

INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

¿Porqué Bioestadística?

A pesar de ser una disciplina científica relativamente nueva, el uso de la Estadística se expandió rápidamente a la mayoría de todos los campos del conocimiento humano, ello a partir de la instalación de sus primeras Universidades a fines del siglo XIX y principios del XX en Inglaterra y Estados Unidos; en la actualidad el área de la salud cada vez se familiariza más con la aplicación de la metodología estadística en todos sus campos del conocimiento, hoy por hoy cualquier médico o enfermera, no nada más podemos hacer uso de esa maravillosa herramienta, sino que cada vez más se convierte en un recurso indispensable tanto para nuestra vida de relación con la sociedad, como también para facilitar nuestra actualización científica y profesional.

La estadística aplicada a los aspectos de la salud también es un instrumento necesario y determinante en la investigación médica así como en todas sus disciplinas científicas, esto lo facilita el hecho de que la estadística utiliza el mismo lenguaje en cualquier área del conocimiento en el que interviene, lo que contribuye a dar una mayor apertura hacia la comprensión de tópicos médicos, aun por la gente que no ha estudiado medicina.

En forma paralela, la computación al igual que el uso de un esfigmomanómetro, de un electrocardiógrafo o de un simple estetoscopio, se ha presentado en su debido momento tanto en el campo clínico como en el epidemiológico como otra herramienta más, la cual junto con el acceso a la metodología estadística y a la metodología científica favorecen a pasos agigantados la comprensión de este antiguamente oscuro tema.

Conviene comprender que el desarrollo entero de la civilización ha requerido de estándares objetivos de medición indispensables en todos los órdenes de la vida; es decir, los patrones de medición ya sea de longitud, o de masa o de cualquier otra índole se han mantenido como piedras angulares del desarrollo científico, tecnológico, industrial, etc., ello no le resta a la subjetividad esa extraordinaria potencialidad en el progreso de las matemáticas, de la filosofía misma o de las actividades artísticas humanas, sin embargo, es menester pensar en el papel que han asumido



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

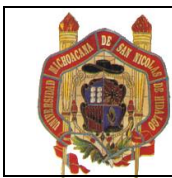
todos los patrones de medición en nuestro mundo para comprender que la estadística ha acompañado al hombre desde sus primeros pasos, ejemplos sobran, pero donde verdaderamente asumió una importancia total, fue desde el momento en que ha permitido establecer inferencias y predicciones acerca de un determinado fenómeno. Esa extraña fascinación del hombre de asomarse al mismo tiempo al pasado y al futuro, se ha convertido en una necesidad moderna, viva y ampliamente presente.

En el cuadro subyacente se muestra una breve y simplificada cronología del empleo de la estadística históricamente.

Época	Hechos	Autores
sXV ?	Aritmética de asuntos de Estado	Roma (imperio)
1532	Registro de mortalidad en Londres	Sir W. Petty
1539	Registro de mortalidad bodas y bautismos en Francia	Archivos históricos
sXVII	Estudio de la probabilidad	sXVIII
1730	La distribución normal	Bernovilli-De Moivre
sXIX	Astronomía, Sociología, Educación	Laplace-Gauss
sXIX	Herencia, Eugenesia, Psicología, Antropometría, Correlación.	Galton-Pearson
sXIX	Cátedras (James McKeen Cattell, E.L. Thorndike)	EUA, Inglaterra.
sXX	Metodología en muestras pequeñas (Biología-Agricultura)	R.A. Fisher

¿Qué puede hacerse con la Estadística?

La metodología empleada por la estadística se ha desarrollado a la par con el uso de la metodología científica, difícilmente se puede imaginar a una cosa sin la otra; es indispensable recordar que detrás de cada conocimiento sólo tenemos un determinado nivel de certeza, y que un paradigma científico dura exclusivamente el tiempo que tarda en producirse otro. De tal manera que en la actualidad, más que hablar de cualquier tipo de certeza, preferimos hablar de



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

una probabilidad, pero de una probabilidad probada. El uso de la metodología probabilística en el campo de la salud y de las disciplinas afines se reconoce con el nombre de **bioestadística**, sin embargo, no deja de ser simplemente estadística.

En todo caso, al trabajar buscando un principio de certidumbre debemos definirle un nivel de certeza o de probabilidad; es decir, un nivel de fiabilidad y de validez, de precisión y de exactitud, pero siempre en términos de probabilidad. Independientemente de que los procedimientos operativos para comprobar la precisión y exactitud de una afirmación son generalmente lentos y dilatados, la mayor o menor reproducibilidad de sus resultados es un elemento indispensable para poder hablar de su validez y de su fiabilidad. Además, una afirmación estadística acerca de un resultado cualquiera no es fiable y no puede ser aplicativa si no se incluyó el análisis estadístico desde las premisas, ¿qué quiere decir esto?

Quiere decir, que el análisis estadístico no se aplica a los resultados de un estudio o de un experimento como una forma de calificarlos, avalarlos o descartarlos a posteriori. De hecho, la aplicación de la estadística debe estar presente desde el nacimiento de los proyectos de estudio o de investigación, desde la elección de los métodos más idóneos para cualquier medición, en fin, la estadística debe ayudarnos a determinar cada uno de los pasos que se planean ejecutar, de otra manera se incurre en el error más frecuentemente cometido entre las personas que desean hacer estudios de investigación, y es la pérdida o el desperdicio de información por ausencia de una metodología adecuada, cuando esto sucede, cuando la pérdida de información tiende a superar a la obtenida, el análisis estadístico final pierde importancia, pierde poder y deja de ser tal.

En ese pequeño paso arriba descrito se pierde no nada más información importante, se puede perder fuerza de trabajo, entusiasmo e inventiva por parte de los estudiantes y los investigadores jóvenes, en suma, se puede convertir en un esfuerzo contraproducente si no se opta por hacer cada cosa **simplemente bien**. Sin importar el grado de simplicidad o el grado dificultad, e independientemente de los costos económicos, nunca se debe olvidar que finalmente sólo hay



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

dos tipos de investigación en todos los campos del conocimiento humano: *la investigación que se hace bien y la que no se hace bien.*

¿Con qué fin se aplica la metodología estadística?

El objetivo de una investigación persigue impecablemente al indicador de certeza o de probabilidad de un resultado, el cual recibe el nombre de **estadímetro** cuando hablamos de la información de una muestra menor representativa de una población, y recibe el nombre de **parámetro** cuando se trata de la información completa de toda una población o universo de cosas.

Al hablar del estadímetro o parámetro, estamos hablando de cualidades, características o simplemente de cualquier cosa mensurable (esto es, que se puede medir), y cuya dimensión puede variar en mayor o menor grado entre un caso y otro dentro de una misma población dependiendo de la uniformidad de sus componentes. Dada dicha variación en sus dimensiones se le reconoce con el nombre de **variable** a todo aquello que se esté midiendo.

