

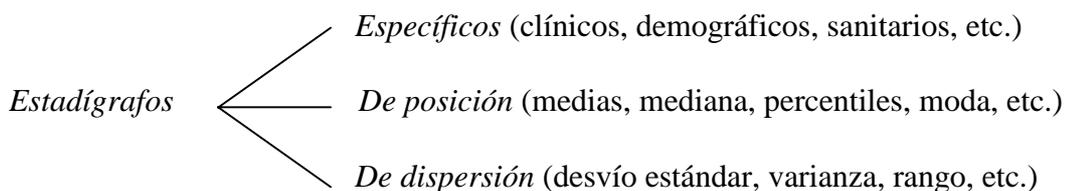
4

Índices clínicos y estadígrafos

En este capítulo se describen diferentes índices clínicos que se usan para juzgar el comportamiento de un método de diagnóstico en Medicina y los estadígrafos clásicos. Un diagnóstico es una predicción acerca del estado del paciente. Al mismo, se puede llegar a través de una serie de análisis tales como: estudio de los síntomas, tests clínicos, radiografías, electrocardiogramas, etc. El resultado final de todo este proceso es binario (sano /enfermo), al igual que la decisión final del médico como: se lo interna al paciente o no, con o sin medicación, cirugía, quimioterapia, etc. Si el resultado es dudoso, esto significa que serán necesarios más estudios y la decisión final solo se pospone para un futuro cercano; pero esta es siempre dicotómica o binaria. En los comienzos, el concepto de calidad se asociaba estrechamente a las ideas de *precisión y exactitud*. La idea clásica era: cuanto más exacto y preciso sea un método, mejor es la calidad del mismo. El problema es que estas ideas solo pueden aplicarse a magnitudes de tipo continuo, donde pueden determinarse los errores de tipo sistemático y casuales asociados a estos conceptos. Por lo que hubo que buscar una nueva manera de entender la calidad en los análisis clínicos y para abarcar a todo tipo de magnitud que se pueda presentar. Lo que conduce al caso dicotómico, pues toda magnitud se puede transformar en una binaria, adoptando un *punto de corte* apropiado en la magnitud que se está analizando. Por ejemplo, la temperatura es una magnitud continua, si se adopta un punto de corte como 37° C para separar un caso normal de un caso de fiebre, entonces se la ha transformado en binaria: con o sin fiebre. La solución a este problema fue propuesta por Galen y Gambino en 1975, en su obra “Beyond Normality”, y desde entonces se ha ido refinando esta forma de encarar la cuestión, tal como se verá a continuación. Básicamente las objeciones son dos: (a) Los análisis clínicos se usan mayormente para detectar enfermedades. Se necesita que ayuden a efectuar un buen diagnóstico. De nada sirve que una técnica, sea muy precisa y exacta si diagnostica mal. (b) Cuando las magnitudes son cualitativas, el criterio de calidad tradicional de precisión y exactitud es inaplicable.

4.1. Clasificación de estadígrafos.

Los estadígrafos se clasifican como se muestra a continuación:



Los estadígrafos de posición y dispersión están estrechamente vinculados con los conceptos de precisión y exactitud. Si se mide muchas veces, en forma repetida, un valor patrón cualquiera de una magnitud como un suero control de concentración conocida o *estándar*, se tiene un conjunto de datos para juzgar el comportamiento de la técnica clínica de medición. Una manera de saber la exactitud de la misma es tomar un valor índice de tendencia central, como el promedio, y compararlo con el valor de referencia. Si el estándar está muy alejado del valor promedio, entonces se puede pensar que la técnica no es muy exacta. Caso contrario, se piensa que la técnica está *calibrada en exactitud*. Para juzgar si la diferencia observada entre ambos valores es por causalidad, o bien se trata de un error de tipo sistemático, no se puede usar un criterio *cualitativo* sino las técnicas estadísticas apropiadas, que se verán más adelante, para poder tener un criterio *cuantitativo*. Esto significa no juzgar las diferencias con criterio humano, subjetivo, sino con un criterio matemático de tipo estadístico: es el adoptado en la ciencia como un *criterio objetivo*. Análogamente, en el caso de la precisión, si la técnica clínica tiene un índice de dispersión muy pequeño respecto de los valores aceptables de referencia en la bibliografía, se puede pensar que es bastante precisa. Nuevamente, sólo un método estadístico permite juzgar imparcialmente si la diferencia entre el índice hallado y el valor de referencia se debe al azar.

Se debe resaltar: *Toda medición científica, de magnitudes continuas, debe ir siempre acompañada con el error de medición, que cuantifica y engloba ambos conceptos*. Un informe presentado sin la aclaración de error de método empleado, no es serio. Cuando se trata de un recuento de unidades enteras, no hay “error” sino equivocaciones. En cambio, en las magnitudes cualitativas, al no ser numéricas, no tienen sentido los conceptos de precisión y exactitud usados tradicionalmente para hablar de *calidad*. Es costumbre en los laboratorios de física, química y relacionados, decir que si un método tiene buena precisión y exactitud, entonces es de buena calidad. Esa costumbre se extendió a los laboratorios de análisis clínicos, donde para optar entre dos técnicas se empleaba ese criterio. Sin embargo, a mediados de la década del cuarenta algunas críticas empezaron a aparecer en la bibliografía. El primer trabajo precursor fue el de Yerushalmi y otros en 1947, junto con los de Galen y Gambino en 1975. Básicamente las objeciones son dos:

1. Los análisis clínicos se usan mayormente para detectar enfermedades. Se necesita que ayuden a efectuar un buen diagnóstico. De nada sirve que una técnica, sea muy precisa y exacta si diagnóstica mal.
2. Cuando las magnitudes son cualitativas, el criterio de calidad tradicional es inaplicable.

La respuesta para ambas críticas, fueron los llamados *índices clínicos*.

4.1.1. Clasificación de enfermedades por diagnosis

Debe tenerse muy en cuenta el tipo de enfermedad que se está analizando y el grado de avance de la misma en el paciente, porque de acuerdo a ello se deberá puntualizar en ciertos aspectos, más que en otros, como se explicará más adelante. Por ahora, basta con efectuar una primera clasificación de las enfermedades en tres tipos básicos de acuerdo al siguiente esquema:

Tipo I : Son todas aquellas enfermedades donde la peor equivocación que se puede cometer en el diagnóstico es un falso negativo.

Tipo II : Son todas aquellas enfermedades donde la peor equivocación que se puede cometer en el diagnóstico es un falso positivo.

Tipo III : Son las restantes, donde no se puede clasificar claramente como una de las anteriores.

Las enfermedades que se pueden clasificar como Tipo I son:

- Una enfermedad que es curable si se detecta a tiempo.
- Una enfermedad que es grave y no puede dejarse pasar inadvertida.
- Los falsos positivos no trauman ni psicológicamente, ni económicamente al paciente, pero los diagnósticos falsos negativos sí lo hacen.

Ejemplos: infarto de miocardio, enfermedades venéreas y otras enfermedades infecciosas curables, la fenilcetonuria, el feocromocitoma (curable en el 100% de los casos si es detectado a tiempo), cáncer de mama o de útero, etc.

En cambio, una enfermedad donde para el paciente, un diagnóstico falso positivo es más peligroso que un falso negativo, se considera como del Tipo II, como ser:

- La enfermedad es grave, pero difícilmente curable o sin remisión.
- Es muy importante para el paciente o para la población él saberse un verdadero negativo.
- Los falsos positivos trauman seriamente al paciente, en cambio los falsos negativos no.
- El tratamiento de los falsos positivos ocasiona serios perjuicios al paciente.

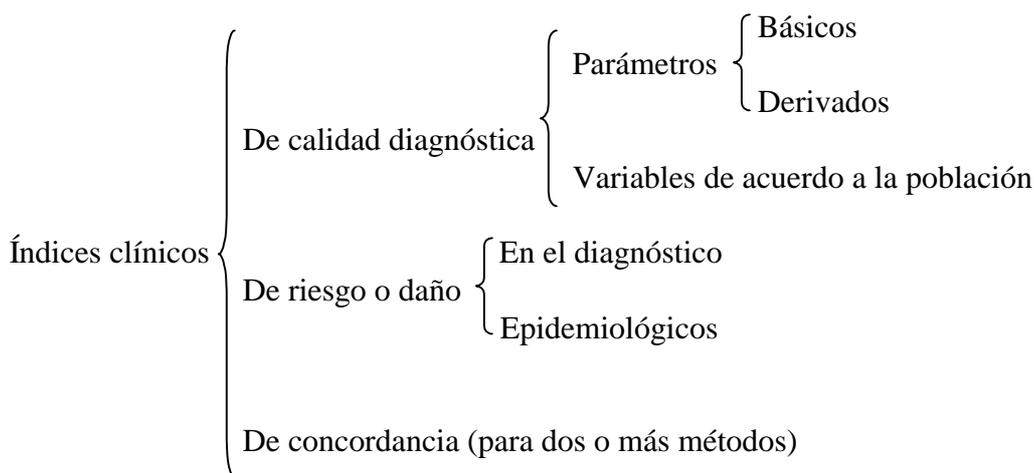
Ejemplos: estado terminal o sin remisión de la enfermedad, cáncer oculto, esclerosis en placas, tratamientos como radioterapia, quimioterapia, lobectomía, cirugía innecesaria, estado IV de la sífilis cuando ya no tiene remedio, etc.

Cuando la enfermedad no puede ser encasillada en ninguno de los dos casos anteriores, entonces se la considera como del Tipo III, como por ejemplo: lupus eritomatoso, ciertas formas de leucemia o linfoma, diabetes, etc. En el caso del SIDA un falso positivo puede ser muy dañino para el paciente, pero un falso negativo sería muy peligroso para la sociedad y como ambos peligros son graves, lo mejor es clasificarla como del Tipo III desde un punto de vista ético. Notar que una misma enfermedad puede ser clasificada de maneras diferentes, de acuerdo al estado de avance que tiene en el paciente. Un cáncer que pueda ser curado si se detecta a tiempo es un claro caso de Tipo I, pero una vez que se produjo una metástasis en el paciente pasa a ser del Tipo II. Lo mismo ocurre con los diferentes estados de la Sífilis, que se puede curar si es detectada a tiempo, pero se torna incurable cuando está muy avanzada.

4.2. Índices clínicos

El objetivo principal de una técnica clínica es que sirva para diagnosticar bien. Si para efectuar el mismo trabajo hay muchas técnicas disponibles, entonces se debe optar por aquella que mejor sirva a ese propósito, y si además es precisa y exacta, mejor. Al realizar un diagnóstico se está “dicotomizando” a la magnitud clínica medida, no importa del tipo que sea. Se puede postular que siempre habrá dos resultados posibles: *Sano* o *Enfermo*. En principio, se pueden descartar a los casos intermedios, donde no está muy claro el estado real del paciente, para simplificar

el problema. La práctica clínica ideal sería una que discrimine perfectamente a los sanos de los enfermos, una que no cometa equivocaciones, una utopía. La realidad actual es bastante diferente; para poder diagnosticar bien se necesita, además de la revisión personal, varias determinaciones de laboratorio que confirmen la diagnosis. Aún así, todavía habrá un cierto grado de incertidumbre, mayor cuanto menos avanzada esté la enfermedad. Y entonces, se puede sospechar que la calidad del método, estará relacionada con la distribución en la población del grado de avance de la enfermedad, con la forma de seleccionar las muestras, además de la calidad intrínseca de la técnica clínica. Por otra parte, de acuerdo a los objetivos del estudio clínico los índices clínicos derivados se pueden clasificar con:



Los índices clínicos para un solo método se dividen en dos partes, los relacionados con el método de diagnóstico (la *calidad* del diagnóstico hecho y el *riesgo* que implica hacer un diagnóstico) y los relacionados con el *riesgo* o *daño* potencial de un factor, que se cree relacionado con la causa de la enfermedad que se está estudiando. Pero cuando se trata de comparar dos o más métodos entre sí, se pueden comparar los índices anteriores si las muestras son independientes, o bien se puede analizar la *concordancia* entre ellos cuando las muestras son apareadas. Los tipos de investigaciones clínicas se detallaron en el capítulo 1, donde se mostraron las tres maneras más sencillas de presentar los datos en tablas de 2 x 2 (Tabla 1.1, 1.2 y 1.3). Ahora, se mostrará como calcular los índices relacionados a cada una de ellas:

4.3. Índices de calidad diagnóstica.

En su forma más simplificada el problema de hacer una predicción o diagnóstico se puede esquematizar con la Tabla de diagnóstico tal como la vista en el primer capítulo Tabla 1.1. Esta forma de organizar los datos se llama tablas doble dicotómicas, donde los cuatro casos posibles de la tabla son: dos verdaderos (vp y vn) y dos falsos (fp y fn) en diagnósticos donde se verifica la presencia de la enfermedad o síntoma buscado (positivo) o la ausencia del mismo (negativo). Los totales marginales de la tabla son:

- TE = vp + fn : Total de Enfermos = verdaderos positivos + falsos negativos
- TS = vn + fp : Total de Sanos = verdaderos negativos + falsos positivos
- TP = vp + fp : Total de positivos = verdaderos positivos + falsos positivos
- TN = vn + fn : Total de negativos = verdaderos negativos + falsos negativos

Resultados del test	Resultados verificados		Total
	Enfermos	Sanos	
Positivo (+)	<i>vp</i> <i>verdadero positivo</i>	<i>fp</i> <i>falso positivo</i>	TP
Negativo (-)	<i>fn</i> <i>falso negativo</i>	<i>vn</i> <i>verdadero negativo</i>	TN
Total	TE	TS	N

La técnica clínica ideal sería aquella que no cometa equivocaciones en sus pronósticos, o sea una con $fn = fp = 0$. Las consecuencias de las equivocaciones no son igualmente graves para el paciente, todo depende del tipo de enfermedad de que se trate. Y por ese motivo se realizó la clasificación de las enfermedades en el punto anterior, teniendo en cuenta el tipo de equivocación cometida en el diagnóstico. Una manera de analizar la *calidad* de un método de diagnóstico es tener en cuenta siempre los dos tipos de equivocaciones, o sea se trata de un problema de tipo dual. Y cuantificar su calidad de acuerdo a la capacidad intrínseca del mismo en equivocarse.

4.3.1 Sensibilidad, Especificidad e Índice de Youden.

El comportamiento de un método de diagnóstico al pronosticar, puede ser analizado en forma dual de la manera siguiente:

- *Sensibilidad*: Es la capacidad del método para diagnosticar como positivo a un enfermo.
- *Especificidad*: Es la capacidad del método para diagnosticar como negativo a un sano.

Por ejemplo, si el método es capaz de detectar a tres casos de positivos cada cuatro enfermos que analiza, entonces su capacidad para identificar correctamente a los que pronostica como positivo es del 75%. Aquí se dice que su sensibilidad $S = 0,75$. Por otro lado, si el método es capaz de detectar a ocho casos negativos, cada diez individuos sanos, entonces su capacidad para identificar correctamente como negativo es del 80%. Se dice aquí que su especificidad $E = 0,8$. Aplicando estos conceptos al caso general presentado en la Tabla de diagnóstico, se pueden deducir las fórmulas que siguen:

Sensibilidad: $S = vp / TE = (\text{verdaderos positivos}) / (\text{total de enfermos})$

Especificidad: $E = vn / TS = (\text{verdaderos negativos}) / (\text{total de sanos})$

Ambos valores están acotados entre cero y uno. Cuando la sensibilidad es máxima $S = 1$, quiere decir que no hay falsos negativos, ya que $TE = vp + fn$, y esa sería la cualidad buscada para una enfermedad del Tipo I. En cambio, para una enfermedad del Tipo II, lo deseable es que $fp = 0$ y para ello deber ser máxima la especificidad: $E = 1$, pues $TS = vn + fp$.

La calidad puede ser expresada con los criterios de sensibilidad y especificidad, por lo tanto una prueba clínica es ideal cuando tiene una capacidad para diagnosticar $S = E = 1$, o sea cuando no comete equivocaciones en sus pronósticos.

