**Estrategia didáctica 3.3.4.2. Proporciones muestrales**

Imagina que se selecciona una muestra de 20 obreros textiles de la población dada en la Estrategia 1.2.2.2. Sólo que en este caso no se calculará la media de la muestra sino que se calculará la proporción de obreros bien pagados en la muestra. Si decimos que si un obrero gana al menos 240 dólares se considera bien pagado, entonces podemos calcular tanto la proporción de los obreros bien pagados en la muestra como en la población, y luego comparar ambas cifras. Comencemos con el cálculo de la proporción de obreros bien pagados en la población (que, como se comprenderá, se trata de un parámetro que hemos representado con la letra π):

π = 68/250 = 0.272 (27.2%)

Lo anterior significa que la proporción de obreros bien pagados en la población es de un 27.2%. Ahora, seleccionemos una muestra de tamaño 20 (usando números aleatorios) de la misma población y calculemos la proporción de obreros bien pagados en la muestra (que como se ve será un estadístico), la cual fue 212.44, 213.16, 216.35, 218.27, 220.95, 222.96, 226.72, 230.12, 230.63, 231.87, 232.59, 233.90, 235.21, 237.80, 238.10, 239.50, 241.00, 244.13, 252.55, 270.25. Notemos que hay 4 salarios que fueron mayores a 240, lo que significa que:

p = 4/20 = 0.2 (20%)

Es decir, el 20% de los datos muestrales son mayores a 240.

Puede notarse que se seleccionamos más muestras de tamaño 20, entonces el estadístico p se identificará como una variable, porque claramente su valor no será siempre 20%. Por tanto p es una nueva variable estadística que habrá que caracterizar para hacer predicciones con ella. Es posible que el lector ya se haya dado cuenta que el camino para hacer esto es el mismo que se siguió para caracterizar a la variable , es decir tomar varias muestras de cierto tamaño y calcular varias proporciones muestrales p, para después calcular el histograma, la media de las proporciones muestrales y la desviación de las proporciones muestrales, para que con ello nos demos una idea del valor de los parámetros de la proporción muestral p, para luego hacer cálculos probabilísticos tal y como hicimos para . Sin embargo, ya que se tiene una idea del TCL para , ya no es necesario hacer toda esta labor, porque se sabe que existe un TCL para la proporción, al igual que se obtuvo para ,con el cual vamos a caracterizar la nueva variable p:

TEOREMA CENTRAL DEL LIMITE: Si se selecciona una muestra de tamaño n de una población en la que se conoce π, y se calcula p, esta es una variable aleatoria normal, siempre que nπ ≥ 4 y n (1- π) ≥ 4, con parámetros

 (1)

Nuevamente puede observarse que el TCL nos dice cómo se distribuye p y cómo calcular sus parámetros. Para ello sólo hace falta conocer π, y verificar que se cumplen las dos desigualdades enunciadas en el TCL para garantizar la normalidad de p. De lo contrario, la normalidad de p no se garantiza y no se podrá usar el TCL para calcular probabilidades para p.

De la misma manera que como ocurrió con , pa p se le puede aplicar la regla empírica porque es normal, bajo las condiciones del TCL y por tanto también se podrá estandarizar para realizar cálculos probabilísticos con ella. Sólo hace falta conocer su ecuación de estandarización, lo cual no será difícil porque debe seguir la misma regla que las dos variables aleatorias previas. Por tanto su ecuación de estandarización será:

 (2)

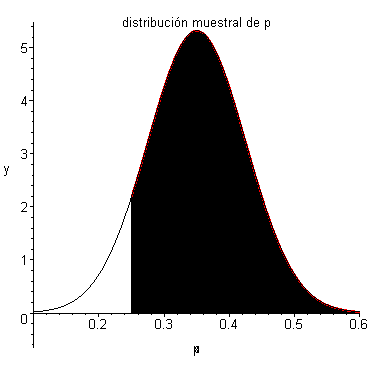
Demos un ejemplo:

Ejemplo. Se sabe que el 35% de los alumnos que ingresan al CCH terminan su bachillerato en 3 años. ¿Cuál es la probabilidad de que una muestra de 40 alumnos recién ingresados, al menos el 25% termine su bachillerato en 3 años?

La proporción muestral es p = 0.25, la proporción poblacional es de 0.35 y la desviación es de 0.075, luego:

1. P(p ≥ 0.25) = Φ(-1.33) = P(z ≥ -1.33) = 1- 0.0918 = 0.9082





Lo que significa que el 90.82% de las muestras 40 alumnos al calcularles p, en ella habrá al menos 10 alumnos egresados en 3 años.

