

ANUARIO DE PSICOLOGÍA
Núm. 36/37 - 1987 (1-2)

ALGUNAS CONSIDERACIONES SOBRE
LA UTILIZACIÓN DEL COEFICIENTE r
DE PEARSON COMO ÍNDICE DE ACUERDO
ENTRE OBSERVADORES

ARTURO SILVA RODRÍGUEZ
Escuela Nacional de Estudios Profesionales
Universidad Nacional Autónoma de México

Arturo Silva Rodríguez
Escuela Nacional de Estudios Profesionales
Universidad Nacional Autónoma de México
IZTACALA. C.P. 54090 México

Uno de los temas recientemente abordados es el de los principios psicométricos de las técnicas de evaluación conductual. Aunque si bien es cierto que existía ya una teoría psicométrica muy desarrollada a partir de modelos estructural-diferenciales, se empezó a cuestionar si todas estas técnicas desarrolladas bajo la tutela teórica de la psicometría clásica eran adecuadas para la evaluación conductual, desarrollada ésta bajo marcos teóricos totalmente diferentes.

Aunque algunas técnicas de evaluación conductual estructuralmente parecen similares a las técnicas tradicionales, difieren significativamente en sus supuestos y objetivos, puesto que como señalan Nelson y Hayes (1981), los métodos de la evaluación conductual consisten en la identificación de unidades de respuesta significativas y las variables que las controlan (tanto orgánicas como ambientales), con el propósito de entender y modificar la conducta; mientras que la evaluación tradicional generalmente sostiene que la conducta es el resultado de variables intraorgánicas, más aún, supone que estas variables son mentales en lugar de físicas.

Este trabajo tiene como objetivo el hacer un análisis del coeficiente de correlación r de Pearson como una técnica utilizada para estimar las propiedades psicométricas de algunos instrumentos de evaluación en situaciones de observación conductual directa.

EL COEFICIENTE DE CORRELACIÓN r DE PEARSON COMO PUNTUACIÓN ESTÁNDAR

El coeficiente r de Pearson es el estimador por excelencia de la confiabilidad de los instrumentos en la psicometría clásica. Dos son las suposiciones básicas de confiabilidad desde este punto de vista. La primera se refiere a la obtención de puntajes idénticos, y la segunda se refiere a la obtención de idénticas posiciones relativas en las distribuciones de los puntajes obtenidos por los individuos en medidas sucesivas.

La forma de ejemplificar claramente estas dos suposiciones es calculando el coeficiente de correlación r de Pearson a través de la tipificación de los puntajes de los individuos, ya que por medio de este procedimiento es posible determinar "...hasta qué punto los mismos individuos o sucesos ocupan

la misma relación relativa respecto a dos variables" (Haber y Runyon, 1973, p. 121).

Supongamos que aplicamos el mismo instrumento al mismo grupo después de cierto tiempo, y encontramos los siguientes resultados.

Tabla 1. Resultados hipotéticos de la aplicación de un mismo test en dos ocasiones diferentes a un grupo de cinco sujetos.

Ss	1ª MEDICIÓN (X)	2ª MEDICIÓN (Y)	Z _x	Z _y
1	20	20	-1.29	-1.29
2	25	25	-0.75	-0.75
3	30	30	-0.22	-0.22
4	45	45	1.40	1.40
5	40	40	0.86	0.86

Como se puede observar en la Tabla 1, los cinco sujetos en ambas ocasiones de medida obtienen el mismo puntaje absoluto y, por consiguiente, al obtener las puntuaciones tipificadas para el cálculo del coeficiente de correlación, cada individuo obtiene exactamente la misma calificación estándar en ambas aplicaciones del test.

Como en este caso hipotético se encontró un coeficiente $r = 1.00$, se puede afirmar que existen evidencias de que dicha prueba es confiable, puesto que presenta estabilidad temporal a través de aplicaciones sucesivas.

Sin embargo, como se mencionó anteriormente, dentro de este concepto de confiabilidad el interés se centra principalmente en la posición relativa de los puntajes de los individuos en las dos distribuciones, independientemente de las diferencias en los puntajes absolutos obtenidos en las distintas situaciones de medida.

Para ejemplificar esto, supongamos que administramos durante la misma sesión dos formas equivalentes de un test y obtenemos los siguientes resultados:

Tabla 2. Resultados hipotéticos de la aplicación de dos formas paralelas de un test a un grupo de cinco sujetos.

Ss	1ª FORMA (X)	2ª FORMA (Y)	Z _x	Z _y
1	20	24	-1.29	-1.29
2	25	29	-0.75	-0.75
3	30	34	-0.22	-0.22
4	45	49	1.40	1.40
5	40	44	0.86	0.86

En la Tabla 2 se observa que los cinco sujetos en las dos formas del test obtienen puntajes absolutos diferentes, sin embargo, las puntuaciones tipificadas en ambas formas son las mismas, por lo que al igual que en el ejemplo anterior, la confiabilidad estimada a través del coeficiente r de Pearson es igual a 1.00, pudiéndose afirmar que existen evidencias que llevan a suponer que ambas formas del test son equivalentes.