A las técnicas clínicas de laboratorio, o bien a los métodos de diagnóstico que se parecen mucho a este ideal se los denomina: métodos de referencia (o “golden standard”). La realidad es bastante diferente a esta utopía, son muy raros los casos donde ambos índices se acercan al máximo a la vez, aún en una biopsia se pueden cometer equivocaciones. Entonces el problema se reduce a elegir aquella prueba que tenga máxima sensibilidad cuando hay disponibles varias técnicas para las enfermedades del Tipo I, y máxima especificidad para las del Tipo II. Para el caso de las enfermedades del Tipo III, se necesita otro índice de calidad que mezcle ambos conceptos.

Este índice clínico fue propuesto por Youden para analizar la capacidad del método de diagnóstico, usando un único valor en reemplazo de la forma dual de hacerlo. La idea es mezclar los dos índices anteriores para hacer el estudio de calidad. Se define como:

Índice de Youden: $Y = S + E - 1 = (\text{sensibilidad}) - (\text{especificidad}) - 1$

En el caso ideal, este índice vale 1 ya que $S = E = 1$. En el caso de lanzar una moneda al aire para efectuar el diagnóstico, es decir puro azar, será $S = E = 0,5$ y por lo tanto este índice es nulo. La condición evidente para tener una calidad aceptable es que no sea negativo. La idea para seleccionar un método de diagnóstico en las enfermedades del Tipo III, es elegir aquel que tenga el máximo valor de este índice, pues eso equivale a tener un mínimo de resultados falsos, prorrateados entre el total de los casos analizados. Estos tres índices dan un valor que cuantifica los conceptos básicos de calidad diagnóstica, de acuerdo al tipo de enfermedad analizada. Sin embargo, puede verse que el último es una consecuencia de los dos primeros, a los que se considera los índices básicos o *parámetros* de calidad, de los cuales se derivan todos los demás.

4.3.2 Eficiencia diagnóstica (“diagnostic accuracy”)

La otra manera de ver la misma idea es usando otro índice, sugerido por muchos autores como la manera de evaluar el “poder discriminatorio” del método clínico. Una especie de índice general de la calidad, porque lo que se analiza aquí es el porcentaje de éxitos logrados por el método en sus pronósticos. Es como el clásico concepto del rendimiento, ver que tan bien trabaja el método de diagnóstico o test clínico. Por ejemplo, si se imagina al método como una “máquina de hacer predicciones”, el rendimiento del mismo se puede expresar con:

Eficiencia: $A = (vp + vn) / N = (\text{verdaderos positivos} + \text{verdaderos negativos}) / (\text{total de casos})$

La idea de eficiencia es simplemente la del porcentaje de éxitos obtenidos al diagnosticar, se puede descomponer en dos tipos de éxitos: el término (vp / N) es el porcentaje de éxitos en positivos y el término (vn / N) es el porcentaje de éxitos en negativos. Cuanto mayor sea la eficiencia de una prueba clínica, mejor será su poder discriminatorio. La prueba clínica ideal tendrá una eficiencia máxima (100%) $A = 1$, mientras que cuando $A = 0$ significa que no tuvo ningún acierto de sus pronósticos. Notar que si se lanza una moneda al aire para hacer pronósticos, la eficiencia sería del 50%, o sea $A = 0,5$. También se puede usar para seleccionar el mejor método en el caso de las enfermedades del Tipo III. De hecho, es el recomendado en la bibliografía. Sin embargo, hay una diferencia conceptual importante entre estos dos índices, y es que el índice de Youden no depende de la población en donde se aplica el método, sino que es una característica intrínseca propia del método, en cambio la eficiencia depende del lugar donde se está usando el

método, o bien del momento en que se lo aplica. Esto es, la eficiencia depende de la *prevalencia* de la población mientras que el índice de Youden no, como se puede ver a continuación:

Prevalencia: $p = TE / N = (\text{total de enfermos}) / (\text{total de individuos de la población estudiada})$

Teniendo en cuenta que la eficiencia se puede expresar como:

$$A = (vp / N) + (vn / N) = [(S \cdot TE) / N] + [(E \cdot TS) / N]$$

Reemplazando los valores verdaderos con los índices sensibilidad y especificidad respectivamente, y luego haciendo las simplificaciones necesarias resulta:

$$A = S \cdot p + E (1 - p) = (\text{sensibilidad} \times \text{prevalencia}) + [\text{especificidad} \times (1 - \text{prevalencia})]$$

Por ejemplo, en una población donde la enfermedad es muy rara o casi inexistente, la prevalencia sería prácticamente nula ($p = 0$) y entonces la eficiencia resulta igual a la especificidad. En cambio en un caso extremo de epidemia, donde casi toda la población está enferma, la prevalencia es prácticamente del 100% ($p = 1$) y entonces la eficiencia resulta igual a la sensibilidad. Notar que, con el transcurso del tiempo una enfermedad puede propagarse y empezar a aumentar la prevalencia de la misma, lo que originaría un cambio en la eficiencia. Pero esto no ocurre con el índice de Youden, porque este no depende de la prevalencia. En forma matemática esto sería $A = f(p)$ mientras que $Y \neq f(p)$. Otro punto importante a tener en cuenta es que, en una misma población geográfica, se pueden tener distintos valores de prevalencia. Por ejemplo, el número de pacientes con problemas cardíacos que concurren a hacerse atender en un centro asistencial de la zona rural de Posadas, es mucho menor que los que concurren al hospital, los que a su vez son menores que los concurrentes a la unidad coronaria del mismo, derivados de otros centros asistenciales. Por lo tanto, el mismo método de diagnóstico aplicado en estos tres puntos tendrá un valor diferente de eficiencia (A) a pesar de que todos están en una misma zona geográfica. Debe remarcar aquí, que la población de referencia para una investigación clínica, no es la de la zona geográfica circundante, sino de aquellos pacientes que concurren a hacerse atender en el centro asistencial.

Para explicar mejor esta diferencia entre parámetros y variables de la calidad del método de diagnóstico, se puede usar un símil de la Física. Calor y frío son dos conceptos opuestos que aparecen como una dualidad. Sin embargo se trata de la misma magnitud física: la temperatura, o sea de la misma cosa. En un extremo de la escala cuando la temperatura es baja se asocia con uno de los extremos, una sensación: el frío. En el otro extremo de la escala, cuando la temperatura es alta se la asocia con la sensación opuesta: el calor. Aparecen como cosas contrarias, cuando en el fondo se trata de lo mismo. Análogamente en Medicina, la *eficiencia* de una prueba diagnóstica es la magnitud clínica variable, en un extremo de la escala cuando la prevalencia es muy baja se la asocia con el concepto de *especificidad*. Mientras que en el otro extremo de la escala, cuando la prevalencia es muy alta se la asocia con el concepto de *sensibilidad*. Es decir, que sensibilidad y especificidad no son otra cosa que la eficiencia de una prueba clínica, pero expresada en sus dos casos extremos.

Para entender estas ideas se puede usar la noción de *relatividad* tal como se estudia en Física. Por ejemplo: la diferencia conceptual entre masa y peso, es que la primera es una cualidad *intrínseca* del cuerpo que no depende del sistema de referencia en el cual se halla. En cambio el peso varía con la influencia de la atracción de la gravedad del planeta en donde se encuentre el

cuerpo. Se dice por ese motivo que el peso es una cualidad *extrínseca* del cuerpo. Expresado en otros términos, a la *magnitud* masa le corresponde una variable llamada peso. Este es un *parámetro* en el planeta Tierra porque todo individuo está afectado por la misma fuerza de gravedad, pero es una *variable* dentro del sistema solar.

Análogamente, si se supone que la capacidad de un método clínico es constante en toda la población humana, entonces la sensibilidad y especificidad se mantendrán fijas para todos los individuos, por lo que pueden ser considerados como cualidades *intrínsecas* del método. Y lo mismo ocurre con el Índice de Youden. En cambio la eficiencia varía (cualidad *extrínseca*) según el sistema de referencia o población en donde se usa el método, de acuerdo a la *prevalencia* de la enfermedad existente en la población. Es decir, sensibilidad y especificidad son los principales parámetros de un método clínico mientras que la eficiencia es una variable.

Si se considera a un método de diagnóstico, o bien a un test clínico, como si fuese una máquina para hacer predicciones, entonces se puede explicar este concepto con otro símil: Un Ferrari es un excelente auto para correr en un premio de Fórmula I, pero a pesar de su alta calidad puede que no sea la mejor opción para correr en Indianápolis, como también puede llegar a ser un desastre para competir en las arenas del desierto en el Rally de Dakar.

4.4 Relatividad de los tests clínicos.

Todo esto muestra que la eficiencia debe ser informada al lector con una curva, en lugar de usar un solo valor, como se acostumbra hacer en la práctica actual. Para obtener estas curvas fácilmente se puede usar un algoritmo desarrollado para computadoras en Excel, disponible en la página web de la cátedra. Pero también se lo puede obtener con un procedimiento que se muestra a continuación:

Ejemplo: Sea un test clínico que tiene una calidad cuantificada por $S = 0,9$ y $E = 0,75$. Se desea averiguar el efecto de la relatividad en la eficiencia.

Como S y E son los parámetros del método, se puede obtener el otro parámetro con $IY = 0,65$. Y para ver como afecta la prevalencia de la enfermedad a la eficiencia, se pueden simular diferentes valores de la misma tomando varios valores de p , como ser $p = 0,1; 0,2; 0,3; \dots; 0,9$ y un tamaño muestral cualquiera como por ejemplo $N = 1000$.

Paso 1: Para la primera simulación se adopta el valor $p = 0,1$, es decir 1 enfermo cada 10 casos, y se calcula el número de verdaderos positivos y negativos:

Como $S = vp / TE$ resulta $vp = S \cdot TE = S \cdot N \cdot p = 0,9 \cdot 1000 \cdot 0,1 = 90$

Como $E = vn / TN$ resulta $vn = E \cdot TN = E (N - TE) = E (1 - p) N = 0,75 (1 - 0,1) 1000 = 675$

Paso 2: Se calculan ahora los valores de falsos negativos y positivos con:

$TE = N \cdot p = 100 = vp + fn$; o sea $fn = 100 - vp = 100 - 90 = 10$

$TS = N (1 - p) = 900 = vn + fp$; o sea $fp = 900 - vn = 900 - 675 = 225$

Paso 3: Con los cuatro valores obtenidos en las dos etapas anteriores se puede armar la Tabla de diagnóstico para el caso de una prevalencia del 10% y así calcular el total de positivos y negativos como se muestra a continuación:

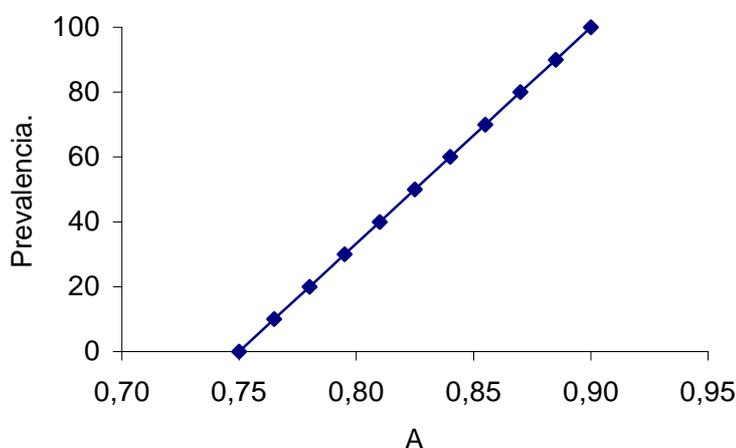
Test	Resultado verificado		Total
	Enfermo	Sano	
(+)	90	225	TP = 315
(-)	10	675	TN = 685
Total	TE = 100	TS = 900	1000

Paso 4: Se calcula la eficiencia para estos valores de la Tabla completa:

$$A = (vp + vn) / N = (90 + 675) / 1000 = 0,765$$

Y se tiene estimado el primer valor de la eficiencia para una prevalencia del 10%. Análogamente, se toman diferentes valores de la prevalencia y se repiten los 4 pasos anteriores para obtener los respectivos valores de la eficiencia, resultando la curva siguiente:

Gráfico 4.1: Variabilidad de la eficiencia con la prevalencia (S = 0,9 y E = 0,75).



En el Gráfico 4.1 se muestra un método de diagnóstico con los parámetros S = 0,9 y E = 0,75. En una población donde la enfermedad es muy rara, la prevalencia será muy baja tal como es el caso de la meningitis, y la eficiencia será muy parecida al valor de la especificidad (si $p \rightarrow 0$ entonces $A \rightarrow E$). En cambio, para otra población donde la enfermedad es muy frecuente, como puede ser una epidemia, la eficiencia del mismo método clínico cambia bastante y se acerca a la sensibilidad (si $p \rightarrow 1$ entonces $A \rightarrow S$). Este ejemplo muestra que si se mide la capacidad de un método de diagnóstico, con el índice de la eficiencia (A), este valor será diferente en poblaciones diferentes. Y por lo tanto, no es aconsejable seguir ese criterio para informar la calidad o poder discriminatorio del método como se recomienda usualmente en alguna bibliografía. O por lo menos, no usar un único valor para presentar el dato, sino que se debe mostrar la curva para que el lector sepa como trabajará ese método en su propia población de referencia. Notar que esta curva no depende del tamaño de la muestra (N) usada para determinarla.

4.5 Valores Predictivos.

Otra manera de analizar la capacidad de un método es analizando el porcentaje de éxitos logrados sobre la base de sus pronósticos. La idea es estudiar que tan bien predice el método. Otra vez se trata de una dualidad, porque hay que ver los dos aspectos principales: como trabaja en los casos positivos y como lo hace en los casos negativos. Se puede definir entonces:

- *Valor Predictivo de Positivos:* Es la capacidad del método para identificar correctamente a un caso positivo, de entre todos los resultados positivos obtenidos.
- *Valor Predictivo de Negativos:* Es la capacidad del método para identificar correctamente a un caso negativo, de entre todos los resultados negativos obtenidos.

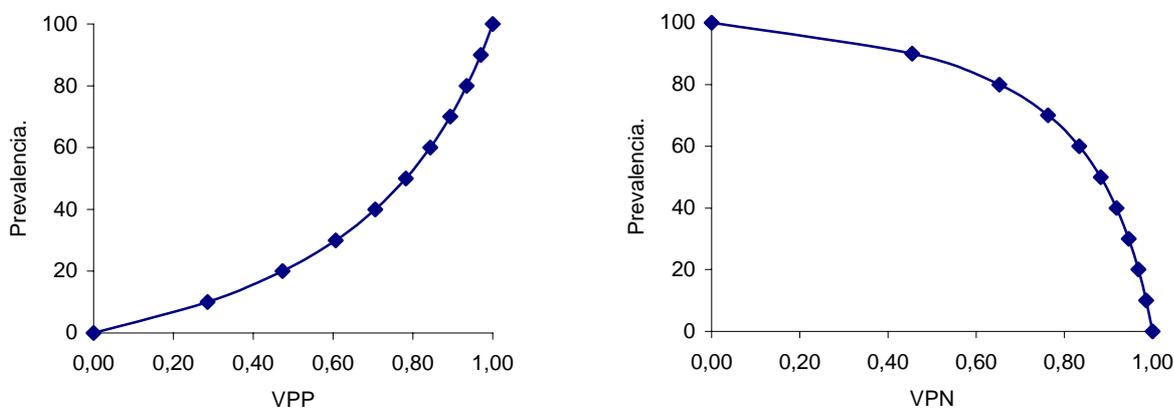
Por ejemplo, si el método es capaz de detectar a cuatro casos verdaderos de positivos cada cinco resultados positivos encontrados, entonces su capacidad para identificar correctamente a los que pronostica como positivo es del 80%. Aquí se dice que su Valor Predictivo de Positivos es de $VPP = 0,8$. Por otro lado, si el método es capaz de detectar a nueve casos verdaderos de negativos, cada diez casos que resultaron negativos, entonces su capacidad para identificar correctamente como negativo es del 90%. Se dice aquí que su Valor Predictivo de Negativos resulta: $VPN = 0,9$. Aplicando estos conceptos al caso general presentado en la Tabla de diagnóstico, se pueden deducir las fórmulas que siguen:

Valor Predictivo de Positivos: $VPP = vp / TP = (\text{verdaderos positivos}) / (\text{total de positivos})$

Valor Predictivo de Negativos: $VPN = vn / TN = (\text{verdaderos negativos}) / (\text{total de negativos})$

La relación matemática entre estos índices relativos, con los dos índices básicos S y E se expresa con el Teorema de Bayes, que se explicará más adelante. Por ahora, se puede visualizar esta idea con el gráfico siguiente, obtenido de la misma manera que la explicada para la eficiencia, pero ahora se calculan además los valores predictivos:

Gráfico 4.2: Variabilidad de los valores predictivos con la prevalencia (S = 0,9 y E = 0,75)



En este ejemplo se analiza la variabilidad de los valores predictivos con la prevalencia, para un método clínico de parámetros $S = 0,9$ y $E = 0,75$. Por ejemplo, para una prevalencia del 10% resultan: $VPP = vp / TP = 90 / 315 = 0,286$ y $VPN = vn / TN = 675 / 685 = 0,985$

Mientras los VPP varían desde cero a uno, a medida que aumenta la prevalencia, los VPN lo hacen de forma inversa. Si la prevalencia aumenta, aumentan los valores predictivos de positivos y bajan los negativos. Cuando el $VPP = 1$, significa que $vp = TP$, es decir no hay falsos positivos, la condición ideal para una enfermedad del Tipo II. En cambio, si $VPP = 0$, esto implica que no hay aciertos en positivos y todos los pronósticos de esa índole tienden a ser falsos. Notar que esto ocurre cuando la enfermedad es rara y su prevalencia muy baja. Por otra parte, cuando $VPN = 1$, significa que $vn = TN$, o sea que no hay falsos negativos, la condición buscada para las enfermedades del Tipo I. Viendo la gráfica anterior, se puede observar que cuando la prevalencia de la enfermedad sea muy alta (epidemias, unidades coronarias, etc.), el $VPN = 0$ y que casi no van a existir aciertos en negativos. En cambio, si la enfermedad es muy rara, casi todos los resultados negativos serán falsos.

