

Muestreo

Introducción

En este documento ofrecemos un resumen sobre el concepto de muestreo, y los tipos de muestreo existentes. Además, adjuntamos una hoja para el cálculo de tamaños muestrales en auditorías de historias clínicas (muestreo.xls). No pretendemos, ni mucho menos, ser exhaustivos. Simplemente ofrecemos una pequeña herramienta que pueda servir de apoyo en la evaluación de los distintos indicadores de calidad.

Concepto de muestreo

El muestreo es una herramienta de la investigación científica. Su función básica es determinar que parte de una realidad en estudio (población o universo) debe examinarse con la finalidad de hacer inferencias sobre dicha población. El error que se comete debido a hecho de que se obtienen conclusiones sobre cierta realidad a partir de la observación de sólo una parte de ella, se denomina error de muestreo. Obtener una muestra adecuada significa lograr una versión simplificada de la población, que reproduzca de algún modo sus rasgos básicos.

Terminología

Población objeto: conjunto de individuos de los que se quiere obtener una información.

Unidades de muestreo: número de elementos de la población, no solapados, que se van a estudiar. Todo miembro de la población pertenecerá a una y sólo una unidad de muestreo.

Unidades de análisis: objeto o individuo del que hay que obtener la información.

Marco muestral: lista de unidades o elementos de muestreo.

Muestra: conjunto de unidades o elementos de análisis sacados del marco.

Muestreo probabilístico

El método otorga una probabilidad conocida de integrar la muestra a cada elemento de la población, y dicha probabilidad no es nula para ningún elemento.

Los tipos de muestreo probabilístico son:

- Muestreo aleatorio simple.
- Muestreo sistemático.
- Muestreo aleatorio estratificado.
- Muestreo por conglomerados (clusters)

Muestreo aleatorio simple

Muestreo equiprobabilístico: Si se selecciona una muestra de tamaño n de una población de N unidades, cada elemento tiene una probabilidad de inclusión igual y conocida de n/N .

Ventajas:

- Sencillo y de fácil comprensión.
- Cálculo rápido de medias y varianzas.
- Se basa en la teoría estadística, y por tanto existen paquetes informáticos para analizar los datos.

Desventajas:

- Requiere que se posea de antemano un listado completo de toda la población.
- Cuando se trabaja con muestras pequeñas es posible que no represente a la población adecuadamente.

Muestreo sistemático

Procedimiento:

- Conseguir un listado de N elementos.
- Determinar un tamaño de muestra n .
- Definir un intervalo de salto k ; $K=N/n$.
- Elegir un número aleatorio, r , entre 1 y k (r =arranque aleatorio).
- Seleccionar los elementos de la lista.

Ventajas:

- Fácil de aplicar.
- No siempre es necesario tener un listado de toda la población.
- Cuando la población está ordenada siguiendo una tendencia conocida, asegura una cobertura de unidades de todos los tipos.

Desventajas:

- Si la constante de muestreo está asociada con el fenómeno de interés, se pueden hallar estimaciones sesgadas.

Muestreo aleatorio estratificado

El azar no es una garantía de representatividad. Este muestreo pretende asegurar la representación de cada grupo en la muestra. Cuanto más homogéneos sean los estratos, más precisas resultarán las estimaciones.

Ventajas:

- Tiende a asegurar que la muestra represente adecuadamente a la población en función de unas variables seleccionadas.
- Se obtienen estimaciones más precisas.

Desventajas:

- Se ha de conocer la distribución en la población de las variables utilizadas para la estratificación.
- Los análisis son complicados, en muchos casos la muestra tiene que ponderarse (asignar pesos a cada elemento).

Muestreo por conglomerados

- Los conglomerados se caracterizan porque la variación en cada grupo es menor que la variación entre grupos.
- La necesidad de listados de las unidades de una etapa se limita a aquellas unidades de muestreo seleccionadas en la etapa anterior.

Ventajas:

- Es muy eficiente cuando la población es muy grande y dispersa. Reduce costes.
- No es preciso tener un listado de toda la población, sólo de las unidades primarias de muestreo.

Desventajas:

- El error estándar es mayor que en el muestreo aleatorio simple o estratificado.
- El cálculo del error estándar es complejo.

Cálculo del tamaño muestral

Cada estudio tiene un tamaño muestral idóneo, que permite comprobar lo que se pretende con una seguridad aceptable y el mínimo esfuerzo posible. Para el cálculo del tamaño muestral en cada tipo de estudio existe una fórmula estadística apropiada. Se basan en el error estándar, que mide el intervalo de confianza de cada parámetro que se analiza (media aritmética, porcentaje, diferencia de medias, etc.). La precisión estadística aumenta (el error estándar disminuye) cuando el tamaño muestral crece. Para cada caso en concreto, será necesario consultar un libro de estadística, o más fácil, utilizar algún programa como Epiinfo.

