

Evaluación de las técnicas diagnósticas: Análisis estadístico

Margarita Núñez

***Escuela Universitaria de Tecnología Médica
UdelaR, Montevideo, Uruguay
Comité de Tecnólogos de ALASBIMN***

2008

Evaluación de las técnicas diagnósticas: Análisis estadístico

1. EL RAZONAMIENTO CIENTÍFICO.

La razón de existir de la ciencia, su propósito, es *descubrir la “verdad”*. Pero, ¿cómo decidimos si algo es verdad? Existen básicamente dos herramientas para la búsqueda científica:

- *Nuestros sentidos*, que nos conectan con el mundo y nos permite realizar *observaciones*.
- La **capacidad de razonar**, lo que nos permite llegar a hacer *inferencias lógicas*.

Obviamente, necesitamos *ambas herramientas* pues toda la lógica del mundo no produce observaciones, ni éstas tienen sentido si no dan lugar a una teoría.

Existen dos tipos de relacionamiento de la ciencia con el mundo, esto es, la *inferencia deductiva* y la *inferencia inductiva*.

a) **En la inferencia deductiva**, elaboramos una *teoría* y a partir de ella intentamos realizar una serie de *predicciones* acerca de sus consecuencias, o sea, predecimos el tipo de observaciones que se deberían dar. Por ejemplo, podemos tener una teoría del aprendizaje basada en que es mejor la recompensa que el castigo, previendo la observación de que los estudiantes recompensados tendrán un mejor desempeño que los castigados. En este caso, vamos *de lo general* (la teoría) *a lo específico* (las observaciones).

b) **En la inferencia inductiva**, vamos *de lo específico* a lo *general*, es decir que las *observaciones* nos llevan a inducir una *teoría*. Por ejemplo, en el siglo XIX muchas mujeres morían por infección puerperal en un hospital docente y se observó que este hecho era más común entre las que eran examinadas por estudiantes que provenían de la sala de disección sin higienizarse. Esto desarrolló la teoría de que la infección podría prevenirse mediante el lavado de manos antes de atender un parto, lo cual efectivamente se logró al aplicarlo en la práctica.

A su vez, puede darse que la teoría se vea *invalidada por nuevas observaciones*, lo cual da lugar a una *modificación* de la teoría original y así sucesivamente hasta irse acercando a la verdad, la cual probablemente nunca se alcanza. Por tanto, el progreso de la ciencia requiere tanto de la *inferencia inductiva* como de la *deductiva*, íntimamente correlacionadas.

Las teorías pueden usarse para *predecir observaciones*, pero éstas pueden no ser exactamente como las predijimos debido al *error* y a la *variabilidad* inherente a los fenómenos naturales. Si las observaciones son *demasiado diferentes* a nuestras predicciones, entonces deberemos *abandonar o modificar* nuestra teoría. ¿Cómo podemos *medir la discordancia* entre nuestras predicciones basadas en la teoría y la realidad de las observaciones? La manera es mediante un *test estadístico* o de probabilidad, llamado el *test de la hipótesis nula*, piedra fundamental de la inferencia estadística.

2. LA NECESIDAD DEL ANÁLISIS ESTADÍSTICO.

Un *fenómeno natural* puede estar basado en un *modelo determinístico* o en un *modelo probabilístico*.

a) **En el modelo determinístico**, cuando se dan determinadas condiciones, *siempre* ocurre el mismo resultado. Un ejemplo es la ley de Boyle que establece que al aumentar la temperatura de un gas, aumenta su presión si el volumen se mantiene fijo. Lo único que puede variar el resultado, es un *error* en la medida. Muchos fenómenos físicos y químicos pertenecen a esta categoría.

b) **En el modelo probabilístico**, el fenómeno ocurre con una *determinada frecuencia*, no es constante. Por ejemplo, existe una determinada probabilidad de que un hombre adulto mida 175 cm en una población dada – pero obviamente no todos los hombres llegan a medir la misma estatura. La mayoría de los *fenómenos biológicos* caen en esta categoría, así como los fenómenos psicológicos, sociológicos y otros influenciados por muchos factores a su vez variables y aún por factores no identificados.

Cuando el modelo es *predominantemente probabilístico*, se necesitan *técnicas estadísticas* para alcanzar el conocimiento científico – *la presencia de variabilidad requiere análisis estadístico*.

Los *métodos estadísticos* son *herramientas objetivas* por las cuales se pueden definir las *tendencias de los grupos* a partir de cierto número de *observaciones individuales* (también llamadas *muestras*). Los promedios, porcentajes, desvíos estándar y demás representan una pequeña parte del método, importante pero solamente restringido a *describir las características* de la población bajo estudio. En realidad, en el razonamiento estadístico elaboramos *inferencias inductivas* de lo particular (muestras) a lo general (población), dado que no es posible medir un parámetro específico en la *totalidad* de la población investigada.

