Socioestadística I
Análisis estadístico en Sociología

Capítulo 4

TEORÍA DE LA PROBABILIDAD Y SUS PRINCIPIOS. ESTADÍSTICA INFERENCIAL

1. INTRODUCCIÓN A LA ESTADÍSTICA INFERENCIAL

En los capítulos anteriores, hemos utilizado diversas técnicas para obtener y presentar de forma resumida la información estadística con el fin de facilitar la interpretación y análisis de los datos. Esto es lo que se llama Estadística Descriptiva. Estas técnicas no se pueden utilizar para hacer generalizaciones que sean aplicables a individuos, características o ítems que no hayan sido observados.

En muchas ocasiones estamos interesados en determinar o explicar un mayor número de individuos que sería imposible hacerlo de uno en uno por lo que utilizamos las muestras (subgrupos) de esa población. Ósea, la obtención de generalizaciones estadísticas sobre un población determinada, a partir del estudio de las características extraídas de una muestra de dicha población o universo.

También con la estadística inferencial, podemos hacer predicciones sobre el comportamiento de poblaciones a partir del estudio directo de muestras que pertenecen a esa población. Conociendo el comportamiento de la muestra, podremos predecir el comportamiento de la población.

Si las técnicas se utilizan tan solo para resumir datos, se trata entonces de técnicas descriptivas. Si se utiliza para estimar parámetros de una población a partir de los datos de una muestra, se utilizará la estadística inferencial.

Si hablamos de parámetros, nos estamos refiriendo a características de una población y si hablamos de indicadores o estadísticos, nos referiremos a características de una muestra, la muestra de esa población.

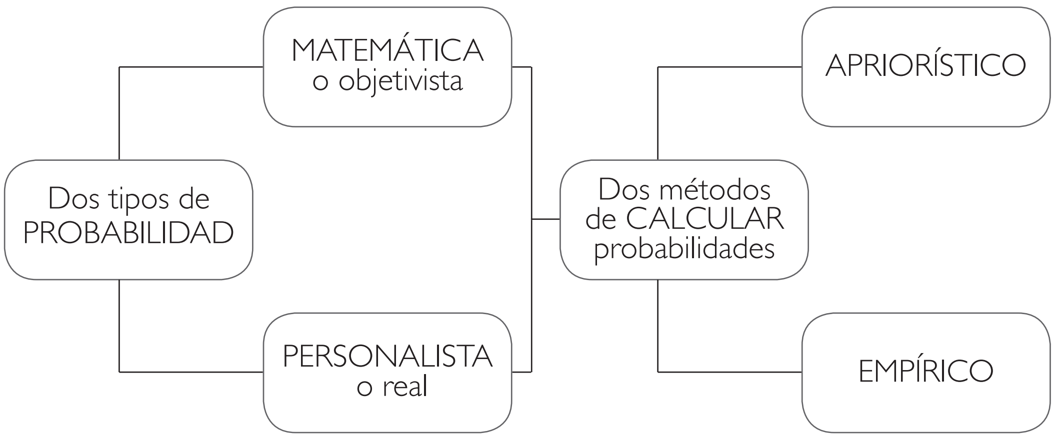
Para distinguir las características de unos y otros, utilizaremos las letras griegas (si es para la población) y las letras latinas (si es para la muestra).

Los parámetro son valores fijos de la población y suelen desconocerse y los estadísticos que varía de muestra a muestra, se utilizan para estimar los parámetros.

El proceso de estimación, eje de la estadística inferencial, se basa en la teoría de las probabilidades y en la teoría del muestreo. Por ello, vamos a empezar a examinar lo que significa la probabilidad en la estadística.

2. DEFINICIÓN, PROPIEDADES Y PROCESOS.

2.1. DEFINICIÓN.



Se puede definir el concepto de probabilidad de forma estricta desde dos puntos de vista que aparentemente resultan contradictorios:

Por un lado tenemos la probabilidad matemática: que se refiere a sucesos repetidos bajo condiciones determinadas y constantes. Aquí la probabilidad matemática tiene muy poco de probable. La **Probabilidad Matemática** u objetiva se refiere al resultado medio de un gran número de apariciones del suceso u ocurrencias. Hace falta, por tanto, una perspectiva más a largo plazo, donde no permita establecer un estudio de la evolución del fenómeno que nos ayude a establecer el valor que más se repite o que más veces se da. Es pura matemática y como tal solo afecta a conceptos que sólo se pueden estudiar desde esa perspectiva (juegos de azar, tirada de dados, los errores de una medición repetida, etc). Las demás probabilidades, simplemente no lo son tal, solo, o son correctas o son incorrectas, pero nunca será una probabilidad.

