
REVISIONES

Instituto Nacional de Angiología y Cirugía Vascular

LA EVALUACIÓN ESTADÍSTICA DE LA PRECISIÓN DE LAS MEDICIONES EN UN LABORATORIO CLÍNICO

Dr. Armando H. Seuc Jo¹ y Dr. Alfredo Aldama Figueroa²

RESUMEN: La evaluación estadística de la confiabilidad de las mediciones en un laboratorio clínico resulta de gran importancia para lograr brindar un servicio de calidad, y al mismo tiempo es un tema que ha sido objeto de controversias a nivel internacional. En este trabajo presentamos una revisión de los principales métodos estadísticos empleados para este tipo de evaluaciones. Puntualizamos las principales ventajas y desventajas de cada método, e ilustramos su aplicación e interpretación con datos provenientes de las mediciones de la presión arterial (sistólica) en el brazo. Para la presión arterial se empleó un esfigmomanómetro convencional y un medidor de velocidad de flujo ultrasónico tipo Vasflo 3C de la firma Sónica; el método empleado se corresponde con el definido según *Riva-Rocci*. Los resultados de este trabajo serán especialmente útiles para todos los laboratorios clínicos, tanto en servicios de angiología como en otros servicios, que en el marco del autocontrol de su calidad requieran evaluar la confiabilidad de sus mediciones.

DeCS: BIOMETRIA; LABORATORIOS/estadística & datos numéricos; CONTROL DE CALIDAD; QUIMICA CLINICA; PRESION SANGUINEA; MEDICIONES.

Uno de los objetivos fundamentales del control de la calidad en un laboratorio clínico es lograr consistencia en los resultados, es decir, lograr que la medición reiterada de la variable de interés produzca casi siempre el mismo resultado cuando se analiza la misma muestra o paciente, y se supone que el verdadero valor de la variable en esa mues-

tra o paciente no cambia durante el período de tiempo que dura la realización de esta serie de mediciones. Además, resulta necesario o al menos conveniente que como promedio el resultado que consistentemente se obtenga sea o se acerque suficientemente al valor verdadero. Por supuesto, nos estamos refiriendo a los clásicos conceptos

¹ Doctor en Ciencias Matemáticas. Investigador Auxiliar.

² Especialista de II Grado en Fisiología Normal y Patológica.

de precisión y exactitud de nuestras mediciones, respectivamente.¹ Numerosos trabajos describen distintas técnicas, tanto para el control interno del laboratorio como para el control externo interlaboratorio.²⁻⁵

En el Instituto Nacional de Angiología y Cirugía Vasculár el control de calidad en el laboratorio clínico es tarea prioritizada. Como parte de este esfuerzo nos propusimos revisar la evaluación estadística de la precisión de las mediciones en una de sus secciones, hemodinámica. En particular, nos propusimos revisar la evaluación estadística del grado de concordancia o consistencia entre dos técnicos de este laboratorio en la medición de la presión arterial sistólica en distintas partes del cuerpo. Para la medición de la presión arterial se empleó un esfigmomanómetro convencional y un medidor de velocidad de flujo ultrasónico tipo Vasoflo 3C de la firma Sonicaid; el método empleado se corresponde con el definido según *Riva-Rocci*, aunque en este caso los ruidos de *Korotkoff* se obtienen por medio de un detector de ultrasonido.

En este trabajo discutiremos distintos enfoques estadísticos para el análisis de este tipo de datos. A manera de ilustración utilizaremos sólo los datos de presión arterial en el brazo a partir de las mediciones hechas por los dos técnicos ya mencionados.

Distintos enfoques estadísticos para la evaluación de la concordancia entre las mediciones de dos técnicos

Antes de analizar un estudio, primero hay que diseñarlo, el cual debe tener las siguientes características esenciales:

- Reclutar un número suficiente de muestras de pacientes, preferiblemente no menos de 30.