Ejemplo 1: Se ha realizado un estudio para ver la capacidad de un test clínico que se usa habitualmente en el laboratorio. Las historias clínicas de los pacientes se separan en dos grupos. En el primero se agrupan los casos verificados de enfermedad “a posteriori”, es decir se hace un tipo de estudio retrospectivo una vez que se está seguro del estado real del paciente. Así se obtiene el número de casos identificados correctamente como positivos (vp) y las falsas predicciones (fp). En el segundo conjunto se agrupan los casos verificados de ausencia de la enfermedad, y con el estudio retrospectivo se obtienen los números de casos verdaderos negativos y falsos negativos (vn y fn). Los casos dudosos se descartan para eliminar variabilidad indeseada. Se seleccionaron al azar 400 casos de enfermedad y otros tantos de no-enfermedad. Los resultados obtenidos se muestran en la tabla. Se desean conocer los índices clínicos correspondientes a este test.

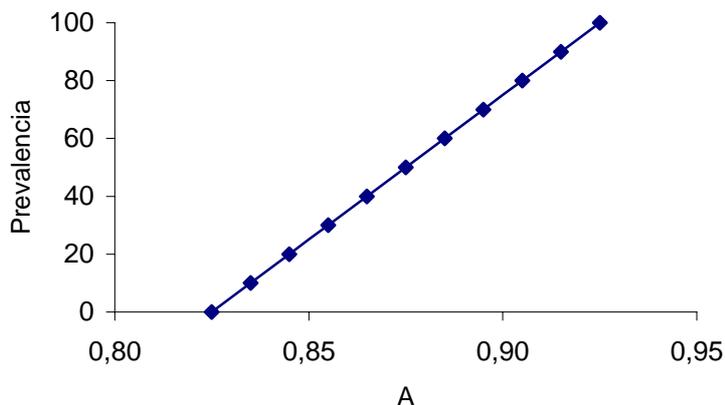
Test	Resultado verificado		Total
	Enfermo	Sano	
(+)	370	70	440
(-)	30	330	360
Total	400	400	800

Lo primero es calcular los parámetros principales del método:

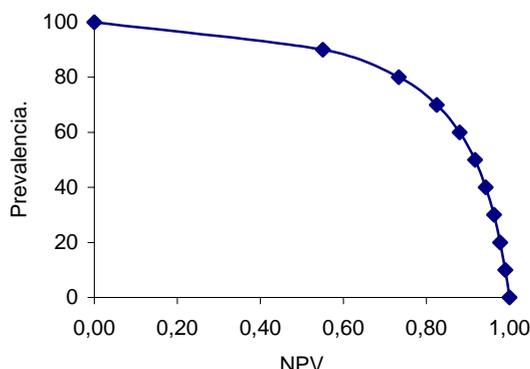
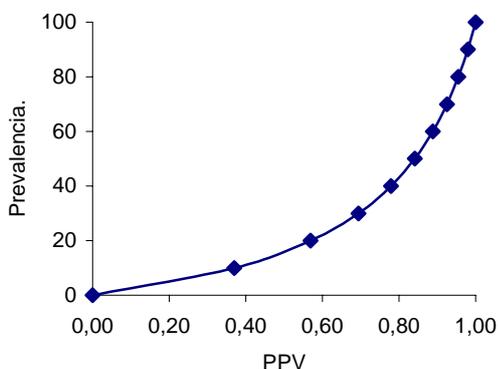
$$S = 370 / 400 = 0,925 \quad \text{y} \quad E = 330 / 400 = 0,825$$

$$\text{Con estos valores se obtiene } Y = 0,925 + 0,825 - 1 = 0,75$$

Y para informar la eficiencia y los valores predictivos se muestran las curvas calculadas con el método que se vio en el ejemplo anterior. La eficiencia (A) varía como se grafica a continuación:



Los valores predictivos varían con la prevalencia como se ve más abajo:



Ejemplo 2: De la sumatoria total de casos encontrados por los investigadores del ejemplo anterior se determina que la prevalencia de la enfermedad para ese laboratorio es $p = 0,4$ (solo se usan los casos totales verificados). Se pide calcular como son los valores relativos para ese laboratorio en particular, de acuerdo a su población de referencia.

De las curvas anteriores se obtiene: $A = 0,865$; $VPP = 0,779$ y $VPN = 0,943$. Esto significa que el porcentaje de éxitos de ese test, en esa población, será del 86,5%. Además predecirá correctamente a los positivos en un 77,9% de los casos, y a los negativos en el 94,3% de los casos.

4.6 Likelihood Ratios.

Habitualmente, estos índices suelen informarse en los trabajos de investigación junto a los parámetros básicos. Se trata de un par de valores que no varían con la prevalencia de la población, sino que son características propias del método clínico. Su uso se relaciona con las probabilidades “a posteriori” del test como se explicará más adelante. Por ahora, basta saber como se los define:

- *Likelihood Ratio de Positivos:* $LR+ = S / (1 - E)$

- *Likelihood Ratio de Negativos*: $LR- = (1 - S) / E$

Por ejemplo, para el caso visto más arriba donde $S = 0,925$ y $E = 0,825$ se pueden obtener:

$LR+ = 5,286$ y $LR- = 0,091$

A mediados de la década del ochenta, se los comienza a emplear para facilitar el cálculo de la compleja fórmula de Bayes en los trabajos de Medicina. Un ejemplo desarrollado por Simel y publicado por The Lancet en 1985 muestra como determinar si una prueba es o no necesaria para mejorar el diagnóstico. Los LR no son una probabilidad, ni dependen de la prevalencia, es decir *son magnitudes o valores intrínsecos* de la prueba clínica. Sin embargo, se calculan como un cociente de probabilidades. El LR+ es el cociente entre la probabilidad de predecir correctamente la enfermedad y la probabilidad de predecir incorrectamente a dicha enfermedad. Da una indicación de cuanto puede aumentar o bajar la probabilidad del resultado del método de diagnóstico, antes de efectuarlo (probabilidad pre-test). Cuando el LR+ = 1 indica que si se le realiza el test al paciente, no se obtendrá información adicional y por lo tanto lo mejor es no someter al paciente a gastos inútiles. En cambio cuando LR+ > 1 indica que es conveniente efectuar el test porque se obtendrá información adicional. Pero si LR+ < 1 entonces el test no incrementa la probabilidad de asesorar bien al clínico, sino que disminuye. Para profundizar estos conceptos hay que tener una noción de probabilidades, cosa que se hará más adelante.

4.7 Influencia de los puntos de corte.

Hasta ahora se vio el caso de resultados dicotómicos del método de diagnóstico que es el caso más general. Sin embargo, muchas mediciones clínicas se hacen con otro tipo de magnitudes como las continuas. La solución para estos casos es transformar una magnitud cualquiera en una binaria adoptando un *punto de corte* adecuado. Por ejemplo, el peso es una magnitud continua y si se toma el valor 70 kg para diferenciar a gordos de flacos, entonces cuando un individuo pesa hasta 70 kg se lo considera flaco y en caso contrario gordo. Este valor de 70 kg es el punto de corte adoptado para separar dos zonas bien definidas en la variable continua. El problema es que no existe en los casos reales, un valor que permita discriminar claramente a los sanos de los enfermos en magnitudes continuas. Se encuentran tres zonas: una para los sanos, una intermedia donde los casos son dudosos y la tercera donde los resultados indican claramente la presencia de la enfermedad. Para ilustrar estos conceptos se analizará el caso de infartos de miocardio del capítulo anterior, en los Cuadros 3.1 y su resumen en el Cuadro 3.3. Pero antes de proseguir hay que hacer notar una diferencia conceptual importante:

Sensibilidad, especificidad y sus índices derivados no varían con la prevalencia, pero sí lo hacen con el punto de corte que se adopta en magnitudes continuas. Los otros índices variables como los valores predictivos y la eficiencia, pueden variar por ambas cosas.

Por ejemplo, tomando el caso de los pacientes con supuesto infarto de miocardio visto en el Cuadro 3.3, y definiendo como punto de corte un valor de CPK de 20 UI/l, entonces cuando un paciente tenga un valor menor o igual de CPK se lo supondrá sano (-), o sea una falsa alarma. En cambio, si el valor medido es mayor se lo internará de inmediato en la unidad coronaria, pues se lo supondrá infartado, es decir como enfermo (+). Así, de los infinitos valores posibles que puede

arrojar la medición de una variable continua, al adoptar un punto de corte, este delimitará dos zonas, “dicotomizando” la variable en cuestión. Analizando los histogramas del Gráfico 3.12, se puede ver que muchos puntos de corte arrojarán falsos diagnósticos. Puede verse esta superposición entre 20 y 120 UI/l, creando una zona de incertidumbre muy amplia. Adoptando como punto de corte 20 no habrá ningún caso (-) que sea falso. Viceversa, si se adopta 120 como punto de corte entonces no habrá ningún falso (+). En cambio, una técnica clínica ideal, separaría ambas zonas sin problemas y así no se cometerían equivocaciones. Pero, en la actualidad es muy raro el disponer de una prueba que de certidumbre total. Aún una biopsia puede fallar. El desafío es encontrar una prueba lo más cercana posible a la ideal, llamada el “golden standard”. Los investigadores están probando combinaciones de dos o más pruebas usando diferentes modelos en serie o paralelo, el Análisis Multivariado, etc.

El problema es encontrar un criterio, basado en la ética, que le permita al clínico elegir el mejor punto de corte posible. Luego de adoptarlo, la variable continua se transforma en una binaria y se le puede aplicar todo lo visto en los puntos anteriores. Para definir un criterio conviene analizar *la variabilidad de los índices de acuerdo al punto de corte* adoptado en la magnitud clínica. En el Gráfico 3.12 se observa que, tomando a 120 UI/l como punto de corte, no habrá falsos positivos. En cambio, tomando a 20 UI/l como valor referente, entonces $fn = 0$. Adoptando diferentes puntos de corte se puede deducir que si se aumentan los fp , bajan los fn y viceversa. La pregunta más difícil de contestar es: *¿Cuál será el valor óptimo a elegir para poder discriminar entre sanos y enfermos?* Hay varias respuestas. Si se trata de una enfermedad mortal, en caso de no ser atendida a tiempo, salta a la vista que lo más importante es tener $fn = 0$, como los infartos, ciertos tipos de tumores, etc. En cambio, si se trata de una enfermedad terminal, sin cura posible, puede resultar muy grave informarle al paciente que la padece cuando en realidad está sano. Aquí se requiere de $fp = 0$, como el caso un cáncer terminal, metástasis de ciertos tumores, etc. Por otro lado, hay muchos casos donde esta decisión no es tan sencilla como por ejemplo en el caso del SIDA: desde el punto de vista del paciente lo peor es un fp por los temores y angustias que le producirá creerse enfermo cuando en realidad está sano, pero desde el punto de vista de la sociedad lo peor es un fn porque al no tomar las prevenciones del caso la enfermedad se propagará en más personas. Como antes, para *decidir* no hay que perder de vista el tipo de enfermedad que se está analizando.

En el caso visto de mediciones de CPK se tienen los archivos de pacientes que llegan a una unidad coronaria con un diagnóstico presunto de infarto, el cual fue confirmado o rechazado a posteriori. Es decir, de todas las historias clínicas de esa unidad se descartan los casos dudosos y solo se usan aquellos confirmados en dos clases: los que tuvieron un infarto y los que no lo tuvieron pues fue una falsa alarma. De entre todos los casos disponibles se eligen al azar una cierta cantidad N lo más grande posible (por lo menos 50 casos), clasificadas un 50% para los de enfermedad confirmada y el otro 50% para los sanos. En el ejemplo visto se tomaron 400 pacientes, 200 con infarto confirmado y 200 sin infarto. Para poder estudiar la influencia del punto de corte, que se adopte para la medición de la CPK se emplearán los datos del Cuadro 3.1 y las frecuencias del Cuadro 3.3. Como se puede clasificar a los infartos como una enfermedad del Tipo I, se deduce que el principal criterio a tener en cuenta va a ser la sensibilidad. La estrategia será adoptar un punto de corte que maximice dicho índice, para evitar la mayor cantidad de falsos negativos posibles. Sin embargo, si se tratase de otra enfermedad habría que maximizar la especificidad para las enfermedades del Tipo II, o el índice de Youden para las del Tipo III. Para este ejemplo, en la Tabla 4.1 siguiente, se toman a los límites superiores de cada intervalo de clase como puntos de corte y así se pueden efectuar los cálculos como se muestra a continuación:

Tabla 4.1: Obtención de los diagnósticos posibles con datos del Cuadro 3.1 y 3.3

Clases	Sin Infarto		Con Infarto		Puntos de Corte	TN	TP	vn	fn	vp	fp
	Fr.	Fr. ac.	Fr.	Fr. ac.							
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
0-20	104	104	0	0	20	104	296	104	0	200	96
21-40	36	140	8	8	40	148	252	140	8	192	60
41-60	28	168	16	24	60	192	208	168	24	176	32
61-80	18	186	32	56	80	242	158	186	56	144	14
81-100	10	196	46	102	100	298	102	196	102	98	4
101-120	4	200	66	168	120	368	32	200	168	32	0
121-140	0	200	23	191	140	391	9	200	191	9	0
141-160	0	200	6	197	160	397	3	200	197	3	0
161-180	0	200	3	200	180	400	0	200	200	0	0

El procedimiento para obtener los valores de la Tabla 4.1 es como sigue:

Paso 1 : En la primer columna (1) se colocan las clases adoptadas con un ancho de 20 UI/l.

Paso 2 : En las columnas (2) y (4) se ponen las frecuencias obtenidas del Cuadro 3.3.

Paso 3 : Se calculan las frecuencias acumuladas para ambos casos y se colocan en (3) y (5).

Paso 4 : En la columna (6) se colocan los diferentes puntos de corte elegidos.

Paso 5 : En la columna (7) se calcula el TN como la suma de todos los datos menores que el punto de corte adoptado. O sea, la suma de las frecuencias acumuladas correspondientes. En este caso resulta la suma de las columnas (3) y (5).

Paso 6 : En la columna (8) se obtiene el TP como la diferencia entre el total de datos y TN de la columna (7). O sea $TP = N - TN$.

Paso 7 : En la columna (9) se calculan los vn como el número de pacientes cuyo diagnóstico resulta (-) por el punto de corte adoptado, y que realmente son no infartados. Entonces, será el valor de la frecuencia acumulada de los no infartados. Así resulta: columna (3) = columna (9).

Paso 8 : En la columna (10) se obtiene el valor de los fn con los datos de la columna (5). Pues todos los valores menores que el punto de corte corresponden a la frecuencia acumulada de los infartados, estos serán diagnosticados como (-) cuando en realidad tienen un infarto.