En conclusión, de los dos ejemplos anteriores se desprende que el supuesto fundamental de la confiabilidad en la psicometría clásica, es el comparar en qué medida un grupo de sujetos mantienen la misma posición relativa en las dos distribuciones, cuando es aplicado un instrumento de evaluación en diferentes situaciones de medida.

De esta forma, en los diferentes diseños para evaluar la confiabilidad en la psicometría clásica, lo que interesa es comparar la posición relativa de cada individuo con respecto al grupo, debido a que siempre el análisis del puntaje absoluto obtenido por un individuo dependerá del status relativo que guarde este puntaje con respecto al grupo "...todos los procedimientos psicométricos clásicos dependen de la consistencia en las diferencias entre individuos para demostrar lo adecuado de la medida..." (Cone, 1981, p. 42).

En la evaluación conductual, sin embargo, el acuerdo entre observadores como indicador de confiabilidad, sólo supone la existencia de medidas absolutas idénticas entre los observadores en los diferentes momentos m_i de observación, independientemente de la posición relativa que tengan las medidas obtenidas en los diferentes m_i de cada observador.

Supongamos que dos observadores han registrado la ocurrencia de la conducta de lavarse las manos de un paciente obsesivo compulsivo, a lo largo de cinco días, habiendo anotado cada uno la frecuencia de aparición de tal segmento conductual durante las ocho horas de la jornada laboral, obteniendo los siguientes resultados:

Tabla 3. Datos hipotéticos de la ocurrencia de lavarse las manos de un paciente obsesivo-compulsivo durante cinco días.

Día	OBSERVADOR 1 (X)	OBSERVADOR 2 (Y)	RANGO (x)	RANGO (y)	Z_x	Z_y
1	30	30	4	4	0.71	0.71
2	15	15	1	1	-1.41	-1.41
3	25	25	3	3	0.00	0.00
4	35	35	5	5	1.41	1.41
5	20	20	2	2	-0.71	-0.71

Como se puede ver en la Tabla 3, en todos los días los observadores concuerdan en el número de veces que el segmento conductual aparece durante la jornada laboral; por consiguiente, la posición ordinal de aparición de la conducta evaluada a través del rango, durante los cinco días para cada una

de las distribuciones de los observadores, es la misma, así como también la posición relativa de los puntajes dentro de cada distribución (puntaje z). De tal manera, que al calcular la confiabilidad de acuerdo entre observadores por medio del r de Pearson el valor será igual a 1.00, el cual indicará un acuerdo perfecto entre los observadores.

Sin embargo, cuando existen cambios en las medidas absolutas entre los observadores, manteniéndose constante la posición ordinal y la relativa, el coeficiente r de Pearson produce índices de acuerdo falsos. Por ejemplo, supongamos que en lugar de haberse obtenido los datos de la Tabla 3, se hubieran encontrado los siguientes resultados:

Tabla 4. Datos hipotéticos de la ocurrencia de lavarse las manos de un paciente obsesivo-compulsivo durante cinco días.

Día	OBSERVADOR 1 (X)	OBSERVADOR 2 (Y)	RANGO (x)	RANGO (y)	Z_x	Z_y
1	30	35	4	4	0.71	0.71
2	15	20	1	1	-1.41	-1.41
3	25	30	3	3	0.00	0.00
4	35	40	5	5	1.41	1.41
5	20	25	2	2	-0.71	-0.71

La Tabla 4 muestra que no hubo acuerdo entre los dos observadores en ningún día sobre el número de veces que apareció el segmento conductual, puesto que el observador 2 estuvo sobreestimando consistentemente la ocurrencia de la conducta. No obstante, tanto la posición ordinal como la posición relativa de las medidas dentro de cada distribución son las mismas, por lo que si se calcula el r de Pearson indicará un acuerdo perfecto entre los observadores, con un valor estimado igual a 1.00; esto es debido a que como se mencionó anteriormente, el coeficiente de correlación sólo compara las posiciones relativas de los puntajes con respecto a las dos distribuciones, independientemente de los valores absolutos y, como en este caso, las posiciones relativas son iguales, el r de Pearson sobreestima el acuerdo entre observadores.

De esto se deriva, que si bien es cierto lo que menciona Hartman (1977), de que una r de Pearson igual a 1.00 indica perfecto acuerdo entre observadores en el sentido de puntajes estándar idénticos, esta afirmación no tiene ninguna validez, puesto que en la evaluación conductual el concepto de confiabilidad no supone igualdad en los puntajes estándar, sino en los puntajes absolutos de ambos observadores; esto es, $Y = X$.