- Tratar de que los resultados de las mediciones en estos pacientes recorran el rango de valores que usualmente se encuentra en la práctica clínica; al menos, cerca de la mitad de los pacientes deben tener valores “normales” y los restantes valores “patológicos”.
- Cada uno de los dos técnicos hará la medición de la variable en cada uno de los pacientes de manera independiente, es decir, el técnico B hará la medición en cada paciente sin saber el resultado de la medición del técnico A en ese paciente.
- Si el caso lo amerita, buscar que en cerca de la mitad de los pacientes la primera medición sea realizada por el técnico A, y en el resto de los pacientes la primera medición la realice el técnico B.

ANÁLISIS CON EL “ENFOQUE CLÁSICO USANDO LAS OBSERVACIONES ORIGINALES”

Este enfoque se utilizó, y todavía se utiliza⁶ en la comparación de métodos analíticos. En estos estudios se trata de dilucidar si los dos métodos analíticos que se comparan son intercambiables o equivalentes, de manera que uno de ellos puede ser sustituido por el otro en la práctica clínica sin que ello cause problemas.

En este enfoque se calcula el coeficiente de correlación (usualmente el de *Pearson*) entre las dos series de observaciones, y se calcula la recta de regresión de una de las dos series de observaciones, digamos la serie B, con respecto a la serie de observaciones A.

Si el coeficiente de correlación es alto y positivo (cercano a +1), y si el intercepto y la pendiente de la recta de regresión son 0 y 1 respectivamente, se dice que hay equivalencia entre los dos métodos, en este caso

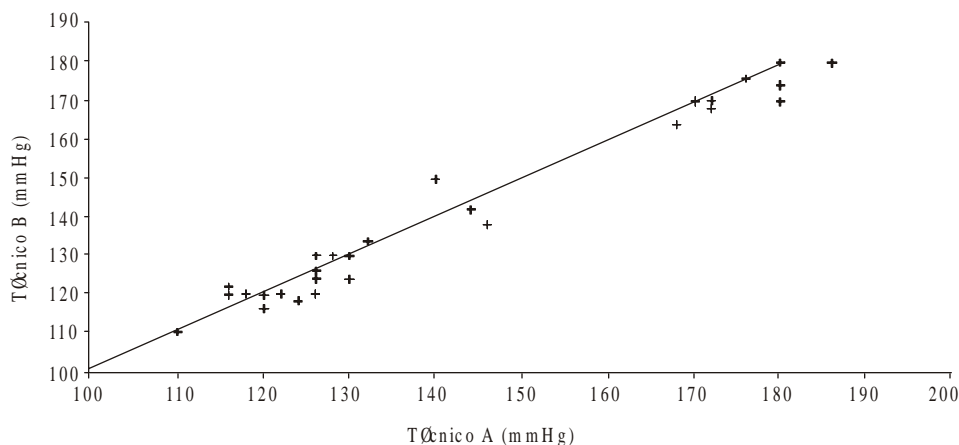


FIG. 1. Comparación de las mediciones de tensión arterial en el brazo entre los dos técnicos. Enfoque clásico. La recta mostrada tiene intercepto 0 y pendiente 1; la recta de mejor ajuste (no mostrada) tiene pendiente 0,94 e intercepto 7,4.

entre los dos técnicos de laboratorio, el técnico A y el técnico B. En esta situación los datos se dispondrían aproximadamente a lo largo de la bisectriz del primer cuadrante, en un gráfico de dispersión con las observaciones del técnico A en el eje horizontal y las del técnico B en el eje vertical (fig. 1).

La primera dificultad importante de este enfoque es que resulta poco intuitivo y demasiado lastrado por la significación estadística, en el sentido de que resulta necesario evaluar estadística y simultáneamente tres parámetros para llegar a la conclusión deseada, es decir, determinar si hay concordancia alta o no entre las dos series de observaciones. El hecho de que el coeficiente de correlación sea alto por sí solo no implica que la concordancia entre las observaciones sea alta. Por otro lado, puede que el intercepto sea suficientemente cercano a cero pero que la pendiente esté muy alejada de 1, o viceversa.