3. DISEÑO DE ESTUDIOS.

Mientras que una hipótesis puede originarse en observaciones más o menos anecdóticas, la *prueba de una hipótesis* debe ser realizada mediante *observaciones controladas*, lo cual se puede lograr mediante *estudios bien diseñados*. Existen dos tipos de estudios:

a) **Observacionales**, en los cuales la *selección de grupos* existe *naturalmente*, y el investigador intenta establecer si existe una *asociación* entre dos o más factores. Esta asociación, a su vez, puede o no reflejar una relación causa-efecto (*causalidad*). Existen 3 diseños diferentes de estudios observacionales:

- **Transversales** (cross-sectional), en que las medidas son tomadas en un determinado *momento temporal*. Por ejemplo, a todos los pacientes con diagnóstico de infarto de miocardio, se investiga la presencia de hipertensión arterial. Puede demostrarse una asociación entre infarto e hipertensión al medir ambos parámetros, sin embargo no puede decirse si uno es consecuencia del otro, o si ambos son producto de una tercera causa común.
- **Retrospectivos**, o con *casos control*. En este diseño, el investigador compara pacientes con un factor de interés contra otros sin ese factor, y observa la *influencia de ese factor sobre la ocurrencia de un fenómeno*. Por ejemplo, la relación entre el hábito de fumar y el cáncer de pulmón, analizando cuántos con la enfermedad eran fumadores y cuántos no lo eran. El hecho de ser estudios retrospectivos, o sea *recolectando datos pasados*, expone este diseño a varias

inexactitudes, aunque los resultados son más consistentes que los de estudios transversales.

- **Prospectivos**, también llamados *longitudinales* o de *cohorte*. Aquí el investigador comienza con un **grupo controlado** (cohorte) de pacientes que **presentan la variable** en estudio (p.ej., estudios de perfusión miocárdica positivos) y uno que **no la presentan** (estudios normales), y sigue la evolución de ambos grupos para determinar la **frecuencia** con que los integrantes de cada grupo **desarrollan un evento** (p. ej., infarto de miocardio) transcurrido un cierto tiempo. Estos estudios proporcionan **evidencia más sólida** que los anteriores pero son más difíciles, costosos y prolongados.

b) Experimentales. En general se diseñan para **verificar el resultado** de una intervención, fármaco, método diagnóstico, etc. y en medicina se denominan **ensayos clínicos**. El investigador **asigna** los pacientes o participantes a uno u otro grupo, generalmente de manera **aleatoria**, procurando **mantener constantes otras variables** que puedan influir en el resultado. Por ejemplo, si se desea saber la eficacia de dos métodos para evaluar la viabilidad miocárdica en pacientes con insuficiencia cardíaca se **asignan al azar** a estudios de medicina nuclear o a ecocardiografía. A continuación, todos los pacientes son sometidos a revascularización miocárdica. Los que mejoran, tenían viabilidad por lo tanto se puede saber cuál de los métodos fue más eficaz para identificar esos pacientes. Este diseño proporciona la **evidencia más rigurosa de causalidad**.

4. LA HIPÓTESIS NULA.

Es una manera **objetiva** y adecuada de **verificar una hipótesis**. Supongamos que deseamos saber si los estudios de perfusión son mejores que el ecocardiograma para detectar enfermedad coronaria. Dividimos en dos grupos a los pacientes con dolor torácico, y a cada grupo lo sometemos a uno de los estudios. Finalmente, enviamos a todos los pacientes a coronariografía. Si bien nuestra **hipótesis inicial** es que los estudios de perfusión son más eficaces (y es lo que esperamos como especialistas en medicina nuclear), realizamos un planteamiento diferente. La **nueva hipótesis** dice que **no existe diferencia** entre ambos métodos, y **esperamos demostrar durante nuestra investigación que esta hipótesis es falsa**. Esta nueva hipótesis se denomina **hipótesis nula**. Si demostramos que una **diferencia observada** entre ambos métodos **no es debida al azar, rechazamos la hipótesis nula y aceptamos la hipótesis alternativa**, esto es, que los estudios de perfusión son más eficaces que el ecocardiograma (aunque en este caso, una segunda hipótesis alternativa sería que el eco demuestra ser mejor que la perfusión). Si **no observamos una diferencia**, la **hipótesis nula** será **verdadera** y por tanto nuestra **hipótesis inicial** (luego transformada en alternativa) de que el estudio de perfusión es mejor, será **falsa**.

La **prueba de una hipótesis** está sujeta a **dos tipos de error**: el error de **rechazar** la hipótesis cuando en realidad es verdadera, y el error de **aceptarla** como verdadera cuando en realidad es falsa. El propósito de un protocolo de investigación bien diseñado es **minimizar** la ocurrencia de tales errores.

5. EL ANÁLISIS ESTADÍSTICO.

a) La estadística como herramienta para enfrentar la variabilidad.

La investigación de un tema científico requiere del *análisis estadístico*, especialmente si existen características de interés que presentan *variabilidad*. Ésta puede provenir de dos fuentes:

- el hecho a investigar puede *presentarse en forma variable* en la población, y/o
- puede haber *errores en las medidas* efectuadas.

La *estadística* es un componente necesario de un estudio experimental ya que constituye una herramienta capaz de *evaluar correctamente* la evidencia obtenida por el mismo. Si bien no es posible generalmente alcanzar la certeza absoluta sobre algo, la estadística es capaz de asignar un valor que describa el *grado de incertidumbre*.

b) Variables dependientes y variables independientes.