De la misma manera se habrá observado que p es una estimación de π. Por ejemplo, en el caso de los obreros textiles, resultó que en la muestra de 20 obreros, el 20% de ellos estaban bien pagados, mientras que π = 0.272. Esto significa que habrá un error de estimación de

|0.20 – 0.272| = 0.072

Y por lo tanto, todo el procedimiento para calcular la probabilidad de que el error máximo de estimación, cuya expresión será semejante a la de la media muestral dada en el boletín anterior, será muy parecida a la manera en que se hizo anteriormente. Luego:



Por lo tanto

 (3)

y si despejamos *e* de (3) se tendrá:

 (4)

La ecuación (4) nos permite calcular el error de estimación cuando se conozca el tamaño de la muestra n; la proporción poblacional y el valor de z que se propone según el valor que se proponga previamente para el nivel de confianza, tal y como se hizo para la media muestral.

Ejercicio: ¿Cuánto debe valer el error de estimación si se va a tomar una muestra de n = 40 de la población de los salarios de los obreros textiles, para un nivel de confianza del 95%?

Sustituyendo



Esto significa que cuando se toma una muestra de 40 salarios y se calcula la estimación p, entonces 95 de cada 100 veces esta proporción muestral se alejará como máximo una distancia de 13.7% con respecto de π.

Aquí también podemos darnos cuenta que se están haciendo predicciones acerca de lo que ocurrirá cuando se realice un muestreo. Recordemos que esto es justamente lo que se hace en las encuestas: realizar predicciones. Recuerda el ejemplo de la aprobación de los habitantes de Cuautitlán Izcalli dado en el boletín anterior.

Una segunda fórmula igualmente importante se deduce de (4). Es la siguiente

 (5)

Con la que podemos calcular el tamaño de muestra para estimar π.

Ejemplo: ¿De qué tamaño debe tomarse una muestra de la población de obreros textiles, con un nivel de confianza del 95% de manera que el error máximo de estimación no sea mayor de 5%?

Sustituyendo



Lo que significa que la muestra debe ser de 305 obreros para que al calcular p de la muestra, este valor no se aleje más de 5 % de la verdadera proporción π. Esto sucederá 95 de cada 100 veces que se realice el muestreo en las mismas condiciones.

**EJERCICIOS**

1. Deduce las ecuaciones (3), (4) y (5)
2. Calcula el tamaño de muestra para ejercicio final del boletín usando los siguientes errores de estimación: 20%, 15%, 10%, 7%, 3% y 1%. ¿Qué observas ¿porqué ocurre esto? ¿qué concluyes cuando realices un muestreo para estimar π?
3. Observa que en la ecuación (5) para calcular n es necesario que conozcas π. Esto es contradictorio y por lo tanto esta ecuación no se puede usar realmente ¿Cuál es la dificultad en la ecuación?
4. Escribe la función de densidad para el problema de los obreros textiles y dibuja la curva normal que hayas obtenido usando un software graficador.

1. Calcula la probabilidad de que al extraer una muestra de la población A (250 datos dados en clase) de tamaño 35 y calcules *p* (proporción de obreros con buen salario), la proporción muestral de obreros bien pagados que se obtenga sea menor al 50%. Considera que un obrero de esta población estará bien pagado si gana 240 o más dólares. (*Simula el fenómeno: toma una muestra aleatoria de la población de 35 obreros y calcula p. ¿Está el valor de p dentro de las probabilidades previstas?)*
2. En junio del 2000, se estimó que la proporción de habitantes del país que habían sido asaltados era de un 29%. ¿Cuál es la probabilidad de que en un grupo de 150 personas más del 40% hayan sido asaltadas? ¿de que menos del 25% hayan sido asaltadas? ¿de que más del 50% lo hayan sido? *(Con base en estos resultados: ¿puedes idear un procedimiento para rebatir las cifras oficiales de la disminución del crimen? ¿en qué consistiría?)*
3. De los tabulados del INEGI, se dan cifras acerca del número de emigrantes por estado. Particularmente en el DF, se puede calcular la proporción de emigrantes de los estados de Michoacán, Guerrero, Oaxaca, Puebla y México entre otros estados además del total. Si se toma una muestra aleatoria de 300 personas del DF, ¿cuál es la probabilidad de que la proporción de emigrantes de los estados citados en el DF sea mayor al 60%?
4. ¿De qué tamaño debe tomarse una muestra de alumnos del CCH para que la proporción muestral de los alumnos que trabajan se aleje a lo más 0.04 de la proporción poblacional con un nivel del 90%?
5. Calcula el error máximo de estimación de la proporción de alumnos que trabajan, cuando se toma una muestra de 200 alumnos del CCH, obteniéndose una proporción de 0.22. Hazlo para un nivel de confianza del 98%. Explica el resultado.