Paso 9 : En la columna (11) se usa la relación $vp = TE - fn$. O sea, $vp = 200 - fn$. Esto es: columna (11) = 200 - columna (10).

Paso 10 : En la columna (12) se usa la relación $fp = TS - vn = 200 - tn$. Otra forma de obtener estos valores es usando: $fp = TP - vp$. Así, columna (12) = 200 - columna (9).

Se puede notar que, para cada punto de corte adoptado, se puede formar una Tabla de diagnóstico. Analizando los resultados de la Tabla 4.1, puede verse que los fn son nulos sólo cuando el punto de corte se adopta en 20 UI/l. Esto es, a todo paciente que ingrese al servicio con

una CPK mayor que 20 UI/l se lo internará porque se sospecha que está infartado. Naturalmente, esto arrojará el número más alto de $fp = 96$ pacientes internados sin que sea necesario. A medida que se aumenta el valor del punto de corte, la situación se invierte. Así van aumentando los fn mientras disminuyen los fp , hasta el punto 120 UI/l; allí resulta $fp = 0$. O sea, no habrá pacientes internados sin necesidad, pero la contrapartida, son los $fn = 32$ pacientes diagnosticados como sanos, o peor aún, dados de alta, a pesar de su infarto. A la luz de esta información, se puede contestar la pregunta planteada anteriormente: ¿Cuál es el mejor punto de corte a ser adoptado desde el punto de vista ético ?

Para el caso de infartos de miocardio (Tipo I) el valor óptimo es aquel que reduce al mínimo la posibilidad de un fn . (para este ejemplo se debe tomar como punto de corte 20 UI/l).

Claro que si al responsable del servicio, no le interesa minimizar el número posible de muertos por falta de atención, sino reducir sus costos evitando las falsas alarmas, entonces para una persona así el valor óptimo sería de 120 UI/l, y sólo internaría a pacientes cuyas CPK sean superiores a tal valor. Pero, si al responsable lo que más le interesa es no cometer errores, por una cuestión de imagen, mejorará su rendimiento adoptando como punto de corte 60 UI/l, pues allí la suma de fp y fn es mínima. La adopción de un punto de corte está influenciada por la subjetividad de las personas y por los condicionantes externos producto de las presiones laborales y económicas de la cuestión. *El deber es siempre usar la ética como criterio básico de decisión.*

De todos los índices propuestos hay dos básicos de los cuales derivan todos los demás, que son la Sensibilidad y la Especificidad. *Conviene tener la máxima Sensibilidad en aquellas enfermedades del Tipo I.* Es decir, en enfermedades de consecuencias graves si no se las detecta a tiempo. Además del infarto de miocardio, el caso más común son las enfermedades venéreas. La sífilis en su fase IV ya no tiene remisión, pero en una etapa más temprana igual puede tener secuelas nefastas, como la sífilis congénita que transmite la embarazada, la neurosífilis y la sífilis cardiovascular. En las fases tempranas hay mucha dificultad para detectarla cuando aún no se observan reacciones epidérmicas y es fundamental usar una técnica con el mínimo fn posible. Por lo tanto, si se está usando la VDRL para detectarla y aparece una nueva técnica, el profesional debe optar por aquella con máxima sensibilidad entre ambas. Hay otros casos como el feocromocitoma, fatal si no se lo detecta a tiempo, pero curable con tratamiento; el hipotiroidismo, la toxoplasmosis en especial entre las embarazadas, el cáncer de mama o de útero, etc.

Conviene tener máxima especificidad en las enfermedades del Tipo II. Este caso se busca cuando es muy grave informarle al paciente que está enfermo, si en realidad está sano. Esto le podría ocasionar serios daños en lo psicológico, moral o económico. Por ejemplo, el caso de un cáncer incurable, la esclerosis múltiple, etc. Cuando una enfermedad es muy grave y entró en una etapa de no-remisión, se requiere mucha especificidad para minimizar los falsos positivos.

Cuando la magnitud clínica es continua, la especificidad varía con el punto de corte adoptado por el clínico. En la Tabla 4.2 se ilustra esta situación para el caso de los pacientes con diagnóstico presunto de infarto de miocardio. La especificidad es mínima cuando el corte adoptado es 20 UI/l y máxima luego de los 120 UI/l. Análogamente para el índice de Youden, este será máximo cuando se igualen sensibilidad y especificidad, como es en el punto de corte 60 UI/l. Los demás índices pueden obtenerse a partir de los anteriores, y de acuerdo al punto de corte adoptado, como se puede apreciar en la tabla siguiente:

Tabla 4.2: Obtención de los Índices Clínicos (continuación de la Tabla 4.1).

Clases	Puntos de corte	S %	E %	Y	A	VPP %	VPN %	LR+	LR-
1	6	13	14	15	16	17	18	19	20
0-20	20	100	52	0,52	0,76	67,57	100	2,08	0
21-40	40	96	70	0,66	0,83	76,19	94,59	3,2	0,057
41-60	60	88	84	0,72	0,86	84,62	87,5	5,5	0,143
61-80	80	72	93	0,65	0,825	91,14	76,86	10,29	0,301
81-100	100	49	98	0,47	0,735	96,08	65,77	24,5	0,52
101-120	120	16	100	0,16	0,58	100	54,35	∞	0,84
121-140	140	4,5	100	0,045	0,523	100	51,15	∞	0,955
141-160	160	1,5	100	0,015	0,508	100	50,38	∞	0,985
161-180	180	0	100	0	0,5	***	50	∞	1

Paso 11 : Para calcular la sensibilidad S en forma porcentual de la columna (13), se emplea la relación: $S = (vp / TE) 100$ y en este caso resulta ser $S = vp/2$, pues $TE = 200$.

Paso 12 : Para calcular la especificidad E en forma porcentual de la columna (14), se aplica la relación: $E = (vn / TS) 100$ y en este caso resulta ser $E = vn/2$, pues $TS = 200$.

Paso 13 : El Índice de Youden de la columna (15) se obtiene con: $Y \% = [S + E - 1] \cdot 100$. O sea se suman las columnas (13) y (14) y se le resta 100.

Paso 14 : La Eficiencia de la columna (16) se la calcula con la relación $A = (vp + vn) / N$ O sea, el número total de diagnósticos acertados sobre el número total de casos.

Paso 15 : En la columna (17) se calcula el Valor Predictivo de Positivos en forma porcentual con: $VPP = (vp / TP) 100$

Paso 16 : En la columna (18) se saca el Valor Predictivo de Negativos en forma porcentual con: $VPN = (vn / TN) 100$

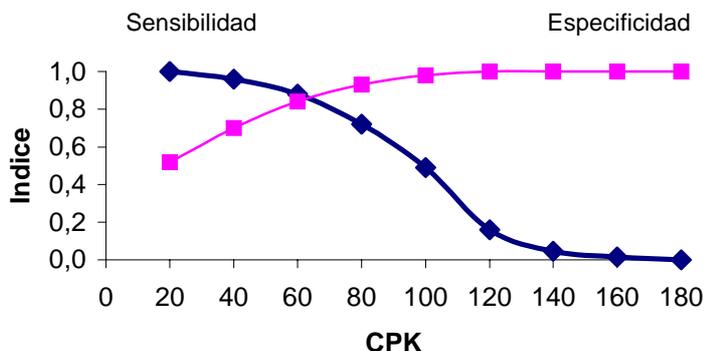
Paso 17 : En la columna (19) se saca el Likelihood Ratio de la enfermedad dado un resultado positivo con: $LR+ = S / (1 - E)$ usando los valores respectivos de las columnas (13) y (14).

Paso 18 : En la columna (20) se saca el Likelihood Ratio de la enfermedad dado un resultado negativo con: $LR- = (1 - S) / E$ usando los valores respectivos de las columnas (13) y (14).

Se pueden graficar los valores obtenidos, para ilustrar acerca de la variabilidad de los índices clínicos con el punto de corte. Notar que, todos los índices varían con el punto de corte adoptado, no así con la prevalencia de la enfermedad que es otra de las causas de variabilidad. Todo esto indica la dificultad de tener situaciones estables en la práctica. El supuesto principal es que la sensibilidad y la especificidad no varían con la prevalencia. Sin embargo se puede dar el caso de encontrar una variación en ambas y las explicaciones más comunes son: (a) el método no es estable. (b) el avance de la enfermedad en los pacientes (se supone que cuanto más avanzada esté la enfermedad, hay menores posibilidades de equivocarse en el diagnóstico).

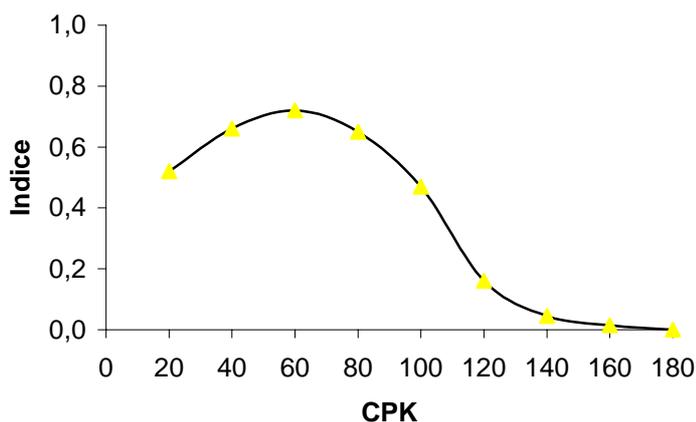
Gráfico 4.3: Variación de Sensibilidad y Especificidad con el punto de corte en la CPK.

El hecho más notable es que *si la sensibilidad aumenta, la especificidad disminuye y viceversa*.



Por eso, el problema real es decidir cual de ellas debe ser optimizada en cada caso, adoptando un punto de corte apropiado al tipo de enfermedad estudiada. Si la enfermedad es del Tipo I conviene elegir el punto de corte que maximiza la Sensibilidad ($X = 20$ UI/l). Pero si la enfermedad es del Tipo II conviene elegir al punto que maximiza la Especificidad ($X \geq 120$ UI/l). Cuando la enfermedad sea del Tipo III conviene maximizar el índice de Youden, como se puede ver en el Gráfico 4.5 siguiente. Notar que esto ocurre para el valor $X \approx 60$ UI/l de la magnitud clínica, que es el punto donde se cortan las curvas de sensibilidad y especificidad aproximadamente.

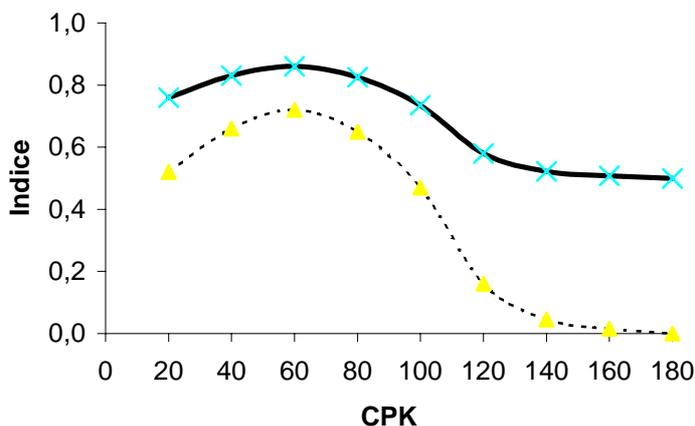
Gráfico 4.4: Variación del Índice de Youden con el punto de corte en la CPK



Esto muestra otra de las ventajas del índice de Youden. Advierte el peligro de tomar puntos de corte tales que el diagnóstico no se diferencie del azar. Por ejemplo, si se toma como punto de corte el valor 180 UI/l, en la Tabla 4.2, se obtiene $E = 1$ y $S = 0$, o sea un $Y = 0$ y el diagnóstico resultará como al azar, esto es como si se utilizase el lanzar una moneda al aire para efectuar el diagnóstico. En resumen, este índice debe acompañar siempre el análisis de sensibilidad y especificidad de los diagnósticos basados en una prueba clínica de laboratorio.

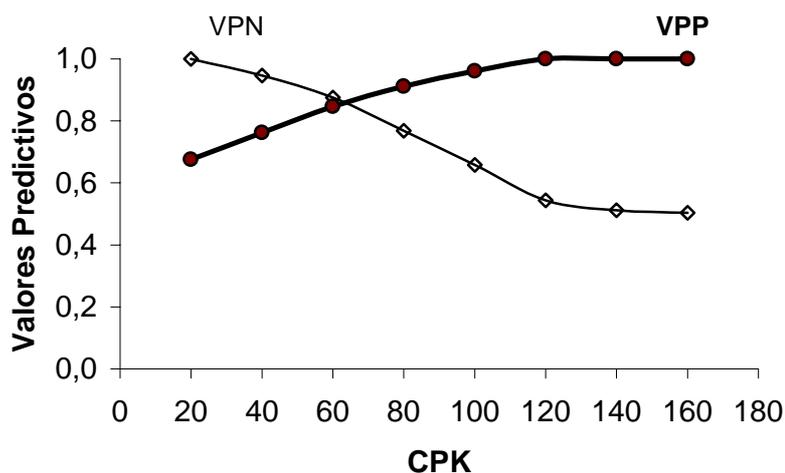
En el Gráfico 4.5 se muestra la variabilidad de la eficiencia con el punto de corte. Además se ha vuelto a mostrar la variabilidad del Índice de Youden en línea punteada, para visualizar lo similar de ambas curvas en su forma. En ambos casos el punto más aconsejable para una enfermedad del Tipo III es donde se maximizan ambas curvas ($X \approx 60$ UI/l).

Gráfico 4.5: Variación de la eficiencia (A) con el punto de corte en la CPK



Para ilustrar la variabilidad de los valores predictivos de acuerdo al punto de corte adoptado por el clínico, se presenta el Gráfico 4.6 donde se puede ver que: (a) el punto de corte ideal para una enfermedad del Tipo I corresponde al valor máximo del VPN ($X = 20$ UI/l), al igual que ocurrió en el caso de la sensibilidad. (b) cuando $X \geq 120$ UI/l el VPP se maximiza y corresponde al punto ideal para las enfermedades del Tipo II, como en el caso de la especificidad. (c) ambas curvas se cortan en $X \approx 60$ UI/l que es el punto ideal para las del Tipo III.

Gráfico 4.6: Variación de los valores predictivos con el punto de corte en la CPK



En resumen:

- 1) Sensibilidad, Especificidad, Índice de Youden y Likelihood Ratios varían con el punto de corte, pero no lo hacen con la prevalencia de la población (son parámetros del método clínico).
- 2) Eficiencia y Valores Predictivos varían con el punto de corte y con la prevalencia (no son parámetros sino variables del método clínico).

4.8 Otros Índices diagnósticos

Además de los índices ya vistos, en la bibliografía aparecen algunos otros que no agregan mucho en el ámbito conceptual; son variantes lógicas derivadas de los anteriores. Por ejemplo, en lugar de calcular el porcentaje de aciertos con la eficiencia, también se puede definir su complemento como:

$$\text{Porcentaje de errores} : \quad PE \% = \{ (fn + fp) / N \} \cdot 100$$

$$PE \% + A \% = 100\%$$

Se busca minimizar a este porcentaje para acercarse a la técnica ideal para diagnosticar. A veces, también se lo llama *error combinado*. Otro índice es el llamado *Positivo presunto*: se calcula como el porcentaje de pruebas, cuyo resultado dio positivo, respecto del total N de pruebas realizadas, es decir:

$$\text{Positivo presunto} : \quad PP\% = (TP / N) \cdot 100$$

Otra forma parecida es calcular el porcentaje de aciertos en positividad definido con :

$$\text{Índice de detección} : \quad ID = (vp / N) \cdot 100$$

La proporción de error se define como el cociente entre el número total de errores y los diagnosticados como enfermos correctamente, esto es :

$$\text{Proporción de error} : \quad P\%E = \{ (fp + fn) / TN \} \cdot 100$$

Hay confusiones en la bibliografía, acerca de los términos: *Incidencia*, *Predominancia* y *Morbilidad*. Para acotar este problema se toma el aporte de Galen definiendo:

Incidencia : Es la cantidad de enfermos que aparecen en un período dado de tiempo (un año), en la población, sobre el tamaño de la población expresado cada 100.000 habitantes. Es una medida epidemiológica tomada en períodos fijos de tiempo.