De acuerdo con lo anterior, muy pocas variables o ninguna podrá alcanzar a ser considerada realmente como un parámetro de las grandes poblaciones actuales, supuesto que para definir las se precisa medirlas en cada uno de sus componentes lo cual es verdaderamente difícil, a veces es casi imposible y en la mayoría de las veces totalmente injustificable, en cambio, a partir de un excelente estadímetro obtenido en tan solo una pequeña muestra representativa de la misma población se pueden obtener nociones más o menos precisas de lo que es el parámetro respectivo.

La forma de medir algo puede variar infinitamente, pero en un proyecto de investigación debe ser muy puntual, explícita y tan universalmente reconocida que cualquiera que tenga acceso a los resultados de un estudio pueda comprender la manera de reproducirlo puntualmente.



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

Dependiendo de la modalidad de que se trate, la estadística puede emplear tanto datos numéricos como datos no numéricos, ello es importante ya que los procedimientos empleados para el análisis de cada tipo de datos son necesariamente distintos so pena de asumir errores metodológicos o sistemáticos muchas veces menospreciados indebidamente.

El tipo de datos depende exclusivamente de la manera de medir la variable de que se trate, sólo trataremos de **dos tipos de variables**, una **continua** (necesariamente numérica) y una **no continua** (discontinua o discreta y también, necesariamente numérica). Las variables continuas siempre son datos numéricos y regularmente se emplean en entidades tales como la longitud, la masa y el tiempo, y sus valores pueden ser expresados en unidades entre las cuales siempre hay al menos un valor evaluable. Las variables discontinuas pueden ser datos numéricos pero también pueden ser datos no numéricos, sin embargo los datos no numéricos terminan siendo contabilizados y expresados en unidades entre las cuales no hay un valor evaluable.

Cada variable poseerá un grupo de datos extraído de los actos de medición. Cuando queremos saber cuál es el comportamiento de cada variable, organizamos cada grupo de datos según su escala. Para ello sabemos al menos de 4 tipos diferentes de escalas:

1) *La escala cardinal o nominal*: nombres, categorías, cualidades, etiquetas. No maneja datos numéricos y si acaso se usaran números, éstos se emplearían simplemente como letras o etiquetas. Esta escala solo permite contar la frecuencia con la que aparece una categoría o cualidad, también puede ser expresada como un porcentaje o una fracción de un total. Un ejemplo sencillo sería decir que en un grupo escolar hay 30 mujeres y 20 hombres, también se podría decir que el 60 % del total son mujeres y que 30/50 (léase 30 de 50) son mujeres.

2) *La escala ordinal*: Ordenamiento en base a conceptos como $<$: menor que, $>$: mayor que, $=$: igual que, y no requiere el uso del cero. Es la escala numérica más rudimentaria, simplemente sirve para ordenar de menor a mayor una serie de valores, sin que esto signifique que entre una y otra unidad exista la misma magnitud, un ejemplo es catalogar una variable como el dolor en



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

términos de ligero, moderado e intenso o bien como dolor una cruz, dos cruces y tres cruces: el hecho de que la calificación de 2 cruces sea mayor que una cruz no significa que sea el doble.

3) *La escala de intervalos*: Mediciones cuantitativas con uso relativo de cero (cero no significa ausencia de una magnitud), las unidades no son equivalentes en todos los puntos de la escala (un ejemplo es la temperatura en grados centígrados o Fahrenheit). La escala de intervalos soporta operaciones matemáticas con uso de números racionales, sin embargo su aplicación está limitada al entorno de su cero relativo, en ocasiones se abusa del uso de intervalos donde ya se cuenta con una escala de razón.

4) *La escala de razón*: Mediciones cuantitativas, las unidades son equivalentes en todos los puntos de la escala, utiliza el cero en forma absoluta. Es aplicable en todas las variables continuas propiamente dichas, es decir, en aquellas que miden tiempo, masa o longitud, dineros y la temperatura absoluta.

La prioridad en todo experimento es contar con variables lo más cuantificables y continuas posibles ya que la escala de razón permite mayor exactitud en las determinaciones, mejor representatividad y fidedignidad en la reproducibilidad de la información, de manera que las variables principales, centrales o determinantes en todo experimento siempre tienden a ser medidas en escala de razón.

Con tal objetivo, son innumerables los avances en la instrumentación y en el desarrollo tecnológico de equipos diversos de medición prácticamente en todos los órdenes del quehacer científico. Primero, tratando de hacer mensurable todo aquello que no lo es, y segundo, perfeccionando la forma de medición de las distintas variables a escalas cada vez más cuantificables.

De manera que, la metodología de la estadística está fundamentada operativamente en el uso de *variables* continuas y/o discontinuas medidas en *escalas* predeterminadas con el objetivo de cuantificar la magnitud de un fenómeno en cualquier campo del conocimiento. Por un lado se



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

busca describir cuantitativamente o en el peor de los casos describir cualitativamente al fenómeno de que se trata, y por otro lado, se busca determinar parámetros o estadímetros que permitan en lo sucesivo controlar y anticipar sus probables variaciones.

Procedimiento descriptivo: El primer paso es construir una base de datos confiable, para ello se debe preferir un método que elimine los errores de interpretación y que respete al máximo la naturaleza de las mediciones. En una base de datos, éstos se ordenan en una forma en la que los casos inician cada renglón y las variables encabezan a cada columna. Cuando una variable es medida más de una vez, se puede construir una tabla aparte en la que las columnas sean ocupadas por cada una de las veces que se haya medido dicha variable.

No todos los datos son números pero la inmensa mayoría de las veces se trabaja con datos numéricos, en tal caso luego de construir la tabla respectiva se procede a obtener *valores de tendencia central* y *valores de dispersión* que permitan describir en detalle nuestro conjunto de datos y extraerles el máximo de información.

Entre los **valores de tendencia central** están: la media o promedio aritmético, la mediana y la moda, en forma ortodoxa la primera se usa para las variables continuas (medidas con escalas de razón o al menos de intervalo), y las dos últimas se emplean tanto en las variables continuas como en las discontinuas (medidas con escalas nominales, ordinales y de intervalo).