Luego tenemos la **Probabilidad Real** o personalista, en donde se emplea la probabilidad como el grado de creencia que tenemos sobre si un suceso vaya o no a ocurrir. También está la posibilidad que distintas personas tengan opiniones distintas sobre lo que va ha suceder.

Con todo ellos podemos definir la PROBABILIDAD (definición clásica) del siguiente modo:

$$\text{Probabilidad de un suceso} = \frac{\text{Sucesos o casos favorables}}{\text{Casos totales o posibles}}$$

$$P(A) = \frac{a}{n} = \frac{\text{Casos favorables}}{\text{Casos posibles}}$$

La probabilidad A PRIORI es el cociente entre el número de casos favorables y el número de casos posibles, teniendo en cuenta que todos los casos son igualmente probables. Aquí se determina la probabilidad en base a la lógica y a la naturaleza del suceso en lugar de la experiencia o experimentación. Tiene la dificultad de que se basa en el supuesto de igualdad o igual probabilidad o sucesos igualmente probables y en los fenómenos sociales esto no suele ocurrir así.

El método EMPÍRICO se basa en el supuesto de que la proporción de aparición de los sucesos observada en el pasado persistirá en el futuro. Cuanto mayor sea el número de casos observados en el pasado, más precisa será la estimación.

2.2. PROPIEDADES MATEMÁTICAS DE LA PROBABILIDAD.

1. La probabilidad nunca puede ser mayor a la unidad ni menor de cero.

$$0 \leq P(A) \leq 1$$

2. La segunda regla se puede considerar como un caso especial de la regla de adición. Si A y B son sucesos mutuamente excluyentes, la probabilidad de obtener A o B, que se denota por "P(A o B)" es igual a la probabilidad de A más la probabilidad de B.

$$P(A \text{ o } B) = P(A) + P(B)$$

Excluyente significa que no puede ocurrir simultáneamente en el mismo experimento. Ejemplo: no se puede obtener en la misma tirada una cara y una cruz de una sola moneda. Si sale un resultado, excluye o eliminada la probabilidad del otro resultado.

La regla de la adición se puede extender al caso de más de dos sucesos.

Hay que tener en cuenta, que la suma de todos los sucesos posibles de un fenómeno ha de ser la unidad:

$$P(A \text{ o } B \text{ o } C) = P(A) + P(B) + P(C) = 1$$

Y que la probabilidad de que no suceda un sucedo A será igual a la suma de todas las probabilidades de los restantes (mutuamente excluyentes) sucesos. Si sustraemos P(A) a la unidad, tendremos la probabilidad de no obtener A:

$$1 - P(A) = P(B) + P(C)$$

Si los resultados no son mutuamente excluyentes la ecuación quedará de la siguiente manera:

$$P(A \text{ o } B) = P(A) + P(B) - P(AB)$$

en donde P(AB) representa la probabilidad de obtener simultáneamente los sucesos A y B. Ejemplo: Si tenemos una región de España que el 75% de la población ha votado en las elecciones, que el 54% de esa población son mujeres y que el 40% de esas mujeres además han votado en la elecciones. La probabilidad de que un residente de esa región sea mujer o haya votado en las elecciones se calculará:

$$P(A \text{ o } B) = P(A) + P(B) - P(AB)$$

$$\boxed{P(A \text{ o } B) = 0{,}75 + 0{,}54 - 0{,}40 = 0{,}89}$$

3. La probabilidad de que dos o más sucesos ocurran simultáneamente *(la regla de la multiplicación)*: si A y B son dos sucesos cualesquiera, la probabilidad de obtener simultáneamente A y B es igual a la probabilidad de obtener uno de ambos sucesos

multiplicada por la probabilidad condicional de obtener el otro suceso una vez ha ocurrido el primer suceso.

$$P(AB) = P(A)\ P(B/A) = P(B)\ P(A/B)$$

$P(B/A)$ y $P(A/B)$

Estas son la probabilidades condicionales. La probabilidad del suceso A puede depender de la ocurrencia de otro suceso B y viceversa, en este caso son sucesos dependientes.