En los estudios de investigación, deseamos medir la *relación entre una o más variables* que en principio creemos puedan ser *predictoras o determinantes* de un cierto *resultado o efecto*. Las primeras son llamadas *variables independientes* y el último, *variable dependiente*. Por ejemplo, para una cierta población de pacientes, el nivel de colesterol, los trastornos del ST en una prueba de esfuerzo y los defectos de perfusión en un estudio de cardiología nuclear, son *variables independientes* y la incidencia de infarto es una *variable dependiente* de los factores antes mencionados.

c) Variables discretas y variables continuas.

Para evaluar la eficacia de un medicamento o un método diagnóstico, es necesario contar con métodos para *cuantificar* las variables. Éstas pueden clasificarse en 2 tipos: *continuas* (p.ej. el nivel de colesterol, que puede mostrar cualquier valor, digamos desde 50 hasta 500 mg/dl), o *discretas* (p.ej. el resultado de una ergometría, que puede ser positiva o negativa sin probabilidades intermedias, a lo sumo dudosas). A su vez, las variables continuas se pueden *agrupar en rangos* y transformarse en discretas (p.ej., colesterol menor a 200 mg/dl, de 200 a 250, y mayores de 250). Existen técnicas apropiadas para analizar estos tipos de variables tales como el test de chi-cuadrado, el test Z, el test de t, los intervalos de confianza, el análisis de varianza, la correlación y la regresión, que veremos más adelante.

d) Concepto de probabilidad.

La *probabilidad* se define como *la posibilidad de que un evento ocurra, dadas determinadas condiciones*. La probabilidad está dada por un número *entre 0 y 1*: si la probabilidad es 0, el evento *nunca* ocurrirá (al menos bajo las circunstancias mencionadas), mientras que si es 1, es *seguro* que tendrá lugar. Una probabilidad de 0,5 es obviamente intermedia, o sea que el evento ocurrirá la mitad de las veces en que las condiciones se den (p. ej., sacar cara al echar una moneda al aire).

En muchos casos, sin embargo, no podemos enumerar todas las formas posibles en que un evento puede ocurrir y por tanto usamos el concepto de *frecuencia o probabilidad relativa*. Esta se define como el *número de veces que un evento tiene lugar* dividido el *número de oportunidades que ha tenido de ocurrir* (experimentalmente, sería por ejemplo el número de veces que se produce una reacción química sobre el número de veces que se realiza la prueba de poner en contacto dos sustancias determinadas). Sin embargo, es necesario tener en cuenta que la probabilidad funciona para *grandes números*. Por ejemplo, si tiramos una moneda al aire 10 veces,

podemos obtener 6 caras y 4 números (probabilidad relativa de 0,6 para cara), pero si lo hacemos 10.000 veces, la probabilidad será 0,5 o muy cercana, y tanto más cuanto más veces hagamos la prueba.

Esto lleva al concepto de *frecuencia esperada*, definida como $P \times N$, en que P es la probabilidad y N el número de pruebas. Si echamos la moneda al aire 10.000 veces, la frecuencia esperada de que salga cara será $0,5 \times 10.000 = 5.000$.

e) Probabilidad y chances (odds).

Sabemos que de 5 pacientes de sexo femenino con nódulos hipocaptantes o “fríos” únicos, sólo uno es maligno. Podemos decir entonces que un nódulo tiroideo “frío” tiene una *chance* de 4 a 1 de ser benigno, lo cual significa que la probabilidad es $4/5 = 0,80$. Para *convertir* una “chance” en “probabilidad”, simplemente *sumamos las posibilidades* ($4 + 1 = 5$) y las colocamos como denominador de las *chances* ($4/5 = 0,8$). Del mismo modo, decimos que las chances de que el nódulo sea maligno son 1 en 5 = probabilidad de 0,20. Multiplicados por 100, estos valores pueden expresarse en %.

f) Razón de chances (“odds ratio”) y riesgo relativo.

En medicina es frecuente referirse a la relación o *razón de chances* (“odds ratio”), por ejemplo, si la chance de que un nódulo frío sea maligno es de 0,20 y de un nódulo “caliente” es de 0,02, la relación de chances es de $1/5 : 0,1/5 = 5/0,5 = 10$. Una relación de chances de 10 significa, en este caso, que un nódulo tiroideo frío tiene 10 veces más probabilidad de ser maligno que uno caliente.

Sin embargo, la razón de chances no nos permite saber la *probabilidad absoluta* de un cierto nódulo de ser maligno sino solamente cuántos nódulos fríos son malignos en relación a los calientes. Se dice que la razón de chances nos proporciona una *estimación del riesgo relativo* de ocurrencia de un evento.

g) Razón de verosimilitud (“likelihood ratio”).