EL COEFICIENTE DE CORRELACIÓN r DE PEARSON COMO UN MODELO LINEAL

En la psicometría clásica el concepto de confiabilidad, además de la igualdad entre las mediciones absolutas y/o la igualdad entre posiciones relativas en las dos distribuciones, supone que existe una relación lineal entre ambos conjuntos de medida. "El coeficiente de correlación indica entonces el grado en que los puntajes en una de las variables mantiene una relación lineal sistemática con los puntajes en la otra". (Magnusson, 1975, p. 47).

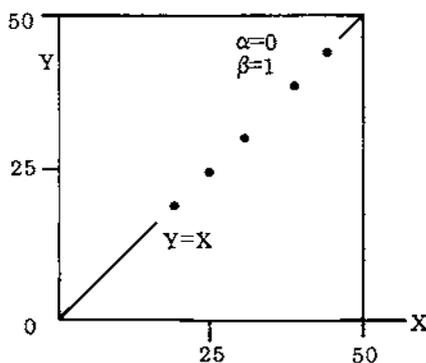
De esta manera, el supuesto de linealidad entre los puntajes de medidas sucesivas estará representado por el modelo matemático de:

$$Y = \alpha + \beta X$$

Conforme a la suposición de linealidad, la forma de evaluar la confiabilidad es analizando la dispersión de los puntos de ambas mediciones con respecto a un modelo lineal estimado a partir de los datos obtenidos. Cuando los puntos están ampliamente dispersos alrededor del modelo lineal estimado, el coeficiente de correlación r de Pearson es pequeño, mientras que cuando todos los puntos del diagrama de dispersión caen sobre el modelo estimado el coeficiente r es igual a 1.00.

Hagamos ahora una interpretación del coeficiente de correlación r de Pearson, calculando para los datos hipotéticos de las Tablas 1 y 2, el modelo matemático lineal.

Figura 1. Resultados hipotéticos de la aplicación de un mismo test en dos ocasiones diferentes a un grupo de cinco sujetos.

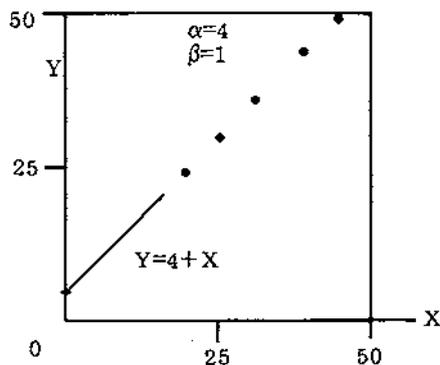


Al evaluar la relación lineal de las dos aplicaciones del test, a través de los mínimos cuadrados, los estimadores de los parámetros α y β son cero y uno respectivamente. El modelo lineal estimado quedaría:

$$Y = 0 + 1(X); \text{ es decir, } Y = X$$

El diagrama de dispersión de la Tabla 2 que representa los resultados hipotéticos de la aplicación de dos formas paralelas de un test, se muestra a continuación:

Figura 2. Resultados hipotéticos de la aplicación de dos formas paralelas de un test a un grupo de cinco sujetos.



En la Figura 2 se ve que al igual que los puntos de la Figura 1, éstos también caen sobre una línea recta, sólo que aquí los puntos se han desplazado un poco más hacia arriba en el eje de la ordenada. Este es el efecto característico que sucede en el modelo lineal cuando los valores absolutos en ambas distribuciones no son los mismos.

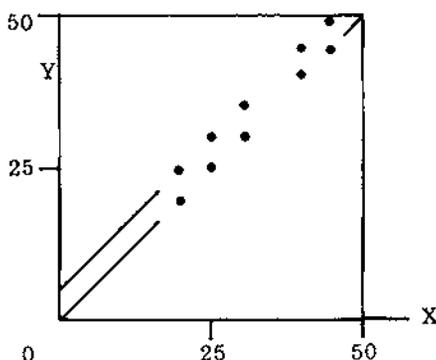
Sin embargo, puesto que el coeficiente r de Pearson visto como un modelo lineal, lo que evalúa es el grado en que los pares de valores dan como resultado puntos que caen sobre una recta, y en el diagrama de dispersión de la Figura 2 todos los puntos caen sobre la recta, el coeficiente de correlación es también igual a 1.00.

Si estimamos los valores de α y β para la relación lineal de la Figura 2, éstos serán 4 y 1 respectivamente. El modelo lineal estimado quedaría:

$$Y = 4 + X$$

En la Figura 3 se representan conjuntamente ambas relaciones lineales:

Figura 3. Representación conjunta de las relaciones lineales obtenidas de las Tablas 1 y 2.