Por último, y quizás la principal dificultad metodológica de este enfoque, es que la recta de regresión que usualmente se ajusta para estimar el intercepto y la

pendiente parte del supuesto que de una de las dos series, digamos la serie B (en este caso el técnico B) representa la variable dependiente, y la otra serie A (el técnico A) representa la variable independiente; por lo tanto, estaríamos suponiendo que las observaciones del técnico B sí están sujetas a errores de medición pero las observaciones del técnico A no, lo que es totalmente falso para el caso que nos ocupa.

ANÁLISIS CON EL “ENFOQUE DE COMPARACIÓN DE MÉTODOS ANALÍTICOS”

Este enfoque surgió en buena medida como solución a los problemas del enfoque clásico descrito en la sección anterior. Su esencia consiste en el cálculo de las diferencias entre las observaciones de los dos técnicos para cada uno de los pacientes seleccionados para el estudio.⁷

Estas observaciones diferentes son en primer lugar representadas en un gráfico

donde en el eje horizontal están los promedios de las mediciones entre los dos técnicos, y en el vertical están las observaciones diferentes. Si hay concordancia perfecta entre los dos técnicos, los puntos del gráfico de dispondrán a lo largo de una línea recta paralela al eje horizontal y que cruza el eje vertical en cero. Este gráfico permite detectar rápidamente si la concordancia entre los dos técnicos es la misma a lo largo del rango de valores de la variable, o si por el contrario varía según sean los valores (bajos o altos) de la variable (fig. 2).

En el gráfico resulta útil representar la recta de mejor ajuste para la estimación de las observaciones diferentes a partir del “verdadero valor” de la presión arterial. El cálculo de esta recta se justifica, pues la evaluación de la precisión de las mediciones del laboratorio es compatible con la estimación de la diferencia en las mediciones de los dos técnicos a partir del “verdadero valor” de la presión arterial. Si la recta coincide con el eje horizontal entonces como promedio los dos técnicos coinciden, cualquiera que sea el “verdadero valor” de la presión arterial; rectas con pendientes e interceptos diferentes se interpretan correspondientemente.

Estas observaciones diferentes son finalmente la base para el cálculo de la media

y la desviación estándar, y para el cálculo de intervalos de confianza para las observaciones diferentes. Si la media de las observaciones diferentes no están muy alejadas de cero y si el intervalo de confianza (para las observaciones) es suficientemente pequeño, podemos concluir que la concordancia entre los dos técnicos es adecuada. Por supuesto, esta conclusión deberá tomar en cuenta las características de la medición realizada, qué enfermedad o entidad médica están involucradas, y el posible efecto en los pacientes de un error en el resultado, entre otros aspectos.

ANÁLISIS CON EL “ENFOQUE DE COMPONENTES DE LA VARIANZA”

El análisis de componentes de la varianza se ha utilizado en diferentes circunstancias, en particular para la cuantificación del error de medición.^{8,9} El modelo básico plantea que la medición realizada en el sujeto “i”, X_i , es igual a:

$$X_i = A_i + e_i$$

Donde A_i representa el verdadero valor de la variable en el “i-ésimo” sujeto, y e_i re-

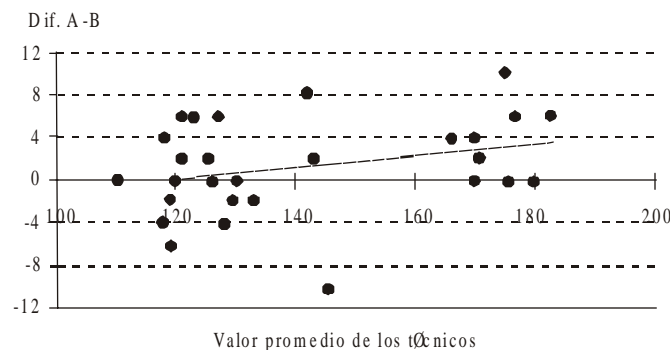


FIG. 2. Diferencias entre las tensiones arteriales en el brazo, Téc. A-Téc B. Media = 1,27. IC 95 % para las dif. = (-7,35-9,89).