La relación o *razón de verosimilitud* nos dice *qué tan probable* es que un determinado evento sea *consecuencia* de *un grupo de circunstancias*, en relación a la probabilidad de que aparezca como resultado de *otro* grupo diferente de circunstancias. Por ejemplo, si se presenta un paciente con pérdida de memoria, deseamos saber si puede ser síntoma de una enfermedad de Alzheimer (EA), aún conociendo que puede ser consecuencia de otras patologías. Puesto de otro modo, lo que deseamos saber es la probabilidad de que la pérdida de memoria *sea producida por una EA* en relación a la probabilidad de que refleje *otro tipo de enfermedad*. En cuanto a técnicas diagnósticas, podría decirse que intentamos conocer la probabilidad relativa de que un cierto defecto de perfusión cerebral sea consecuencia de una u otra enfermedad. La razón de verosimilitud (LR) representa un cociente entre probabilidades (P) condicionadas:

$$LR = P(\text{defecto de perfusión, EA presente}) : P(\text{defecto de perfusión, EA ausente}).$$

Naturalmente, para calcular la razón de verosimilitud necesitamos tener una *estimación de las probabilidades* involucradas en la ecuación, o sea: qué proporción de pacientes con EA tiene defectos de perfusión cerebral, y qué proporción de personas sin EA presentan defectos. Este último dato es difícil de conseguir, pues deberíamos saber la *prevalencia* de defectos de perfusión no solamente en pacientes con otras condiciones patológicas sino también en la población general.

6. TÉCNICAS DE ANÁLISIS ESTADÍSTICO.

a) Técnicas aplicables a variables discretas: el chi cuadrado.

El chi cuadrado es un método estadístico usado para determinar si el resultado de un experimento *puede ser o no causa del azar*, estando involucradas variables discretas (o sea, del tipo “todo o nada”). Consideremos por ejemplo la presencia de áreas hipocaptantes reversibles en el miocardio en estudio de perfusión. Creemos que este hallazgo puede ser predictor de un infarto de miocardio, por lo tanto establecemos la *hipótesis nula*:

- No existe diferencia en la ocurrencia de infarto entre los pacientes con y sin defectos de perfusión.

Y la *hipótesis alternativa* dirá:

- La incidencia de infarto es mayor en el grupo con defectos de perfusión.

Registremos nuestros datos en una tabla de 2x2, en la cual cada paciente es asignado a una de las cuatro celdas:

	Con defectos	Sin defectos	Total
Con infarto	223	89	312
Sin infarto	39	40	79
Total	262	129	391

La incidencia de infarto en el grupo *con defectos* es $223/262 = 85\%$, mientras que en el grupo *sin defectos* de perfusión es $89/129 = 69\%$. Pero, ¿podría esta diferencia haberse dado *por azar*? Para saberlo, necesitamos conocer la *frecuencia esperada* de un evento en la población total y compararla con los datos obtenidos, así como *medir la diferencia* eventualmente encontrada entre esos valores.

Por ejemplo, en nuestro caso 312 pacientes de un total de 391 tuvieron infarto, lo que hace una frecuencia de $312/391 = 80\%$. Podría *esperarse* que esa frecuencia se diera tanto para el grupo *con* defectos de perfusión como el de aquellos *sin* defectos. Sin embargo, vemos que no es así pues el grupo con defectos tuvo una frecuencia de infarto de $223/391 = 57\%$, y para el grupo sin defectos la frecuencia fue de $89/391 = 23\%$. Para ver si la *diferencia* entre *valores recogidos* y *esperados* es producto del azar, calculamos un término llamado **chi-cuadrado**. Este se computa tomando el *valor observado en cada celda de la tabla, sustrayéndolo del valor esperado en cada celda, llevando esa diferencia al cuadrado y dividiendo el resultado por el valor esperado en cada celda*. Finalmente, **sumamos** los cuatro valores obtenidos lo cual da el **chi-cuadrado**.

En nuestro caso, el valor calculado es 13,94. Existen *tablas* que indican el *nivel de significación* de los valores de chi-cuadrado, siendo el *valor crítico* (por encima del cual el hallazgo es significativo) = 3,84. Entonces, nuestro valor indica que la *probabilidad* de que la diferencia encontrada entre ocurrencia de infarto en el grupo con y sin defectos de perfusión sea producida por el *azar*, es **tan remota** que nos permite **rechazar la hipótesis nula y confirmar la hipótesis alternativa**. Vemos así cómo al **comparar la frecuencia observada de un evento** (infarto vs. no infarto) en subgrupos de población (con defectos vs. sin defectos), **con la frecuencia esperada** en la población total, podemos estimar si el resultado es *producto del azar* o si existe una *relación de causalidad*.

b) Técnicas de análisis estadístico aplicables a variables continuas.

• *Descripción de una población: la distribución normal y el desvío estándar.*

Cuando deseamos *describir una población* con respecto a cierta *característica cuantitativa*, usamos la *media* o el *promedio* como indicador de una *tendencia central* de los datos. Otros valores descriptivos son la *mediana* (valor por encima y por debajo del cual se encuentra el restante 50% de los datos) y el *modo* (valor que se da con mayor frecuencia en la población estudiada).

Sin embargo, la media no proporciona una descripción adecuada de la población pues se necesita algún *indicador de la variabilidad* de los datos alrededor de este valor. El indicador más común es el *desvío estándar (DS)*, que representa algo así como el *promedio de las diferencias* de los datos respecto a la media.