Predominancia : Es la cantidad de enfermos, en un momento dado de la población, dividido el número total de habitantes y expresado cada 100.000. O sea, es una medida de tipo censal.

Morbilidad : Se usa en Demografía, y puede adaptarse a los dos casos anteriores, pero sin usar un nombre específico que los distinga.

La predominancia o prevalencia de una enfermedad en la población afecta grandemente a los Valores Predictivos y sus índices relacionados, pero no lo hacen con la Sensibilidad, la Especificidad y el Índice de Youden. Por lo tanto, es muy importante tener en cuenta este valor a la hora de seleccionar un punto de corte buscando optimizar el diagnóstico. Se podrían seguir agregando otras proporciones e incrementando la lista de índices clínicos, más allá de los vistos. Sin embargo, a modo de resumen se presentan los cuatro postulados de Galen y Gambino, para la elección práctica de las pruebas clínicas en el diagnóstico:

(1) *Se elige alta Sensibilidad cuando las enfermedades son del Tipo I*

- La enfermedad puede curarse si es detectada a tiempo.
- La enfermedad es grave y no puede dejarse pasar inadvertida.
- La enfermedad es tratable.
- Los fp no trauman ni psicológicamente, ni económicamente al paciente.
- Ejemplos: Cáncer de útero o de mama, enfermedades venéreas y otras enfermedades infecciosas curables, la fenilcetonuria, el feocromocitoma

(2) *Se elige alta Especificidad cuando las enfermedades son del Tipo II*

- La enfermedad es grave, pero difícilmente curable o sin remisión.
- Es muy importante para el paciente o para la población él saberse un vn.
- Los fp trauman seriamente al paciente.
- Ejemplos: Cáncer oculto, esclerosis en placas, etc.

(3) *Se elige un alto Índice de Youden cuando las enfermedades son del Tipo III :*

- Tanto los fp como los fn, sean igualmente graves para el paciente o para la población.
- Ejemplo: SIDA, lupus eritomatoso, ciertas formas de leucemia o linfoma, diabetes, etc.

4.9 Índices de riesgo o daño.

En el primer capítulo se explicaron las cuatro cuestiones clínicas fundamentales que son: terapia, daños, diagnóstico y pronosis. Para tratar la primera cuestión de la terapia, se mostró el diseño de los RCT para hacer investigaciones. Para tratar la segunda cuestión se usan las tablas diagnóstica y los índices detallados en los puntos anteriores. Para la tercera y cuarta cuestión se presentaron los diferentes diseños de investigación del riesgo o daño, que se explicaron más detalladamente en el segundo capítulo (estudios por cohortes, caso-control y RCT). La forma más sencilla de presentar los datos es con una tabla doble dicotómica llamada tabla de riesgo (ver Tabla 1.2) donde se generaliza el cuadro para los diferentes tipos de diseño experimental para el estudio del riesgo:

Factor de riesgo	Resultados observados		Total
	Enfermos	Sanos	
Expuesto	<i>a</i>	<i>b</i>	TEx
No expuesto	<i>c</i>	<i>d</i>	TnEx
Total	TE	TS	N

Usando la información de esta tabla se pueden calcular los dos índices clínicos básicos para poder cuantificar el concepto de riesgo o daño:

Odds Ratio: $OR = (a \cdot d) / (b \cdot c)$

Riesgo Relativo: $RR = [a (c + d)] / [c (a + b)]$

El Odds Ratio se define como el cociente entre dos Odds, el numerador es el Odds de un evento en el grupo de individuos expuestos, dividido el Odds del mismo evento entre los individuos no expuestos. El evento en este caso es la ocurrencia de la enfermedad o la aparición del síntoma (efecto) buscado en la investigación. Un Odds se define como el cociente entre la probabilidad de que ocurra un evento, dividida la probabilidad de que no ocurra. Por ejemplo, si la probabilidad de contraer una cierta enfermedad es del 20%, el Odds de ese evento será el cociente entre ese 20% y el 80% de que no ocurra. Esto es, $Odds = 20\% / 80\% = 1 / 4$; lo que significa que en 5 casos hay una sola posibilidad de contraer la enfermedad y cuatro de no hacerlo. El significado del OR en Medicina representa la proporción de pacientes que presentan la condición buscada, dividida por la proporción que no la presentan. Muchas veces se toma el concepto de Odds y de riesgo como si fuera la misma cosa, o aproximadamente iguales. Muchos autores calculan los Odds relativos (otra forma de denominar los OR) y entonces informan sus resultados como si hubiesen calculado el riesgo relativo. Por lo tanto conviene mirar esta cuestión más de cerca, por ejemplo si se sabe que 1 / 5 de los pacientes estudiados sufren un golpe de presión, los Odds de que sufran el golpe es de 1 / 4 como se vio antes, o bien 0,25. Para convertir los Odds en riesgo se puede aplicar la relación siguiente:

$$Riesgo = (Odds) / (1 + Odds)$$

En este caso sería, $Riesgo = 0,25 / 1,25 = 0,2$, que es lo mismo que 1 / 5 o 20% que fue el punto de partida. Análogamente se puede expresar la relación inversa con:

$$Odds = Riesgo / (1 - Riesgo)$$

Notar que el riesgo siempre es un valor entre 0 y 1, en cambio los Odds siempre son positivos entre dos casos extremos 0 e infinito. La siguiente es una tabla con los valores más comunes:

Tabla 4.3: Relación entre el riesgo y los Odds

<i>Riesgo</i>	<i>Odds</i>
80%	4
60%	1,5
50%	1
40%	0,67
33%	0,5
25%	0,33
20%	0,25
10%	0,11
5%	0,053

Ejemplo: A 360 pacientes que tuvieron cesáreas en sus partos, 170 de ellas tuvieron una cateterización umbilical durante la cirugía y las demás no. Entre ellas se observaron 60 que desarrollaron una bacteremia con 50 cateterizadas. Los datos se presentan en la tabla siguiente:

Factor de riesgo	Resultados observados		Total
	Infección	No-infección	
Cateterización	50	120	170
No cateterización	10	180	190
Total	60	300	360

$$OR = (50 \cdot 180) / (120 \cdot 10) = 7,5 \text{ mientras que } RR = (50 \cdot 190) / (10 \cdot 170) = 5,6$$

En este caso el Odds de enfermarse en las cateterizadas es de 50 versus las 120 que no se infectaron, o sea $50/120 = 0,42$. Mientras que el Odds de infectarse en las no cateterizadas es de 10 versus las 180 que no enfermaron, esto es $10/180 = 0,056$. Por lo tanto, el cociente entre estos dos Odds resulta $(50/120) / (10/180) = OR = 7,5$. Si se hablase en términos de riesgo o daño, se podría denominar al cociente de Odds como Odds relativos, pero el término impuesto por la Epidemiología es OR. Un médico tiene una buena intuición acerca del riesgo y aún del concepto de un cociente de riesgos como RR. Un jugador tiene buena intuición sobre los Odds, pero nadie (excepto tal vez algunos estadísticos) puede intuitivamente captar el concepto de OR. A pesar de esto, el OR es la medida de asociación más empleada en Medicina, porque tiene ventajas estadísticas en cuanto al hecho de que es independiente de la selección arbitraria entre una comparación de riesgos entre enfermos y sanos, lo cual no es verdadero con el RR. Los clínicos tienen una tendencia natural a sustituir el concepto intuitivo de RR por el concepto poco intuitivo del OR, lo cual puede ser correcto bajo ciertas condiciones: Por ejemplo si $a \lll b$ y $c \lll d$. En otras palabras, la enfermedad tiene que ser mucho menos frecuente que la no-enfermedad, tanto entre los expuestos como entre los no expuestos.

Por su parte el RR muestra la proporción del riesgo original (en este ejemplo el riesgo de enfermarse al ser cateterizadas), que está aún presente cuando las pacientes no fueron expuestas (en este caso sin cateterización). En el ejemplo, el riesgo de enfermarse al ser cateterizadas es de 50 casos entre las 170 en total ($50/170$), mientras que el riesgo de enfermarse sin haber sufrido una cateterización es de 10 casos entre las 190 en total ($10/190$). Por lo tanto el cociente entre ambos riesgos es: $RR = (50/170) / (10/190) = 5,6$. Esto significa que el hecho de sufrir una cateterización durante la cesárea incrementa el riesgo de infectarse 5,6 veces. Intuitivamente, uno puede pensar que no conviene la cateterización en cesáreas, pero clínicamente se sabe que hay otros factores que pueden producir una infección en una cirugía, los que deben ser analizados cuidadosamente si lo que se busca es reducir el número de infectadas.

4.9.1 Índices de riesgo derivados.

Se suelen usar otros índices de riesgo conceptualmente derivados de los anteriores como por ejemplo los siguientes:

Riesgo absoluto: Es el riesgo de contraer la enfermedad habiendo estado expuesto, o el de enfermarse sin haber estado expuesto.

En el ejemplo anterior sería 29,4% (50/170) el riesgo de infectarse habiendo sufrido la cateterización, mientras que el riesgo absoluto de enfermarse sin haber sido cateterizada es de un 5,3% (10/190). A este último valor se lo denomina también proporción del evento en el grupo control (CER) o bien *riesgo básico*. En cambio al riesgo absoluto de enfermarse habiendo estado expuesto se lo denomina (EER) proporción del evento experimental en el grupo expuesto. En otras palabras, se puede considerar al riesgo relativo como el cociente entre los dos riesgos absolutos:

$$RR = EER / CER = [a / (a + b)] / [c / (c + d)]$$

Reducción del riesgo absoluto (ARR): Es la diferencia entre el riesgo básico y el experimental tomado en valor absoluto.

$$ARR = | CER - EER |$$

En el ejemplo visto se calcula como: $ARR = | 5,3\% - 29,4\% | = 24,1\%$. Cuando la diferencia da negativa como en este caso, significa que el exponerse a una cateterización aumenta el riesgo de infectarse en un 24,1% en términos relativos. En cambio cuando la diferencia da positiva significa que el riesgo disminuye, y corresponde al caso de aplicarse una terapia de cura, o al efecto de un medicamento que en efecto funcionan.

Reducción del riesgo relativo (RRR): Es una estimación de la proporción de riesgo básico, que es removida por la terapia, o bien por la exposición. Se calcula como el cociente entre:

$$RRR = ARR / CER = | 1 - RR |$$

En el ejemplo es $RRR = 24,1\% / 5,3\% = 4,6$; o bien $| 1 - 5,6 | = 4,6$. Y significa, usando un lenguaje no técnico que la cateterización incrementa el riesgo relativo de enfermarse 4,6 veces, en comparación a la cirugía sin la cateterización.

4.10 Concordancia clínica

Además de la validez interna de un método, dada por su prestación medida por los índices clínicos antes descriptos, se necesita otra cosa más: *fidelidad*. Esto es, no debe ser influenciado por la persona que lo ejecuta. No debe ocurrir que el mismo test, hecho por dos bioquímicos diferentes al mismo paciente, en el mismo tiempo y lugar, arrojen resultados opuestos o muy diferentes. Significa, que se necesita la reproducibilidad de los resultados de la prueba, independientemente de quien la ejecuta. Se define un índice para cuantificar esta situación, con el nombre de grado o *Nivel de Concordancia* (λ %):

$$\lambda \% = (NC / N) \cdot 100$$

Donde NC: es el número de diagnósticos que concuerdan
N : número total de casos

Un test perfectamente reproducible tendría un índice $C\% = 100\%$. Para estudiar la concordancia entre dos métodos clínicos, se arma la llamada Tabla de Concordancia que consta de 2 filas y 2 columnas (Tablas 2×2). A su vez cada fila o columna puede estar subdividida en clases o categorías. Se habla de una tabla de $k \times k$ cuando hay k clases en filas y columnas. Esto se esquematiza en el Cuadro 4.1, donde por ejemplo hay dos bioquímicos A y B, que pueden diagnosticar k casos diferentes, a partir de la misma prueba clínica. Entonces, se puede armar una tabla como la siguiente:

Cuadro 4.1: Tabulación cruzada en un diagnóstico.

		Bioquímico A					Σ
		1	2	3	...	k	
Bioquímico B	1	n_{11}	n_{12}	n_{13}	n_{1k}	$n_{1.}$	
	2	n_{21}	n_{22}	n_{23}	n_{2k}	$n_{2.}$	
	3	n_{31}	n_{32}	n_{33}	n_{3k}	$n_{3.}$	
		
	k	n_{k1}	n_{k2}	n_{k3}	n_{kk}	$n_{k.}$	
Σ	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{.k}$	N		

$$\lambda \% = \{ (n_{11} + n_{22} + n_{33} + \dots + n_{kk}) \} \cdot 100 / N$$

$$\lambda \% = (100 \cdot \Sigma_k n_{kk}) / N$$

La concordancia observada, es la sumatoria de la cantidad de diagnósticos coincidentes para cada caso n_{kk} , dividida por el número total de casos, expresada porcentualmente. En el gráfico, se suman filas y columnas, colocando un punto como subíndice para subrayar este hecho. Así, el total de la 2ª columna es $n_{.2}$ y el total de la primera fila $n_{1.}$. En el ejemplo anterior se obtuvo un índice del 50% lo que significa una mala concordancia entre ambos bioquímicos.

La manera más usual de aplicar este concepto es para comparar dos test clínicos entre sí, cuando se aplican al mismo paciente (es decir muestras apareadas), como fuera presentado en el primer capítulo (Tabla 1.3) como la siguiente:

Método 2	Método 1		Total
	(+)	(-)	
(+)	a	b	$a + b$
(-)	c	d	$c + d$
Total	$a + c$	$b + d$	N

Donde el número total de concordancias está dado por la cantidad (a + d) y el número total de discordancias por la cantidad (b + c). Siendo N el número total de casos analizados. Los índices de concordancia más usados son dos:

Nivel de concordancia: $\lambda = (a + d) / N$ (expresado en forma porcentual: $\lambda\% = 100 / [1 + DO]\%$)

Odds de discordancia: $DO = (b + c) / (a + d)$ y su relación con $\lambda\%$ es: $DO = (100 / \lambda\%) - 1$

El nivel de concordancia indica el porcentaje de éxitos (concordancias) obtenido por la comparación de las pruebas. La idea es compararlo contra un valor recomendado en la literatura médica para que la concordancia sea considerada aceptable. Por ejemplo, un nivel de concordancia del 90% es suficiente para muchas enfermedades, pero no alcanza para casos como el Sida o ciertos tipos de cáncer.

Los Odds de discordancia son otra manera de expresar el mismo concepto y es usado por aquellos que prefieren el concepto de Odds al de probabilidad. Por ejemplo, un $DO = 1/9$ significa que habrá 1 discordancia y 9 concordancias en 10 casos. O que se tendrá un desacierto cada 9 aciertos entre ambas pruebas. La concordancia ideal se da cuando $DO = 0$ ($\lambda = 100\%$), en cambio cuando la concordancia sea $DO = 1$ significa que habrá el mismo número de concordancias y de discordancias, lo cual clínicamente está en el límite de lo aceptable ($\lambda = 50\%$). Cuando DO sea mayor que 1, o bien $\lambda < 50\%$, la discordancia entre ambos métodos será inaceptable desde un punto de vista clínico.

Ejemplo: Comparación de diagnósticos hechos por dos individuos diferentes.

A 200 pacientes internados en un sanatorio, se les efectuó un análisis de sangre en heces. Dos bioquímicos proceden a efectuar el test al mismo grupo de individuos. Los resultados fueron:

Diagnósticos:	Sangre en heces			Σ
	0	+	++	
Bioquímico A	120	56	24	200
Bioquímico B	120	56	24	200

De acuerdo al cuadro anterior, la coincidencia entre ambos parece perfecta. Sin embargo, si se efectúa una tabulación cruzada para evitar que el mismo paciente aparezca dos veces en la tabla y lograr así independencia. Se obtiene:

		Bioquímico A			Total
		0	+	++	
Bioquímico B	0	80	40	0	120
	+	28	12	16	56
	++	12	4	8	24
Total		120	56	24	200

De donde se puede calcular:

$$\lambda \% = 100 \cdot (80 + 12 + 8) / 200 = 50 \%$$

4.11 Estadígrafos clásicos

En este punto se describen los diferentes estadígrafos, llamados también *números índices*, *estadísticas*, *estadísticos*, etc. Clásicamente, se trata de valores que se calculan para magnitudes clínicas cuantitativas, y no cualitativas. Se trata de valores que aportan mucha más información que la propia, como datos en sí mismos. Así como un histograma puede ser considerado como una manera clara y práctica de resumir la información contenida en la “nube” de datos, los estadígrafos la condensan aún más, reduciéndola a unos pocos valores que toma el lector para darse una idea de todo el conjunto. El trabajo de analizar un gran número de datos siempre es engorroso; en la actualidad, con las computadoras la tarea está mucho más aliviada, pero en los inicios era una tarea muy pesada. Por eso, desde siempre se buscó la manera de unificar criterios y técnicas para homogeneizar la manera de presentar datos en los informes científicos. Eso evitó la anarquía que se hubiese tenido, si cada investigador los tratase a su peculiar manera, para luego presentarlos a consideración de los demás en congresos, simposios, publicaciones, etc.