presenta el error de medición para el “i-ésimo” sujeto. Para cada sujeto “i” el término e_i es una variable aleatoria que toma distintos valores si repetimos la medición en el “i-ésimo” sujeto un número grande de veces. Estamos suponiendo que la distribución de e_i es la misma para todo “i”.

En este modelo se supone que las variables A_i y e_i son independientes, por lo que la varianza de las observaciones X_i es la suma de las varianzas de A_i y de e_i , es decir:

$$\sigma^2 = \sigma_A^2 + \sigma_e^2$$

De lo que se trata entonces es de estimar los componentes de la varianza, s_A^2 y s_e^2 , y de evaluar su contribución relativa a la variabilidad total en términos porcentuales.¹⁰⁻¹²

En el ejemplo que nos ocupa el error de medición, representado por s_e^2 , comprende sólo la variabilidad entre los dos técnicos (variabilidad intertécnicos) pues la variabilidad intratécnico no fue contemplada en el diseño del estudio (para ello tendríamos que haberle pedido a cada uno de los técnicos que hicieran al menos dos mediciones de la presión arterial en cada uno de los pacientes).

ANÁLISIS CON EL “ENFOQUE DE CONCORDANCIA ENTRE JUECES”

Este enfoque supone que las mediciones de los técnicos sobre las muestras (o pacientes) se han categorizado (por ejemplo en “normal”, “patológico” y “dudoso”), y se evalúa hasta qué punto coinciden las clasificaciones hechas por los técnicos (dos en este caso) mediante estadígrafos como el conocido coeficiente *Kappa*.

Los datos a los que se les aplica el coeficiente *Kappa* se representan en una ta-

bla como la que sigue (usando las tres categorías antes mencionadas):

		Técnico B			total
		“N”	“P”	“D”	
Técnico A	“N”	n11	n12	n13	n1.
	“P”	n21	n22	n23	n2.
	“D”	n31	n32	n33	n3.
Total		n.1	n.2	n.3	n

“N”: normal

“P”: patológico

“D”: dudoso

El coeficiente *Kappa* cuantifica la concordancia observada por encima de la concordancia que es de esperar sólo por azar; la fórmula más sencilla^{13,14} para *Kappa* es:

$$Kappa = (p_0 - p_e) / (1 - p_e)$$

donde p_0 es la proporción de sujetos en los cuales hay concordancia entre los dos evaluadores, y p_e es la proporción de concordancia esperada por azar, dada por la expresión:

$$p_e = \sum_{i=1, k} (p_i \cdot x \cdot p_i)$$

$$p_i = n_{i.} / n$$

$$p_i = n_{.i} / n$$

La pérdida de información que representa este enfoque (al categorizar una variable originalmente cuantitativa) puede ser una limitante importante, por lo que debe valorarse como una herramienta generalmente complementaria a alguna de las restantes opciones.

Resultados

Se reclutaron 30 pacientes, en cada uno de los cuales se midió la presión arterial en el brazo (TBRAZO). Esta variable se midió en cada uno de los 30 pacientes por separado

y de forma independiente por los dos técnicos objeto de evaluación (técnico A y técnico B). Como resultado, la matriz de datos inicial tenía 30 filas (los 30 pacientes) y 2 columnas (la medición del técnico A y la del técnico B), todo lo cual se puede ver en los datos del anexo.