La media aritmética es el resultado de dividir la suma de todos los valores entre su número total, en algunas ocasiones podremos observar valores promedio que pueden representar a un grupo de datos aun cuando su valor no aparezca en ninguno de ellos, un ejemplo rápido: el promedio de 4 y 6 es igual a sumar ambos números lo cual nos da la cifra de 10 y luego lo dividimos entre 2 que es el número total de datos en este caso, de tal manera que la media es de 5. *La mediana* como el valor que corresponde al rango intermedio, también puede en ocasiones asumir un valor inexistente en el grupo de datos que se describe aun cuando la escala de medición lo impida, explícitamente la mediana de los numerales 4 y 6 es 5 justo lo mismo que la media, sin embargo, en una serie tan irregular por ejemplo de 3 numerales como el 4 el 6 y el 20, la media sería el



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

10, pero la mediana sería el 6. Por otra parte, *la moda*, que se define como el dato más frecuente en una serie numérica, siempre va a estar presente entre los datos que se describen.

Valores de dispersión.- Al lado de los valores de tendencia central regularmente aparece algún valor de dispersión, los valores de dispersión más usados para el caso de variables continuas son la desviación estándar y el error estándar de la media, aunque para ciertos propósitos se prefiere manejar a la varianza o a la desviación media, todos ellos se basan en las diferencias que existen entre cada uno de los datos para con su promedio, a mayor diferencia mayor dispersión y más variabilidad, a menor diferencia menor dispersión y menor variabilidad. Para el caso de las variables discontinuas se utilizan otros valores de dispersión como son *los valores del recorrido y los valores cuantilares*.

El recorrido suele ser llamado indebidamente como rango, rango se le llama al lugar que ocupa cada dato en un ordenamiento de todos los datos desde el menor hasta el mayor, pero el recorrido no es otra cosa que la magnitud que hay entre el valor mínimo y el valor máximo, de hecho el recorrido se representa con ambos valores y la letra a o un guión horizontal separándolos (mínimo a máximo). Los valores cuantilares o cuantiles pueden ser expresados de muchas maneras, la más frecuente se refiere a la distribución de un grupo de datos en percentiles, es decir, en un orden porcentual a partir de los valores de menor magnitud hasta alcanzar a los de mayor magnitud. En un ejemplo de 300 datos los 3 valores más chicos constituyen el percentil uno o pc1 y los 3 valores más grandes representan al percentil cien ó pc100. Los cuantiles que describen los valores de los pc10, pc20, ... pc100, son llamados por esa razón como deciles, pero también hay quintiles, cuartiles, etc...

Una vez descritas las variables en razón de su comportamiento como valores de tendencia central y como valores de dispersión, pueden interesarnos diversas cuestiones, las cuales suelen ser el motivo principal de un estudio, en todos los casos se trata de comparaciones ya sea de un antes versus un después, o bien comparar un grupo A contra un grupo B, etc. En toda comparación habrá que definir los criterios de decisión desde antes de hacer las comparaciones,



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

esos criterios dependerán del grado de exigencia pertinente para aceptar o no un probable error por azar.

El haber realizado la estadística descriptiva nos brinda información necesaria acerca de cómo están distribuidos los elementos de un conjunto o población estudiada, lo cual es determinante para poder tomar decisiones de tipo estadístico.

Desarrollo de una estadística descriptiva

El primer paso es construir una base de datos confiable, para ello se debe preferir un método que elimine los errores de interpretación y que respete al máximo la naturaleza de las mediciones. En una base de datos, los datos se ordenan en una forma en la que los casos inician cada renglón y las variables encabezan a cada columna. Cuando una variable es medida más de una vez, se puede construir una tabla aparte en la que las columnas sean ocupadas por cada una de las veces que se haya medido dicha variable en orden cronológico.

	Edad	Sexo	Peso	...
Caso 1				
Caso 2				
Caso 3				
...				

En el caso de los datos numéricos, después de contar con las tablas respectivas se procede a obtener **valores de tendencia central** y **valores de dispersión** que permitan describir ambos al conjunto de datos incluidos.

Entre los valores de tendencia central está la *media* o promedio aritmético, la *mediana* y la *moda*, sólo la primera se usa para las variables continuas (medidas con escalas de razón o al menos de



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

intervalo), y las dos últimas se emplean en las variables discontinuas (medidas con escalas nominales, ordinales y de intervalo), pero también para las continuas.

En la tabla siguiente se utilizan los mismos numerales para expresar tanto una variable continua como dos discontinuas, nótese, no representan a la misma cosa pero en aras de detallar la diferencia entre los tres valores de tendencia central se utilizan así.

Casos	Metros	Naranjas	Zapatos
1	5	5	5
2	6	6	6
3	7	7	7
4	8	8	8
5	9	9	9
6	9	9	9
7	10	10	10
	Media: 7.71	Mediana: 8	Moda: 9

Por otra parte los valores de dispersión son variados y distintos, empleando la misma serie numérica describiremos algunos:

Caso	Valor	Deltas	Deltas ²
1	5	-2.71	7.34
2	6	-1.71	2.92
3	7	-0.71	0.50
4	8	0.29	0.08
5	9	1.29	1.66
6	9	1.29	1.66
7	10	2.29	5.24
	Media: 7.71	Σ: 0.0	Σ: 19.4

Deltas: $(x - \bar{x})$, $\bar{x}$: media o promedio, deltas²: $(x - \bar{x})^2$ también conocidos como “cuadrados”. El número de casos se representa como una n o N.



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

El primer intento con los **deltas**, resulta infructuoso para hablar de valores de dispersión como se observa al momento de sumarse (igual a cero), en cambio la **suma de los deltas cuadrados** ya nos ofrece una cantidad la cual puede ser representada por su división entre el número de datos, solo que en todos los casos con $n < 30$, se recomienda realizar un ajuste que consiste en que en lugar de dividir entre n se divide entre $n-1$, en el ejemplo de la tabla precedente $19.4/n-1 = 19.4/6 = 3.23$ tal es el valor de la **varianza**, pero el estadímetro más usado por ser el más ajustado a los valores reales es la desviación estándar (d.e.), la **d.e.** = $\sqrt{\text{varianza}}$, en este caso la raíz cuadrada de la varianza es igual a 1.798. Otro dato de dispersión es el error estándar de la media (e.e.m.), el cual resulta de dividir la d.e. entre $\sqrt{n-1}$, en este caso **e.e.m.** = 0.734

Como se observó antes, otra modalidad de los valores de dispersión lo constituye el **recorrido** de los datos, en el caso previo decimos que el recorrido va de 5 a 10, esto se logra ordenando todos los valores del mínimo al máximo, en ocasiones suele llamársele rango a esta modalidad, sin embargo el **rango (R)** es cada uno de los lugares que ocupan los datos ordenados: R1, R2,...R7 Este ordenamiento suele ser aprovechado para realizar la descripción de los datos por **cuantiles** o por **percentiles**. Si tuviésemos una base de cien datos cada uno de ellos ordenados del menor al mayor ocuparía un percentil (pc): pc1, pc2, pc3...pc100, tal sería la descripción percentilar de dichos datos. Ante un número distinto a 100, los datos pc deben ser determinados en forma de rangos porcentuales.

