DISTRIBUCIONES DISCRETAS

BINOMIAL - HIPERGEOMÉTRICA

Se acerca el período electoral y existe interés en conocer las posibilidades del candidato local, Mr. X....

El mejor amigo del candidato afirma que el 60 % de las personas incluidas en el padrón electoral están a favor de Mr. X. Se resuelve hacer una encuesta y tomar una muestra aleatoria de n=10 persona incluidas en el padrón electoral y observar Z (variable aleatoria) que es el número de encuestados que prefieren a Mr. X.

¿Cuál es el modelo de probabilidad adecuado para resolver este problema?
Razonemos detenidamente y realicemos los supuestos que sea necesario hacer en los dos casos especiales que se proponen:

I.- Se cuenta con una ***gran población*** de persona incluidas en el padrón electoral.
II.- Se tiene un ***n*** (tamaño de muestra) que es ***relativamente grande con respecto a N*** (tamaño de la población) es decir, ***no se tiene una gran población de...*** con ***relación al número de personas encuestadas.***

Veamos el caso I

a) Se encuestará a ***n=10*** personas de la población, esto es se harán 10 selecciones o pruebas o ***repeticiones del experimento básico***.
b) Si las selecciones se hacen sucesivamente y en forma aleatoria, tendremos que ***en cada elección*** (o experimento básico o elemental) *obtendremos dos posibles respuestas:* prefiero a Mr. X o no prefiero a Mr. X, las que evidentemente son dos respuestas ***mutuamente excluyentes*** y que forman una clase exhaustiva.
c) La variable aleatoria Z es entonces el número de encuestados que prefieren a Mr. X, entre las 10 personas seleccionadas. Dicho en otros términos, es ***Z: el número de éxitos en las 10 repeticiones del experimento básico***. Además, el menor valor posible de Z es cero, esta es ninguna persona encuestada prefiere a Mr. X y el mayor es diez, o sea las diez personas encuestadas están a favor de Mr. X, esto es p(E) es 0,6 → p(E)= 0,6, lo que hace que p(E) + p(F) = 1, (expresión que coincide con lo manifestado en el párrafo b) y que no esté a favor del Mr. X, este es p(É) = p(F) es 0,4→p(F) =0,4

Pero ¿cuál es la probabilidad de que el segundo encuestado esté a favor de Mr. X? ¿Es también p(E la segunda vez)= 0,6?
Si estamos aceptando que se tiene una gran población de personas incluidas en el padrón electoral, seguramente coincidiremos en que

n=10 personas encuestadas, es una parte muy pequeña de la población, y entonces el hecho de haber sacado una persona de la población anteriormente, no cambiará la proporción de quienes prefieren a Mr.X, en tal forma que se advierta numéricamente dicho cambio.
De acuerdo con este razonamiento, es p(E) = 0,6 también cuanto se considera al segundo encuestado, y esto se mantendrá aproximadamente igual (casi constante) para los n=10 encuestados.

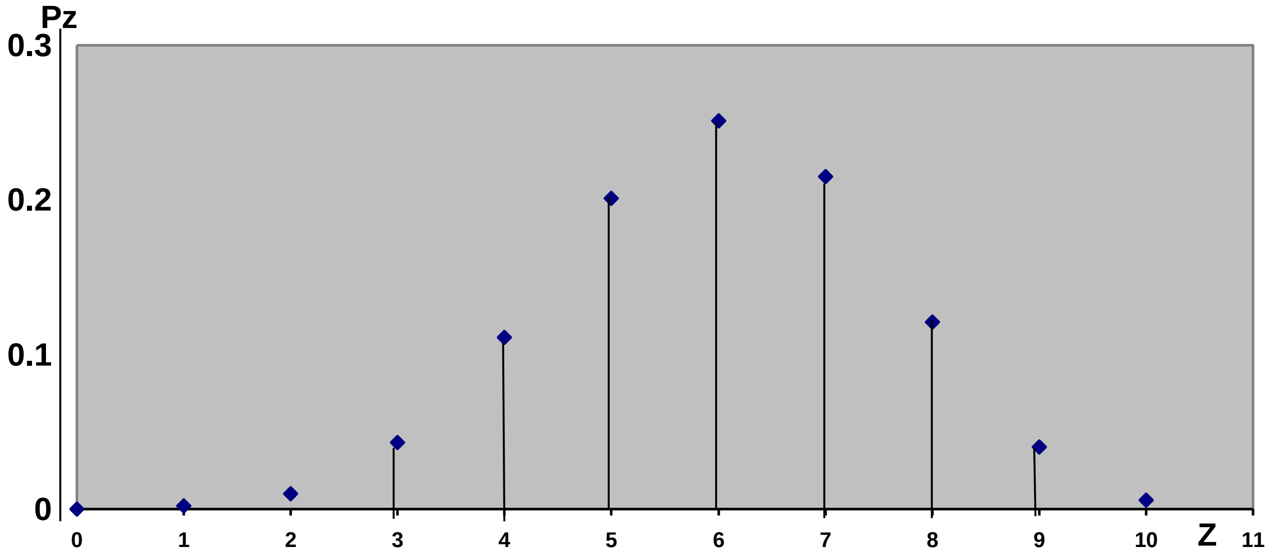
Estamos admitiendo que, en general, cuando la muestra es pequeña con relación a la población o dicho de otro modo, ***cuando la población es lo suficientemente grande o se puede suponer que es tan grande que el hecho de hacer selecciones sucesivas,*** sin reposición, ***no modifica la situación sustancialmente,*** la ***probabilidad de obtener un éxito se mantendrá aproximadamente constante o casi igual, de selección en selección,*** esto es ***no importan los resultados de las selecciones anteriores*** y ***las selecciones*** (pruebas) ***sucesivas serán aproximadamente independientes.***