$P(B/A) = P(A)$

Si no afecta B a la ocurrencia de A, tenemos sucesos independientes

Vamos a ver un ejemplo en la que no existe independencia:

	Izquierda	Derecha	Neutro	Total
Conflictivo	150	300	150	600
No Conflictivo	300	50	50	400
Total	450	350	200	1000

Vamos a calcular la probabilidad de que un joven que seleccionemos sea conflictivo e ideológicamente neutro.

1º Aplicando la probabilidad:
Casos favorables = conflictivo y neutro = 150
Casos posibles = Total de jóvenes = 1000

$$P(A) = \frac{\text{Casos favorables}}{\text{Casos posibles}} = \frac{150}{1000} = 0{,}15$$

2º Aplicando la propiedad de la multiplicación:

$$P(AB) = \underset{\text{neutro}}{P(A)}\ \underset{\substack{\text{ne/conf}\\ \text{neutro}}}{P(B/A)} = \underset{\text{conflictivo}}{P(B)}\ \underset{\substack{\text{ne/conf}\\ \text{conflictivo}}}{P(A/B)}$$

$$P(AB) = \frac{200}{1000} \times \frac{150}{200} = \frac{600}{1000} \times \frac{150}{600}$$

$$P(AB) = 0{,}2 \times 0{,}75 = 0{,}6 \times 0{,}25 = 0{,}15$$

También tenemos que tener en cuenta los estudios longitudinales o de series temporales. Si disponemos de un modelo que explica como se desarrollan los sucesos de un fenómeno a lo largo del tiempo tenemos lo que se llama PROCESOS y si está regido por las leyes de la probabilidad de denomina PROCESOS ESTOCÁSTICOS.

Si los dos sucesos son independientes entre sí, ósea que el suceso A no depende que le ocurra el suceso B, entonces tenemos que P(B/A) = P(B) y P(A/B) = P(A). La regla de la multiplicación se simplifica, ya que la probabilidad de la ocurrencia conjunta de los sucesos independientes es igual al producto de las probabilidades por separado.

$$P(AB) = P(A)\,P(B)$$

$$P(ABC) = P(A)\,P(B)\,P(C)$$

2.3. VARIACIONES Y PERMUTACIONES.

Ya hemos dicho que para calcular las probabilidades de un suceso nos basamos en el calculo de los casos favorables partido los casos posibles, pero muchas veces, no podemos determinar las distintas posibilidades que puede tomar un suceso o por lo menos sin utilizar una operación matemática. Para ello, vamos a utilizar las siguientes operaciones:

VARIACIONES: Se refiere a los distintos grupos que se pueden formar con m elementos tomados de n en n (siendo n < m), con la condición de que dos grupos serán distinto si difieren en el orden o en la naturaleza de sus elementos. ($V_{m,n}$).

$$V_{m,n} = m\,(m-1)\,...\,(m-n+1) = \frac{m!}{(m-n)!}$$

Por ejemplo, si tenemos un grupo de cuatro elementos y los tenemos que agrupar de 2 en 2 el número de grupos que podríamos formar sería:

$$V_{4,2} = 4 \cdot 3 = \frac{4!}{(4-2)!} = 12$$

Si los cuatro elementos del grupo fueran a - b - c - d y los tuviéramos que agrupar de dos en dos, tendríamos: (a,b), (a,c), (a,d), (b,a), (b,c), (b,d), (c,a), (c,b), (c,d), (d,a), (d,b), (d,c). Estos doce grupos o tienes elementos distintos o bien los tienes en orden distinto.

Cuando pueden darse repeticiones de los elementos, tenemos las VARIACIONES CON REPETICIÓN y la ecuación es:

$$VR_{m,n} = m^n$$

En el mismo ejemplo de antes:

$$VR_{4,2} = 4^2 = 16$$

Tenemos los mismos grupos que antes, más cuatro más en los que se repiten el mismo elemento: **(a,a)**, (a,b), (a,c), (a,d), (b,a), **(b,b)**, (b,c), (b,d), (c,a), (c,b), **(c,c)**, (c,d), (d,a), (d,b), (d,c), **(d,d)**. Difieren en el orden o en la naturaleza de sus elementos, pero hay cuatro grupos en los que coinciden sus elementos.

