

Introducción a la estadística básica para enfermería nefrológica

**Alberto Rodríguez Benot,
Rodolfo Crespo Montero**

Servicio de Nefrología.
Hospital Reina Sofía, Córdoba.

RESUMEN

La estadística es uno de los pilares del método científico una vez alcanzada la fase de análisis de los datos. La *estadística descriptiva* permite organizar y presentar los datos en tablas o gráficos, así como resumirlos con medidas de centralización y de dispersión, simplificando la interpretación de los mismos. La *estadística inferencial* estudia las variables o características que presentan los individuos, generalizando los datos obtenidos a partir de una muestra a un número mayor de individuos (población). La estadística inferencial se basa en la teoría de las probabilidades, ya que la generalización de los datos de la muestra a una población está siempre sujeta a un pequeño margen de error. La muestra debe obtenerse al azar y ser representativa de las características de la población. La mayoría de las variables biológicas (temperatura, glucemia...) siguen una distribución de frecuencias en forma de campana invertida, denominada distribución normal o de Gauss. En otras ocasiones siguen una distribución diferente, como la binomial o la de Poisson. La distribución de frecuencias de una variable en una muestra pasa a ser una distribución de probabilidades cuando se generaliza a una población. Esta es la base para la comparación de grupos de datos (medias, proporciones) utilizando los tests de contraste de hipótesis. Estos tests comparan dos o más grupos de datos entre sí indicando si existen o no diferencias entre ellos, con una pequeña probabilidad de error p . Existen textos paramétricos cuando se comparan variables que siguen una distribución normal, y tests no paramétricos para comparar variables cuantitativas discretas o cualitativas. También es posible conocer el grado de relación o asociación existente entre dos o más variables mediante los tests de correlación. En sentido inverso puede

predecirse el valor de una variable dependiente a partir del valor de otra independiente mediante la regresión. Por último, en medicina es interesante el análisis estadístico de la supervivencia. El método de Kaplan-Meier es uno de los más utilizados para obtener las curvas de supervivencia, comparándose posteriormente si existen diferencias significativas entre ellas mediante el test de Log-Rank. El análisis estadístico se ha simplificado enormemente gracias al desarrollo de potentes paquetes informáticos que se ocupan de la mayoría de las funciones de la estadística descriptiva, inferencial y representación gráfica de los datos de forma automática, facilitando así al investigador la obtención de conclusiones y la presentación de los datos en comunicaciones y publicaciones científicas.

LA ESTADÍSTICA EN CIENCIAS DE LA SALUD

Aunque aparentemente la bioestadística parece una ciencia fundamentalmente teórica, es utilizada en la práctica médica a diario. Cuando hablamos de la dosis media de eritropoyetina administrada en diálisis o el tiempo medio de duración de una sesión de hemodiálisis estamos utilizando la estadística. O cuando decidimos tratar a un paciente con unas cifras de colesterol o de presión arterial elevadas, previamente se ha demostrado estadísticamente que existe un riesgo elevado cuando esas cifras están por encima de un determinado valor. O por ejemplo, cuando queremos comparar si existe diferencias entre un fármaco inmunosupresor u otro a la hora de prevenir el rechazo de un trasplante renal.

El objetivo de este artículo es iniciar y familiarizar a la enfermería con el método científico, y más concretamente con las nociones básicas del análisis estadístico necesario para cualquier estudio de investigación.

La estadística es una ciencia de reciente desarrollo, basada en las matemáticas y en la actualidad resulta relativamente fácil su utilización práctica gracias al apoyo de la informática. El cálculo de complicadas expresiones se simplifica y acelera haciéndolo casi "un juego de niños", algo impensable hace unos pocos años. Lo único que el ordenador no sabe es *qué hay que hacer*, él sólo se encarga del cálculo. Afortunadamente (por ahora) es el investigador, y no la máquina, quien toma la decisión.

1. TIPOS DE ESTADÍSTICA

Definida por Schwartz en 1981 como un método de razonamiento que permite interpretar un conjunto de datos cuyo carácter esencial es la *variabilidad*, la estadística permite estudiar el comportamiento de ciertas características en una población, y es un instrumento fundamental para la medicina actual. La *estadística descriptiva* comprende la presentación, organización y resumen de los datos de una manera científica. Incluye diversos métodos de organizar y representar gráficamente los datos, para dar una idea de lo que nos muestran. Las tablas, los diagramas de barras o los gráficos sectoriales o "tartas" son algunos de los elementos de estadística descriptiva. También incluye varios parámetros numéricos (como la media aritmética) que resumen los datos con muy pocos números clave. Por otra parte, la *estadística inferencial* o inductiva permite generalizar los datos obtenidos a partir de una muestra a un número mayor de individuos (población). La estadística inferencial se basa en la teoría de las probabilidades y trabaja con los datos que le proporciona la estadística descriptiva.

2. POBLACIÓN, MUESTRA, INDIVIDUO, VARIABLES TIPOS DE VARIABLES

Supongamos un estudio hipotético en el que pretendemos comparar si el calibre de la aguja de punción influye en la supervivencia de la fístula arteriovenosa de los pacientes en hemodiálisis. Nuestro estudio lo llevaremos a cabo en unos cuantos pacientes, en los que iremos recogiendo una serie de datos: calibre de la aguja, tipo de fístula, antigüedad, número de punciones fallidas, existencia de diabetes... Si nuestro estudio demuestra que un determinado calibre acorta el tiempo de vida de la fístula en los pacientes estudiados, su importancia realmente radica en que esos resultados son aplicables no sólo a nuestros pacientes, sino a *todos los pacientes en hemodiálisis* con las mismas características que los que hemos estudiado. Nuestro estudio se ha realizado en una *muestra* de *individuos* que forman parte de una *población*, los pacientes en hemodiálisis.

De forma genérica la *población* se define como un conjunto homogéneo de individuos que generalmente es inaccesible para su estudio al ser de un tamaño inabordable. Es también el hipotético (y habitualmente infinito) conjunto de personas a las que se desea aplicar una generalización. La *muestra* es un conjunto menor de individuos, accesible y limitado, sobre el que se realiza el estudio con idea de obtener conclusiones generalizables a la población. Debe ser un conjunto reducido, pero representativo de la población de donde procede. Cada uno de los componentes de la población y de la muestra se denomina *individuo*. Al número de individuos que forman la muestra se llama *tamaño*, y se representa con la letra *n*.

Las *variables* o *caracteres* son las propiedades o características que se estudian en cada individuo de la muestra, como la edad, el peso, la presión arterial, o el tiempo en diálisis, el tipo de aguja empleado, o la intensidad del dolor a la punción. Una variable no es más que lo que está siendo observado o medido. Hay variables de dos tipos:

Variables dependientes: son el objeto de interés, que varía en respuesta a alguna intervención.

Variables independientes: es la intervención, o lo que está siendo aplicado. En nuestro ejemplo, la variable dependiente es el tiempo de supervivencia de la fístula, que depende del calibre de la aguja (variable independiente). Las variables pueden contener datos muy diversos, que están agregados en *categorías*. Por ejemplo, la variable "sexo" tiene dos categorías: masculino y femenino. A su vez, según el tipo de datos que contienen las variables, se pueden clasificar en:

Variables cualitativas, que tienen valores no numéricos (sexo, religión, color de los ojos). Pueden ser: *nominales*, con categorías con nombre: religión, estado civil, especialidades de un hospital... Cuando se les puede ordenar en sentido creciente o decreciente se denominan *ordinales*. Por ejemplo, el dolor medido como leve, moderado o grave. Si las variables cualitativas pueden tomar sólo dos posturas o valores opuestos (vivo/muerto, varón/mujer, sano/enfermo), se llaman *dicotómicas* o *binarias* y son excluyentes entre sí.

Variables cuantitativas, que son aquellas que toman valores numéricos (glucemia, número de hijos, peso, coeficiente intelectual). Pueden ser: *discretas*, cuyos valores son números finitos, generalmente números enteros (pacientes ingresados en un hospital, número de partos, número de dientes con caries) o *continuas*, que pueden tomar cualquier valor de un intervalo determinado. Por ejemplo, la altura, el peso o nivel de colesterol: se pueden fraccionar cuanto se quiera. La única limitación viene dada por el aparato de medida.

3. PROBABILIDAD BÁSICA

La probabilidad se define como el límite de la frecuencia relativa cuando el número de repeticiones de un experimento tiende al infinito. Una definición menos científica, pero más inteligible y práctica es el número de casos favorables dividido por el número de casos posibles. La teoría de la probabilidad es la base de la estadística inferencial.

Ejemplo: En un centro de diálisis con 100 pacientes, 20 han contraído una gastroenteritis. ¿Cuál es la probabilidad de contraer esa enfermedad) $P(\text{enfermedad}) = 20/100 = 0.2$ (o también 20%). La probabilidad de que no ocurra (también llamado suceso complementario) se calcula restando de 1 probabilidad de que ocurra. $P(\text{no enfermedad}) = 1 - P(\text{enfermedad}) = 1 - 0.2 = 0.8$ (80%).