Muchas variables biológicas cuantitativas, como por ejemplo los valores de laboratorio, la fracción de eyección ventricular, la tasa de filtrado glomerular, etc. poseen lo que se denomina una *distribución normal* al graficar la frecuencia relativa con que se encuentran. Esta grafica tiene forma de campana (curva de Gauss) y se denomina también *distribución gaussiana*, donde la mayoría de los valores se encuentran *alrededor de la media*, y la frecuencia de valores es progresivamente decreciente hacia los extremos. En una distribución de este tipo, el 68% de los valores se encuentran dentro de 1 DS a cada lado de la media, el 95% a 2 DS y el 99% a 3 DS de la media. Habitualmente se toma como *rango normal* para una población, *la media más 2 DS hacia cada lado*. Los valores hallados fuera de este rango se consideran *anormales* o al menos sospechosos y requieren evaluación posterior.

• *Límites e intervalos de confianza.*

En el caso mencionado, hablamos de que los *límites de confianza* (para considerar un dato o valor como perteneciente a la población normal) son los ubicados en los *extremos* de la media ± 2 DS (que comprende el 95% de la población normal). El *intervalo de confianza* es el comprendido entre ambos límites. Por ejemplo, para la fracción de eyección ventricular izquierda medida por ventriculografía radioisotópica (MUGA), el intervalo de confianza es de $67 \pm 12\%$, y los límites de confianza son 55% y 79% respectivamente ($67 - 12$ y $67 + 12$). Los valores encontrados fuera de esos límites (o intervalo) son probablemente patológicos.

• *Comparación entre dos grupos: test de t.*

Para *comparar dos grupos de datos* cuando la medida de interés es una *variable continua*, usamos el *test de t*. Supongamos que deseamos comparar los valores de fracción de eyección (FEVI) en pacientes que han recibido la droga A vs. los que han recibido la droga B. Asumiendo que hemos obtenido un apropiado número de muestras de cada grupo, tendremos *dos juegos de valores* de FEVI, cada uno con su promedio, DS, etc.

Nuestra *hipótesis nula* será:

- *No hay diferencia* entre la FEVI de pacientes que ha recibido la droga A vs. la de aquéllos que recibieron la B.

Y la *hipótesis alternativa*:

- Los dos grupos de FEVI *son diferentes* entre sí.

Sabemos que habrá *variaciones de valores* alrededor de la media del grupo A y del grupo B, simplemente por el carácter variable de los parámetros biológicos. Si juntamos todos los valores de ambos grupos, éstos a su vez tendrán una *media general*.

Para conocer si nuestros dos grupos son diferentes entre sí, lo que hacemos es ver si la **variabilidad entre las dos medias grupales** respecto a la **media general**, es **mayor** que la **variabilidad de las muestras** respecto a la **media de cada grupo**. Para esto aplicamos el **test de t**.

Se calcula el **valor de t**, que es igual a la **diferencia entre las medias** de los dos grupos (dividida por el error estándar). Luego se busca el valor de t en una tabla donde está asignada una **probabilidad** de que dicho valor corresponda a la hipótesis nula (o sea, no existe diferencia entre las medias). Si esta probabilidad (**p**) es pequeña (p.ej., <0,05) entonces la hipótesis nula es rechazada y aceptamos que ambos grupos de FEVI son diferentes.

- **Comparación entre varios grupos: análisis de varianza (ANOVA).**

La técnica apropiada para analizar **variables continuas** cuando existen **más de dos grupos** es el análisis de varianza, más conocido como **ANOVA**. Un ejemplo podría ser el valor de FEVI en tres grupos de pacientes tratados con 3 drogas diferentes.

Los **principios** del ANOVA son **los mismos que para el test de t**. Bajo la hipótesis nula, la diferencia entre las medias de cada grupo y la media de la población general debería ser debida solamente al azar. Si demostramos que esto no es así, aceptaremos que las variaciones se deben a que efectivamente son **grupos distintos**, o sea que el efecto de las drogas es diferente. Aquí se debe calcular la **varianza entre los grupos** (intergrupala) y la **varianza dentro de cada grupo** (intragrupala), cuyo cociente da lugar al **valor F**. Este se busca en tablas y se observa el correspondiente valor de **p**, tal como se hace con el test de t.

- **Coefficiente de correlación.**

Un problema común en investigación es determinar si existe una **asociación entre dos variables**. Por ejemplo, queremos saber si el **aumento de la FEVI** es dependiente de la **dosis de cierta droga**. La medida de la asociación entre los valores de dosis del medicamento y los valores de FEVI correspondientes se investiga mediante el **coeficiente de correlación (r)**, número que puede variar entre **-1** y **+1**. Cuando **r=1**, significa que existe una **correlación perfecta** entre ambas variables (el aumento de una lleva al aumento de la otra, como podría suceder en el caso del ejemplo). Esto permite **predecir** aproximadamente el valor de una variable conociendo el de la otra.

Habitualmente, una buena correlación directa demuestra un valor de **r cercano a 1**, p.ej. 0,87. Un coeficiente **r=-1** indica a su vez una **perfecta relación inversa**. Por ejemplo, la tasa de filtrado glomerular (GFR) y el valor plasmático de creatinina (creatininemia) están relacionados inversamente, ya que al caer aquél, éste se incrementa por la insuficiencia renal. Un valor de **r=0** significa que **no existe asociación alguna** entre las variables.