Importancia de la distribución estadística o probabilística

En una población o conjunto cualquiera, sus elementos pueden ser clasificados u ordenados de distintas maneras ya sea por su tamaño, volumen, masa, color, etc. La característica por la cual se determina realizar esa clasificación se convierte en una variable estadística, por lo común en torno de esa variable todos los elementos se distribuyen de menor a mayor magnitud. El cómo se distribuye un conjunto de elementos suele ser una característica típica de cada población. Se habla de una tendencia cuasi universal en la que la distribución de los elementos de una población aglutina a la mayor parte de sus componentes (los más frecuentes) alrededor de los



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

valores medios, en cambio, a medida que se observan sus valores extremos: mínimo y máximo, los elementos tienden a disminuir en frecuencia de aparición hasta el cero.

A este fenómeno se le reconoce como tendencia o regresión a los valores medios, (de donde adopta su nominación la famosa regresión lineal, la cual se tratará en otra parte), y nos explica cómo los elementos de una población tienden a ser semejantes entre sí y cómo las diferencias existentes con otros conjuntos de elementos dependen en gran parte, de cómo están distribuidos ambos conjuntos en torno a sus respectivos valores medios.

Aún más, las diferencias que puedan existir entre dos muestras de un mismo conjunto o población de datos igualmente dependerán de la forma en que ambas muestras estén distribuidas en torno a los valores medios de su conjunto o población.

A partir del concepto de la distribución de valores, se genera uno nuevo que habla acerca de la probabilidad de que en algún lugar comprendido entre el valor mínimo y el valor máximo (del recorrido) encontremos el valor de un elemento determinado. De antemano sabemos que la mayoría de las probabilidades apuntan a encontrar elementos que tienen valores vecinos a los valores medios y que estas probabilidades se aproximan a cero conforme se trate de encontrar valores extremos mínimos y máximos. El azar nos habla de que todos los elementos tienen la misma probabilidad de ser observados, la distribución, en cambio nos habla de los valores más y menos frecuentemente observados.

La distribución de valores vista en base a la probabilidad se conoce como *distribución probabilística*, entre las distribuciones probabilísticas las más estudiadas son la *distribución binomial* y la *distribución normal*, ambas distribuciones llegan a tener un perfil semejante por la forma en la que tienden a aglutinar elementos en base a sus valores, aun cuando ambas manejan distintas escalas de medición.

La distribución binomial se apoya en escalas dicotómicas, esto es, mediciones de solo dos alternativas: si o no, todo o nada, en fin “x” o “y”; no importa en cuántas oportunidades se define



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

una u otra cosa, ambas pueden tener igual o diferente probabilidad de ocurrir, p.e., una moneda de características perfectas en su regularidad presenta el 50 % de probabilidades de obtener una cara y 50 % de probabilidades de obtener la otra cara. A la larga, en un uso persistente las probabilidades indicarán resultados directamente relacionados con el balance o desbalance de la moneda.

Por su parte, la distribución normal utiliza una escala racional, esta característica la identifica estrictamente con variables continuas, sin embargo toda aquella variable medida en escalas de intervalo y aun en escalas ordinarias describen preferentemente distribuciones similares a la normal, tales como la distribución de chi cuadrada y la distribución de Student.

INTRODUCCIÓN A LA TEORÍA DE LA PROBABILIDAD

En un número n de resultados probables existe otro número ns que posee la característica s y su probabilidad (P) de existir se define como $P(s) = ns/n$

De acuerdo con ello $P(s) \geq 0$ y $P(s) \leq 1$ cuando ns tiende a ser igual a 0 y cuando ns tiende a ser igual a n , respectivamente.

Por otra parte la ausencia de s denominada como $\sigma = n - ns$, por lo tanto $P(\sigma) = (n - ns)/n$; sabemos que:

$$[(n - ns) + (ns)] / n = n / n = 1; \text{ luego entonces } P(\sigma) + P(s) = 1$$

En un dado perfectamente regular sabemos que existen 3 números pares (p) y 3 números impares (i), la $P(p)$ es por tanto $= P(i)$; esto es: $np / (np + ni) = ni / (ni + np) = 1 / 2$, ambos casos son independientes y mutuamente excluyentes.

Se dice que dos sucesos son independientes entre sí cuando no hay una relación determinística entre ambos. La probabilidad de que ocurran los dos sucesos independientes al mismo tiempo es igual a la multiplicación entre sí de sus probabilidades aisladas, se dice por ejemplo la



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

probabilidad de que ocurra A y B = $P(A) \times P(B)$. En cambio, cuando A y B son sucesos mutuamente excluyentes para los que se desea saber la probabilidad de que ocurran A ó B, entonces es $= P(A) + P(B)$. Además, aunque no sean mutuamente excluyentes, se puede obtener la $P(A \text{ ó } B) = P(A) + P(B) - P(A \text{ y } B)$.

Otra modalidad en términos de probabilidad es la condicional, un ejemplo es cuando queremos saber la probabilidad de que perteneciendo a B se presente A, esto es: $P(A|B) = P(A \text{ y } B) / P(B)$. Dentro de las probabilidades condicionales encontramos probabilidades marginales, una de ellas puede ser la probabilidad de encontrar mujeres entre 80 y 89 años de edad dentro de una población: $P(80-89|Mujer)$.

REGLA DE BAYES

De acuerdo con los elementos previos, la regla de Bayes nos dice que en una población de casos, alguna característica puede ser predecible en base a la definición de su relación probabilística con sucesos que en mayor o menor grado le acompañan. En términos coloquiales de salud, una sintomatología S caracteriza a una enfermedad D, pero no siempre que sucede S quiere decir que exista D, cuando no hay enfermedad se define como Δ , de manera que la $P(S) = P(S \text{ y } D) + P(S \text{ y } \Delta)$ como sucesos mutuamente excluyentes.

Por otra parte, la $P(S | D) = [P(S \text{ y } D)] / [P(D)]$ y la $P(S | \Delta) = [P(S \text{ y } \Delta)] / [P(\Delta)]$. Despejando $P(S \text{ y } D) = [P(D)] \cdot [P(S | D)]$ y también: $P(S \text{ y } \Delta) = [P(\Delta)] \cdot [P(S | \Delta)]$

Complementariamente: la $P(D | S) = [P(S \text{ y } D)] / [P(S)]$, si sustituimos los términos de la variable dependiente por sus equivalentes:

$$P(D | S) = [P(D)] \cdot [P(S | D)] / [P(D)] \cdot [P(S | D)] + [P(\Delta)] \cdot [P(S | \Delta)]$$

En donde: $P(S | D)$ se define como la probabilidad de estando enfermo presentar la sintomatología y significa la sensibilidad con que la detección de S representa la presencia de D (ver el verdadero positivo adelante). En cambio, $P(S | \Delta)$ se define como la probabilidad de



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

que sin estar enfermo se presente la sintomatología y significa la probabilidad complementaria a la representada por la sensibilidad de la relación entre S y D (véase más adelante el falso positivo).