La probabilidad de un suceso cualquiera está comprendida entre 0 y 1. La probabilidad del suceso imposible es 0, y la del suceso seguro, 1.

PROBABILIDAD CONDICIONADA: Es la probabilidad de que ocurra un suceso (A) habiendo ocurrido otro (B):

$$P(A/B) = \frac{\text{n}^\circ \text{ de veces que ocurre A y B}}{\text{n}^\circ \text{ de veces que ocurre B}} = \frac{P(AyB)}{P(B)}$$

Ejemplo: en el caso anterior, de 100 pacientes, 80 han tomado un bocadillo durante la sesión de diálisis; de ellos, contraen gastroenteritis 20. ¿Cuál es la probabilidad de contraer gastroenteritis (A) habiendo tomado bocadillo (B)?

De los 80 que tomaron bocadillo, sólo 20 presentan gastroenteritis, o lo que es lo mismo, la probabilidad de contraer gastroenteritis (A) y comer bocadillo (B) es 20. La probabilidad $P(AyB)$ es 20% o 0.2 (viene dada en el enunciado); la probabilidad de comer bocadillo $P(B)$ es 80% o 0.8:

$$P(A/B) = \frac{P(AyB)}{P(B)} = \frac{20}{80} = 0.25 \text{ ó } 25\%$$

Cuando dos sucesos son independientes (no pueden suceder juntos), no existe condicionamiento, y:

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

LEY ADITIVA: Representa la probabilidad de que ocurra un suceso o bien que ocurra otro. Si los sucesos son excluyentes (no pueden presentarse simultáneamente, como sacar cara o cruz al lanzar una moneda): $P(A \text{ o } B) = P(A) + P(B)$. Si los sucesos son no excluyentes (pueden darse simultáneamente), $P(A \text{ o } B) = P(A) + P(B) - P(AyB)$.

LEY MULTIPLICATIVA: Representa la probabilidad de que ocurra un suceso y de que ocurra otro a la vez. Para sucesos independientes, $P(AyB) = P(A) \times P(B)$. Para sucesos dependientes (la ocurrencia de uno (B) está condicio-

nado a la aparición de otro (A)):

$P(AyB) = P(A) \times P(B/A)^{(1)}$. En caso de ser A el suceso dependiente o condicionado a B, la expresión es: $P(AyB) = P(B) \times P(A/B)^{(2)}$. Ejemplo: la enfermedad X causa la muerte al 20% de los afectados. Si tenemos 2 pacientes con esa enfermedad, ¿cuál es la probabilidad de que mueran los 2 pacientes?

Son sucesos independientes, por lo que: $P(AyB) = 0.2 \times 0.2 = 0.04 = 4\%$.

Teorema de Bayes.

Es una fórmula derivada de las expresiones anteriores, por la que, siendo A y B dos sucesos dependientes o asociados entre sí, según las expresiones ⁽¹⁾ y ⁽²⁾,

$$P(AyB) = P(A) \times P(B/A) = P(B) \times P(A/B)$$

$$P(A/B) = \frac{P(A) \times P(B/A)}{P(B)} \quad \text{y} \quad P(B/A) = \frac{P(B) \times P(A/B)}{P(A)}$$

El teorema de Bayes hace referencia a aquellas situaciones donde una vez producido un suceso B, se trata de calcular si el mismo es debido a una causa A. En medicina se utiliza con frecuencia la probabilidad condicionada; un ejemplo muy común es la evaluación de un método diagnóstico, como la probabilidad de que un test sea positivo o negativo teniendo realmente una enfermedad. Un ejemplo sería cuál es la probabilidad de que un paciente tenga un cáncer de hígado cuando tiene una alfa-fetoproteína elevada en sangre. Gracias al Teorema de Bayes podemos calcular la especificidad y la sensibilidad, o el valor predictivo positivo o el valor predictivo negativo de un test diagnóstico.

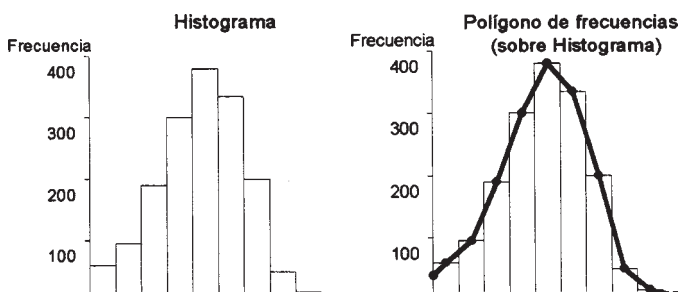
4. ESTADÍSTICA DESCRIPTIVA, PRESENTACIÓN DE LOS DATOS

Una vez obtenidos los datos es preciso mostrarlos de una forma ordenada y comprensible. La forma más sencilla es colocarlos en una *Tabla*, donde se muestran las variables, las categorías de cada variable y el número de eventos de cada categoría. En ciertas ocasiones, especialmente cuando trabajamos con un gran número de datos, las tablas no son prácticas y se hace necesario una mejor visión de los datos con una mirada rápida. Esto se consigue con los gráficos. La selección del gráfico dependerá del tipo de datos empleados. Comenzaremos con los gráficos para datos cuantitativos:

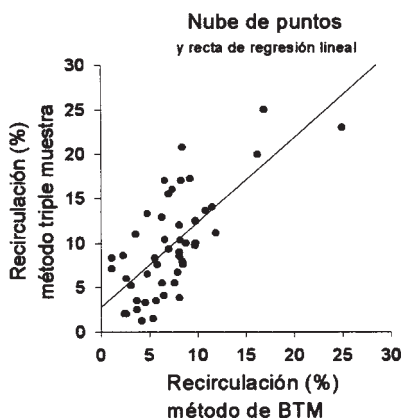
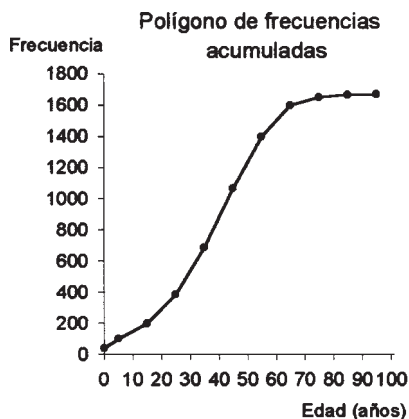
Histograma: Se utiliza para variables cuantitativas continuas. En el eje x se muestran los datos de la variable, que por ser continuos requieren ser agrupados previamente en intervalos, y en el eje y se representa la frecuencia con la que aparece cada dato. La anchura del intervalo

y la altura que alcanza determinan el área de cada intervalo, que es proporcional a la frecuencia de cada intervalo. Da una idea muy aproximada de la forma de la distribución que sigue la variable.

Polígono de frecuencias: Utiliza la misma escala que el histograma, y se construye uniendo los puntos medios de la zona más alta de los rectángulos. También aquí lo más importante es el área existente debajo del polígono, que es igual al área del histograma correspondiente. En el **polígono de frecuencias acumuladas**, la línea representa la frecuencia de cada intervalo sumada a la de los intervalos anteriores. Es un método práctico para determinar percentiles (concepto que veremos más adelante). El ejemplo más típico son las tablas de crecimiento en altura.



Nube de puntos: Es un gráfico donde se muestran dos variables cuantitativas, una en el eje x y otro en el y, mostrando los valores mediante puntos o símbolos.



Para los datos cualitativos:

Diagrama de barras: Se utiliza para variables cualitativas y cuantitativas discretas, y se construyen de forma similar al histograma, pero las barras están separadas entre sí (indicando que la variable no ocupa todo el eje de abscisas, precisamente por ser discreta o cualitativa). El **diagrama de barras compuesto** representa dos o más variables en el mismo gráfico.

Gráfico sectorial o pastel: Es otro método empleado con frecuencia para datos cualitativos, en el que un círculo representa el total, y un segmento o porción del pastel es la proporción o porcentaje de cada categoría de la variable. Es el gráfico adecuado para variables con categorías mutuamente excluyentes (no se puede estar soltero y casado a la vez).

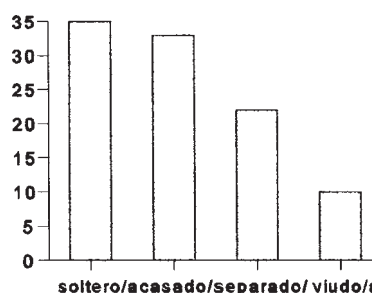


Diagrama de barras

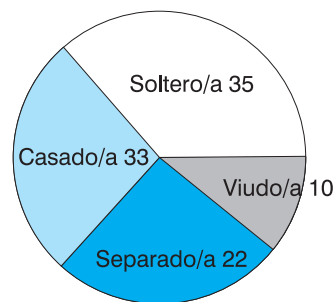


Gráfico sectorial o pastel

5. SÍNTESIS DE LOS DATOS

Una vez organizados los datos en tablas y representados gráficamente, es útil sintetizarlos o resumirlos en medidas o números que permitan trabajar cómodamente y que contengan el máximo de información. Existen dos tipos de medidas que describen las características de la distribución de frecuencias de los valores de una variable: las medidas de *centralización* y de *dispersión*.