ANÁLISIS CON EL “ENFOQUE CLÁSICO USANDO LAS OBSERVACIONES ORIGINALES”

Los resultados siguiendo este enfoque fueron:

- a) Coeficiente de correlación $p = 0,987$ ($p < 001$);
- b) intercepto $a = 7,38$; IC 95 % para $a = (-1,32, 16,08)$; y
- c) pendiente $b = 939$; IC 95 % para $b = (879, 999)$.

En la figura 1 se observa hasta qué punto concuerdan las observaciones de los dos técnicos.

Los anteriores resultados destacan que la pendiente estimada b es significativamente menor que 1 (el IC 95 % no contiene el valor 1), lo que significa que el sesgo (diferencia) entre los dos técnicos no es uniforme a lo largo del rango de observaciones considerado. Se observa también en la figura 1 que para valores altos de la presión arterial el técnico A tiende a dar resultados por encima de los resultados del técnico B.

ANÁLISIS CON EL “ENFOQUE DE COMPARACIÓN DE MÉTODOS ANALÍTICOS”

Según este enfoque, la figura 2 muestra en el eje horizontal el promedio de las

observaciones (se toma como una estimación del “valor verdadero”), y en el eje vertical las diferencias observadas entre las mediciones de los dos técnicos.

Se observa claramente que para valores altos de la presión arterial en el brazo, el técnico A tiende a obtener resultados que sistemáticamente están por encima de las mediciones del técnico B, sin embargo, para valores bajos de la presión arterial en el brazo los resultados de los dos técnicos como promedio coinciden. La variabilidad de las observaciones diferentes están dadas por el IC 95 % = $(-7,35, 9,89)$, es decir, aproximadamente un 95 % de las diferencias en las mediciones de los dos técnicos se encuentran entre $-7,35$ y $9,89$ mmHg. Se observa también que no hay pacientes con tensiones arteriales intermedias (entre 146 y 164 mmHg), lo cual deja abierta una interrogante acerca de la verdadera concordancia entre los dos técnicos para este rango de valores.

En la figura 2 se representa la recta de mejor ajuste para la estimación de las observaciones diferentes a partir de las observaciones promedio. La pendiente positiva que se aprecia en la misma nos indica que el promedio de las diferencias entre las mediciones de los dos técnicos varía (aumenta) con el “verdadero valor” de la presión arterial (que se ha asumido como el promedio de sus mediciones).

ANÁLISIS CON EL “ENFOQUE DE COMPONENTES DE LA VARIANZA”

Para el cálculo de los componentes de la varianza se utilizó el paquete estadístico SPSS/PC for Windows versión 5. La manera más fácil de obtener las estimaciones necesarias con este paquete es mediante el modelo de análisis de varianza con diseño

factorial simple. En este modelo los datos se deben disponer de manera que todas las observaciones estén en una columna, digamos la primera, mientras que en la otra se indica la variable que representa el factor principal, en este caso los pacientes; el resto de la variabilidad se considera en este modelo aleatoria, aspecto evidenciado en la matriz de datos modificada empleada para este enfoque en el anexo.

Los resultados del ANOVA se presentan en la tabla 1. Luego S^2_e (estimación del componente intrapacientes) = 9,8 y s^2_A (estimación del componente entre pacientes) = $(1267,46 - 9,8) / 2 = 628,83$; es decir, la variabilidad intrapacientes (debida a los técnicos) se estima es un $9,8 / (9,8 + 628,83) = 1,5\%$ de la variabilidad total en las mediciones de la presión arterial en el brazo, mientras que la variabilidad entre pacientes contribuye en un $(100 - 1,5)\% = 98,5\%$ a la variabilidad total en las mediciones de la presión arterial en el brazo.

Esta proporción de 1,5 % parece pequeña o despreciable en términos exclusivamente numéricos, pero la verdadera connotación de este resultado deberá siempre obtenerse a partir de otras consideraciones no numéricas tales como las clínicas, prácticas, éticas, etc.