Cálculo del tamaño muestral en una auditoría de historias clínicas

Se trata de una situación especial, en la que se va a determinar la presencia o ausencia de un determinado documento, por ejemplo (variable dicotómica). En este caso, hay que determinar la proporción esperada de la variable de interés, la precisión deseada, y el nivel de confianza. Podemos aplicar las siguientes fórmulas para el cálculo del tamaño muestral (si el muestreo es aleatorio).

Tamaño de la población desconocido

$$\text{Tamaño muestral} = z^2 pq / i^2$$

Donde:

z. Valor de z, 1,96 para $\alpha=0,05$ y 2,58 para $\alpha=0,01$.

p: Prevalencia esperada del parámetro a evaluar. En caso de desconocerse, aplicar la opción más desfavorable ($p=0,5$), que hace mayor el tamaño muestral.

$q=1-p$

i= Error que se prevé cometer. Por ejemplo, para un error del 10%, introduciremos en la fórmula el valor 0,1. Así, con un error del 10%, si el parámetro estimado resulta del 80%, tendríamos una seguridad del 95% (para $\alpha=0,05$) de que el parámetro real se sitúa entre el 70% y el 90%. Vemos, por tanto, que la amplitud total del intervalo es el doble del error que introducimos en la fórmula.

Tamaño de la población conocido

$$\text{Tamaño muestral} = N z^2 pq / (i^2(n-1) + z^2 pq)$$

Donde:

N: tamaño de la población, número total de historias.

z. Valor de z, 1,96 para $\alpha=0,05$ y 2,58 para $\alpha=0,01$.

p: Prevalencia esperada del parámetro a evaluar. En caso de desconocerse, aplicar la opción más desfavorable ($p=0,5$), que hace mayor el tamaño muestral.

$q=1-p$

i= Error que se prevé cometer.

Estas fórmulas están introducidas en la hoja de cálculo *muestral.xls* que se adjunta, realizada en Excel97. En ella basta con introducir los distintos parámetros señalados para obtener el tamaño muestral correspondiente.

Recomendamos calcular tamaños muestrales para $\alpha=0,05$, y un error 0,1, o más precisas.

Algunas consideraciones para el cálculo del tamaño muestral en estudios de calidad del INSALUD.

Cuando el número de pacientes o historias a estudiar no supera los 50, lo más razonable es estudiarlos a todos. En caso contrario, si el estudio supone un esfuerzo considerable, se requerirá seleccionar una muestra, puesto que no tiene mucho

sentido emplear esfuerzos en estudiar de forma exhaustiva una población muy grande. Uno de los aspectos más importantes es seleccionar correctamente la muestra para el estudio. Al igual que en años anteriores, junto con la metodología definitiva se enviará un anexo denominado "Muestreo.doc" para ayudar a aquellas personas que lo precisen en el proceso de selección de la muestra, así como una hoja de cálculo en Excel 97 "muestreo.xls", para la obtención del tamaño muestral. Se recomienda utilizar los siguientes parámetros en la hoja de cálculo:

- Usar la versión "población conocida", introduciendo dicha población.
- Utilizar el nivel del 95% de seguridad ($\alpha=0,05$, $z=1,96$)
- Utilizar una precisión del 5% al 10% (error i , 0,05-0,1). La precisión del 5% es la ideal, pero requiere tamaños muestrales bastante elevados. La precisión del 10% puede ser suficiente en la mayoría de los casos.
- Utilizar la frecuencia esperada del parámetro, utilizando los resultados de estudios pilotos o de años anteriores. Se recomienda que, en caso de no conocer dicha frecuencia esperada, se debe introducir 0,5 en la hoja, que es el caso más desfavorable (obliga al mayor tamaño muestral posible).

Si se siguen estas recomendaciones, se obtendrán resultados finales más comparables entre hospitales.

Este tamaño muestral puede calcularse para obtener un resultado representativo de todos los servicios implicados. En este caso, la población es la total (suma de todos los pacientes o historias de todos los servicios). Este resultado es válido para el INSALUD, pero no es útil para gestionar la calidad a nivel de servicio. Es muy importante conocer el resultado de cada servicio y para ello es necesario conociendo la población del servicio, calcular la muestra para esa población. Lógicamente, el número de historias a estudiar es mayor, y se deberá planificar con cuidado la carga de trabajo.

Bibliografía

Carrasco JL. El método estadístico en la investigación médica. 5ª ed. Madrid. Editorial Ciencia.

Domenech JM. Métodos Estadísticos en Ciencias de la Salud. Signo. Barcelona 1999.

Hulley SB, Cummings SR. Diseño de la investigación clínica. Ed Doyma. Barcelona 1993.

Kelsey IL, Thompson WD, Evans A. Methods in observational epidemiology. New York. Oxford University Press 1986.

Mira JJ, Gómez J, Aranaz J, Pérez E. Auditoría de historias clínicas: ¿cuál es el tamaño adecuado de la muestra?. Todo Hospital 1997; octubre: 58-64.