Si para efectos prácticos, queremos saber al estar presente la sintomatología, con que probabilidad encontraremos a la enfermedad en una población dada, **si la enfermedad tiene una prevalencia histórica del 1%** y las estadísticas indican que el 95% de los enfermos cursan con la sintomatología, al mismo tiempo hasta el 3% de la población no enferma registra la sintomatología; esto es, con 95% de **sensibilidad** y 97% de **especificidad**:

$$P(D | S) = \frac{(0.01)(0.95)}{(0.01)(0.95) + (0.99)(0.03)} = 0.24$$

El resultado nos dice que solo en el 24% de las veces que aparecen los síntomas coincide con la enfermedad en el total de la población.

LA PROBABILIDAD EN PERMUTACIONES Y COMBINACIONES

Existen series de eventos que se presentan en un orden determinado, el cual puede variar en una forma exhaustiva afectando a algunos o a todos sus componentes, tal es el caso de la probabilidad de las permutaciones, que se definen en notación factorial como ${}_kP_r = k! / (k-r)!$

Donde k= número de eventos o casos en total, r= número perteneciente a k que constituye una serie posible de sucesos ordenados (r puede ser igual a k), P= permutaciones.

Para ejemplo queremos saber el número de permutaciones posibles (P) en los 3 primeros lugares (r) de aprovechamiento en un total de 10 alumnos (k). Ello implica que sustituyamos en la ecuación previa:

$${}_{10}P_3 = 10! / (10-3)! = 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 / 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 3628800 / 5040 = 720$$



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

De manera que, si a todos los alumnos se les conceden las mismas probabilidades de quedar en el 1er. lugar, cada uno podría participar en $720/10 = 72$ arreglos distintos como tal, y también cada uno en $72 \times 3 = 216$ permutaciones diferentes podría quedar dentro de los 3 primeros lugares ($216/720 = 0.3$ o el 30% de probabilidad como la premisa inicial).

Por otra parte, cuando no nos interesa el orden que ocupen los r sujetos para el mismo caso previo, estaremos hablando de combinaciones (C) y ya no de permutaciones, luego entonces ${}_kC_r = k! / (k-r)! r!$ En donde después de hacer las sustituciones correspondientes, tenemos:
 ${}_{10}C_3 = 3628800 / 30240 = 120$ El cual es ya un número considerablemente menor al número de permutaciones.

LA PROBABILIDAD EN LA DISTRIBUCIÓN BINOMIAL

Continuando en la descripción de las bases que sustentan el hecho de que podamos hacer inferencias estadísticas, hablaremos de la distribución binomial que es una de las más importantes y más usadas, ésta se refiere a la forma en la que se acumulan las frecuencias de uno de dos eventos mutuamente excluyentes.

Para ilustrar lo arriba escrito, tenemos una enfermedad que históricamente se presenta con una probabilidad de curación de 0.6 (ó sea del 60 %). Sabemos que la probabilidad de curación $P(C)$ se obtiene de dividir el número de curaciones entre el total de los casos, esto es entre el número de casos curados (C) + el número de casos no curados (ΔC): $P(C) = C / (C + \Delta C)$. Suponiendo que nos encontramos con un grupo de 4 pacientes con la enfermedad y queremos conocer cómo se distribuyen todas las probabilidades entre ellos para su curación vemos que pueden existir 5 modalidades de grupo: cuando se curen 4, 3, 2, 1 y 0 respectivamente.

La $P(c) = 0.6$ luego la $P(\Delta c)$ es complementaria de $P(c)$ es decir $= 1 - P(c) = 1 - 0.6 = 0.4$, cada caso tomará la $P(c)$ o la $P(\Delta c)$ que le corresponda en cada **escenario** de los 5 posibles de acuerdo con sus combinaciones:



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

Para 0 curaciones: ${}_kC_r = {}_4C_0 = k! / (k-r)! r! = 4! / (4-0)! 0! = 24 / 24 = 1$ combinación

Para 1 curación: ${}_kC_r = {}_4C_1 = k! / (k-r)! r! = 4! / (4-1)! 1! = 24 / 6 = 4$ combinaciones

Para 2 curaciones: ${}_kC_r = {}_4C_2 = k! / (k-r)! r! = 4! / (4-2)! 2! = 24 / 4 = 6$ combinaciones

Para 3 curaciones: ${}_kC_r = {}_4C_3 = k! / (k-r)! r! = 4! / (4-3)! 3! = 24 / 6 = 4$ combinaciones

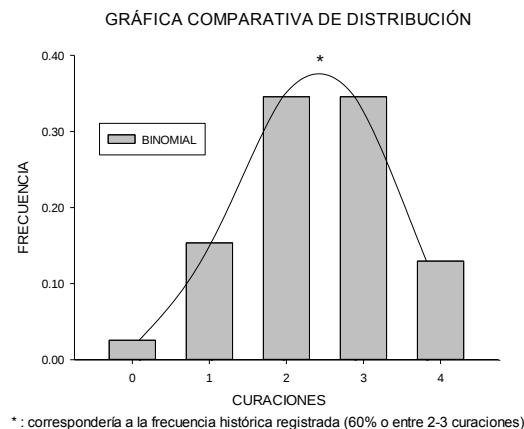
Para 4 curaciones: ${}_kC_r = {}_4C_4 = k! / (k-r)! r! = 4! / (4-4)! 4! = 24 / 24 = 1$ combinación

Cada combinación representa un escenario diferente en el que el número de curaciones posee una probabilidad que se multiplica entre sí, por la probabilidad de no curación y por el número de escenarios posibles, en la tabla siguiente se incluyen todas las probabilidades de curación en 4 pacientes con una enfermedad que regularmente se cura en el 60 % $P(c)$ de las veces.

Escenarios	Combinaciones	(c)	(Δc)	P
0 curaciones	1	$(0.6)^0$	$(0.4)^4$	0.0256
1 curación	4	$(0.6)^1$	$(0.4)^3$	0.1536
2 curaciones	6	$(0.6)^2$	$(0.4)^2$	0.3456
3 curaciones	4	$(0.6)^3$	$(0.4)^1$	0.3456
4 curaciones	1	$(0.6)^4$	$(0.4)^0$	0.1296
TOTAL				1.0

$P(c)$: probabilidad de curación, $P(\Delta c)$: probabilidad de no curación, P: probabilidad.