Medidas de centralización: definen los valores de la variable en torno a los cuales tienden a concentrarse las observaciones. Son: media, mediana, moda y los cuartiles, deciles, y percentiles.

Media: La *media aritmética* es la medida de centralización más conocida y utilizada. Se calcula sumando todos los valores observados y dividiendo por el número de observaciones de la muestra. Se representa como \bar{x} .

Su principal ventaja es su fácil manejo matemático y estadístico. Sin embargo, tiene la desventaja de ser muy sensible a los valores extremos en una muestra que no tenga una distribución normal (veremos más tarde qué significa esto). Si por ejemplo analizamos los días de estancia hospitalaria de los 7 últimos trasplantados renales en nuestro Servicio, y tenemos: 3, 3, 4, 7, 9, 11 y 12 días. Puesto que son 7 datos, $\bar{x} = (3+3+4+7+9+11+12)/7 = 49/7 = 7$; la estancia media de los pacientes es de 7 días. Pero si en lugar de 12 días un paciente permanece ingresado 89, la nueva media sería 18 días, muy alejada de la previa de 7 días. Esto se debe a que un valor extremo (89), muy distante del resto, influye negativamente en la media. En este caso, la *mediana* es una medida mejor de centralización.

La *media geométrica* es un parámetro de centralización que se utiliza para datos exponenciales o del tipo de crecimiento de poblaciones. Se calcula multiplicando los datos entre sí y aplicando después la raíz de orden n . Se utiliza con mucha menor frecuencia que la media aritmética.

Mediana: La *mediana* es la observación equidistante de los extremos, o lo que es lo mismo, el valor que, una vez ordenados los datos, deja igual número de observaciones por encima y por debajo. En el ejemplo anterior, la mediana es el valor 7. Como vemos, la *mediana* es mucho menos sensible a los valores extremos que la *media*, y es la medida de centralización a emplear en las variables cualitativas ordinales, en las que es imposible calcular la media aritmética. Por supuesto, se puede utilizar también con datos interválicos y proporcionales. Gráficamente, en el polígono de frecuencias acumuladas, la mediana es el valor correspondiente al 50% de las observaciones en el eje de abscisas (eje x).

Moda: La *moda* es el valor que se observa con más frecuencia, el más repetido. En el ejemplo anterior la moda es 3 por ser el valor más repetido. Si no se repite ningún valor, la muestra no tiene moda, es *amodal*. Si se repiten varios valores diferentes, puede ser *bimodal*, *trimodal*, o *multimodal*. Gráficamente, la *moda* equivale al valor que alcanza la frecuencia máxima o pico en el polígono de frecuencias.

Cuartiles, Deciles, Percentiles: Son medidas de localización, pero no central, sino que localizan otros puntos de una distribución. Los *cuartiles* dividen los datos en cuatro partes iguales, los *deciles* en diez partes iguales y los *percentiles*, en cien partes iguales. Por definición, el cuartil 2 coincide con el decil 5 y con el percentil 50, y todos ellos con la mediana.

MEDIDAS DE CENTRALIZACIÓN

Una vez definidos los valores de la variable en torno a los cuales tienden a concentrarse las observaciones, el siguiente planteamiento es describir cómo de agrupados o dispersos se encuentran los datos de la muestra en torno a esos valores. Esta información nos la ofrecen las medidas de dispersión: Recorrido o rango, desviación media, varianza, desviación estándar y coeficiente de variación.

Recorrido o rango: Es la diferencia entre los valores máximo y mínimo de la variable. En el ejemplo 3, 3, 4, 7, 9, 11, 12, el rango es $12-3 = 9$. Su principal ventaja es que se calcula con gran facilidad. Pero dado que no tiene en cuenta los valores intermedios, su utilidad es muy limitada. Es útil como medida de dispersión en las variables cualitativas ordinales, o para indicar si nuestros datos tienen algunos valores extraordinarios.

Recorrido intercuartil: Como consecuencia de los problemas que presenta el recorrido, en particular su inestabilidad al considerar muestras diferentes o bien cuando se añaden nuevos individuos, a veces se usa otro índice de dispersión con datos ordinales, el *recorrido intercuartil*, también llamado *media de dispersión*. Se calcula dividiendo en primer lugar los datos (previamente ordenados) en cuatro partes iguales, obteniendo así los cuartiles Q1, Q2, y Q3; la diferencia entre el cuartil Q3 y el Q1 es el *recorrido intercuartil*, y abarca el 50% de los datos. Recordemos que Q2 = mediana. Como el recorrido intercuartil se refiere sólo al 50% central de los datos, se afecta en mucha menor medida por los valores extremos que el recorrido propiamente dicho, lo que la convierte en una medida mucho más útil.

Desviación media, Varianza (S^2) y desviación estándar (S o DE): Son las medidas de dispersión más frecuentemente utilizadas en biomedicina. Se basan en cálculos de la diferencia entre cada valor y la media aritmética ($x-\bar{x}$). Al calcular esta diferencia, debe prescindirse del signo negativo o positivo de cada resultado, por lo que la medida de dispersión se muestra como " \pm " *desviación*. La principal diferencia entre las tres medidas es cómo se prescinde del signo negativo: en la *desviación media*, se toman los valores absolutos $|x-\bar{x}|$; en la *varianza* (S^2 para muestras y σ^2 para poblaciones) se eleva al cuadrado la diferencia: $(x-\bar{x})^2$.

Como en la varianza los datos están al cuadrado, para regresar a las unidades originales basta tomar la raíz cuadrada de la varianza. Obtenemos así la *desviación típica o estándar* (DE), S para muestras y σ para poblaciones.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$$

Cuanto más dispersos estén los valores de la media, mayor será la desviación estándar. Es la medida de dispersión más importante y utilizada.

De esta forma hemos visto cuáles son los índices básicos que describen, de forma resumida, los valores de una muestra (también es aplicable a una población, como veremos):

- El *tamaño* de la muestra, o *n* (el número de observaciones).
- La *media aritmética*: valor alrededor del cual se agrupan los datos.
- La *desviación estándar*, valor que indica la dispersión de los datos alrededor de la media.

Coefficiente de variación: Se emplea para comparar la variabilidad relativa de diferentes distribuciones, partiendo del problema de que las desviaciones estándar no son comparables al estar referidas a distintas medias. Este sería el caso de querer comparar la variabilidad de la presión arterial de un grupo de pacientes con su edad. Se usa con frecuencia para comparar métodos de medida, y es un valor adimensional. Se calcula dividiendo la DE por la media, multiplicando después por 100:

$$CV = \frac{DE \text{ o } S}{\bar{x}} \cdot 100$$

Medidas para variables cualitativas

La mayoría de las medidas anteriores no son aplicables a las variables cualitativas, ya que sus valores no son numéricos, sino que representan recuentos o frecuencias de ocurrencia de un suceso. Existen tres formas básicas de presentar estos datos:

1. *Proporción o frecuencia relativa*, que es el número de casos que se presenta una característica (a) dividido por el número total de observaciones (a+b): $a/(a+b)$. Su valor oscila entre 0 y 1. Si multiplicamos una proporción por 100, obtenemos un *porcentaje*.

2. *Razón o cociente*, que es el número de casos que presentan una característica (a) dividido por el número de casos que no la presentan (b): (a/b) .

3. *Tasa*, que es similar a la proporción, pero multiplicada por una cifra (por ejemplo 1.000, 10.000, 100.000) y se calcula sobre un determinado período de tiempo.

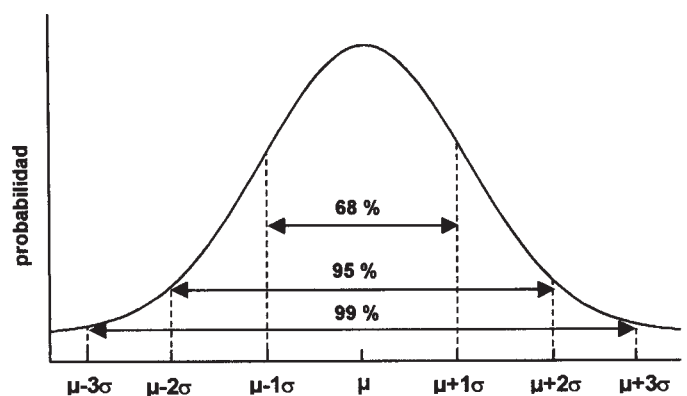
6. DISTRIBUCIONES DE PROBABILIDAD. LA DISTRIBUCIÓN NORMAL

Las distribuciones de frecuencia reflejan cómo se reparten los individuos de una muestra según los valores de la variable. Cuando se trata de poblaciones, el comportamiento teórico de una variable puede conocerse mediante las *distribuciones de probabilidad*, de las que la más conocida es la *distribución normal o de Gauss*. Otras distribuciones de interés en bioestadística son la *binomial* y la *distribución de Poisson*.