- **Regresión.**

Para establecer el **grado de asociación entre variables**, los valores de éstas se grafican de modo que en Y se coloca la variable dependiente (p.ej., la FEVI) y en X la variable independiente (p.ej., la dosis del medicamento). Decimos entonces que la FEVI **es función** de la dosis administrada. Aparte del valor de r, son importantes otras 2 características de la gráfica: la **intercepción** (punto de corte en el origen sobre Y) y la **pendiente**, que muestra la velocidad de variación de un parámetro respecto al otro.

Como los valores no suelen estar perfectamente alineados, la conexión de los puntos dará lugar a una línea quebrada por lo cual debe trazarse una línea que represente

la **tendencia general** de la forma más exacta posible. Esto se llama **regresión lineal** y la forma de realizarla es mediante un cálculo denominado **ajuste por mínimos cuadrados**, intentando que la gráfica resultante se mantenga aproximadamente equidistante de cada valor individual de la gráfica original.

7. CARACTERÍSTICAS DE UN MÉTODO DIAGNÓSTICO.

a) Sensibilidad, especificidad, exactitud, valores predictivos.

El problema de la aplicación de técnicas diagnósticas es acertar con el adecuado **compromiso** entre la **detección de una enfermedad** entre las personas que realmente la padecen y **evitar asignar esa patología** a personas que no la tienen. Un modo importante de ver el desempeño de los tests diagnósticos es mediante el **análisis de sensibilidad**. A continuación definiremos los términos más relevantes de este análisis, partiendo de la tabla 2x2 clásica:

		Presencia de enfermedad	
		SI	NO
Test	positivo	VP	FP
	negativo	FN	VN
		total enfermos	total sanos

VP = verdaderos positivos (diagnosticados con enfermedad y realmente la tienen).

FP = falsos positivos (diagnosticados con enfermedad y no la tienen).

VN = verdaderos negativos (no diagnosticados, no tienen enfermedad).

FN = falsos negativos (no diagnosticados pero tienen la enfermedad).

- **Sensibilidad** de un test: es la **capacidad para detectar las personas enfermas** dentro de una población de enfermos. Dicho de otro modo, representa la proporción de enfermos que el test clasifica como positivos. Se calcula como la relación de test verdaderos positivos entre el total de enfermos, o sea:

$$\text{Sens} = \text{VP} / \text{VP} + \text{FN}$$

- **Especificidad** de un test: es la **capacidad para detectar las personas sanas** dentro de una población sana. Dicho de otro modo, representa la proporción de sanos que el test clasifica como negativos. Se calcula como la relación de test verdaderos negativos entre el total de sanos, o sea:

$$\text{Esp} = \text{VN} / \text{VN} + \text{FP}$$

- **Exactitud** de un test: es la **capacidad para detectar tanto enfermos como sanos**, o sea la proporción de tests con resultado correcto. Se calcula como la relación de tests verdaderos entre el total de pacientes estudiados, ya sean sanos o enfermos.

$$\text{Exac} = \text{VP} + \text{VN} / \text{VP} + \text{VN} + \text{FP} + \text{FN}$$

- **Valor predictivo positivo**: es la **probabilidad de que la enfermedad esté presente cuando el test es positivo**. Se calcula como la relación de tests verdaderos positivos sobre la totalidad de tests positivos.

$$\text{VPP} = \text{VP} / \text{VP} + \text{FP}$$

• **Valor predictivo negativo:** es la *probabilidad de que la enfermedad esté ausente cuando el test es negativo*. Se calcula como la relación entre los tests verdaderos negativos sobre la totalidad de tests negativos.

$$\text{VPN} = \text{VN} / \text{VN} + \text{FN}$$

Algunos conceptos importantes:

• La *sensibilidad* y la *especificidad* suelen ir en *sentidos opuestos*, es decir que cuanto más sube una más baja la otra, y viceversa. Debemos establecer criterios para la interpretación de pruebas diagnósticas que alcancen un *adecuado compromiso* entre sensibilidad y especificidad. Por ejemplo, si para calificar un estudio de perfusión pulmonar como positivo para diagnóstico de TEP establecemos el criterio de que un defecto segmentario es suficiente, probablemente tendremos un diagnóstico correcto en la mayoría de pacientes con TEP (alta *sensibilidad*), pero también habrá estudios positivos en muchos pacientes sin TEP, que presentan un defecto aislado por otras razones (baja *especificidad*). Por otra parte, si nuestro criterio de positividad es la presencia de más de 5 defectos segmentarios, nuestra *especificidad* será alta pero clasificaremos como sin TEP muchos enfermos que tengan solamente 1-4 defectos de perfusión (baja *sensibilidad*). Lo mismo ocurre con valores de laboratorio, en el sentido que la sensibilidad y especificidad dependerá de la ubicación del umbral de referencia (*punto de corte o "cutoff point"*) para calificar un test como normal o anormal.