Cuando se representan en una gráfica las probabilidades del cuadro previo (columna P) se describe una distribución binomial de probabilidades y dicha distribución ajusta su comportamiento a la ecuación: $P(c) = {}_kC_r (c)^r (\Delta c)^{1-r}$



	INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA	Por José Fortino Chávez Carbajal
--	--	-------------------------------------

Para el caso de la distribución normal la ecuación correspondiente se refiere a:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

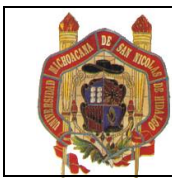
La cual, como es de notar resulta algo más elaborada que la previa, con ella se genera una curva de distribución con características muy peculiares y rígidas o normatizadas comúnmente conocida como campana de Gauss. La base de su construcción se relaciona con:

LOS VALORES TIPIFICADOS (Z), cuando la media = μ y cuando X es cualquier valor de la variable entonces $Z = \mu - X / D.E.$ Sii $\mu - X = D.E.$ tendremos que $Z = D.E./D.E.$ y $Z = 1$. Si los datos presentan una distribución normal (tendencia típica a agruparse alrededor de los valores medios) en un intervalo de $\pm 3 Z$ se encuentra aproximadamente el 99% de los datos, el 95% entre $\pm 2 Z$ y hasta 68% en $\pm Z$, siempre repartidos en forma simétrica respecto al promedio. Distribuciones como la de student, la de chi cuadrada o la binomial tienden a comportarse en forma similar.

La diferencia estadísticamente significativa

Como se dijo anteriormente, la diferencia o incluso la semejanza entre dos grupos de datos van a estar en gran parte relacionadas con la distribución que mantengan todos sus valores. Diferencias que se antojan abismales pueden sin embargo no ser estadísticamente significativas, es decir; pueden ser solo un efecto del azar o la casualidad. No obstante, la diferencia de tipo: “estadísticamente significativa”, no necesariamente quiere decir que sea una diferencia de antemano de gran magnitud. La mayor o menor diferencia siempre recibe su importancia a raíz de su significado teórico o pragmático dentro del fenómeno que se esté estudiando y nunca por la decisión estadística en sí misma.

Pero el término de estadísticamente significativo es determinante sobre todo cuando desde el inicio de un estudio, los practicantes definimos su significado, esto es; el hecho de predeterminar



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

los límites que consideramos importantes para establecer o juzgar a una diferencia como significativa, es lo que define en un momento dado la significancia tanto estadística como pragmática.

Tal vez no sea importante p.e. el que un tratamiento “A” tenga resultados positivos en un 65% de casos si un tratamiento “B” con menor costo y menores efectos adversos, logra resultados positivos hasta en 60% de los casos, aun cuando esta diferencia entre ambos tratamientos dejara sólo una en mil las posibilidades ($p = 0.001$) de que ésta se debiera nada más a la casualidad. En el mismo caso una diferencia de 20 o 30% entre ambos tratamientos seguramente sería suficiente para considerarla definitiva aun cuando la posibilidad de error fuera mucho mayor v.g. $p = 0.01$. Aquí ya introducimos el término de “error” para el cual existen distintas connotaciones como veremos en seguida.

Acerca del error estadístico

Más que un significado relacionado con la equivocación, el error estadístico es una variación o diferencia que se da en un proceso repetido de medición respecto a un valor de referencia, en esa repetición de mediciones las diferencias pueden presentarse azarosamente como menores o como mayores, de manera que al final, la suma de todas las diferencias se aproxima a cero. Este tipo de error se contrapone al error sistemático en el cual las diferencias son consistentemente mayores o bien invariablemente menores al valor de referencia.

El error estadístico introduce el argot del más-menos ($\pm$) cuando se intenta informar acerca de los límites de una variación respecto a un valor central cualquiera que sea. Hay sin embargo otro género de errores relacionados más que otra cosa con conceptos de límite y se utilizan sobre todo en las técnicas de validación de premisas o de la prueba de una hipótesis, estamos hablando de los niveles de exigencia de una prueba estadística así como de los valores de confianza de la misma. El nivel de exigencia es complementario a lo que se denomina error de tipo I o error α , p.e. a un nivel de exigencia de 99% le corresponde un error α de 1% o dicho en forma más adecuada le corresponde una fracción de uno: 0.01 del total de las probabilidades, estamos



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

diciendo que de cada 100 resultados positivos 99 serían verdaderos positivos y solo 1 sería falso positivo. Por otra parte el nivel de confianza es complementario con el error de tipo II o error de tipo β ; generalmente el nivel de confianza se refiere a valores que van del 80 al 90 %, rara vez asume valores mayores al 90 %, el error tipo II a su vez alcanzaría fracciones generalmente entre 0.1 y 0.2. Un nivel de confianza de 90% nos dice que de cada 100 resultados negativos 90 serían verdaderos negativos y por ende 10 resultados serían falsos negativos.

El significado de encontrar una de las cuatro formas de resultado (verdadero positivo y falso positivo, verdadero negativo y falso negativo), va a depender mucho de la naturaleza del estudio que se trate. Por supuesto que un resultado falso negativo en un estudio de Papanicolau pudiera ser mucho más delicado y hasta peligroso que un resultado falso positivo, sin embargo, el nivel de exigencia de una prueba estadística es el sinónimo de la apuesta del estudio, de su propuesta y de la prueba de su hipótesis de trabajo.

Sensibilidad y Especificidad.

Al poner una prueba a prueba (PP) se requiere compararla con un estándar de oro (EO) o un patrón reconocido como referencia, su calificación se expresará en términos tanto de sensibilidad como de especificidad, en el ensayo siguiente se describen sus características: en un grupo de 100 diagnósticos obtenidos por duplicado entre EO y PP se observó:

		E O		E O					
		+	-		+	-			
PP	+	50	10	80	VP	FP	Sensibil idad	α	50/60: 0.83 10/40: 0.25
	-	10	30	20	FN	VN	β	Especific idad	10/60: 0.17 30/40: 0.75
		60	40	100					
				0					

VP: verdadero positivo, sensibilidad, 0.83; VN: verdadero negativo, especificidad, 0.75; FN: falso negativo, error tipo β = 0.17; FP: falso positivo, error tipo α = 0.25



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

Coefficiente de correlación (r) producto momento de Pearson.

Sobre un plano de coordenadas, $y = x$ es una ecuación que describe una línea recta cuya forma general es $y = mx + b$, donde “y” y “x” son dos variables que en este caso correlacionan perfectamente y en forma directa (+), la línea que describen tiene una inclinación de $m = 1$ y pasa cruzando el eje de las “y” por el punto “b” que en este caso es $= 0$, el grado de correlación (r) entre las dos variables aquí es completa y positiva o $= 1$. La r puede ser + o - y su magnitud siempre está entre 0 y -1 o bien entre 0 y 1.