ANÁLISIS CON EL “ENFOQUE DE CONCORDANCIA ENTRE JUECES”

En este trabajo decidimos categorizar las mediciones de la presión arterial en dos categorías:

- a) Hasta 140 mmHg (normal), y
- b) más de 140 mmHg (patológico).

Los resultados de la concordancia entre las clasificaciones de los dos técnicos se presentan en la tabla 2. El coeficiente Kappa para estos datos tiene un valor de:

$$\text{Kappa} = (0,9333 - 0,5200) / (1 - 0,5200) = 0,8611$$

valor que se considera representativo de concordancia “excelente” entre los dos técnicos, según *Fleiss*.¹⁴ Por supuesto, este resultado depende de cómo se haya categorizado la medición objeto de análisis.

Podemos observar en la tabla 2 que aunque en este caso hay concordancia marginal perfecta (ambos técnicos identifican a 12 pacientes como patológicos) la concordancia sujeto a sujeto no es perfecta; en 2 de los 30 sujetos los técnicos obtienen resultados discrepantes. El coeficiente *Kappa* cuantifica convenientemente la concordancia

TABLA 1. Resultados del análisis de varianza

Source of Variation	Sum of Squares	DF	Mean Square
Main Effects	36756,333	29	1267,460
Pacientes	36756,333	29	1267,460
Explained	36756,333	29	1267,460
Residual	294,000	30	9,800
Total	37050,333	59	627,972

60 cases were processed.
0 cases (. 0 pct) were missing.

sujeto a sujeto, que es realmente la que nos interesa.

TABLA 2. **Concordancia entre los técnicos**

	Técnico B		Total
	Normal	Patológico	
Normal	17	1	18
Patológico	1	11	12
Total	18	12	30

En el caso que trabajemos con datos cuantitativos también podemos encontrar una situación similar de concordancia marginal (ó global) pero discordancia a nivel de sujeto, es decir, puede que los dos técnicos obtengan el mismo resultado promedio para la muestra de 30 pacientes, digamos 135 mmHg; sin embargo, a nivel de sujeto podríamos observar una discordancia significativa entre los dos técnicos para todos y cada uno de los pacientes.

Discusión

Los 4 enfoques que hemos considerado en este trabajo para evaluar estadísticamente la precisión en las mediciones de un grupo de dos técnicos pertenecientes a

un mismo laboratorio parten todos del supuesto de que los dos técnicos tienen el mismo nivel o jerarquía. Si por el contrario, uno de los técnicos puede considerarse como “estándar” con respecto al cual se comparan los resultados del otro técnico “de prueba”, la evaluación estadística deberá hacerse en términos de sensibilidad y especificidad. Alternativamente podemos hacerlo mediante el procedimiento de calibración, donde inicialmente se calcula una regresión que haga predicciones de la medición del técnico B (“de prueba”) a partir del conocimiento de la medición hecha por el técnico A (“estándar”). Posteriormente se utiliza esta regresión en sentido inverso, es decir, para predecir el resultado del técnico “estándar” a partir del resultado del “de prueba”.¹⁵

Para situaciones como la descrita en este trabajo, los cuatro enfoques aquí esbozados se han empleado con mayor o menor éxito; en nuestra opinión son adecuados todos, excepto el primero, el “enfoque clásico usando las observaciones originales”. Los restantes tres pueden ser hasta cierto punto complementarios, pero en la mayoría de los casos el “enfoque de comparación de métodos analíticos” debe ser suficiente.

ANEXO

Matriz de datos (modificada).
ANOVA con diseño factorial simple.
(paquete estadístico SPSS for Windows v.5.0.)