• En general, si estudiamos una población de *alta prevalencia* de la enfermedad preferiremos aplicar un *test muy sensible* (o utilizar un criterio de positividad menos estricto) ya que no deseamos que muchos casos escapen del diagnóstico. Por el contrario, si la población posee *baja prevalencia* de la enfermedad, preferiremos un test *más específico*, pues no deseamos sobrediagnosticar demasiados casos. Por ejemplo, se sabe que los estudios de perfusión miocárdica son relativamente más sensibles y el ecocardiograma de estrés es más específico para el diagnóstico de enfermedad coronaria. Si se presenta un paciente de sexo masculino, mayor de 60 años con dolor torácico (perteneciente a un grupo de alta prevalencia de enfermedad coronaria), preferiremos un test muy sensible, mientras que si se trata de una mujer menor de 50 años, preferiremos un test más específico.

b) Teorema de Bayes.

El teorema de Bayes dice que, dada una serie de circunstancias, es posible *cuantificar la probabilidad de ocurrencia* de un evento. En medicina, nos informa acerca de la probabilidad de una enfermedad en particular dada una cierta presentación clínica y *según el resultado de un test*. En realidad, un buen médico clínico aplica constantemente el teorema de Bayes, muchas veces sin saberlo: simplemente se trata de una *deducción diagnóstica* basada en la propia experiencia y en la evidencia científica disponible, a partir del conjunto de síntomas y signos encontrados en un paciente durante la consulta.

En *técnicas diagnósticas*, el teorema de Bayes nos indica *la probabilidad de presencia de una enfermedad dado el resultado de un test* realizado a un paciente, o sea que *representa el valor predictivo positivo y el valor predictivo negativo* del procedimiento. El caso más conocido es el de los estudios de perfusión en cardiología nuclear. El teorema de Bayes se puede representar como un *par de curvas*, una indicando la *probabilidad de enfermedad cuando el resultado es positivo*, y una cuando el resultado es *negativo*. En las ordenadas (y) se indica la probabilidad de enfermedad

luego de realizar el test, y en las abscisas (x) se expresa la probabilidad antes de realizar el test. El concepto básico es que el resultado de un buen test diagnóstico debería **aportar información significativa** para el adecuado manejo de un paciente (**modificar la probabilidad pretest**). La aplicación del teorema de Bayes presupone el manejo de cierta información:

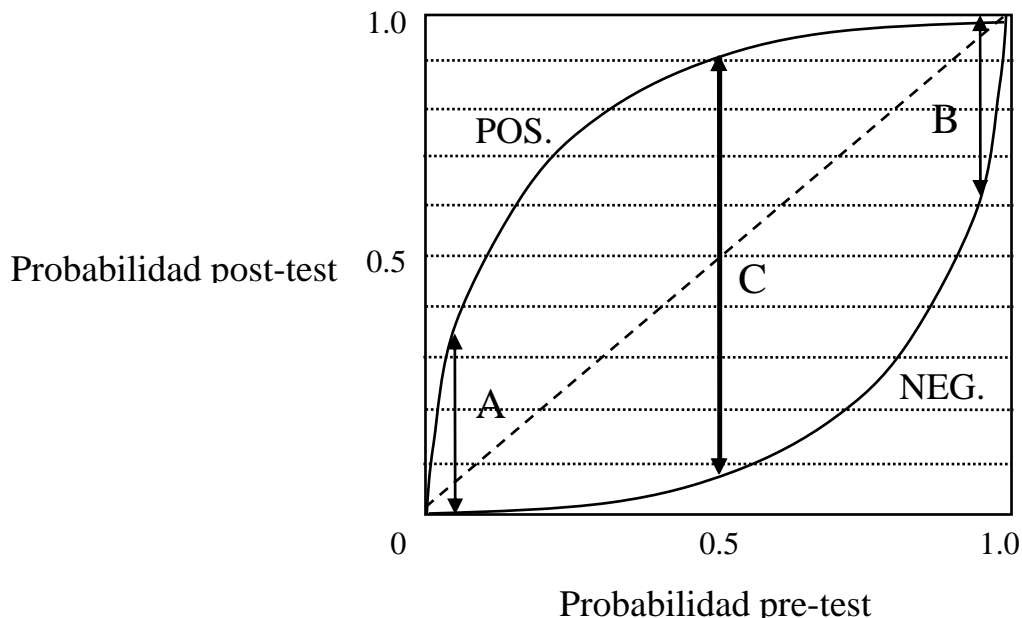
- Primero, debe conocerse la **sensibilidad** y la **especificidad** del test, o sea la capacidad para detectar pacientes con y sin la enfermedad.

- Segundo, se debe saber la **probabilidad pre-test** de que el paciente posea la enfermedad, basándose p. ej. en su edad, sexo, factores de riesgo e incluso en el resultado de otros tests. Esto puede estimarse a través de la **prevalencia** de la enfermedad en el grupo objeto de estudio. La prevalencia es simplemente el **número de personas con la enfermedad** dentro del **total de personas del grupo**, y es un dato que suele conocerse a través de *estudios epidemiológicos* previos.