Vemos entonces que en el caso I propuesto, y de acuerdo con los supuestos que hicimos, se puede considerar que la variable sigue una distribución aproximadamente binomial, y que será (según se vio anteriormente):

$$Pb(n=10;\ p=0,6;\ z_0) = \binom{10}{z_o} 0.6^{Zo}\, 0.4^{10-Zo}$$

para Z: 0,1,2,…,n

y que, gráficamente, se obtendrá de la tabla de distribución de probabilidades, un ***diagrama de probabilidades*** como sigue:



(Probabilidades obtenidas de tabla 1, Mendenhall, Wackerly and Scheaffer)

Veamos el caso II

Al realizar el análisis del caso I, en el Item C se consideró que al ser el número de personas encuestadas una parte muy pequeña de la población, el hecho de seleccionar una segunda persona (sin haber repuesto a la primera en la población) no cambiará la proporción de quienes prefieren a Mr. X (y de quienes no lo prefieren) en tal forma que se advierta numéricamente dicho cambio.

Si por el contrario se tiene que n es relativamente grande con respecto a N o al revés, ***si la población es suficientemente grande con respecto a la muestra, se vería afectada significativamente la probabilidad de que un encuestado prefiera al candidato Mr. X por el hecho de haber seleccionado previamente a otra persona para integrar la muestra, haya o no preferido al candidato mencionado.***

Por ejemplo, si se tuviera un padrón de N=20 personas de la cual se dice que el 60 % prefieren a Mr. X (se tienen k= 12 personas que prefieren a Mr. X), la probabilidad de que el primer encuestado prefiera a Mr. X es 12/20, pero la probabilidad de que el segundo encuestado prefiera a Mr. X es 11/19. (y de que no lo prefiera es 8/19).

Entonces la probabilidad de que ambos encuestados, el primero y el segundo, prefieran a Mr. X es:

$$P(E_1).p(E_2/E_1) = (12/20).(11/19)$$

Y así sucesivamente completaríamos los resultados probabilísticos que nos soliciten.

Estamos confirmando con este razonamiento que en el caso II, ***las probabilidades de éxito*** (y también las de fracaso, ***no se mantendrán constantes y las selecciones sucesivas no serán independientes.***

Detengámonos aquí a pensar que los otros razonamientos seguidos en el desarrollo del caso I, siguen siendo válidos. Nos referimos a lo mencionado en los Items a, b, c y d.

De cualquier manera, al no coincidir con lo razonado y supuesto en el Item e, **no será posible emplear un modelo probabilísticos binomial**.

Recordemos las condiciones que deben cumplirse para poder emplear un esquema binomial y veremos si se cumplen o no en los casos I y II.

	Caso I	Caso II
- El experimento básico se repite n veces	si	si
- En cada experimento básico existen 2 resultados posibles	si	si
Y estos son mutuamente excluyentes , con probabilidades p(E) y p(F), siendo p(E) + p(F) =1		
- Las probabilidades (de éxito y de fracaso) se mantienen constantes	si	no
Es decir que las repeticiones sucesivas son independientes	si	no

En caso de darse las condiciones del caso II ***corresponderá emplear una DISTRIBUCIÓN DE PROBABILIDAD HIPERGEOMETRICA***

$$H(N, k, n, z_o) = \frac{\binom{k}{z_0}\binom{N-k}{n-z_0}}{\binom{N}{n}}$$

Este esquema se logra razonando como sigue:

La probabilidad de que de N personas que figuran en los padrones, de las cuales se sabe que k prefieren al candidato Mr. X, cierto número z_o de los n encuestados, prefiera a Mr. X. será, según el concepto más elemental de probabilidades, el cociente entre el número de casos favorables y el número de casos posibles, si son igualmente posibles (teoría clásica)

¿Qué entendemos por número de casos posibles?

Es el número de formas distintas en que pueden seleccionarse las n personas de la muestra, de un total de N personas. (Si importar si prefieren o no a Mr. X)

Esto es: ***N tomados de a n*** (número de formas diferentes en que pueden formarse grupos de n elementos de un total de N elementos)

$$\binom{N}{n}$$

Y de aquí, la probabilidad de un grupo cualquiera o de una muestra cualquiera, es

$$P = \frac{1}{\binom{N}{n}}$$

Probabilidad de que salga uno cualquiera sea la muestra (resultados equiprobables).

¿Y que entendemos por número de casos favorables?

Es el número de formas en que pueden obtenerse z_o respuestas prefiero a Mr. X y (n - z_o)de respuestas no prefiero a Mr. X, de un total k. que prefieren a Mr. X y de (N -k) que no prefieren a Mr. X respectivamente.
Se tiene entonces ...

$$\binom{k}{z_0} y \binom{N-k}{n-z_0}$$

es decir: número de formas en que pueden selecionarse Zo entre los k que prefieren seleccionarse en el resto de miembros de la muestra entre los que no prefieren a Mr. X.

$$\binom{k}{z_0} * \binom{N-k}{n-z_0}$$

Se pueden escribir esto como

En resumen es

$$\frac{\textit{Número de casos favorables}}{\textit{Número de casos posibles}} = \frac{\binom{k}{z_0}\binom{N-k}{n-z_0}}{\binom{N}{n}}$$

Como se expresaba antes.