Todo esto llevó a encontrar formas simples y concisas para describir el conjunto de datos hallados en experimentos. Formas conocidas y aceptadas por todos los científicos, a manera de convención. Maneras cuantitativas que permitan efectuar comparaciones entre trabajos y caracterizar la población de donde fueron extraídos los datos. Por esa necesidad, aparecieron los estadígrafos, como por ejemplo números para dar una idea al lector de la posición de la nube de datos, o un valor alrededor del cual los datos parecen agruparse, o un número que cuantifica la dispersión de los datos alrededor del anterior. Además, cada profesión aporta más índices propios de su campo específico. En Economía se usan en forma habitual, los índices de costo de vida, de la construcción, de precios mayoristas no agropecuarios, etc. En Demografía se usan las tasas de mortalidad, morbilidad, fecundidad, casamientos, etc. En Sanitarismo se emplean la prevalencia, la predominancia, y otros muchos casos. En Clínica, los hay específicos para el diagnóstico de enfermedades, a partir de los resultados de los análisis del laboratorio. Los estadígrafos clásicos se dividen en dos tipos generales: los llamados estadígrafos de *posición* que dan una idea de alrededor de que valor parecen agruparse los datos, y los de *dispersión* que muestran como se distribuyen los datos alrededor de los de posición.

4.12 Estadígrafos de posición

Se denominan así, a los estadígrafos que indican un valor, alrededor del cual los datos parecen agruparse de cierta manera, como si fuese el “centro de gravedad de los datos”, tal es el caso de los estadígrafos de tendencia central (media, mediana y moda). O bien, dejando una cierta cantidad de valores por debajo o por encima de él (percentiles).

4.12.1 Medias

$$\text{Media aritmética : } \bar{x} = \sum_{i=1}^n (x_i / n) \quad (\text{para datos sueltos})$$

$$\text{Media aritmética ponderada : } \bar{x} = \sum_{k=1}^m (f_k \cdot x'_k / n) \quad (\text{para datos agrupados})$$

Cuando se conocen los datos en forma individual, se puede calcular la *media aritmética* tradicional en forma directa aplicando la primera de las fórmulas de arriba. En cambio, si la información se suministra a través de n datos agrupados en m clases, se usa la *media aritmética ponderada* usando la otra fórmula. Donde la frecuencia de cada clase se expresa con f_k , siendo x'_k la marca de cada clase (es decir el punto medio del ancho de clase, obtenido como la semi-suma de sus límites reales de clase). La expresión de la media aritmética usando las frecuencias de clase no es otra cosa que un promedio ponderado, donde se le da un peso porcentual a cada marca de clase, de acuerdo al tamaño que tiene cada una. Como cada valor de frecuencia está dividido por la frecuencia total, la fórmula resulta la sumatoria de las marcas de clase (o valores individuales) multiplicada por la frecuencia relativa respectiva. Esta fórmula de media ponderada se aplica de muchas maneras en la práctica: es un concepto similar al centro de masa visto en Física (o centro de gravedad), la velocidad promedio en la teoría cinética de los gases, etc.

CUADRO 4.2: Cálculo de la media aritmética.

Caso 1: Datos individuales. (usando los datos para los no infartados del Cuadro 3.1).

$$\bar{x} = \frac{\sum_{i=1}^n (x_i / n)}{200} = \frac{10 + 33 + 20 + \dots + 120 + 92 + 90}{200} = 30,735 \text{ UI/l}$$

Caso 2 : Datos agrupados en clases (usando los mismos datos, pero tomados del Cuadro 3.3).

CPK	f_k	x'_k	$f_k \cdot x'_k$	
1 - 20	104	10,5	1092	La media aritmética aproximada se calcula con :
21 - 40	36	30,5	1098	
41 - 60	28	50,5	1414	
61 - 80	18	70,5	1269	
81 - 100	10	90,5	905	
101- 120	4	110,5	442	$\bar{x} = 6220 / 200 = 31,1 \text{ UI/l}$
121- 140	0	130,5	0	
141- 160	0	150,5	0	
161- 180	0	170,5	0	
Σ	200		6220	

El valor exacto de la media aritmética, está dado por el primer caso, donde se tienen identificados los datos, uno por uno. El segundo valor de la media, obtenido a partir de su agrupación en clases, es solo una aproximación, pues al perder información se pierde exactitud (aumentó un 2,7%).

Caso 3 : Datos y sus frecuencias (valores tomados de un conteo celular - Student, 1907).

Nº de células x cuadrícula	0	1	2	3	4	5
Frecuencias Observadas	213	128	37	18	3	1
$\bar{x} =$	$\frac{\{0(213) + 1(128) + 2(37) + 3(18) + 4(3) + 5(1)\}}{213 + 128 + 37 + 18 + 3 + 1} = \frac{273}{400} = 0,6825$					

En este caso las marcas de clase coinciden con la cantidad de células por cuadrícula de la cámara que tiene en total 400.

Caso 4. Una farmacia maneja 4 familias de productos en su trabajo. Los márgenes de utilidad de cada uno de ellos, durante el período fiscal pasado fueron: Familia A: 8,4%; la B: 11,0%; la C: 14,8% y la familia D: 20,2%. Se pide calcular el margen de utilidad promedio.

Producto	Ventas (miles de \$) (V)	Márgenes (%) (M)	Participación (V.M)
A	300	8,4	25,20
B	200	11,0	22,00
C	50	14,8	7,40
D	<u>30</u>	20,2	<u>6,06</u>
Totales	580		60,66

Si se calculase el promedio de los márgenes en forma directa, sería de: 13,6%, lo que no es correcto. Pero esto *sólo es correcto si las 4 familias tuvieron igual participación en las ventas*. Lo cual es muy raro que ocurra. Por lo tanto, el farmacéutico debe estudiar como se le distribuyeron las ventas y hacer la *ponderación* respectiva como se calcula más arriba. Entonces, el margen promedio ponderado buscado es: $M = \frac{\sum (V_k \cdot M_k)}{\sum V_k} = 60,66 / 580 = 10,46\%$

Caso 5. El tiempo de atención al cliente en una farmacia con varios vendedores fue medido durante un día de trabajo donde se atendieron 47 clientes. El encargado de medir los tiempos informó lo siguiente. El farmacéutico a cargo quiere calcular el tiempo promedio que tiene su negocio a fin de mejorarlo.

Informe recibido		Cálculos efectuados por el farmacéutico			
Tiempo en minutos	Nºclientes	Clases	Medios	Frec. ponderada	Frec. acumulada
Entre 5 y 8	10	5- 7,99	6,5	65,0	10
Más de 8 y menos de 11	17	8-10,99	9,5	161,5	27
Más de 11 y menos de 14	12	11-13,99	12,5	150,0	39
Más de 14 y menos de 17	6	14-16,99	15,5	93,0	45
Más de 17 y menos de 20	2	17-19,99	18,5	37,0	47
Más de 20	<u>0</u>	20 y más	---	---	
Totales	47			506,5	

El tiempo promedio de atención al cliente es: 10,8 minutos = 506,5 / 47

Caso 6. En una industria farmacéutica se fabrican tres tipos de ampollas de un mismo medicamento. Se desea averiguar el porcentaje de fallas de la máquina envasadora. Luego de un mes de pruebas, Control de Calidad obtuvo los datos siguientes:

Ampollas	% defectuosas	Nº de ampollas	Amp. defectuosas
Tipo A	1,1	210.000	2310
Tipo B	1,5	120.000	1800
Tipo C	2,3	<u>50.000</u>	<u>1150</u>
Totales		380.000	5260

El promedio ponderado de ampollas defectuosas es de 1,38 % = 5260/380.000

Caso 7. Se observan 10 pacientes adultos de sexo masculino, que llegan a hacerse a atender al laboratorio Otros 15 pacientes jóvenes y 25 niños. En cambio, de sexo femenino llegan 20 adultas, 5 jóvenes y 15 niñas. Calcular el promedio de cada grupo de edades (grupo etáreo).

Resulta que en promedio llegan: 15 adultos, 10 jóvenes y 20 niños.

En el Cuadro 4.2 se han desarrollado ejemplos de aplicación. En el primer caso se muestra un ejemplo de media aritmética simple aplicado a los 200 casos de los no infartados del Cuadro 3.1. La media es 30,735 UI/l. En cambio, si se tuviesen los mismos datos pero agrupados en clases, solo se podría calcular una estimación de la media a través de una ponderación. Usando los datos del Cuadro 3.3 para los no infartados y la fórmula de media ponderada, se obtiene 31,1 UI/l, o sea un aumento del 2,7% respecto del valor exacto. En el tercer caso se calcula la media ponderada del conteo celular pues se carecen de las mediciones individuales hechas por Student en 1907. Los tres casos siguientes son aplicaciones en Farmacia del concepto de la media. En el Caso 4 se desea calcular el margen promedio de ganancias del negocio, utilizando los márgenes de 4 líneas de productos. En el Caso 5 el farmacéutico desea mejorar la atención al público reduciendo los tiempos de espera de sus clientes. El primer paso para ello es calcular el tiempo promedio de atención al cliente en el mostrador. Toma los tiempos agrupados en clases y sus frecuencias para poder hacer la ponderación del promedio. Descubre que en su farmacia se tarda casi 11 minutos en despachar a un cliente. En el último caso se desea saber el porcentaje de ampollas falladas de una máquina envasadora. Durante un período de un mes se calculan los porcentajes de fallas para tres productos diferentes A, B y C que se envasan en ese equipo, y efectuando la ponderación resulta 1,38% de fallas. En el último ejemplo, se muestran los promedios por grupo etéreo de pacientes que concurren al laboratorio.

La media sólo se usa en variables cuantitativas, pues en las cualitativas carece de sentido. Las principales propiedades de la media se muestran en el Cuadro 4.3., y se explican a continuación por su importancia en los capítulos posteriores:

Cuando a una media se le suma o resta una constante, la nueva media queda sumada (o restada) por la misma constante. Esto es el efecto que tiene un error de tipo sistemático en una serie de mediciones, porque el efecto del mismo es siempre del mismo signo y aproximadamente de la misma magnitud. Todo ocurre como si se trasladase el eje de coordenadas a lo largo del eje de abscisas (a la derecha cuando se le suma una constante y a la izquierda cuando se la resta).

Cuando a la media se la multiplica o divide por una constante, la nueva media queda multiplicada o dividida por dicha constante. Esto es el efecto de cambiar de escala en un gráfico. Otra propiedad de mucha importancia es: La sumatoria de los desvíos de valores medidos en un sistema respecto de su media es nula. Esto significa que de todos los puntos posibles para referir las mediciones, la media aritmética equilibra esos desvíos. Lo que explica que se puede considerar a la media como el “centro de gravedad” de los datos. Una aplicación práctica de la media aritmética ponderada es el cálculo del centro de masa, o el de inercia en Física.

El método de los mínimos cuadrados se basa en la propiedad: Si se suman los cuadrados de los desvíos de todas las mediciones respecto a un punto cualquiera, se demuestra que tal suma se hace mínima cuando el punto elegido es la media aritmética. Por eso, si se divide tal suma por los grados de libertad se tiene un desvío de referencia llamado: *desvío estándar*.

Existen otros tipos de medias como la *media geométrica* y la *media armónica*, usadas en otras disciplinas. Por eso, se deja al lector la inquietud de verlas en la bibliografía. La media geométrica se usa para interpolar datos en funciones exponenciales. El uso clásico de la media armónica es el cálculo de la velocidad promedio de un móvil que recorre una distancia a diferentes velocidades.

CUADRO 4.3: Propiedades de la media aritmética.

Sean n observaciones de una variable clínica cualquiera $(x_1, x_2, x_3, \dots, x_n)$, cuya media aritmética es: \bar{x} , entonces :

1. Si a cada valor medido se le suma (o resta) una constante a , la nueva media aritmética es igual a la anterior, más (o menos) dicha constante. Esto se demuestra así:

$$\text{Si } z_i = x_i + a, \text{ entonces } \bar{z} = \frac{\sum_{i=1}^n z_i}{n} = \frac{\sum (x_i + a)}{n} = \frac{\sum (x_i/n) + n \cdot a}{n}$$

O sea

$$\bar{z} = \bar{x} + a$$

Por comodidad, se le han sacado los subíndices a las sumatorias y así será de ahora en más.

2. Si a cada valor medido se le multiplica (o divide) por una constante a , la nueva media aritmética es igual a la anterior, multiplicada (o dividida) por dicha constante. Esto se demuestra :

$$\text{Si } z_i = x_i \cdot a; \text{ entonces } \bar{z} = \frac{\sum z_i}{n} = \frac{\sum (x_i \cdot a)}{n} = a \cdot \frac{\sum (x_i/n)}{n} = a \cdot \bar{x}$$

3. La sumatoria de todos los *desvíos* es nula. Esto se demuestra así :

Por desvío, se entiende a la diferencia entre cada valor medido y su promedio, es decir, para el dato número i su desvío es $d_i = x_i - \bar{x}$

$$\sum d_i = \sum (x_i - \bar{x}) = \sum x_i - \sum \bar{x} = n \cdot \bar{x} - n \cdot \bar{x} = 0$$

4. La sumatoria de los cuadrados de los desvíos es un mínimo. Más conocida como el principio de los mínimos cuadrados. Esto se demuestra así :

Sea un valor cualquiera a , para saber que valor de a hace mínima la sumatoria, se saca la primer derivada y se iguala a cero:

$$\frac{\delta}{\delta a} (\sum d_i^2) = \frac{\delta}{\delta a} [\sum (x_i - a)^2] = \sum [2 (x_i - a) (-1)] = -2 \sum (x_i - a) = 0$$

El único valor que cumple con la propiedad anterior, es el promedio. O sea, $\sum (x_i - a) = 0$ sólo sí $a = \bar{x}$, de acuerdo a la propiedad **3** vista más arriba. Luego,

$$\frac{\delta^2}{\delta a^2} (\sum d_i^2) = \frac{\delta}{\delta a} [(-2) \sum (x_i - a)] = (-2) \sum (-a) = 2 \cdot n \cdot a$$

Como $(2 \cdot n \cdot a)$ es una expresión siempre positiva, cuando $a = \bar{x}$ hay un mínimo.

4.12.2 Mediana

La mediana de un grupo de datos se define como el valor que es superior al 50% de los mismos, e inferior del 50% restante. Si se ordenan los datos en forma creciente, la mediana debe quedar en el medio. Por ejemplo, la mediana del grupo (1 , 4 , 5 , 6 , 8) es el valor 5, pues deja dos valores por encima y dos por debajo. Cuando la cantidad total de datos es par, se toma como mediana a la semisuma de los dos valores centrales. Si se tiene (1 , 2 , 4 , 8 , 9 , 10), la mediana es el valor 6, que deja igual cantidad de valores a ambos lados y está en el centro de 4 y 8. Cuando los datos vienen agrupados en clases, se usa la fórmula de los percentiles para encontrar más fácilmente el valor de la mediana.

CUADRO 4.4 : Cálculo de la mediana.

Caso 1: Datos individuales. (Usando los datos para los no infartados del Cuadro 3.1).

El valor de la mediana es 19, porque deja la misma cantidad de datos a ambos lados. En este caso, donde hay una cantidad par (200), hay que calcular el promedio entre los dos valores centrales, y como ambos son 19, entonces la mediana adopta dicho valor.

Caso 2 : Datos agrupados en clases (usando los mismos datos, pero tomados del Cuadro 3.3).