Distribución normal o de Gauss: Es la distribución de probabilidad teórica más importante. La mayoría de las variables cuantitativas continuas biológicas siguen una distribución normal, que se define por presentar las siguientes propiedades:

- 1) Está definida por una *función de probabilidad continua*.
- 2) La media, mediana y moda coinciden, y es simétrica respecto a este punto. Es unimodal.
- 3) La función queda suficientemente definida por la media \bar{x} y la desviación estándar S (μ y σ para poblaciones).
- 4) El área comprendida bajo la curva de la distribución es igual a la unidad.
- 5) Es asintótica respecto al eje de abscisas (nunca llega a cortarlo), siendo posible cualquier valor de x entre $-\infty$ y $+\infty$.
- 6) La función tiene forma de campana invertida.

La siguiente figura representa una distribución normal. El intervalo $[\bar{x} \pm S]$ o $[\mu \pm \sigma]$ agrupa aproximadamente al 68%, el intervalo $[\bar{x} \pm 2S]$ agrupa aproximadamente al 95%, y el intervalo $[\bar{x} \pm 3S]$ agrupa aproximadamente al 99% de los valores centrales de la distribución.



Distribución normal o de Gauss

Distribución binomial: Cuando la variable sólo tiene dos valores posibles, se dice que sigue una *distribución binomial*. Para el cálculo de las probabilidades se utiliza el método del *desarrollo binomial*. La distribución binomial posee también una media, varianza y desviación estándar,

que se calculan con expresiones diferentes a las de la distribución normal.

Una característica importante de la distribución binomial es que al ir aumentando el número de sucesos (n), la distribución binomial se va pareciendo cada vez más a la distribución normal. A partir de $n=30$, la distribución binomial se comporta estadísticamente como una normal, por lo que podemos aplicar los tests estadísticos apropiados para esta distribución.

Distribución de Poisson: Es una variante de la distribución binomial en la cual la probabilidad de tomar un valor es muy pequeña y la de tomar el otro valor es muy alta. La distribución discreta de Poisson se utiliza para determinar la probabilidad de que un suceso raro ocurra un determinado número de veces, cuando el número de ensayos es grande y la probabilidad de que aparezca el suceso es pequeña. Esta situación ocurre por ejemplo ante la posibilidad de un parto de sextillizos, o de tener un hijo albino. Otro ejemplo podría ser la posibilidad de presentar una reacción alérgica a una membrana de hemodiálisis.

7. ESTADÍSTICA INFERENCIAL. TÉCNICAS DE MUESTREO. ERRORES Y SESGOS. TAMAÑO MUESTRAL

Como recordamos en el primer capítulo, la estadística inferencial extrae conclusiones para una población a partir de los resultados obtenidos en nuestras:

Técnicas de muestreo.

Para estudiar una característica de una población debemos, en primer lugar, escoger una muestra representativa de esa población, a la que someteremos al estudio deseado. Para que una muestra sea válida requiere dos condiciones: 1), debe ser *aleatoria*: todos los individuos de la población tienen la misma probabilidad de aparecer en la muestra. 2), la muestra debe ser homogénea con la población de la que se ha extraído, manteniendo las mismas proporciones que la población en todas aquellas características que pueden tener influencia en el experimento que se va a realizar (debe guardar la misma proporción de hombres y mujeres, de edad, de fumadores...).

Existen diferentes métodos para captar a los sujetos que constituirán la muestra. Las técnicas de muestreo pueden ser probabilísticas (participa el azar) o no probabilísticas. Entre las primeras, están:

Muestreo aleatorio simple: Cada elemento de la población tiene la misma probabilidad de ser escogido. Para ello, se utilizan números aleatorios extraídos de unas tablas o generadas por ordenadores. Tiene el inconveniente de requerir previamente el listado completo de la población. En muestras pequeñas puede no representar adecuadamente a la población.

Muestreo estratificado: La población se divide en

subgrupos o estratos, y después se obtiene una muestra aleatoria de cada uno de ellos. Si la estratificación se hace respecto a un carácter (hombre/mujer), se denomina muestreo estratificado *simple*, y si se estratifica respecto a dos o más caracteres (sexo, edad, estado civil, posición económica...) se denomina muestreo estratificado *compuerto*.

Muestreo sistemático: Se aplica una regla simple fija para el muestreo, de modo que se divide el total de la población de estudio entre el tamaño de la muestra, hallando así una constante de muestreo, k . El primer individuo se elige al azar, y después, se elige sistemáticamente uno de cada k individuos siguiendo un orden determinado. Como ventajas, es simple de aplicar, no precisa un listado completo de la población, y en poblaciones ordenadas asegura una cobertura de unidades de todos los tipos. La desventaja es que si existe alguna relación entre el fenómeno a estudiar y la constante de muestreo, se pueden cometer sesgos. Ejemplo: seleccionar los individuos por las historias clínicas que acaban en 6.

Muestreo en etapas múltiples: Consiste en seleccionar unidades de muestreo de una población (unidades primarias) y obtener en una segunda etapa una muestra de cada una de ellas (unidades secundarias). Pueden utilizarse tantas etapas como sean necesarias, y en cada una de ellas, una técnica de muestreo diferente. Este método es muy eficaz para poblaciones grandes y dispersas, y es el utilizado generalmente en los estudios multicéntricos.

Muestreo no probabilístico: En este caso se utilizan métodos en los que no participa el azar. El ejemplo típico son los voluntarios que acuden a realizarse algún estudio (acuden por su propia voluntad, pero no por azar, sino estimulados económicamente o ante algún otro beneficio). También es no probabilístico la inclusión de pacientes conforme acuden a una consulta. El inconveniente de este método es que, al no tener todos los individuos la misma probabilidad de ser incluidos en el estudio, no puede asegurarse la representatividad de la muestra respecto a la población de referencia.

Tamaño de la muestra

La muestra debe tener un tamaño que sea apropiado para los objetivos buscados en el estudio y los condicionantes que se está dispuesto a asumir. Un número insuficiente de participantes impedirá encontrar diferencias buscadas, concluyendo erróneamente que no existen, y un número excesivo de sujetos alargará y encarecerá innecesariamente el estudio. Existen diferentes fórmulas para *calcular el tamaño muestral*, que dependerá básicamente de qué estamos buscando con el estudio:

Si tratamos de *estimar parámetros* para una población a partir de una muestra, su tamaño dependerá de la variabilidad del parámetro a estudiar, de la precisión con que

se desee obtener la estimación (es decir, la amplitud deseada del intervalo de confianza, de modo que a mayor precisión deberán estudiarse más sujetos), y del nivel de confianza deseado. La variabilidad del parámetro debe ser previamente conocida (o aproximarse a partir de datos preexistentes o estudios pilotos). La precisión y el nivel de confianza son establecidos por el investigador.

Por ejemplo, si queremos estimar la prevalencia de infección por VHC en hemodiálisis, o la presión arterial de los trasplantados renales.

Cuando se trata de comparar grupos mediante pruebas de *contraste de hipótesis*, el tamaño muestral proporciona una cifra aproximada del número de sujetos necesarios para detectar una diferencia determinada si es que ésta existe, con la aceptación de unos márgenes de error previamente fijados. Para su cálculo se precisa la definición previa de los riesgos a asumir: los errores tipo I o alfa y II o beta, y la magnitud de la diferencia que se desea detectar.

Este sería el caso de comparar la eficacia en depurar urea de dos membranas de diálisis diferentes, o de ver si hay diferencias en el hematocrito cuando se administra eritropoyetina subcutánea o intravenosa. Las fórmulas para calcular el tamaño muestral exceden el contenido de este capítulo. La mayoría de paquetes estadísticos de ordenador la calculan a partir de las condiciones impuestas, dependiendo del tipo de estudio a realizar. Si el estudio es complejo o requiere un muestreo en etapas múltiples, es aconsejable consultar a un estadístico para que nos calcule el tamaño de la muestra.

Errores y sesgos.

Al seleccionar una muestra a partir de una población y estudiarla, podemos cometer dos tipos de errores: los errores aleatorios y los errores sistemáticos o sesgos.

ERROR ALEATORIO: Si comparamos el resultado obtenido de una muestra y de una población habrá una pequeña diferencia; esta diferencia dependerá de cómo escogimos la muestra, de su tamaño y del azar; realmente siempre existirá una diferencia real entre población y muestra. *El error aleatorio* es la diferencia entre el resultado obtenido en la muestra y la realidad de la población. En él siempre interviene el azar y la selección viciada de la muestra realizada por el investigador.

El error aleatorio se debe al azar, sucede en todos los grupos, no afecta a la validez interna del resultado, pero puede disminuir la probabilidad de encontrar relación entre las variables estudiadas. Aunque el error aleatorio no puede ser eliminado, sí puede disminuirse aumentando el tamaño de la muestra y la precisión de las mediciones.