TBRAZO (mmHg)	paciente	TBRAZO (mmHg)	paciente
144	1	146	16
142	1	138	16
116	2	116	17
122	2	120	17
132	3	140	18

ANEXO (continuación)

TBRAZO (mmHg)	paciente	TBRAZO (mmHg)	paciente
134	3	150	18
172	4	172	19
168	4	170	19
186	5	180	20
180	5	180	20
110	6	110	21
110	6	110	21
120	7	118	22
120	7	120	22
130	8	126	23
130	8	126	23
168	9	170	24
164	9	170	24
126	10	126	25
124	10	130	25
124	11	126	26
118	11	120	26
120	12	122	27
116	12	120	27
130	13	128	28
124	13	130	28
176	14	180	29
176	14	170	29
180	15	180	30
174	15	180	30

SUMMARY: The statistical evaluation of the reliability of measurements in a clinical laboratory is very important to render a service of quality and, at the same time, it is a topic that has been discussed at the international level. In this paper we made a review of the main statistical methods used for this type of evaluations. The fundamental advantages and disadvantages of each method are stressed and their application and interpretation are illustrated with data taken from the measurements of the arterial pressure (systolic) in the arm. A conventional sphygmomanometer and a Vasoflo 3C ultrasonic flow speed meter from the Sonicaid firm were used for measuring arterial pressure. The method used matches with the one defined by Riva-Rocci. The results of this paper will be specially useful for all clinical laboratories and for the angiology services and other services that need to evaluate the reliability of their measurements within the framework of the autcontrol of their quality.

Subject headings: **BIOMETRY; LABORATORIES/statistics & numerical data; QUALITY CONTROL; CHEMISTRY, CLINICAL; BLOOD PRESSURE; MEASUREMENTS.**

Referencias bibliográficas

1. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med* 2000;30:1-15.
2. Kelly DT, Kelly ME. Proposals and current practices in quality control. *Am J Med Tech* 1981;47:56-9.
3. Whitehead TP. Quality control in clinical chemistry. London: John Wiley, 1977:206-314.
4. Cekan SZ, Sufi SB, Wilson EW. Internal quality control for assays of reproductive hormones: guidelines for laboratories. Geneva: World Health Organization, 1993.
5. Cekan SZ, Sufi SB, Wilson EW. Principles of external quality assessment: a laboratory manual. Geneva: World Health Organization, 1995.
6. Pollock MA, Jefferson SG, Kane JW, Lomax K, MacKinnon G, Winnard CB. Method comparison - a different approach. *Ann Clin Biochem* 1992;29:556-60.
7. Nexo E. Evaluation of novel assays in clinical chemistry: quantification of plasma total homocysteine. *Clin Chem* 2000;46(8 Pt 1):1150-6.
8. Bartko JJ. The interclass correlation coefficient as a measure of reliability. *Psychol Rep* 1966;19:3-11.
9. Bartko JJ. General methodology II, measures of agreement: a single procedure. *Stat Med* 1994;13:737-45.
10. Healy MJR, Osborn J, Hills M. Manual of medical statistics. London: London School of Hygiene and Tropical Medicine, 1987;vol 2:23-74.
11. Snedecor GW, Cochran WG. Statistical methods. 7 ed. Iowa: The Iowa State University Press, 1980:312-442.
12. Armitage P. Statistical methods in medical research. London: Blackwell Scientific 1994:156-559.
13. Seuc AH. Cochran Q test with Turbo BASIC. *Comput. Methods Prog Biomed* 1995;46:29-34.
14. Fleiss JL. Statistical methods for rates and proportions. 2 ed. New York: John Wiley, 1981:211-36.
15. Miller JN. Basic statistical methods for analytical chemistry. Part 2. Calibration and Regression Methods: a review. *Analyst* 1991;116:3-14.

Recibido: 10 de marzo de 2001. Aprobado: 27 de mayo de 2001.

Dr. *Armando H. Seuc Jo*. Instituto Nacional de Angiología y Cirugía Vascular. Calzada del Cerro No. 1551 esq. Domínguez, Cerro, Ciudad de La Habana. CP 12000. Teléf: 576493.