**LECTURA**

El pionero de la demografía moderna y también de la estadística, fue un comerciante londinense llamado John Graunt (1620-1674). No tenía ningún tipo de preparación formal sobre el tema, pero entró como aprendiz en una tienda de artículos para caballeros y se convirtió en un próspero hombre de negocios. Las numerosas muertes en los años de plaga londinense despertaron en Graunt el interés por la demografía y la estadística. El dato más inquietante relativo a la población británica era el alto índice de mortalidad de los años de plaga; algunas de las cotas más altas se alcanzaron durante la vida de Graunt. En 1625 por ejemplo murió aproximadamente una cuarta parte de la población. Ya en 1527 se habían confeccionado en Londres listas con los muertos, llamadas *bills of mortality*, y en 1592 éstas incluían las causas de la muerte. Graunt no sabía “en virtud de qué accidente”, comenzó a interesarse por las listas de mortalidad. Como hombre práctico que era, le extrañaba que unos datos recogidos de manera tan regular y en gran número se explotaran tan poco. En 1662 escribió un librito de unas 90 páginas*, “Natural and political Observations made upon the bills of mortality*, en el que “...pretendía haber reducido varios confusos volúmenes de las listas de mortalidad a unas pocas tablas inteligibles y haber resumido las observaciones que naturalmente derivaban de ellas en unos pocos párrafos sucintos evitando las largas series de deducciones.”

Graunt agrupó todos los datos similares que se hallaban en las listas de mortalidad, por ejemplo, observó que sólo 2 personas de cada 9 morían de enfermedades agudas, 70 de cada 229 de enfermedades crónicas, 4 de cada 229 de dolencias externas (fracturas, tumores, llagas, etc.). El 7% moría de vejez, mientras que algunas enfermedades y siniestros mantenían una proporción constante. El porcentaje de homicidios en Londres era uno de cada dos mil muertes y sólo 1 entre 4000 moría de hambre. En Inglaterra había más hombres que mujeres y aunque “los médicos tienen dos pacientes mujeres por cada hombre, mueren más hombres que mujeres”. La población rural se multiplicaba cada 280 años mientras que la londinense cada 70 años. La explicación que Graunt ofrecía era que “...la razón que explica esta diferencia es que muchas personas en edad de reproducirse abandonan el campo y llegan a Londres procedentes de todas las regiones del país; las personas que procrean en el campo son casi las que exclusivamente nacieron allí, pero en Londres lo hacen muchas otras”. Tal vez puede afirmarse que era la primera vez que los números ofrecían una explicación de la tendencia social. Además, Graunt aducía, entre muchas otras, las siguientes cifras: El número de nacidos que sobrevivían hasta los 6 años de edad era de 64 de cada 100 y los que sobrevivían hasta los 76 años era de 1 de cada 100.

El mismo año de la publicación de las *Observations,* Graunt fue admitido en la Royal Society. El libro causó buena impresión en el rey Carlos II, y para evitar objeciones de los integrantes de la Royal Society al hecho de que Graunt era comerciante, el rey dijo que “si encontraban algún comerciante más de su estilo, lo admitiesen sin más ceremonia”.

Graunt terminaba su libro defendiendo los datos estadísticos:

“Si todas estas cosas se conocieran con claridad y seguridad (yo no he hecho más que conjeturas), se pondría de manifiesto que sólo una pequeña parte de la gente trabaja en actividades y profesiones necesarias, es decir, ¿cuántas mujeres y niños no hacen nada más que aprender a gastar lo que otros ganan, cuántas personas no son más que meros sibaritas y cuántas meros jugadores de profesión, cuántas viven de engañar al pueblo con ininteligibles doctrinas divinas y filosóficas, cuántas de convencer a personas crédulas, frágiles e inclinadas a pleitear de que sus cuerpos o propiedades están en malas condiciones y en peligro, cuántas luchan como soldados, cuántas viven meramente de oficios placenteros o de adorno, cuántas mediante servicios mal prestados, etc.? Y, por otra parte, cuán pocas están empleadas en el cultivo y la elaboración de los alimentos necesarios y de viviendas; y de los hombres especulativos, cuán pocos son los que realmente estudian la naturaleza de las cosas.”

Entre las consecuencias prácticas de los datos de Graunt están las dos siguientes: en primer lugar, recomienda un salario anual garantizado (recordemos que en esa época Inglaterra aún no era una sociedad industrial), razonando de la siguiente manera: (1) Londres está repleto de mendigos. (2) Casi nadie muere de inanición. (3) Por lo tanto, la riqueza nacional ya los está alimentando. (4) No deberían ser puestos a trabajar pues su producto sería de mala calidad y los holandeses ganarían para sí el comercio británico, y (5) Así que, sin ningún costo extra para la nación deberíamos alimentarlos y evitar que desmerezcan la vía pública a través de su limosneo. Sin embargo, los consejos de Graunt no fueron escuchados: se organizaron casas de caridad en las que se emplearon a londinenses pobres, lo que propició que la manufactura de artículos ingleses resultara de mala calidad, como Graunt lo había previsto.

También estimó por primera vez la población total de Inglaterra: pensó que si se conocía el número de nacimientos y la tasa de fertilidad de las mujeres, entonces podía calcular el número de mujeres fértiles del país y por tanto el número de familias. Si consideraba un promedio de integrantes por familia, un sencillo producto le permitiría calcular la población total de Inglaterra. El método adolecía de muchos defectos, pero la aplicación intuitiva de los fundamentos de la Estadística, que aún ni siquiera eran precisados, era bastante notable.

Graunt murió en 1674 de ictericia.

# Guardar con el nombre nombre-apellido.E3.3.4.2Proporciones muestrales-grupo.doc