ERRORES SISTEMÁTICOS O SESGOS: Son errores que se cometen en el procedimiento del estudio cuando, por

ejemplo, la medición de la variable en estudio es consistentemente desigual entre los distintos grupos. Afectan a la validez interna del estudio y aunque se aumente el tamaño de la muestra, se perpetúa el sesgo introducido, y es prácticamente imposible enmendarlo en la fase de análisis. Pueden ser de varios tipos:

Sesgos de selección: Las muestras no son adecuadamente representativas de la población de estudio, por no reflejar la misma distribución (edad, sexo, efecto de voluntarios...). Se han definido algunos tipos concretos como los siguientes, como el Sesgo de Berkson: las muestras seleccionadas en un medio hospitalario pueden diferir sistemáticamente de las poblaciones generales.

Sesgos de seguimiento: Se cometen cuando no se observan por igual a ambos grupos, o si se pierden más individuos de un grupo que del otro (no al azar) a lo largo del estudio. Por ejemplo, el abandono del tratamiento por parte de los pacientes.

Sesgos de información: Se cometen al recoger las medidas o datos. Podemos incluir en este apartado los sesgos de *observación*, posibles en los ensayos clínicos, que se evitan mediante las técnicas de enmascaramiento o *técnicas de ciego*: *Estudio simple ciego*, cuando la asignación del factor de riesgo es ciega por parte de los participantes (el paciente no sabe si toma el fármaco real o el placebo); *doble ciego*, cuando es ciega también por parte del investigador (no lo sabe el paciente ni el investigador); en el *triple ciego* no lo sabe el paciente, el investigador ni el comité que monitoriza el estudio, incluyendo al estadístico que analiza los datos.

Conforme aumenta el grado de "ceguera", también aumenta la dificultad de aplicar las técnicas de enmascaramiento. Un sesgo de información frecuente es el que cometemos al medir la presión arterial o la temperatura, cuando "redondeamos" las cifras un poco arriba o abajo, influyendo de alguna forma en el resultado del estudio. Esto se puede evitar utilizando tensiómetros digitales, que son "ciegos" para los grupos de estudio.

Sesgos de confusión: Son los producidos por la existencia de factores o variables de confusión. Se trata de variables que son factor o marcador de riesgo para la enfermedad a estudiar, se asocian con la exposición al factor de riesgo que se está valorando (factor de estudio), y no son una variable intermedia en la cadena causal entre la exposición y la enfermedad, dos variables están confundidas en un estudio si aparecen de tal manera que sus efectos separados no pueden distinguirse.

Por ejemplo, en un estudio real, se vacunó a una muestra de niños y a otra se les administró placebo. La incidencia de polio fue menor entre los niños que no se vacunaron (porque sus padres no dieron permiso) que entre los que recibieron el placebo. En este estudio existió un factor

de confusión, pues las familias que no dieron permiso eran de un nivel socioeconómico elevado, por lo que sus niños tenían una menor susceptibilidad a contraer la polio, mientras que los niños vacunados con placebo estaban más expuestos a sufrir la enfermedad por carecer de medidas sanitarias adecuadas. El factor de confusión en este caso fue el nivel higiénico-sanitario.

Los sesgos de confusión pueden prevenirse con las técnicas de muestreo probabilístico, especialmente la *estratificación* (hacer que los dos grupos de estudio sean iguales para la variable de confusión). También pueden evitarse posteriormente, realizando un análisis estratificado, donde los datos son estudiados separadamente para diferentes subgrupos, que se definen en función de los posibles factores de confusión considerados.

Es importante recordar que al aumentar el tamaño de la muestra, los sesgos no se modifican!

8. ESTIMACIÓN DE UNA POBLACIÓN A PARTIR DE UNA MUESTRA

Como ya hemos comentado, la estadística inferencial busca obtener conclusiones válidas para poblaciones a partir de los datos observados en muestras. Dado que la población resulta inaccesible por su elevado tamaño, los datos se obtienen a partir de las muestras, pero podemos generalizarlos y estimar parámetros de esa población.

Por ejemplo, deseamos conocer la glucemia media de los diabéticos de nuestra ciudad. Como la población es muy grande, escogemos una muestra representativa (con la misma proporción de edad, sexo, nivel socioeconómico...) de la población y calculamos la media de glucemia (x_1). Volvemos a escoger otra muestra y calculamos su media (x_2), y así varias (n) veces. Podemos hallar la *media* de las medias: $(x_1 + x_2 + x_3 + \dots + x_n)/n$. A este valor se denomina *media poblacional* (μ), y su desviación estándar se lo conoce como *error estándar de la media* (EEM).

Hay que resaltar la diferencia entre desviación típica y error estándar de la media. La primera, mide la dispersión real de los valores de la muestra: es un índice descriptivo de cómo están agrupados los datos; por el contrario, el error estándar mide la dispersión imaginaria que presentarían las sucesivas medias que se obtendrían ante una hipotética repetición del experimento.

Ya podemos deducir que la nueva campana de la distribución de las medias es más estilizada que la distribución muestral, la cual, por recoger valores individuales, acusa más la dispersión de los datos. Por esta razón, el error estándar (EEM) es siempre mucho menor que la desviación típica, y tanto menor cuanto mayor sea el tamaño de la muestra.

El valor del EEM se calcula según la expresión siguiente, donde S = desviación estándar de la muestra y n = número de individuos de la muestra.

$$EEM = \frac{S}{\sqrt{n}}$$

La nueva curva de Gauss obtenida, aunque sea imaginaria, tiene todas las propiedades de la distribución normal. En consecuencia, podemos estimar la media poblacional de la siguiente manera: ya que $\mu \pm EEM$ contiene aproximadamente el 95% de las medias muestrales, entonces el intervalo $[\bar{x} \pm 2EEM]$ contendrá a μ con una probabilidad del 95%, y el intervalo $[\bar{x} \pm 3EEM]$ contendrá a μ con una probabilidad del 99%. A estos intervalos se les denomina intervalos de confianza de la media poblacional, y, sus límites, límites de confianza para la media. Los intervalos serán más estrechos cuanto mayor sea el tamaño de la muestra.

A la probabilidad de que la media escape del intervalo de confianza se le denomina *probabilidad de error* (p).

9. PRUEBAS DE CONTRASTE DE HIPÓTESIS. ERROR ALFA Y BETA

Muchas investigaciones biológicas comportan estudios comparativos. En la situación más simple, se comparan datos de dos muestras, por ejemplo, el efecto de dos fármacos, o de un placebo y un fármaco. Para evaluar las diferencias obtenidas y estudiar la posibilidad de que se deban a factores distintos del azar, se emplean las pruebas de significación estadística o *test de contraste de hipótesis*.

Elementos de un diseño de contraste de hipótesis.

Hipótesis nula H_0 : supone que no hay diferencias entre los términos comparados. Las diferencias se deben sólo al azar.

Hipótesis alternativa H_1 : la que se acepta si H_0 resulta rechazada. Supone que sí existen diferencias entre los términos comparados. Las diferencias no se deben al azar.

	situación verdadera	
	H_0 es verdadera	H_0 es falsa
H_0 aceptada	Sin error	Error tipo II
H_0 rechazada	Error tipo I	sin error

Error tipo I o α : el que se comete al rechazar la hipótesis nula H_0 , siendo cierta (se acepta que existen diferencias, cuando en realidad no las hay). La probabilidad de cometer este error se conoce como α .

Error tipo II o β : el que se comete al aceptar la hipótesis nula H_0 , siendo falsa (hay diferencia real, pero no se acepta).

Potencia estadística del test $(1-\beta)$: es la probabilidad

de rechazar hipótesis nulas falsas, o bien de detectar hipótesis alternativas correctas.

Al aumentar el tamaño de la muestra, se incrementa la potencia estadística de un test y se reducen ambos tipos de errores (α y β).

Nivel de significación p del estudio: es la probabilidad de que las diferencias se deban simplemente al azar, es decir, que H_0 es cierta. Se llama también *grado de significación estadística α* . Su complementario, $(1-\alpha)$ es el nivel de confianza, o probabilidad de que las diferencias no se deban al azar. Por convenio, suele utilizarse un valor de $p=0.05$ (es decir, del 5%).

- Si p es menor de 0.05, se admite que la probabilidad de que las diferencias se deban al azar son demasiado pequeñas, por lo que se acepta la hipótesis alterna H_1 .

- Si p es mayor de 0.05, la probabilidad de que las diferencias se deban al azar es demasiado grande para aceptar la hipótesis alterna, y por tanto se acepta la hipótesis nula, o que las diferencias entran dentro de las debidas al azar.

El grado de significación estadística *no es una medida de la fuerza de la asociación*, no mide si un tratamiento es más eficaz o mejor que otro; simplemente nos da la probabilidad de que los resultados obtenidos sean fruto de la casualidad o el azar. La p tampoco mide la importancia clínica o la relevancia de las diferencias observadas; ello depende de otros factores, y un estudio puede demostrar diferencias muy significativas entre las variables y carecer de importancia clínica. Por ejemplo, si un fármaco A reduce la presión arterial 10 mmHg y otro B la reduce 9 mmHg, y existen diferencias significativas entre ambos ($p < 0.05$) ello no significa que deba usarse el fármaco A antes que el B: hay que considerar el dudoso beneficio clínico que pueda reportar el reducir la presión arterial 1 mmHg más, los efectos secundarios, la seguridad, o el coste económico.