Si el test posee una sensibilidad y especificidad de 50%, por ejemplo, cualquiera sea el resultado del mismo, no afectará la probabilidad post-test y la curva será una recta que pasa por el origen. En cambio, un test con sensibilidad y especificidad de 80-90% (como es el caso de los estudios de perfusión), afectará la probabilidad, pero especialmente en los pacientes con **probabilidad pre-test intermedia** (digamos, entre 30 y 70%) de presentar la enfermedad.

El teorema de Bayes **ayuda a seleccionar las indicaciones** para las pruebas diagnósticas, **optimizando la relación costo/beneficio** de su aplicación en determinadas circunstancias clínicas.

En la figura se demuestra la aplicación del teorema de Bayes para un estudio diagnóstico con sensibilidad y especificidad cercanas al 90%. La **curva superior** representa la relación entre la probabilidad pre y post-test cuando el resultado de éste es **anormal**, mientras que la **curva inferior** indica dicha relación cuando el resultado es **normal**. Si consideramos un paciente A con **muy baja prevalencia** de la enfermedad (probabilidad pre-test), un resultado anormal modificará escasamente la probabilidad post-test, mientras que si es el estudio es normal, el cambio tampoco será significativo. En el extremo opuesto, un paciente B con **alta probabilidad** de enfermedad pre-test tampoco verá sustancialmente modificada su probabilidad post-test cualquiera sea el resultado del estudio. En cambio, **el mayor beneficio diagnóstico se obtiene con pacientes portadores de probabilidad pre-test intermedia** (C), ya que si ésta es de un 50%, un test positivo la llevará al 90%, mientras que un resultado negativo la descenderá al 10%.



Este concepto reviste gran importancia práctica para una *correcta indicación* del estudio de perfusión referido a pacientes con probabilidad previa en el *rango intermedio* (30% a 70%), ya que es en este grupo donde el resultado del test influirá decisivamente para un adecuado manejo clínico.

c) Curvas ROC.

Como vimos, *modificando el punto de corte* que define el límite entre positividad y negatividad de un test, es posible *cambiar la sensibilidad y especificidad* del mismo. La *elección* del punto de corte puede estar basada, como en los ejemplos mostrados más arriba, en la estimación de las *consecuencias relativas* de tener demasiados estudios *falsos positivos* o demasiados *falsos negativos*. Si la enfermedad en estudio es potencialmente muy *grave*, desearemos *más sensibilidad*, en cambio si investigamos una condición *menos seria* o si la consecuencia de diagnosticar positivos lleva a una *intervención mayor*, por ejemplo una cirugía exploratoria que deseamos evitar a personas que puedan no tener la enfermedad, entonces sacrificaremos la sensibilidad para ganar *especificidad*.

La *relación entre sensibilidad y la proporción de falsos positivos* se expresa en la denominada *curva ROC* (“receiver operating characteristic”) del test. Esta nos ayuda a *establecer el punto de corte* que separa un resultado normal de uno anormal, dependiendo del compromiso que estemos dispuestos a realizar entre la sensibilidad y la proporción de falsos positivos. Con un *punto de corte elevado*, podemos detectar un mayor porcentaje de enfermos pero también tendremos una gran proporción de falsos positivos, mientras que con un *punto de corte más bajo* tendremos menos FP pero nuestra capacidad de detección de enfermos será menor. Idealmente, *para cada test diagnóstico debería disponerse de una correspondiente curva ROC*.

Evaluación de las Técnicas Diagnósticas:
Sensibilidad, Especificidad, Exactitud, Valores Predictivos, Razones de
Verosimilitud, Curvas ROC.

1. EL RAZONAMIENTO CIENTÍFICO.

- a) La inferencia deductiva.
- b) La inferencia inductiva.

2. LA NECESIDAD DEL ANÁLISIS ESTADÍSTICO.

- a) El modelo determinístico.
- b) En el modelo probabilística.

3. DISEÑO DE ESTUDIOS.

- a) Observacionales..
 - *Transversales* (cross-sectional).
 - *Retrospectivos*, o con *casos control*.
 - *Prospectivos, longitudinales o de cohorte*.
- b) Experimentales – ensayos clínicos.

4. LA HIPÓTESIS NULA.

5. EL ANÁLISIS ESTADÍSTICO.

- a) La estadística como herramienta para enfrentar la variabilidad.
- b) Variables dependientes y variables independientes.
- c) Variables discretas y variables continuas.
- d) Concepto de probabilidad.
- e) Probabilidad y chances (odds).
- f) Razón de chances (“odds ratio”) y riesgo relativo.
- g) Razón de verosimilitud (“likelihood ratio”).

6. TÉCNICAS DE ANÁLISIS ESTADÍSTICO.

- a) Técnicas aplicables a variables discretas: el chi cuadrado.
- b) Técnicas de análisis estadístico aplicables a variables continuas.
 - *Descripción de una población: distribución normal y desvío estándar.*
 - *Límites e intervalos de confianza.*
 - *Comparación entre dos grupos: test de t.*
 - *Comparación entre varios grupos: análisis de varianza (ANOVA).*
 - *Coefficiente de correlación.*
 - *Regresión.*

7. CARACTERÍSTICAS DE UN MÉTODO DIAGNÓSTICO.

- a) Sensibilidad, especificidad, exactitud, valores predictivos.
- b) Teorema de Bayes.
- c) Curvas ROC.