La mediana cae dentro de la primer clase, donde los límites reales de clase son 0,5 – 20,5. En este caso hay que aplicar la fórmula de la mediana para datos agrupados, que se verá en el punto siguiente: los fractiles donde el percentil 50 deja el 50% de los datos por encima. O sea, dicho percentil es lo mismo que la mediana.

Caso 3 : Hacer una estimación de cuánto ganan, en promedio, los empleados de una farmacia, con los siguientes datos:

Gerente (1)	\$ 2000	Si se calcula el promedio directo es: $\frac{2000 + 5.(400) + 350}{7} = \$ 621,5$
Vendedores (5)	\$ 400	
Repartidor (1)	\$ 350	

Se puede decir entonces que el sueldo promedio, o “ingreso per cápita” es de \$ 621,5. Sin embargo, este valor no es muy representativo de la realidad, pues de los siete empleados sólo uno de ellos supera dicho monto. En cambio, si se usa la mediana $Md = \$ 400$; este valor se acerca mejor a la idea de “ingreso per cápita”. Esto muestra la conveniencia de usar la mediana, antes que la media aritmética, para estimar los ingresos promedios en un ámbito económico.

La mediana se emplea en lugar de la media aritmética cuando ésta se ve influenciada por valores extremos. Cuando en el grupo de datos obtenidos existen unos pocos valores muy altos o muy bajos, o ambos a la vez. La debilidad de la media aritmética, es esa sensibilidad a los valores extremos, que la corren mucho de la posición centralista que debería tener. Por ejemplo, sea el grupo de datos (1 , 2 , 4 , 5 , 6 , 6 , 7 , 7 , 9 , 9 , 10 , 174) , su valor medio es 20, mientras que su mediana es 6,5 , el cual da una mejor idea acerca del centro de los datos. El promedio está sesgado hacia arriba por el valor extremo de 174. Notar que si este se elimina, entonces la mediana pasa a ser 6 y el promedio pasa a ser también 6, algo más coherente con la noción de centro, o posición central de los datos.

4.12.3 Fractiles

Hay diferentes tipos de fractiles, pero en general son valores que dejan por debajo de él una cierta fracción de los datos ordenados en forma creciente y el resto por encima. Cuando la fracción es la mitad, se trata de la mediana. Los más comunes son:

- *Quartiles* : Las fracciones son cuartas partes del total. El primer cuartil Q_1 deja el 25% de los valores por debajo. El segundo cuartil es igual a la mediana y el tercero Q_3 deja el 75%.
- *Deciles* : Las fracciones son décimas partes del total. El primer decil D_1 deja el 10% de los valores por debajo y el resto por encima. El quinto decil es la mediana.
- *Percentiles* : Las fracciones son centésimas partes del total. Así, el percentil catorce (P_{14}) deja el 14% de los valores por debajo. El percentil cincuenta es la mediana, o el quinto decil.

El método para calcular los percentiles es análogo al de la mediana: primero se ordenan los datos en forma creciente, se calcula la fracción buscada y se lo ubica entre ellos. Pero si los datos vienen agrupados en clases, la cosa se complica y hay que emplear la fórmula siguiente:

$$F_x = LRI_i + (a / f_i) [N (x / 100) - f_{ac_{i-1}}] \quad \text{donde}$$

F_x : Fractil x -avo LRI_i : límite real inferior de la clase i donde se encuentra el fractil
 a : ancho de clase f_i : frecuencia de la clase i
 N : total de datos $f_{ac_{i-1}}$: frecuencia acumulada de la clase anterior a la del fractil

Cuadro 4.5: Obtención de diversos fractiles (con datos de la Tabla 3.3).

Caso 1 : Obtención de la mediana (segundo cuartil, quinto decil o cincuentavo percentil)

Límites reales	f_i	x'_i	f_{ac_i}	
0,5 - 20,5	104	10,5	104	La mediana debe dejar 100 datos por debajo, como en la primer clase ya hay 104, significa que ahí se halla el valor buscado. Por lo tanto es : $F_{50} = 0,5 + (20 / 104) [200 (50/100) - 0]$ $F_{50} = 19,73 \text{ UI/l} = Q_2 = D_5 = P_{50}$
.20,5 - 40,5	36	30,5	140	
40,5 - 60,5	28	50,5	168	
60,5 - 80,5	18	70,5	186	
80,5 - 100,5	10	90,5	196	
100,5- 120,5	4	110,5	200	
Σ	200			

Caso 2 : Obtención del tercer cuartil Q_3 .

El 75% del total de valores es 150. Luego el valor buscado está en la tercer clase .

$$Q_3 = F_{75} = 40,5 + (20 / 28) [200 (75/100) - 140] = 47,64 \text{ UI/l}$$

Caso 3 : Obtención del percentil cinco y del noventa y cinco : P_5 y P_{95} .

El 5% del total de valores es 10. Luego el valor buscado está en la primer clase.

$$P_5 = F_5 = 0,5 + (20/104) [200 (5 /100) - 0] = 2,42 \text{ UI/l}$$

El 95% del total de valores es 190. Luego el valor buscado está en la quinta clase.

$$P_{95} = F_{95} = 80,5 + (20/10) [200 (95/100) - 186] = 88,5 \text{ UI/l}$$

De esto último, se puede concluir que el 90% de los pacientes sin infarto, pero con diagnóstico presunto de infarto de miocardio, sacó una CPK comprendida entre 2,42 y 98,5 UI/l.

Caso 4 : Calcular la mediana para el Caso 5 del Cuadro 4.2.1 y los percentiles 5 y 95.

Mirando la columna de frecuencias acumuladas de este caso, se nota que la mediana cae en el segundo intervalo de clases. Siendo $N = 47$; $F_{ac1} = 10$; $LRI_2 = 8$; $F_2 = 17$ y $a = 2,99$. Entonces,

$$P_{50} = F_{50} = 8 + (2,99/17) [(47/2) - 10] = 10,4 \text{ minutos para atender a un cliente.}$$

Notar que en este caso, la mediana se parece al valor medio 10,8 minutos calculado más arriba. Esto porque, al no haber valores extremos, como en el caso de los salarios del Caso 3 del Cuadro 4.4; los dos valores se pueden emplear indistintamente en los cálculos econométricos.

Análogamente, el percentil 5 cae en el primer intervalo.

$$P_5 = F_5 = 5 + (2,99 /10) [(0,05 \cdot 47 - 0)] = 5,7 \text{ minutos}$$

Por su parte, el percentil 95 cae en el anteúltimo intervalo, muy cerca del LRS_4 .

$$P_{95} = F_{95} = 14 + (2,99 /6) [0,95 \cdot 47 - 39] = 16,8 \text{ minutos}$$

Esto significa, que en el 90% de los casos, se tardará entre 5,7 y 16,9 minutos en atender a un cliente. Este dato, es de mayor utilidad para el farmacéutico que el tiempo promedio de atención (10,8). Le permitirá calcular mejor el número de vendedores que necesita a medida que vaya variando el número de clientes que entra a su farmacia, para mantener el mismo ritmo de atención.

4.12.4 Moda

Es el valor de la variable de mayor frecuencia. El más salidor, el que se repite más veces en el grupo de datos. Se lo usa en lugar de la media o mediana, cuando se desea señalar el valor más típico o común del total de observaciones efectuadas. En términos matemáticos es un máximo relativo. Es más ambiguo que la mediana, muy inestable y no permite tratamientos algebraicos sencillos. La moda no queda influenciada por valores extremos como la media y es el más sencillo de comprender. A su antítesis, se lo llama *antimoda*, o sea el valor menos frecuente. Cuando la distribución de datos presenta más de un máximo relativo, se la llama bimodal por ese motivo. Para los datos del Cuadro 4.4 anterior, la moda pertenece a la primer clase que tiene el valor máximo de frecuencia: 104; para definirlo se toma el centro de la clase, o sea, la marca de clase que en este caso es:

$$x'_1 = 10,5 \text{ UI/l} = \text{Moda}$$

4.13 Estadígrafos de dispersión

Se denominan así a los estadígrafos que cuantifican la separación de los datos entre sí, respecto de un punto de referencia central, como la media aritmética. Es una medida de la dispersión de los valores obtenidos al medir una variable clínica. Cuanto menor es la dispersión, tanto mayor será la precisión del sistema de medición. Por eso se usan a estos estadígrafos como índices de la calidad de las técnicas clínicas. Si los estadígrafos de posición vistos en la sección 4.3 se relacionan con el concepto de exactitud, los de dispersión se relacionan con la precisión de las técnicas de laboratorio. Aquellos con el error sistemático, estos con los errores casuales. Hay varias formas de cuantificar esta dispersión.

La más simple de todas es el *rango*, definido como la diferencia entre el valor máximo y mínimo del grupo de datos. De fácil cálculo y comprensión, tiene la desventaja de ser la medida más grosera de la dispersión. Dos grupos de datos, con muy distinta dispersión pueden llegar a tener rangos similares. Uno de ellos puede tener el 99% de los valores junto al mínimo y el otro el 99% junto al máximo, pero al tener extremos iguales, sus rangos resultarían iguales a pesar de ser tan disímiles intrínsecamente.

Una medida un poco más refinada estaría dada por los desvíos, o sea la diferencia entre cada valor y el promedio. La sumatoria de todos estos desvíos sería una medida de la dispersión, pero resulta ser nula porque las desviaciones se compensan, tal como se vio en las propiedades de la media. Ahora bien, si se toman en valor absoluto el problema de los signos se resuelve, pero aún así no es suficiente. Tal sumatoria tendría el defecto de ser mayor, cuantos más datos se tomen en cuenta, utilizando el mismo sistema de medición en condiciones similares. Por tal motivo, se define al *desvío medio absoluto* como la sumatoria del valor absoluto de todos los desvíos, dividida el número total de mediciones. Este índice es mucho mejor que el rango, pero tiene un defecto: es muy difícil de manejar algebraicamente.

Otra forma de eliminar el signo de las desviaciones es elevándolas al cuadrado. O sea, la sumatoria de los cuadrados de los desvíos, que es un mínimo, ponderada con el número total de mediciones. Esta es la idea de la *varianza* de un grupo de datos. Como las unidades de semejante índice quedarían al cuadrado, conviene tomar la raíz cuadrada para poder compararla con la media y otras medidas análogas. Esta es la idea del *desvío estándar* o desviación típica. También es la fórmula planteada por Gauss en su teoría de errores casuales. Como se verá más adelante, la curva de Gauss tiene dos puntos de inflexión simétricos ubicados a una distancia del centro igual al desvío estándar.

Una medida de posición y otra de dispersión, son suficientes para dar una idea de la forma en que se agrupan los datos. En particular, si se supone que los datos fueron tomados de una población “normal” o “gaussiana”, entonces la media y el desvío estándar son los valores necesarios y suficientes para definir la curva de Gauss involucrada en forma unívoca. Por regla general, los investigadores saben que mediciones repetidas de la misma magnitud, en condiciones análogas, tienen una distribución gaussiana. Separando a los pacientes por sexo y edad, las distribuciones de las magnitudes clínicas continuas en dichos segmentos suele ser gaussiana. Si bien algunos autores disienten en algunos aspectos, es indudable que separando a los pacientes sanos, agrupados por misma edad y sexo, con un número muy grande de casos, las distribuciones obtenidas se parecen mucho a la normal. Esto se discutirá más adelante con más detención.

4.13.1 Desvío estándar y Varianza

Se definen a estos estadígrafos con las fórmulas siguientes,

$$\text{Varianza poblacional: } \sigma^2 = \sum (d_i^2) / N = \sum (x_i - \mu)^2 / N$$

$$\text{Varianza muestral: } DS^2 = \sum (d_i^2) / (n - 1) = \sum (x_i - \bar{x})^2 / (n - 1)$$

Donde :

\bar{x}	es la media de la muestra	n	es el tamaño de la muestra
μ	es la media de la población	N	es el tamaño de la población
x_i	es la i-ésima medición o valor	d_i	desvío de un valor respecto de su media

A su vez, el *desvío estándar poblacional* (σ) y *el muestral* (DS) se obtienen con la raíz cuadrada de las respectivas varianzas. Para poder conocer o calcular tanto la media como la varianza poblacionales, se necesita conocer la población completa. Esto es imposible en el caso de mediciones repetidas porque estas son infinitas desde el punto de vista teórico. Análogamente, la cantidad total de muestras que se le puede extraer a un paciente, para hacerle una determinación en el laboratorio, es tan grande que puede ser considerada infinita. Por lo tanto, para todos los efectos prácticos se consideran a las mediciones de magnitudes clínicas continuas, como provenientes de una población de tamaño infinito. Es decir, que nunca se podrán conocer con certeza sus parámetros poblacionales μ y σ . Necesariamente, habrá que estimarlos a partir de sus respectivas medidas muestrales obtenidas con \bar{x} y DS .

Cuando los datos vienen agrupados en clases se emplea la fórmula:

$$DS^2 = \sum (x'_j - \bar{x})^2 f_j / (n - 1)$$

Donde f_j es la frecuencia de la clase j
 x'_j su respectiva marca de clase

Debe considerarse al desvío estándar como un estadígrafo y a la varianza como a otro. Aquí se desarrollan en conjunto para simplificar el texto, aprovechando la gran similitud de sus fórmulas. Pero, mientras el primero se puede comparar directamente con cualquiera de los valores medidos o de sus estadígrafos de posición, la varianza no. Así, la media y el desvío se emplearán para construir intervalos donde se espera encontrar adentro al verdadero valor de la población, llamado a veces *parámetro*. Eso se estudiará en la estadística inferencial. En cambio el *análisis de la varianza*, llamado ANOVA, es uno de los capítulos más importantes de la Estadística aplicada a experimentos. Se usa fundamentalmente para testear supuestos que se hacen acerca de poblaciones, como detectar si un medicamento sirve, o si un fertilizante es útil a la tierra, o si un alimento balanceado es mejor que otros y muchos usos más.

La fórmula del desvío estándar muestral dada más arriba, debe ser corregida por un factor presentado por Gurland y Tripathi. Si bien DS^2 es un estimador insesgado de σ^2 , no ocurre lo mismo con su raíz cuadrada DS que en forma consistente subestima a σ . El factor de corrección presentado por estos dos autores corrige ese defecto. Cuando $n \geq 30$ se calcula con la relación $Gn \approx 1 + 1 / [4(n-1)]$. Para los demás valores ver en Cuadro 4.6.

CUADRO 4.6: Cálculo de desvío estándar y varianza.

Caso 1: Datos individuales. (usando los datos para los no infartados del Cuadro 3.1).

$$DS^2 = \sum (x_i - \bar{x})^2 / (n-1) = \frac{(10 - 30,3)^2 + (33 - 30,3)^2 + \dots + (90 - 30,3)^2}{200 - 1}$$

$$DS = 27,77 \text{ UI/I}$$

Caso 2 : Datos agrupados en clases (usando los mismos datos, pero tomados del Cuadro 3.3)

CPK	f_j	x'_j	$(x'_j - \bar{x})^2$	$f_j \cdot (x'_j - \bar{x})^2$	
1 - 20	104	10,5	392,04	40772,16	$DS^2 = 143256 / 199$
21 - 40	36	30,5	0,04	1,44	
41 - 60	28	50,5	408,04	11425,12	$DS^2 = 719,88 \text{ (UI/I)}^2$
61 - 80	18	70,5	1616,04	29088,72	
81 - 100	10	90,5	3624,04	36240,40	$DS = 26,83 \text{ UI/I}$
101- 120	4	110,5	6432,04	25728,16	
Σ	200			143256,00	

El valor exacto del desvío y de la varianza está dado por el primer caso, donde se tienen identificados los datos uno por uno. A partir de su agrupación en clases, los valores aproximados resultan ser un 3,5% menores, pues al perder información se pierde exactitud en el cálculo.

Caso 3 : Datos y sus frecuencias (valores tomados de un conteo celular - Student, 1907).

Nº de células x cuadrícula	0	1	2	3	4	5
Frecuencias Observadas	213	128	37	18	3	1

$$DS^2 = \{213 \cdot (0-0,68)^2 + 128 \cdot (1-0,68)^2 + 37 \cdot (2-0,68)^2 + 18 \cdot (3-0,68)^2 + 3 \cdot (4-0,68)^2 + (5-0,68)^2\} / 399$$

$$DS^2 = 324,68 / 399 = 0,8137 \quad \text{y el desvío estándar resulta } DS = 0,902$$

En este caso las marcas de clase coinciden con la cantidad de células por cuadrícula de la cámara que tiene en total 400.

