Intercensal 2015.
* PIB per cápita, es el PIB per cápita no minero de cada estado en 2018. Para obtener el PIB no minero, al PIB total se le restó el PIB por actividades mineras. El dato se obtuvo del Sistema de Cuentas Nacionales.
* Población urbana, representa el porcentaje de la población de cada estado que vive en el área urbana, de acuerdo con la Encuesta Nacional de la Dinámica Demográfica 2018.
* Pobreza, representa el porcentaje de la población de cada estado que vivía en condición de pobreza patrimonial en 2018, de acuerdo con las estimaciones del CONEVAL, realizadas con los Módulos de Condiciones Socioeconómicas de la ENIGH 2018.
* Sin salud, representa el porcentaje de la población de cada estado que no cuenta con acceso a los servicios de salud, de acuerdo con las estimaciones del CONEVAL, realizadas con los Módulos de Condiciones Socioeconómicas de la ENIGH 2018.
* Informalidad, representa el porcentaje de la PEA de cada estado que trabajaba en el sector informal, de acuerdo con los cálculos de la ENOE 2018.
* Turismo, es el porcentaje que representan los servicios de alojamiento temporal y de preparación de alimentos y bebidas en el PIB total de cada estado, de acuerdo con los datos de 2018 del Sistema de Cuentas Nacionales.

# Coeficientes probit vs valores ajustados de la regresión final

A close up of a map

Description automatically generated

# Efecto de los días desde las 10 muertes

La siguiente gráfica muestra el efecto de los días que han pasado desde que los estados acumularon 10 muertes sobre el valor probit calculado con la regresión, suponiendo que ln(PIB)=0.

A screenshot of a cell phone

Description automatically generated

**ANEXO**

A continuación, se presentan los cuadros con los resultados. Aunque estén por separado, los cuadros son parte de un mismo análisis. La variable de días entre primer síntoma e ingreso está asociada a una disminución en 0.1 puntos porcentuales de fallecer, lo cual podría ser causalidad inversa: las personas con síntomas más fuertes acuden rápidamente a una unidad medica mientras que las personas con síntomas leves se abstienen a ir rápidamente. Este comportamiento podría tener como consecuencia el signo negativo en el efecto marginal.

Cuadro 1: Efectos marginales por comorbilidad/rangos de edad

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | dy/dx | Error Estándar | z | P>|z |
| SEXO | -0.0078 | 0.010 | -0.795 | 0.427 |
| DÍAS ENTRE SÍNTOMAS/INGRESO | -0.0011\*\*\* | 0.000 | -7.182 | 0.000 |
| Rango Edades 20-30 | -0.0275\*\*\* | 0.008 | -3.605 | 0.000 |
| Rango Edades 31-40 | 0.0293\*\*\* | 0.007 | 4.136 | 0.000 |
| Rango Edades 41-50 | 0.0982\*\*\* | 0.007 | 14.098 | 0.000 |
| Rango Edades 51-60 | 0.1585\*\*\* | 0.007 | 22.799 | 0.000 |
| Rango Edades 61-70 | 0.2127\*\*\* | 0.007 | 30.394 | 0.000 |
| Rango Edades >70 | 0.2614\*\*\* | 0.007 | 37.107 | 0.000 |
| DIABETES | 0.0354\*\*\* | 0.002 | 21.805 | 0.000 |
| EPOC | 0.0171\*\*\* | 0.004 | 4.119 | 0.000 |
| ASMA | -0.0137 | 0.005 | -2.623 | 0.009 |
| INMUSUPR | 0.0396\*\*\* | 0.005 | 7.710 | 0.000 |
| HIPERTENSION | 0.0098\*\*\* | 0.002 | 6.053 | 0.000 |
| CARDIOVASCULAR | 0.0023 | 0.004 | 0.635 | 0.525 |
| OBESIDAD | 0.0339\*\*\* | 0.002 | 20.526 | 0.000 |
| RENAL\_CRONICA | 0.0646\*\*\* | 0.004 | 18.036 | 0.000 |
| TABAQUISMO | 0.0008 | 0.002 | 0.348 | 0.728 |
| OTRA | 0.0259\*\*\* | 0.004 | 6.986 | 0.000 |
| Número de observaciones | | 296,024 | | |

\*\*\* Se refiere un valor p < 0.001

Cuadro 2: Efectos marginales interacciones sexo/rangos de edad

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | dy/dx | | Error Estándar | z | P>|z |
| SEXO\_Rango Edades\_20-30 | -0.0238 | | 0.011 | -2.119 | 0.034 |
| SEXO\_Rango Edades\_31-40 | -0.0495\*\*\* | | 0.010 | -4.765 | 0.000 |
| SEXO\_Rango Edades\_41-50 | -0.0572\*\*\* | | 0.010 | -5.637 | 0.000 |
| SEXO\_Rango Edades\_51-60 | -0.0551\*\*\* | | 0.010 | -5.455 | 0.000 |
| SEXO\_Rango Edades\_61-70 | -0.0454\*\*\* | | 0.010 | -4.455 | 0.000 |
| SEXO\_Rango Edades\_>70 | -0.0388\*\*\* | | 0.010 | -3.775 | 0.000 |
| Número de observaciones | | 296,024 | | | | |