Cuando no contamos con una ecuación como la previa, su determinación nos debe informar el grado de relación positiva o negativa entre las dos variables y la forma en que se correlacionan, tenemos a 2 variables (“y” y “x”):

	Y	X	Y ²	X ²	YX
	1	5	1	25	5
	2	4	4	16	8
	4	6	16	36	24
	4	8	16	64	32
	5	8	25	64	40
	5	9	25	81	45
	7	9	49	81	63
	9	10	81	100	90
	9	11	81	121	99
	10	12	100	144	120
Σ	56	82	398	732	526
medias	5.6	8.2			

Existen operaciones matemáticas que generan la información necesaria a partir de estos datos:
 $LYY = N(\Sigma Y^2) - (\Sigma Y)^2$; sustituyendo: $10(398) - (3136) = 844$

$LXX = N(\Sigma X^2) - (\Sigma X)^2$; sustituyendo: $10(732) - (6724) = 596$

$LYX = N(\Sigma YX) - (\Sigma Y)(\Sigma X)$; sustituyendo: $10(526) - (56)(82) = 5260 - 4592 = 668$

Para obtener **r**, aplica que $r = LYX / (\sqrt{LYY})(\sqrt{LXX})$

Para obtener **m**, aplica que $m = LYX / LXX$, igual que la función tangente en trigonometría.

Para obtener **y**, aplica que $y = m(X - \bar{X}) + b'$ en este caso $\bar{X}$ es la media o promedio de la variable X y $b' = a$ la $\Sigma Y/N = 5.6$ (media o promedio de la variable Y).

$$r = 668 / (29.052)(24.413) = 668 / 709.25 = \mathbf{0.942}$$

$$m = 668 / 596 = 1.12$$



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

$y = 1.12(X - 8.2) + 5.6 = 1.12X - 9.191 + 5.6 = 1.12X - 3.591$; en este proceso -3.591 se transforma en el punto “b” o el sitio donde la línea cruza al eje de las Y, y la ecuación es:

$Y = 1.12X - 3.591$ lo cual corresponde a la fórmula general ya anotada: $Y = mX + b$.

Con estos datos la línea puede ser trazada en el plano de coordenadas un punto es el de “b” (-3.59), el otro puede ser cuando **$X = 10$** , entonces $11.2 - 3.59 = \mathbf{7.61 = Y}$

Finalmente antes de aplicar la ESTADÍSTICA INFERENCIAL, es necesario considerar que con ella se trabaja una prueba de hipótesis, y para probar una hipótesis se deben establecer premisas en cada proyecto implicado, las premisas más usadas se refieren a la comparación de 2 promedios o bien a la comparación de 2 proporciones. En ambos casos, se debe establecer cuál es el número mínimo de casos por comparar por medio de un procedimiento verificable y aceptado, en el primer caso se puede aplicar el siguiente procedimiento:

$$N = 2 \left(\frac{(Z_{\alpha} - Z_{\beta})\sigma}{\mu_1 - \mu_2} \right)^2$$

El valor de Z_{α} es decir para el error científico tipo I al nivel de una probabilidad de 0.05 es = 1.96, el valor de Z_{β} para un nivel de confianza de 0.8 es = -0.84, para un nivel de confianza de 0.9 es = -1.28; por otra parte σ es la D.E. mayor relacionada con μ_1 o con μ_2 tomadas de un estudio previo validado o confiable. Como ejemplo:

$N = 2 \left(\frac{(1.96 + 0.84)15}{30 - 15} \right)^2$ hablamos de una diferencia de promedios (30 y 15) con una D.E. =

15, resolviendo: $N = 2 \left(\frac{(2.8)15}{15} \right)^2 = 2(1764/225) = 15.7$, esto es: 16 por grupo.

Para la comparación de 2 proporciones una alternativa es la siguiente:

$$N = 2 \left(\frac{z_{\alpha} \sqrt{\left(\frac{P_1 + P_2}{2}\right)\left(1 - \left(\frac{P_1 + P_2}{2}\right)\right)} - z_{\beta} \sqrt{P_1(1 - P_1) + P_2(1 - P_2)}}{P_1 - P_2} \right)^2$$

En un caso hipotético con $z_{\alpha} = 1.96$, $z_{\beta} = -0.84$, $P_1 = 0.5$, $P_2 = 0.8$ el resultado del tamaño de la muestra es de 80.5, cada grupo deberá tener 41 sujetos para cubrir la muestra con casos completos. En ambos tipos de comparaciones suele agregarse entre un 10 y 15 % de sujetos extra para prevenir las bajas por exclusión o eliminación.



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

En la Estadística inferencial con **pruebas no paramétricas** o independientes de la distribución normal se describirán 2 procedimientos, uno de ellos es la prueba de **suma de rangos de Wilcoxon**, con la comparación de 2 grupos de datos de variables no continuas: un grupo A con 2, 3, 5, 7, 8, 10 y un grupo B con 7, 10, 11, 12, 13, 14; se ordenan todos los valores de menor a mayor independientemente del grupo de pertenencia y se les asigna un número de rango creciente, en los casos de valores iguales se repartirán los rangos involucrados en forma equitativa:

A	B	orden	rango	Rangos A	Rangos B
2	7	A 2	1	1	4.5
3	10	A 3	2	2	7.5
5	11	A 5	3	3	9
7	12	A 7	4, 4.5	4.5	10
8	13	B 7	5, 4.5	6	11
10	14	A 8	6, 6	7.5	12
		A 10	7, 7.5		
		B 10	8, 7.5		
		B 11	9		
		B 12	10		
		B 13	11		
		B 14	12		
		Σ		24	54

La Σ de rangos total **ΣRT** es igual a $N(N+1) / 2 = 12(13) / 2 = 156 / 2 = 78$

La diferencia entre la **ΣRA** y la **ΣRB** es de 30, 38.5% del total (78), existen tablas especializadas para esta prueba, sin embargo, en la mayoría de los casos diferencias $> 33\%$ son estadísticamente significativas, por lo que se puede asumir que la diferencia entre ambos grupos es significativa para cualquier propósito.