Si al aplicar un test de contraste de hipótesis se acepta la hipótesis alterna, se tiene plenas garantías de ello con un pequeño error conocido (α), y el experimento finaliza. Pero si se acepta la hipótesis nula, no se tiene plenas garantías de esto ya que no se conoce el error β ; en este caso, el experimento no finaliza y será necesario aumentar el tamaño de la muestra para contrastar nuevamente las hipótesis.

Existe una interdependencia entre el grado de significación (p o α), el poder estadístico ($1-\beta$), el número de individuos estudiados y la magnitud de la diferencia observada, de tal forma que conociendo tres de estos parámetros, se puede calcular el cuarto. Por ejemplo, antes de iniciar un estudio, podemos determinar el número de individuos necesarios para detectar una diferencia determinada, fijando a priori el nivel de significación y el poder estadístico deseado.

10. PRUEBAS DE SIGNIFICACIÓN ESTADÍSTICA O DE CONTRASTE DE HIPÓTESIS

Todas las pruebas de significación estadística intentan rechazar o no la hipótesis nula, calculando la probabilidad de que los resultados sean debidos al azar: nos dan, por tanto, el grado de significación estadística "p". Existen dos tipos de pruebas: las paramétricas y las no paramétricas.

Las pruebas paramétricas se utilizan con variables cuantitativas continuas que siguen una distribución normal. Son las pruebas estadísticas que aportan mayor cantidad de información. En ciertas circunstancias, si las variables no cumplen estrictamente los requisitos (por ejemplo, siguen una distribución binomial), pero el tamaño de la muestra es suficientemente grande (mayor de 30), pueden aplicarse estas pruebas.

Las pruebas no paramétricas son las que se aplican a las variables cualitativas, o cuantitativas que no siguen una distribución normal. Suelen estar basadas en la comparación de los rangos de las variables previamente ordenadas, con la consiguiente pérdida de información. Son, en general, menos potentes y precisas que las paramétricas. Si las muestras son mayores de 30, no existe inconveniente en utilizar pruebas paramétricas, aunque la distribución de los datos no sea normal.

A continuación se detallan las pruebas paramétricas y no paramétricas más utilizadas, representadas esquemáticamente en la tabla XXX. La prueba a aplicar dependerá básicamente de cuántos grupos estamos comparando, del tamaño de los grupos y del tipo de variable que estudiamos.

11. PRUEBAS PARA COMPARAR DOS MEDIAS

La prueba más utilizada para este tipo de estudios es la *t de Student-Fisher*, aunque también existe una *prueba de la Z de comparación de medias*. La *t de Student-Fisher* se emplea para comparar las medias de dos muestras. Para que se pueda aplicar deben cumplirse previamente unas condiciones: los datos deben ser independientes, la variable debe seguir una distribución normal en ambas muestras (no es obligatorio si $n > 30$), y las varianzas de los dos grupos deben ser similares en ambos grupos (homocedasticidad), siendo esta condición importante cuando los tamaños de las muestras son diferentes. Para comparar las varianzas empleamos la prueba de la *F de Snedecor*. En caso de que las varianzas no fuesen iguales se aplica el test de *Welch*, una modificación de la *t de Student* para datos independientes cuando las varianzas son distintas.

El valor hallado de la *t* se busca en una tabla para un grado de significación alfa (generalmente 0.05) y con un

número de grados de libertad (se calcula como $n-1$); según el valor calculado, se acepta o se rechaza la hipótesis nula.

Cuando la *t* de Student no es aplicable por incumplirse alguna de las condiciones previas, puede aplicarse la prueba no paramétrica *U* de Mann-Whitney, también llamada prueba de la suma de rangos. Es útil especialmente en muestras pequeñas.

Si los datos son *apareados*, es decir, se comparan dos observaciones realizadas en un mismo grupo de sujetos, puede aplicarse la *t* de Student para datos apareados si se cumple la condición de que las diferencias individuales de cada par de valores deben seguir una distribución normal (aunque esta limitación es menos necesaria si se han estudiado más de 20 sujetos).

Si no se cumplen las condiciones de aplicación de la *t* de Student para datos apareados, puede recurrirse a la prueba no paramétrica de los rangos con signo o prueba de Wilcoxon.

A) Pruebas para comparar dos medias:

Estas pruebas se utilizan para comparar las medias de dos muestras para una variable cuantitativa continua, como por ejemplo, la comparación del efecto de dos fármacos sobre la presión arterial. La prueba paramétrica más utilizada para este tipo de estudios es la *t* de Student-Fisher, aunque también existe una prueba de la *Z* de comparación de medias.

t de Student-Fisher: Se utiliza para comparar las medias de dos grupos de datos independientes. Para poder aplicarse, la variable debe seguir una distribución normal en ambas muestras (no es obligatorio si $n > 30$), y las varianzas deben ser similares en ambos grupos (homocedasticidad), siendo esta condición importante cuando los tamaños de las muestras son diferentes. Para comparar las varianzas empleamos la prueba de la *F* de Snedecor. El valor hallado de la *t* se busca en una tabla para un grado de significación alfa (generalmente 0.05) y con un número de grados de libertad (se calcula como $n-1$); según el valor calculado, se acepta o se rechaza la hipótesis nula. El test de Welch es una modificación de la *t* de Student para datos independientes cuando las varianzas son distintas. Sin embargo, con el uso del ordenador, los programas estadísticos realizan todos estos cálculos automáticamente mostrando directamente el valor de la *p*.

Cuando la *t* de Student no es aplicable por no seguir las variables una distribución normal se utiliza la prueba no paramétrica *U* de Mann-Whitney, también llamada prueba de la suma de rangos. Es útil especialmente en muestras pequeñas. Si los datos son *apareados*, es decir, se comparan dos observaciones realizadas en un mismo grupo de sujetos, puede aplicarse la *t* de Student para datos

apareados. Si no se cumplen las condiciones de aplicación de la *t* de Student para datos apareados, puede recurrirse a la prueba no paramétrica de los rangos con signo o también conocida como prueba de Wilcoxon.

B) Pruebas para comparar tres o más medias.

Análisis de la varianza (ANOVA): Es la prueba paramétrica a la que se recurre para comparar tres o más medias para datos independientes. Es una prueba global que determina si existe alguna diferencia entre el conjunto de las medias consideradas de modo que, si se obtiene un resultado estadísticamente significativo a favor de la diferencia, se concluye que no todas las medias son iguales, pero *no define cuál de ellas es la que difiere*. En este caso, se utiliza posteriormente algún método de comparaciones de dos medias a un mismo tiempo, como el de Tukey, Scheffé, Newman-Keuls o la corrección de Bonferroni.

Para poder aplicarse, el ANOVA exige que los datos sean *independientes* y que sigan una distribución *normal* en cada grupo, con *varianzas iguales*. Si no se cumplen estas condiciones, se recurre a un análisis de la varianza no paramétrico conocido como prueba de Kruskal-Wallis.

Para datos *apareados*, existe un ANOVA para medidas repetidas. La prueba no paramétrica correspondiente es la prueba de Friedman.

12. PRUEBAS ESTADÍSTICAS PARA COMPARAR PROPORCIONES

A) Comparación de dos grupos:

Las pruebas a aplicar son diferentes según se trate de comparar medidas realizadas en grupos independientes o bien se trate de datos apareados. En el primer caso, las pruebas más utilizadas son la *Z* de comparación de proporciones y la *Chi-cuadrado*. En el caso de datos apareados puede utilizarse la prueba de McNemar. En todos los casos estas pruebas no son paramétricas, y pueden aplicarse tanto a variables cualitativas como cuantitativas.

Prueba de Chi-cuadrado: La prueba de chi-cuadrado, en sentido amplio, es aplicable al contraste de variables cualitativas (nominales u ordinales), cuantitativas discretas o cuantitativas continuas distribuidas en intervalos de clase. Es una prueba frecuentemente utilizada, aplicándose para comprobar: a) Si dos características cualitativas están relacionadas entre sí. Por ejemplo, buscar si existe relación entre el color de los ojos y el color del pelo, o infección por VHC y tipo de diálisis (peritoneal y hemodiálisis). b) Si varias muestras de carácter cualitativo proceden de igual población (ejemplo: comparar si dos muestras determinadas de pacientes proceden de poblaciones con igual distribución de grupos sanguíneos). c) Si los datos observados siguen una determinada distribución teórica (por ejemplo, para saber si nuestros datos siguen o

no una distribución normal).

Para su cálculo, se recogen los datos en forma de tablas de frecuencia (las llamadas *tablas de contingencia*), y se calculan el número de casos que se esperaría encontrar en cada casilla de la tabla si no existiese diferencia de los grupos (es decir, se calculan los *esperados* de cada casilla, suponiendo que la hipótesis nula es cierta). Posteriormente, se calcula el valor de chi-cuadrado, que se compara en una tabla, para un número de grados de libertad (se calcula multiplicando las filas -1 por las columnas- 1). Si alguna de las frecuencias *esperadas* es menor a 5, debe aplicarse la *corrección de Yates*; si alguna frecuencia esperada es menor a 2, no puede aplicarse la Chi-cuadrado, debiéndose utilizar entonces la *prueba exacta de Fisher* (sólo es aplicable para tablas de 2x2). Otros autores aceptan aplicar la prueba exacta de Fisher en caso de que los esperados sean inferiores a 5.

