

Ejercicio de la estimación del total de casos COVID-19 en México

Claudia González González
Jorge Becerril Cejudo

1 de mayo de 2020

Objetivo

Dar a conocer un ejercicio de estimación del total de casos COVID-19 en México con algunas de las consideraciones utilizadas en muestreo a través de factores de expansión o ponderadores, mediante la información disponible dada a conocer por fuentes oficiales.

Vigilancia Centinela

El método de “Vigilancia centinela” utilizado en México es un método de seguimiento continuo en el tiempo donde es posible visualizar el comportamiento generalizado de diferentes enfermedades. El monitoreo de enfermedades respiratorias se lleva a cabo a través de una muestra de 375 unidades monitoras(UM) de las 26,000 unidades médicas distribuidas en toda la República Mexicana.

Diseño de muestreo

El modelo de vigilancia centinela contempla un diseño tipo encuesta con una muestra de representación nacional de (375 UM) a nivel nacional para determinar la confirmación por laboratorio de casos y conocer características de la salud particulares o identificar factores de riesgo en la población Mexicana respecto a las enfermedades de interés. Tiene un corte o ciclo de notificación de una semana por lo que la estimación de diferentes indicadores que contempla su diseño se realiza con dicha periodicidad.

Debido a las particularidades del sistema de salud en sus diferentes Entidades y las características de la enfermedad se vuelve necesaria y razonable

la consideración técnica de estratificación del diseño por entidad federativa y grupo de edad. Al considerar un tipo de estratificación suponen un tipo de diseño que son denominados diseños complejos y que en diferentes ocasiones contemplan la selección en etapas, estratificación, formación de conglomerados o empleo de probabilidades de selección diferenciadas.

Factor de expansión o ponderadores

El muestreo probabilístico se conocen claramente las probabilidades específicas de selección de una unidad en la muestra y estas dependen de sus etapas de selección y diseño o asignación diferenciadas por patrones observados de no respuesta. Los factores de expansión se definen como el inverso de la probabilidad de ser seleccionado y se interpretan como el número de personas en la población que representa cada unidad en la muestra por ejemplo si la población total es N y la muestra n el factor de expansión representará a $f = \frac{N}{n}$ personas de la población. La forma de estimar el total de una característica sería $\sum_{i=1}^N Y_i \approx \sum_{i=1}^n f_i y_i$, el caso del diseño por estratificación la estimación es $\sum_{i=1}^N Y_i \approx \sum_{k=1}^K \sum_{i=1}^{n_k} f_{ki} y_{ki}$.

Consideraciones en el ejercicio de Estimación

Este ejercicio contempla por tanto, la estratificación por 32 Entidades de la República y 14 grupos de edad (0-4,5-10,.....65-más). La muestra considerada se determina por quienes ya están en seguimiento día con día al 30 de abril 87372 por lo tanto la muestra considerada no es fija sino se constituye día a día con los que van llegando a consulta o piden consulta de manera ambulatoria. (Es importante recordar que la consolidación de la información se desarrolla de manera semanal, sin embargo al ser un monitoreo continuo es posible realizar la estimación día con día).

En este ejercicio se define la probabilidad de selección como

$$p_{ij} = \frac{Pob.quinquenal_i.de.alguna.entidad_j}{Pob.Entidad_j} \quad (1)$$

por lo que el factor de expansión es el inverso de dicha probabilidad

$$f_{ij} = \frac{1}{p_{ij}} \quad (2)$$

Ejemplo:

| NOM_ENT | EDAD_QUIN | POB | AguasCalientes | Prob | Factor de expansión |
|--------------------|------------|--------|----------------|-------------------|---------------------|
| 1 Aguascalientes | pobm_00_04 | 124139 | 1434635 | 0.086530023316035 | 11.556682428568 |
| 2 Aguascalientes | pobm_05_09 | 130125 | 1434635 | 0.090702513182796 | 11.0250528338136 |
| 3 Aguascalientes | pobm_10_14 | 133222 | 1434635 | 0.092861250422581 | 10.7687544099323 |
| 4 Aguascalientes | pobm_15_19 | 133976 | 1434635 | 0.093386819644021 | 10.708149220756 |
| 5 Aguascalientes | pobm_20_24 | 130345 | 1434635 | 0.0908558622925 | 11.0064444359201 |
| 6 Aguascalientes | pobm_25_29 | 124826 | 1434635 | 0.087008890763156 | 11.4930783650842 |
| 7 Aguascalientes | pobm_30_34 | 112277 | 1434635 | 0.078261718137366 | 12.7776392315434 |
| 8 Aguascalientes | pobm_35_39 | 99211 | 1434635 | 0.069154175103772 | 14.4604428944371 |
| 9 Aguascalientes | pobm_40_44 | 90383 | 1434635 | 0.063000693556201 | 15.8728411316287 |
| 10 Aguascalientes | pobm_45_49 | 83981 | 1434635 | 0.058538234463818 | 17.0828520736833 |
| 11 Aguascalientes | pobm_50_54 | 72491 | 1434635 | 0.05052922868883 | 19.7905257204343 |
| 12 Aguascalientes | pobm_55_59 | 60288 | 1434635 | 0.04202323239012 | 23.7963608014862 |
| 13 Aguascalientes | pobm_60_64 | 47438 | 1434635 | 0.033066250300599 | 30.2423162865214 |
| 14 Aguascalientes | pobm_65_mm | 91933 | 1434635 | 0.064081107738205 | 15.6052233691928 |
| 15 Baja California | pobm_00_04 | 297372 | 3634868 | 0.081810948843259 | 12.2233027991876 |

Es posible observar que el factor de expansión no es único sino que tendremos (448 factores diferenciados $32Ent * 14grup.edad$) para algún estrato específico varía desde 9.04 y puede llegar a 35.59. Lo que significa que la pertenencia a cada uno de estos grupos genera un peso diferenciado en cada elemento en muestra. Cabe mencionar que es de esperarse que no existe un único valor o ponderador, sino que dependerá de la entidad y grupo de edad al que pertenece la persona en muestra. Si existen más personas infectadas en una entidad específica su probabilidad de infectarse se vuelve más alta para la persona que reside en la misma entidad o grupo de edad.

De acuerdo a nuestra información específica y estratificación por Entidad y grupo de edad se tendría lo siguiente:

Cabe mencionar que pasado el tiempo si la muestra incrementa mucho en un estrato en particular el factor de expansión decrecerá hasta que cada individuo se represente así mismo, esto significa que el actual factor decrece en el tiempo hasta llegar a uno.

Por otra parte la estimación considerada se realiza hacia la población de confirmados que podrían ser ambulatorios y no hospitalizados ya que suponemos que todos los confirmados que requieren hospitalización están registrados al asistir a las unidades médicas y ya forman parte de la muestra utilizada, por lo que a este universo particular se les asigna el factor de expansión de uno. En resumen a las personas estudiadas en muestra que son ambulatorios o no especificados que pertenecen al grupo estudiado $m_{estrato_{ik}}$ se les da el peso de $m_{estrato_{ik}} * f_{ij}$ y a su vez a los que están hospitalizados sería $muestra_{estrato_{ij}} * 1$.

Estimación

La estimación se realiza con la suma de quienes resultan confirmados del universo estudiado por su factor de expansión a través del paquete enfocado

a muestreo en R **survey**.

El cuál genera remuestreos con el diseño especificado para generar un error estándar para la estimación puntual y por intervalo.

```
svydesign(id=pobQuinquenal, strata=pobEstado, data=base, weights=FACTOR)
```

```
svytotal(Variable, base, diseño)
```

```
svyby(Variable, pobEstado ,base, diseño)
```