PERMUTACIONES: Cuando los grupos varían tan solo en el orden de los elementos que los integran, es una caso particular de las variaciones en las que m=n.

$$P_{n,n} = n!$$

Siguiendo con los ejemplos anteriores:

$$P_{4,4} = 24$$

Con los cuatro elementos del ejemplo anterior, podemos formar 24 grupos que difieren entre sí en el orden de sus elementos: (a,b,c,d), (a,b,d,c), (a,d,c,b)... los grupos solo difieren en el orden de los elementos.

Si admito las repeticiones, tenemos las PERMUTACIONES CON REPETICIONES que son los distintos grupos que se pueden formar con n elementos, dentro de los cuales se repiten $n_1, n_2, ..., n_k$ elementos con la condición de que sean distintos en el orden.

$$PR_n = \frac{n!}{n_1!\, n_2! \ldots n_k!}$$

2.4. COMBINACIONES.

Hasta ahora, los grupos se consideraban distintos si variaban en ellos el orden o la naturaleza de los elementos. Pero podemos estar interesado en obtener grupos que solo difieran entre si por la naturaleza de los elementos y no por su orden. Por ello las combinaciones son los distintos grupos que se pueden formar com m elementos tomados de n en n (siendo $n < m$) con la condición de que dos grupos sean distinto si difieren en la naturaleza de alguno de sus elementos.

$$C_{m,n} = \binom{m}{n} = \frac{m!}{n!\,(m-n)!} = \frac{(m)(m-1)\ldots(m-n+1)}{n!}$$

Ejemplo: Si tenemos un población de 4 elementos, las combinaciones de dos en dos que podemos formar son:

$$C_{4,2} = \binom{4}{2} = \frac{4!}{2!\,(4-2)!} = \frac{4\cdot 3\cdot 2\cdot 1}{2!\cdot 2!} = \frac{4\cdot 3\cdot \cancel{2}\cdot 1}{2\cdot \cancel{2}} = \frac{4\cdot 3}{2} = 6$$

Las combinaciones que podríamos formar si difieren solamente en la naturaleza de los elementos de 4 elemento tomados de 2 en 2 son: (a,b), (a,c), (a,d), (b,c), (b,d), (c,d).

Las propiedades de las combinaciones, son:

1. Las combinaciones de n elementos tomados de n en n es igual a la unidad.

$$C_{n,n} = \binom{n}{n} = \frac{n!}{n!\,(n-n)!} = \frac{n!}{n!\,0!} = \frac{n!}{n!} = 1$$

2. También cumple:

$$\binom{n}{0} = 1; \quad \binom{n}{1} = n\,; \quad \binom{n}{2} = n\,(n-1)\,/\,2$$

En el caso de que tengamos en cuenta las repeticiones en las combinaciones, la ecuación será la siguiente:

$$CR_{m,n} = \binom{m+n-1}{n} = \frac{(m+n-1)!}{n!\,(m-1)!}$$

Las combinaciones con repetición de m elementos tomados de n en n ($m \geq n$), son los distintos grupos formados por n elementos de manera que: no entran todos los elementos, no importa el ordeny sí se repiten los elementos.

Ejemplo: En una bodega hay en un cinco tipos diferentes de botellas. ¿De cuántas formas se pueden elegir cuatro botellas?. Hay que tener en cuenta que no entran todos los elementos, sólo elijimos 4 botellas. No importa el orden. Da igual que elija 2 botellas de anís y 2 de ron, que 2 de ron y 2 de anís y sí se repiten los elementos. Puede elegir más de una botella del mismo tipo. De esta forma tenemos la ecuación como sigue:

$$CR_{5,4} = \binom{5+4-1}{4} = \frac{(5+4-1)!}{4!\,(5-1)!} = \frac{8!}{4!\,4!} = 70$$

BIBLIOGRAFÍA

- ALCAIDE INCHAUSTI, Angel: *Estadística aplicada a las Ciencias Sociales,* Madrid, Pirámide, 1976.
- AMÓN, Jesús: *Estad´sitica descriptiva para psicologos,* Madrid, 1973.
- GARCÍA FERRANDO, M.: *Socioestadística. Introducción a la estadística en sociología,* Madrid, Alianza Editorial. 1989.