Caso 4 : Factor de corrección de Gurland y Tripathi (1971).

Este factor puede ser importante cuando el tamaño muestral es pequeño ($n < 30$).

$$G_n = [(n-1/2)^{1/2} \cdot \Gamma (n-1/2)] / \Gamma (n/2)$$

Y para $n \geq 30$ se reduce a :

$$G_n \approx 1 + 1 / 4 (n-1)$$

Para el caso de Student, una mejor estimación de σ es el valor $DS_{\text{corregido}}$.

$$DS_{\text{corregido}} = G_n \cdot DS = [1 + 1 / 4 \cdot (400-1)] \cdot 0,902 = 0,9026$$

CUADRO 4.7: Propiedades de la varianza:

Sean n observaciones de una variable clínica cualquiera $(x_1, x_2, x_3, \dots, x_n)$, cuya media aritmética es: \bar{x} y su desvío estándar DS , entonces:

1. Si a cada valor medido se le suma (o resta) una constante a , la nueva varianza no se altera.

$$\text{Si } z_i = x_i + a \Rightarrow DS^2_z = \sum (z_i - \bar{z})^2 / (n - 1) = \sum [(x_i + a) - (\bar{x} + a)]^2 / (n - 1)$$

O sea, $DS^2_z = DS^2_x$

Esto significa que la presencia de un error sistemático en el sistema de medición no altera la varianza, es decir, el error casual no se afecta.

2. Si a cada valor medido se le multiplica (o divide) por una constante a , la nueva varianza es igual a la anterior, multiplicada (o dividida) por el cuadrado de dicha constante.

$$\text{Si } z_i = x_i \cdot a \Rightarrow DS^2_z = \sum (z_i - \bar{z})^2 / (n - 1) = \sum [(x_i \cdot a) - (\bar{x} \cdot a)]^2 / (n - 1)$$

O sea, $DS^2_z = a^2 \cdot DS^2_x$

3. La varianza de la suma (o resta) de dos variables se obtiene con :

Sea la variable $z = x + y$

$$DS^2_z = \frac{(n_x - 1) DS^2_x + (n_y - 1) DS^2_y}{(n_x + n_y - 2)} \Leftrightarrow \mu_x \approx \mu_y$$

Las propiedades más importantes de la varianza se dan en el Cuadro 4.7, recordando que al ser mínima la suma del cuadrado de los desvíos, como se vio para la media, entonces la varianza respecto de la media aritmética, será un mínimo. Es decir, la medida de la dispersión de los valores alrededor de la media usando la varianza, será menor que si se toma respecto de otra medida de posición cualquiera.

4.13.2 Coeficiente de Variación

Es una relación entre la media y el desvío estándar de una población, o de una muestra extraída de ella. Da una idea del error relativo que se comete en el sistema de medición. No es un estadígrafo de dispersión en un sentido absoluto, sino una relación entre dos de ellos: el de dispersión respecto al de posición. Una especie de coeficiente, que mide la dispersión en términos de la media. El CV no da una buena idea de la “bondad” o “calidad” de dos series de mediciones entre sí, sino una primera aproximación. Sin embargo, en la bibliografía clínica es común encontrarlo usado como índice para decidir entre dos métodos. Al respecto, se debe remarcar que si bien es condición necesaria, no es suficiente para optar entre técnicas clínicas; la forma correcta de hacerlo se verá más adelante en la sección dedicada a ello.

Coefficiente de variación poblacional $v = \sigma / \mu$

Coefficiente de variación muestral $CV = DS / \bar{x}$

Se suele expresar en forma porcentual (CV %). Mientras en la Física, un sistema de medición con coeficientes superiores al 10% es inaceptable, en Clínica, donde las variabilidades inherentes al material de trabajo son mucho más grandes, no se puede ser tan estricto. Por eso se sugiere usar :

10% < CV% < 50% poco aceptables

1% < CV% < 10% aceptable

CV% < 1% muy aceptable

CUADRO 4.8: Cálculo del CV%.

Caso 1: Datos individuales. (usando los datos para los no infartados del Cuadro 3.1).
Del Cuadro 4.6 anterior es $DS^2 = 771 (UI/l)^2$ y el $DS = 27,77 UI/l$ con una media $\bar{x} = 30,3 UI/l$. Por lo tanto, se puede obtener el CV% con la relación:
 $CV\% = (27,77 / 30,3) \cdot 100 = 91,65 \%$

Caso 2 : Datos y sus frecuencias (valores tomados de un conteo celular - Student, 1907).
 $DS = 0,902$ y $\bar{x} = 0,68$ Entonces: $CV\% = (0,902/0,68)100 = 132,6\%$

Caso 3 : El precio de 2 acciones comunes de un Laboratorio en la Bolsa de valores tuvo un promedio, durante un mes, de \$ 1500 para las del tipo A y de \$ 5000 para las B. Sus respectivos desvíos estándar fueron de \$ 500 y \$ 300 respectivamente. Comparando en valores absolutos, la variabilidad de las acciones tipo A fue mayor que las del B por tener mayor desvío estándar. Pero, con respecto al nivel de precios, hay que comparar ambos CV. Resulta, $CV_A = 500/1500=0,033$ y $CV_B = 300/5000= 0,06$. Se deduce que el precio de la acción del tipo B fue casi dos veces más variable que las del tipo A.

Sea el caso de un experimentador que mide tres longitudes diferentes, usando la misma regla milimetrada: $A = 2 \text{ mm}$, $B = 2 \text{ cm}$ y $C = 20 \text{ cm}$. Suponiendo que hace varias mediciones, siguiendo el mismo método en condiciones análogas; es de esperar que los valores promedios se acerquen a los valores reales y que el desvío estándar se mantenga constante en los tres casos, con un valor no mayor de 1 mm. Entonces:

Caso A $CV = 1 \text{ mm} / 2 \text{ mm} = 0,5$ es decir un $CV\% = 50\%$ (inaceptable)

Caso B $CV = 1 \text{ mm} / 20 \text{ mm} = 0,05$ es decir un $CV\% = 5 \%$ (aceptable)

Caso C $CV = 1 \text{ mm} / 200 \text{ mm} = 0,005$ es decir un $CV\% = 0,5\%$ (muy aceptable)

Análogamente en análisis clínicos, si la misma técnica se usa para medir concentraciones muy diferentes, puede ocurrir que el CV% varíe demasiado y ya no sea confiable como índice para la calidad de la técnica.

4.14 Problemas propuestos

1) Decidir si son Verdaderas o Falsas las siguientes afirmaciones:

- | | | |
|--------------------------------------------------------------------------------------------------|-------|-------|
| 1) Todos los índices diagnósticos se pueden deducir con sensibilidad y especificidad. | V | F |
| 2) Los índices clínicos forman parte de los estadígrafos de dispersión. | V | F |
| 3) Lo ideal para diagnosticar es tener mucha exactitud y precisión. | V | F |
| 4) Sensibilidad y especificidad varían con el punto de corte y con la prevalencia. | V | F |
| 5) La eficiencia varía con la prevalencia, lo mismo que el Índice de Youden | V | F |
| 6) Si aumenta la sensibilidad con los puntos de corte, también lo hace la especificidad. | V | F |
| 7) Punto de corte significa lo mismo que valor referente. | V | F |
| 8) Los pacientes Sanos y con resultado (+) son los verdaderos positivos. | V | F |
| 9) Las características: Sano, enfermo, (+) y (-) son una partición del conjunto pacientes. | V | F |
| 10) Donde se cortan las curvas de especificidad y sensibilidad, el Youden es máximo. | V | F |
| 11) Ídem anterior para la eficiencia. | V | F |
| 12) La eficiencia de una prueba clínica es el cociente entre sus aciertos y el total de casos. | V | F |
| 13) Si el VP de positivos aumenta, el VPN disminuye con los puntos de corte. | V | F |
| 14) El concepto de calidad puede ser interpretado como una dualidad | V | F |
| 15) Las enfermedades del Tipo I requieren el menor número posible de fn | V | F |
| 16) Las enfermedades del Tipo II requieren el menor número posible de fp | V | F |
| 17) Las enfermedades del Tipo III son las que se curan si se detectan a tiempo. | V | F |
| 18) A veces, la clasificación de una enfermedad depende del grado de avance de la misma | V | F |
| 19) Clasificar las enfermedades en tres tipos y explicar en que consiste cada tipo | | |
| 20) Los Likelihood Ratios varían con la prevalencia de la población. | V | F |
| 21) Para una enfermedad del tipo I hay que maximizar la sensibilidad | V | F |
| 22) Para una del tipo II la especificidad y del tipo III el Youden | V | F |
| 23) Clasificar los índices clínicos y explicar cuales varían con la prevalencia y cuales no..... | | |
| 24) Incidencia y prevalencia son lo mismo. | V | F |
| 25) Los índices de riesgo y daño son los mismo. | V | F |
| 26) El Odds Ratio es el cociente entre dos tipos de Odds, el de expuestos y no expuestos. | V | F |
| 27) Hay una relación matemática entre los Odds y l riesgo, y viceversa. | V | F |
| 28) El RR es la proporción de riesgo original presente en pacientes no expuestos. | V | F |
| 29) Los riesgos absolutos son la proporción de enfermos expuestos y de no expuestos. | V | F |
| 30) El riesgo básico es la proporción de enfermos en los expuestos. | V | F |
| 31) El CER es la proporción de enfermos no expuestos. | V | F |
| 32) El EER es la proporción de enfermos expuestos. | V | F |
| 33) El RR se calcula dividiendo EER por CER. | V | F |
| 34) La reducción del riesgo absoluto (ARR) es la diferencia entre CER y EER (sin el signo). | V | F |
| 35) La reducción del riesgo relativo (RRR) es la diferencia entre 1 y RR (sin el signo). | V | F |
| 36) El RRR también se puede calcular como el cociente entre ARR y CER. | V | F |
| 37) Explicar el concepto de fidelidad o reproducibilidad..... | | |
| 38) Explicar los conceptos de nivel de concordancia y Odds de discordancia | | |
| 39) La concordancia se usa para ver si se puede reemplazar un método por otro. | V | F |
| 40) El análisis de concordancia se hace para saber cual método clínico es el mejor. | V | F |
| 41) Se puede analiza la concordancia de dos factores, y cada uno con muchos subgrupos. | V | F |
| 42) El error de apreciación de una regla milimetrada es de 0,5 mm. | V | F |
| 43) El error casual no puede ser corregido por el investigador. | V | F |
| 44) Clasificar los índices clínicos y a los estadígrafos en general | | |
| 45) Los estadígrafos de tendencia central son: | | |
| 46) La suma de los desvíos es nula. | V | F |

- | | | |
|---------------------------------------------------------------------------------------------|-------|-------|
| 47) La mediana es más inmune a valores extremos que la media aritmética. | V | F |
| 48) La suma de los cuadrados de los desvíos respecto a la media es un máximo. | V | F |
| 49) La media ponderada se usa en Física para calcular el centro de masa. | V | F |
| 50) Si a un grupo de datos se le suma una constante la varianza no cambia, pero la media sí | V | F |
| 51) El segundo cuartil, es lo mismo que el 50vo percentil y que la mediana. | V | F |
| 52) La moda es el valor de la variable que tiene mayor frecuencia. | V | F |
| 53) Si se multiplica a un grupo de datos por una constante, la varianza no se altera. | V | F |
| 54) Explicar las propiedades de la media aritmética..... | | |
| 55) Explicar las propiedades de la varianza..... | | |
| 56) El valor de la mediana coincide con uno de los datos cuando el número total es impar. | V | F |
| 57) El coeficiente de variación puede medir la calidad de un test clínico. | V | F |
| 58) El coeficiente de variación muestral es lo mismo que el poblacional. | V | F |
| 59) El rango es un índice muy pobre en comparación a la varianza. | V | F |
| 60) Explicar el concepto de fractiles | | |

2) Para los siguientes datos calcular:

			Sensibilidad =	Especificidad =
	(+)	(-)	Eficiencia =	VPN =
Sano	50	100	Índice de Youden =	Likelihood ratio (+) =
Enfermo	120	30	VPP =	Likelihood ratio (-) =

¿ En cual tipo de enfermedad conviene usar este método clínico ?

3) Para los siguientes datos calcular:

	Protegido	No-protegido	OR =	CER =
Sano	46	36	RR =	ERR =
Enfermo	18	29	ARR =	RRR =

4) Para los siguientes datos calcular:

	(+)	(-)	
(+)	400	30	Nivel de concordancia =
(-)	10	200	Odds de discordancia =

5) Para los datos siguientes, encontrar el coeficiente de variación la media, mediana y el desvío:

Frecuencia	10	20	30	15	8
Valor	2	4	6	8	10

6) Los salarios semanales de los operarios de una industria medicinal son:

Salario (\$)	50-59,9	60-69,9	70-79,9	80-89,9	90-99,9	100-109,9	110-119,9
Nº operarios	8	10	16	14	10	5	2

- Hallar: a) Los límites reales de cada clase y sus frecuencias acumuladas (ancho = \$10).
 b) La marca de cada clase.
 c) La media, mediana y moda.
 d) El desvío estándar, la varianza y el Coeficiente de variación.
 e) El percentil 5 y el 95.

f) Representar con un histograma y un polígono de frecuencias acumuladas, para calcular en forma gráfica, todos los valores anteriormente pedidos.

7) Se lanza una moneda al aire 1000 series, de 5 veces cada una, para medir el número de caras. Los resultados obtenidos se muestran en la tabla siguiente:

Número de caras (X)	0	1	2	3	4	5
Número de series (frecuencia)	38	144	342	287	164	25

Se pide: a) Graficar los datos con un diagrama de bastones y con un histograma.

b) Calcular los índices de tendencia central.

c) Calcular los índices de dispersión.

d) Dibujar el polígono de frecuencias acumuladas porcentual.

8) Para hacer un test de Sífilis se emplea la técnica 1 en 400 pacientes y se encuentran 220 pacientes que dieron un resultado (+). Los aciertos en (+) fueron 180 y 160 en (-). Con estos datos calcular la Sensibilidad y Especificidad de esta técnica. A las mismas pacientes se les aplicó la técnica 2 y resultaron 240 pacientes con (+), mientras el número de aciertos en (+) fue de 190 y en (-) 150. Decidir de manera cualitativa cual de ambas técnica es la mejor para adoptar en nuestro Laboratorio y explicar el porque.

9) Se midieron 50 alícuotas de un suero con el Kit de Glucosa marca A, y otros 50 con la marca B. Los resultados encontrados fueron:

Marca A : Entre (80-80,9) 8 veces, (81-81,9) 22 veces, (82-82,9) 15 veces y el resto (83-83,9)

Marca B : Entre (80-80,9) 6 veces, (81-81,9) 30 veces, (82-82,9) 10 veces y el resto (83-83,9)

Decidir en forma cualitativa, si hay diferencia en la exactitud y en la precisión entre las marcas. Si además se sabe que el suero utilizado tenía una concentración de 83 hallar en forma cualitativa el error sistemático de cada marca. Estudiar la precisión si en la bibliografía se dice aceptable la dispersión de la glucosa cuando llega hasta un 8%.

10) Se sabe que la concentración de coloides en el agua de un río es del 5,4 /ml. Se tomó una muestra del mismo a la que se le agregó un reactivo. Luego de un rato, se extrajo una pequeña cantidad para analizarla en una cámara de recuento. Los resultados fueron:

Nº de coloides:	2	3	4	5	6	7 (Nº /ml)
Frecuencia:	10	20	170	100	90	10

Decidir en forma cualitativa si hubo algún cambio significativo en la concentración del agua, por la acción del reactivo (Sulfato de Aluminio).

11) Los clientes de una farmacia gastaron en remedios y son otros artículos los datos que se muestran en el cuadro siguiente expresados en \$:

Remedios: Entre (10-14,9) 5 veces, (15-15,9) 25 veces, (20-25,9) 10 veces.

Otros: Entre (10-14,9) 3 veces, (15-15,9) 20 veces, (20-25,9) 12 veces, (26-30) 5 veces.

Decidir en forma cualitativa si hay diferencia entre los promedios de ventas. Determinar el desvío estándar de cada grupo y su moda.