\*\*\* Se refiere un valor p < 0.001

Cuadro 3: Efectos marginales interacciones sexo/comorbilidades

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | dy/dx | | Error Estándar | z | P>|z | |
| SEXO\_DIABETES | 0.0127\*\*\* | | 0.003 | 4.946 | 0.000 | |
| SEXO\_EPOC | 0.0081 | | 0.006 | 1.308 | 0.191 | |
| SEXO\_ASMA | 0.0010 | | 0.007 | 0.142 | 0.887 | |
| SEXO\_INMUSUPR | 0.0019 | | 0.007 | 0.255 | 0.798 | |
| SEXO\_HIPERTENSION | 0.0155\*\*\* | | 0.003 | 6.073 | 0.000 | |
| SEXO\_CARDIOVASCULAR | 0.0055 | | 0.006 | 0.961 | 0.337 | |
| SEXO\_OBESIDAD | -0.0039 | | 0.003 | -1.504 | 0.132 | |
| SEXO\_RENAL\_CRONICA | 0.0096 | | 0.006 | 1.725 | 0.085 | |
| SEXO\_TABAQUISMO | -0.0167 | | 0.005 | -3.406 | 0.001 | |
| SEXO\_OTRA\_COM | -0.0064 | | 0.005 | -1.190 | 0.234 | |
| Número de observaciones | | 296,024 | | | |

\*\*\* Se refiere un valor p < 0.001

Cuadro 4: Efectos marginales por sector

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | dy/dx | Error Estándar | z | P>|z |
| SECTOR CRUZ ROJA | 0.0329 | 0.053 | 0.617 | 0.537 |
| SECTOR ESTATAL | 0.0195\*\*\* | 0.004 | 5.553 | 0.000 |
| SECTOR IMSS | 0.0849\*\*\* | 0.001 | 73.043 | 0.000 |
| SECTOR ISSSTE | 0.0510\*\*\* | 0.002 | 22.123 | 0.000 |
| SECTOR MUNICIPAL | -0.0875\*\*\* | 0.023 | -3.858 | 0.000 |
| SECTOR NO ESPECIFICADO | 0.0411\*\*\* | 0.006 | 6.751 | 0.000 |
| SECTOR PEMEX | 0.0236\*\*\* | 0.004 | 5.485 | 0.000 |
| SECTOR PRIVADA | -0.0420\*\*\* | 0.004 | -10.556 | 0.000 |
| SECTOR SEDENA | -0.0073 | 0.007 | -1.000 | 0.317 |
| SECTOR SEMAR | -0.0190 | 0.007 | -2.678 | 0.007 |
| SECTOR UNIVERSITARIO | 0.0020 | 0.017 | 0.121 | 0.903 |
| Número de observaciones | | 296,024 | | |

\*\*\* Se refiere un valor p < 0.001

Cuadro 5: Efectos marginales por entidad federativa

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Entidad Federativa | | dy/dx | Error Estándar | z | P>|z| | |
| AGUASCALIENTES | | -0.0562\*\*\* | 0.007 | -8.388 | 0.000 | |
| BAJA CALIFORNIA | | 0.0457\*\*\* | 0.003 | 17.735 | 0.000 | |
| BAJA CALIFORNIA SUR | | -0.0614\*\*\* | 0.008 | -7.284 | 0.000 | |
| CAMPECHE | | -0.0147 | 0.005 | -2.689 | 0.007 | |
| CHIAPAS | | 0.0062 | 0.004 | 1.565 | 0.118 | |
| CHIHUAHUA | | 0.0471\*\*\* | 0.004 | 11.217 | 0.000 | |
| COAHUILA | | -0.0678\*\*\* | 0.005 | -14.706 | 0.000 | |
| COLIMA | | -0.0078 | 0.010 | -0.810 | 0.418 | |
| DURANGO | | -0.0417\*\*\* | 0.007 | -6.182 | 0.000 | |
| GUANAJUATO | | -0.0727\*\*\* | 0.004 | -19.880 | 0.000 | |
| GUERRERO | | 0.0138\*\*\* | 0.003 | 4.191 | 0.000 | |
| HIDALGO | | 0.0389\*\*\* | 0.004 | 10.195 | 0.000 | |
| JALISCO | | -0.0300\*\*\* | 0.003 | -8.945 | 0.000 | |
| MÉXICO | | -0.0196\*\*\* | 0.004 | -4.900 | 0.000 | |
| MICHOACÁN | | 0.0403\*\*\* | 0.004 | 9.367 | 0.000 | |
| MORELOS | | 0.0353\*\*\* | 0.002 | 19.156 | 0.000 | |
| NAYARIT | | -0.0225\*\*\* | 0.006 | -3.533 | 0.000 | |
| NUEVO LEÓN | | -0.0593\*\*\* | 0.004 | -15.279 | 0.000 | |
| OAXACA | | 0.0008 | 0.004 | 0.226 | 0.821 | |
| PUEBLA | | 0.0195\*\*\* | 0.003 | 7.434 | 0.000 | |
| QUERÉTARO | | 0.0153 | 0.005 | 2.846 | 0.004 | |
| QUINTANA ROO | | 0.0325\*\*\* | 0.004 | 8.118 | 0.000 | |
| SAN LUIS POTOSÍ | | -0.0340\*\*\* | 0.005 | -6.224 | 0.000 | |
| SINALOA | | 0.0148\*\*\* | 0.003 | 5.248 | 0.000 | |
| SONORA | | -0.0197\*\*\* | 0.003 | -6.554 | 0.000 | |
| TABASCO | | 0.0120\*\*\* | 0.003 | 4.375 | 0.000 | |
| TAMAULIPAS | | -0.0328\*\*\* | 0.004 | -8.860 | 0.000 | |
| TLAXCALA | | 0.0179\*\*\* | 0.005 | 3.674 | 0.000 | |
| VERACRUZ | | -0.0028 | 0.003 | -1.098 | 0.272 | |
| YUCATÁN | | -0.0177\*\*\* | 0.004 | -4.252 | 0.000 | |
| ZACATECAS | | -0.0139 | 0.008 | -1.662 | 0.096 | |
| Número de observaciones | 296,024 | | | | |

\*\*\* Se refiere un valor p < 0.001