Para comparar dos variables cualitativas observadas en los mismos individuos en dos ocasiones (datos apareados) puede utilizarse la *Chi-cuadrado de McNemar*.

B) *Comparación de 3 o más grupos:*

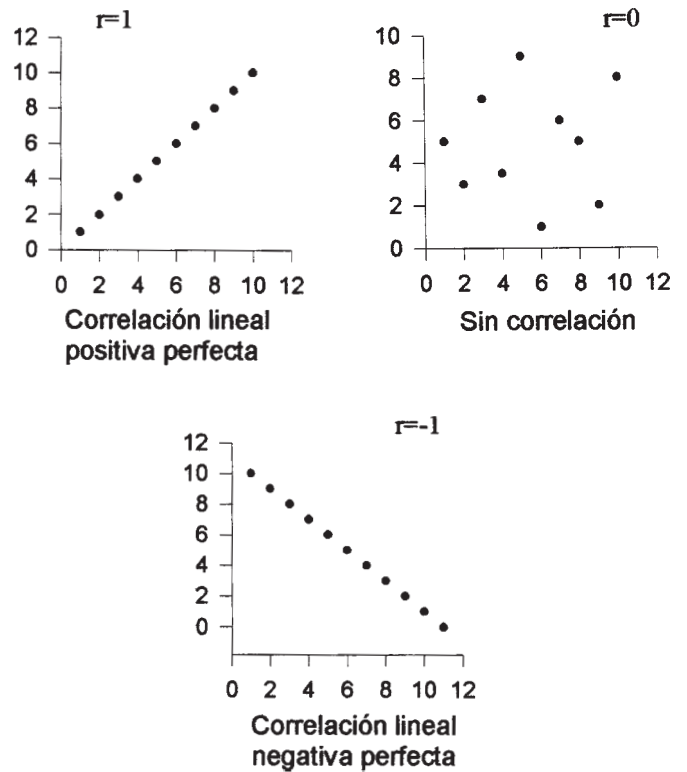
Para datos independientes se utiliza la *Chi-cuadrado de Mantel-Haenszel*. Si los esperados son pequeños, no existe ningún test aplicable, y debemos reagrupar los datos agregando categorías. Para datos apareados, utilizamos la *Q de Cochran*.

13. MEDIDA DE LA ASOCIACIÓN ENTRE VARIABLES: CORRELACIÓN

Hasta ahora, hemos visto una serie de pruebas que, basándose en los tests de significación estadística, nos indican si hay o no diferencias entre grupos; pero estas pruebas no nos informan sobre el grado de asociación, es decir, no dicen si un tratamiento es mejor o peor: sólo indican si es igual o no. Para conocer el grado de asociación entre dos variables cuantitativas, se utilizan los tests de correlación: el *coeficiente de correlación de Pearson* si las distribuciones de las variables son normales, y en caso contrario, se aplica el test no paramétrico de *Spearman*.

Coeficiente de correlación de Pearson: La relación entre dos variables cuantitativas puede representarse gráficamente por una nube de puntos. El *coeficiente de correlación de Pearson (r)* es una prueba estadística que mide numéricamente la existencia de asociación entre esas variables, mediante una fórmula. Existe una relación entre el valor del coeficiente r y la forma de la nube de puntos.

El coeficiente de correlación r es un número comprendido entre -1 (relación lineal negativa perfecta) y 1 (relación lineal positiva perfecta); véase fig. posterior. La asociación es más fuerte cuanto mayor es el valor de r; valores



superiores a 0.7 indican una relación muy fuerte, y 1 es la correlación perfecta. El 0 indica ausencia total de relación. Para poder aplicar el coeficiente de correlación de Pearson se requiere que ambas variables sigan una distribución normal y que la relación entre ambas variables sea lineal.

Coeficiente de correlación de Spearman: Se aplica cuando las variables no siguen la distribución normal. También se emplea para describir la relación entre dos variables ordinales o entre una ordinal y una cuantitativa. El coeficiente de Spearman puede tomar valores entre -1 y +1, y se interpreta de forma parecida al de Pearson.

14. PREDICCIÓN DE UN VALOR DE UNA VARIABLE A PARTIR DEL VALOR DE OTRA VARIABLE: REGRESIÓN

La regresión es un análisis estadístico que se utiliza cuando el objetivo es predecir o explicar el valor de una variable (variable dependiente) a partir del valor de otra (independiente). Cuando las dos variables son cuantitativas continuas, se estudia la posible *relación* entre ellas mediante la *regresión lineal simple*. En ella se calcula un *coeficiente de regresión*, que representa la magnitud del cambio de la variable dependiente por cada cambio de una unidad de la variable independiente. Para que sea aplicable la regresión lineal, debe cumplirse que la relación entre las variables siga una línea recta del tipo $y=ax+b$, y que los valores de y sean independientes unos de otros. Un ejemplo práctico sería cómo aumenta la

recirculación de la fístula arteriovenosa conforme aumenta el flujo sanguíneo.

Para generalizar el análisis a un modelo con más de una variable independiente se emplea la *regresión lineal múltiple*, elaborando una ecuación del tipo $y = a + b_1x_1 + b_2x_2 + b_3x_3 \dots$

Por último, cuando la variable independiente es continua y la dependiente dicotómica, se utiliza la *regresión logística*. Este sería el caso por ejemplo de conocer la relación entre el número de cigarrillos fumados y el desarrollo de cáncer de pulmón; la variable independiente, los cigarrillos fumados, es continua, mientras que la dependiente, tener o no tener cáncer, es dicotómica. En la tabla siguiente se resumen los tests a aplicar en las cuatro posibles situaciones en función del tipo de variable dependiente e independiente.

		Variable dependiente (resultado)	
		Dicotómica	Continua
Variable independiente (predictora)	Dicotómica	Chi-cuadrado	t de Student
	Continua	Regresión logística	Regresión lineal

Al igual que la regresión lineal, la regresión logística posee su *coeficiente de regresión logística* y el cálculo de las estimaciones se basa en procesos matemáticos complejos mediante el *método de la máxima verosimilitud*.

15. ANÁLISIS MULTIVARIANTE

Hasta ahora hemos visto las pruebas estadísticas para analizar dos variables (salvo la regresión lineal múltiple). Si nos planteamos el estudio simultáneo de más de dos variables en cada sujeto, las pruebas se complican. Por ejemplo, si estudiamos la asociación de una determinada enfermedad y 20 factores ambientales. Para estas situaciones existen técnicas estadísticas basadas en modelos matemáticos de cálculo muy complejo, denominadas *análisis multivariante*, en su mayoría derivados del análisis de regresión múltiple. Sin embargo, su utilización es cada vez más frecuente gracias a potentes programas de ordenador que realizan estos cálculos. Su inconveniente es que conforme aumentan las variables a estudiar, también aumenta la probabilidad de encontrar un falso positivo.

16. ANÁLISIS DE LA SUPERVIVENCIA

En numerosas ocasiones se diseñan trabajos de investigación con el fin de conocer la supervivencia de un paciente a lo largo del tiempo ante una enfermedad o un trasplante, o de comparar la eficacia de diferentes trata-

mientos sobre la supervivencia de los pacientes. Para poder extraer conclusiones útiles de estos estudios se requiere aplicar el método estadístico conocido como *análisis de supervivencia*. Con este análisis podemos conocer la probabilidad de sobrevivir a lo largo del tiempo ante la presencia de una enfermedad, estimar las tasas de supervivencia en una población o comparar con seguridad estadística la eficacia de distintos tratamientos sobre dicha supervivencia. El principal problema del análisis de la supervivencia es la irregularidad cuantitativa y cualitativa de los pacientes en la muestra: el tiempo que permanece cada paciente en el estudio es diferente, y debido a causas distintas: algunos fallecen, otros se pierden en el seguimiento y otros siguen vivos al final del estudio.

Los diferentes métodos que se emplean para analizar la supervivencia son el método *directo*, el método de *Kaplan-Meier* y el método *Actuarial*. El objetivo común de estos métodos es estudiar el tiempo que transcurre desde la ocurrencia de un determinado suceso (comenzar un tratamiento, diagnóstico de un cáncer, trasplantarse) hasta la ocurrencia de otro (curación de la enfermedad, morir, perder el trasplante). Con ellos se obtienen una *curva de supervivencia* en la que en el eje horizontal se representa el tiempo y en el vertical la probabilidad de que un paciente sobreviva a un tiempo dado. La curva tiene una morfología en forma de escalera, con ligeras diferencias según el método que se trate.