Otra de las pruebas no paramétricas muy usada es la **prueba de χ^2** (chi cuadrada), en la modalidad de comparación entre proporciones o frecuencias en el modelo del cuadrado latino de 2×2 . En este caso puede haber comparación entre más de 2 grupos, aquí trabajamos solo con dos grupos de pacientes con osteoarticular a los que denominaremos como “A” y como “B”,



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

al primero se le aplicó un tipo de analgésico (A) y al segundo otro tipo (B), en el cuadrado latino siguiente se describen los resultados en términos de alivio o no alivio:

Valores observados				Valores esperados		
	A	B		A	B	
Alivio	a 33	b 40	g 73	a' 36.5	b' 36.5	g' 73
No alivio	c 17	d 10	h 27	c' 13.5	d' 13.5	h' 27
	e 50	f 50	i 100	e' 50	f' 50	i' 100

La magnitud del estadímetro χ^2 determinará la presencia de diferencia estadísticamente significativa o no, para obtenerlo el primer paso es determinar los valores esperados (a' , b' , c' y d'): $a' = (e \times g)/i = (50 \times 73)/100 = 36.5$

$$b' = (f \times g)/i = (50 \times 73)/100 = 36.5$$

$$c' = (e \times h)/i = (50 \times 27)/100 = 13.5$$

$$d' = (f \times h)/i = (50 \times 27)/100 = 13.5$$

El segundo paso es aplicar en 4 ocasiones la fórmula: $\mathbf{Xa} = (O-E)^2/E$, $\mathbf{Xb} = (O-E)^2/E$, $\mathbf{Xc} = (O-E)^2/E$ y $\mathbf{Xd} = (O-E)^2/E$, en cada caso O es el valor observado y E el valor esperado, sustituyendo valores y resolviendo sucesivamente: $Xa = (33-36.5)^2/36.5 = 12.25/36.5 = \mathbf{0.34}$, $Xb = (40-36.5)^2/36.5 = 12.25/36.5 = \mathbf{0.34}$, $Xc = (17-13.5)^2/13.5 = 12.25/13.5 = \mathbf{0.91}$ y $Xd = (10-13.5)^2/13.5 = \mathbf{0.91}$, por lo tanto $0.34+0.34+0.91+0.91 = \mathbf{2.50 = \chi^2}$.

Al consultar la distribución de probabilidad de χ^2 con un grado de libertad para alcanzar una probabilidad <0.05 de error tipo I o α , vemos que se requeriría un valor de 3.84, por lo tanto la diferencia entre tratamientos no es estadísticamente significativa y puede deberse simplemente al azar.

La estadística inferencial aplicada con pruebas paramétricas, exige el cumplimiento de varias premisas, mismas que cuando no se cumplen llevan a la pérdida de confianza y de poder de las pruebas; por esa razón se les conoce como pruebas dependientes de la distribución normal, otro aspecto importante a considerar es la semejanza o no entre las varianzas.



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

Para comparaciones entre grupos con tamaño considerado como grandes muestras (≥ 30) se utiliza la **prueba de Z** (valor tipificado), estadímetro que se define como:

$$Z = \frac{(\mu_1 - \mu_2)}{\sigma_{\mu}}$$

Donde σ_{μ} representa a la desviación típica de la diferencia de promedios, se obtiene con la sustitución de los valores de los errores estándar de la media tanto del grupo 1 como del grupo 2: eem_1 y eem_2 así como de r en la ecuación:

$$\sigma_{\mu} = \sqrt{eem_1^2 + eem_2^2 - 2(r)(eem_1)(eem_2)}$$

Cuando el coeficiente de correlación (r) es = 0, la fórmula previa se reduce a solo:

$$\sigma_{\mu} = \sqrt{eem_1^2 + eem_2^2}$$

Luego entonces, esta última fórmula se utiliza cuando no existe correlación entre los dos grupos de valores. En cambio, cuando sí existe una correlación ($r > 0$), se usará la fórmula previa y el valor de r podrá ser ≤ 1 ; entre más grande resulte el valor de Z mayor será la probabilidad de que exista una diferencia estadísticamente significativa; con una $Z \geq 1.96$ se obtiene una diferencia estadísticamente significativa entre los promedios de ambos grupos a nivel una probabilidad < 0.05 de error tipo I o α .

Cuando se trata de comparar dos promedios de grupos considerados como muestras pequeñas (< 30), la prueba más usada es la de **t de Student**, fundamentalmente se usa en dos modalidades: cuando el número de casos de ambos grupos es igual ($N_1 = N_2$), o bien, cuando son diferentes ($N_1 \neq N_2$), en el primer caso el estadímetro de t de Student se define como:

$$t = \frac{(\mu_1 - \mu_2)}{\sigma_{\mu}}$$



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

en este caso el valor de $\sigma\mu$ se define como:

$$\sigma\mu = \sqrt{\frac{\Sigma x_1^2 + \Sigma x_2^2}{N(N-1)}}$$

Se debe contar con cada uno de los datos que conforman a ambos grupos para obtener las x^2 y la Σ de cada grupo, la N = número de pares de datos y no es la suma de los dos grupos; además se debe determinar el número de grados de libertad (GL) para con él y el valor de la t , buscar el valor crítico de la significancia o no sobre una tabla de distribución de probabilidades de t de Student al criterio mínimo de $\alpha < 0.05$. $GL = N-1$.

En el caso de comparaciones por la t de Student cuando $N_1 \neq N_2$ se emplea la misma fórmula primaria:

$$t = \frac{(\mu_1 - \mu_2)}{\sigma\mu}$$

para este otro caso el valor de $\sigma\mu$ se define como:

$$\sigma\mu = \sqrt{\frac{(\Sigma x_1^2 + \Sigma x_2^2) \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}{(N_1 + N_2 - 2)}}$$

Y los GL es igual a N_1+N_2-2 .

Generalmente los libros de estadística todos cuentan con apéndices donde muestran las tablas de distribución de probabilidades tanto para t de Student, para χ^2 , para la distribución normal (valores de Z) y otras (como la de Wilcoxon), en ellas se pueden comprobar los valores críticos de cada uno de los estadímetros aquí tratados.



INTRODUCCIÓN A LA ESTADÍSTICA BIOMÉTRICA o BIOESTADÍSTICA

Por
José Fortino Chávez Carbajal

Bibliografía recomendada:

1. Remington RD, Schork MA. **Estadística Biométrica y Sanitaria**. Ed. Del Castillo S.A. Madrid, Esp. 2ª. Ed.1977.
2. Daniel WW. **Bioestadística. Base para el análisis de las ciencias de la salud**. Ed. Limusa S.A. de C.V., México, D.F. 1999.
3. Dawson-Saunders B, Trapp RG, **Bioestadística Médica**. Ed. El Manual Moderno, S.A. de C.V., México, D.F. 1993.
4. Montesano-Delfín JR. **Manual del Protocolo de Investigación**. Ed. Auroch, S.A. de C.V. Méx. 1999.
5. Siegel S. **Estadística no Paramétrica**. Ed. Trillas, S.A. de C.V. México, D.F. 1990.
6. Downie NM, Heath RW. **Métodos estadísticos aplicados**. Ed. De Castillo S.A. Madrid, Esp. 1971.