El método de *Kaplan-Meier* es uno de los más utilizados y se diferencia del actuarial en que utiliza para los cálculos el momento *exacto* en que se produce la muerte, mientras que el actuarial sitúa este momento en un intervalo arbitrario. De esta forma, los "escalones" de la curva de supervivencia de Kaplan-Meier son irregulares, mientras que los del método actuarial son todos iguales. Una vez representadas las curvas de cada grupo, se utiliza un test para compararlas; uno de los más utilizados es el test de *Log-Rank* o *rango logarítmico de Mantel-Cox*, una variante de la Chi-cuadrado de Mantel-Haenszel. Aparte de la significación estadística, puede calcularse además el riesgo relativo o probabilidad de riesgo de un grupo respecto a otro. Por último, existe la posibilidad de analizar el efecto de las posibles covariables y de estimar la *magnitud* de la diferencia entre los grupos mediante el *modelo de los riesgos proporcionales de Cox*.

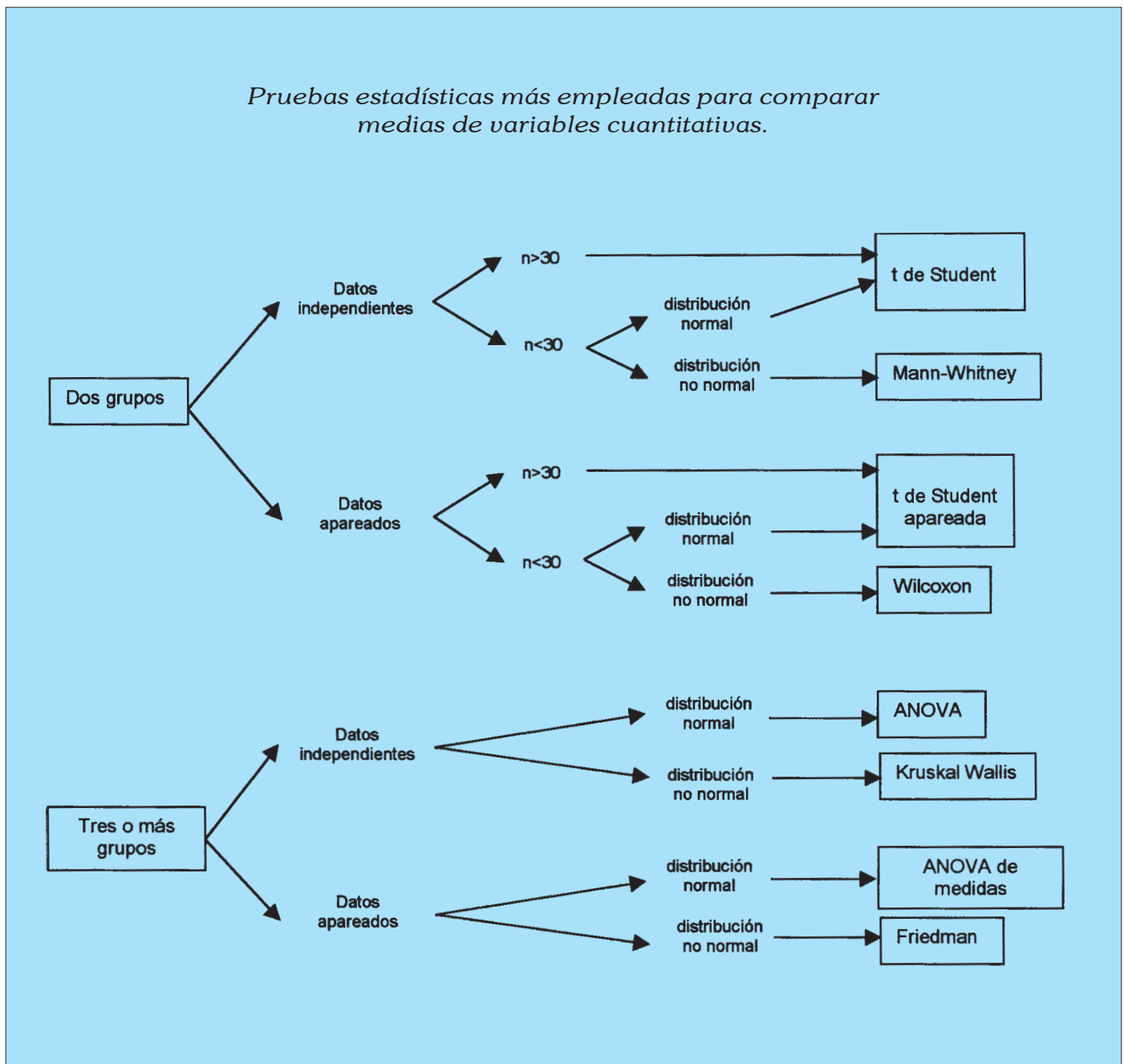
BIBLIOGRAFÍA

- 1.- Bobenrieth MA, Burgos R, Calzas A, Chicharro JA, Ocaña R, Perea-Milla E, Sánchez-Cantalejo E. Metodología de Investigación y Escritura Científica en Clínica. Ed. R. Burgos, Escuela Andaluza de Salud Pública, Granada. 2ª edición, 1998.

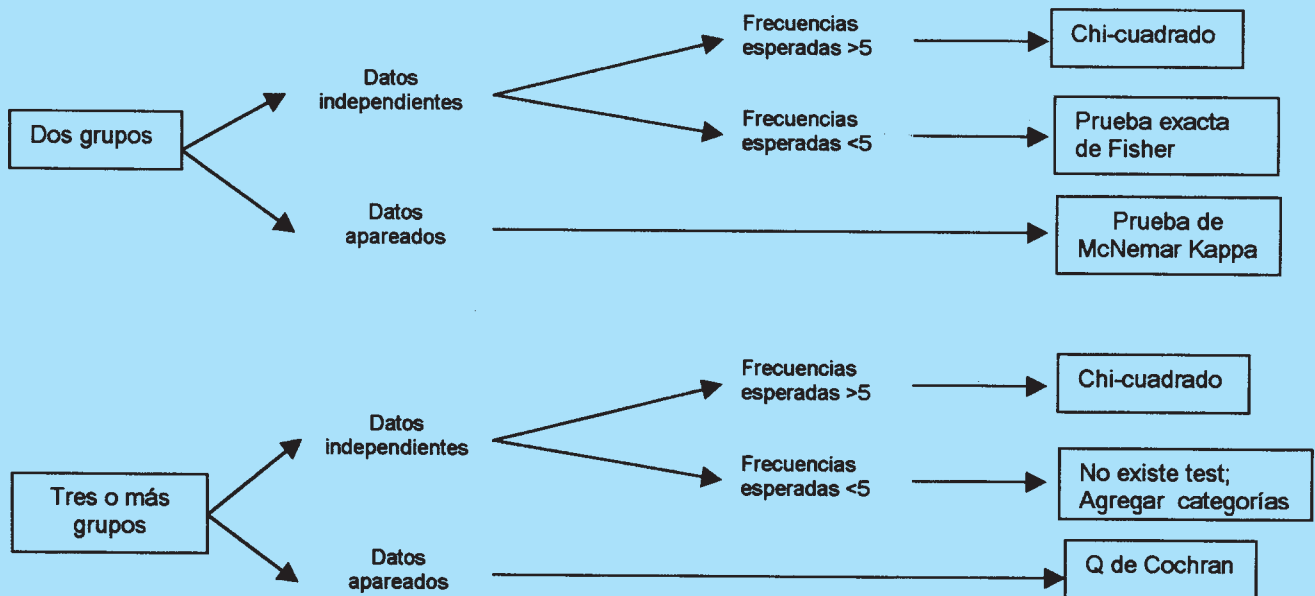
2.- Carrasco de la Peña JL. El Método Estadístico en la Investigación Médica. Ed. Ciencia 3, Madrid, 1995.
 3.- Norman GR, Streiner DL. Bioestadística. Ed. Mosby, Madrid, 1996.
 4.- Martín Andrés A, Luna del Castillo JD. 50±10 horas de Bioestadística. Ed. Norma, Madrid, 1995.
 5.- Laporta, JR. Principios Básicos de Investigación Clínica. Ed. Ergón, Madrid, 1993.
 6.- Huth, EJ. Cómo escribir y publicar trabajos en Cien-

cias de la Salud. Ed. Masson-Salvat, Barcelona, 1992.
 7.- Gil Guillén, VF, Merino Sánchez J, Orozco Beltrán D, Quirce Andrés, F. Curso de Epidemiología Clínica. Estadística Básica, 2ª parte. Ed. Dupont-Pharma, Madrid, 1995.
 8.- Martín Zurro A, Cano Pérez JF. Manual de Atención Primaria (3ª ed.). Ed. Doyma, Barcelona, 1995.
 9.- Mora Temprano, MA. Estadística para enfermería. Ed. Pirámide S.A., Madrid, 1984.

Pruebas estadísticas más empleadas para comparar medias de variables cuantitativas.



Pruebas estadísticas más utilizadas para comparar variables cualitativas, ordinales o proporcionales



Pruebas estadísticas utilizadas con mayor frecuencia para determinar si existe asociación entre dos variables cuantitativas u ordinales