En esta Figura se observa que la única diferencia en estas dos relaciones lineales es en el punto en donde cruzan el eje de las ordenadas, puesto que ambas tienen el mismo valor estimado de β , que es de 1.00. En la primera relación lineal la ordenada al origen es igual a 0 y en la segunda igual a 4.

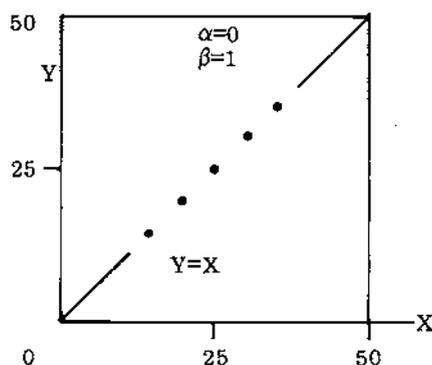
De esta manera, se puede concluir que si bien es cierto que en la psicometría clásica el concepto de confiabilidad también presupone una relación lineal entre ambos conjuntos de mediciones, ésta puede ser cualquier tipo de relación lineal positiva, puesto que como señala Magnusson (1975) la confiabilidad puede tomar valores entre cero y uno, pero jamás puede tener valores negativos. De esta forma, el supuesto de linealidad en la confiabilidad, presupone cualquier tipo de recta con pendiente positiva.

Dentro del campo de la evaluación conductual, algunos autores (Anguera, 1983; Hartmann, 1977; Kent y Foster, 1977) han señalado que es necesario que exista una linealidad en los puntos correspondientes a los valores obtenidos en las sesiones de observación. De esta manera, al igual que en la psicometría clásica, el concepto de confiabilidad dentro de la evaluación conductual supone también la existencia de una linealidad entre el conjunto de datos de los dos observadores.

Igualmente, la relación lineal específica que se presupone es del tipo positiva, puesto que como Hartmann (1977) menciona, coeficientes de confiabilidad de acuerdo entre observadores negativos no son posibles.

Similarmente, la forma de evaluar la linealidad de los datos de los dos observadores es por medio del diagrama de dispersión, obtenido a partir de la representación de los pares de datos. Los datos hipotéticos de la Tabla 3, de la ocurrencia de lavarse las manos de un paciente obsesivo compulsivo, se muestran en la siguiente figura:

Figura 4. Datos hipotéticos de la ocurrencia de lavarse las manos de un paciente obsesivo-compulsivo durante cinco días.

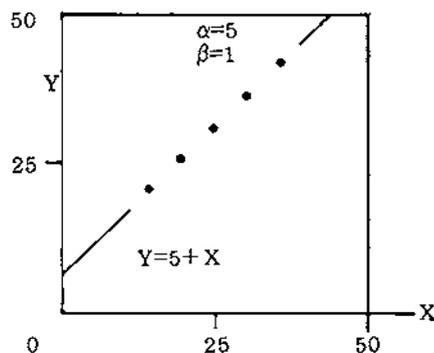


El diagrama de dispersión de la Figura 4 muestra que, debido a que ambos observadores acordaron perfectamente sobre la frecuencia de ocurrencia de la conducta durante los cinco días, es decir, $Y = X$, estos pares de valores dan como resultado que todos los puntos estén sobre una línea recta; además, que la correlación entre los pares de valores sea igual a 1.00.

Los parámetros estimados de esta relación lineal dan como resultado un $\alpha = 0$ y $\beta = 1$, de lo cual se desprende que para afirmar la existencia de un perfecto acuerdo entre observadores, es necesario que los parámetros del modelo lineal tengan una ordenada al origen igual a cero y una pendiente igual a uno.

Con respecto al diagrama de dispersión de los datos hipotéticos de la Tabla 4, en donde a simple vista se observa que no hubo absoluto acuerdo entre los observadores, es decir, $Y \neq X$, sin embargo los pares de valores dan también como resultado una nube de puntos que caen sobre una línea recta.

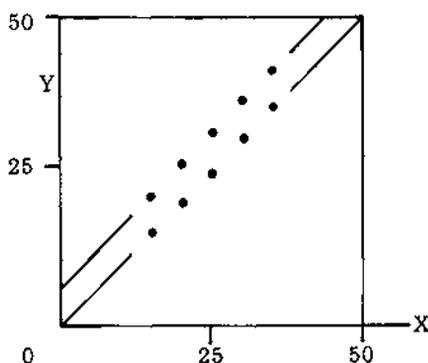
Figura 5. Datos hipotéticos de la ocurrencia de lavarse las manos de un paciente obsesivo-compulsivo durante cinco días.



De esta forma se puede explicar el porqué al calcular el coeficiente de correlación r de Pearson, éste es igual a 1.00.

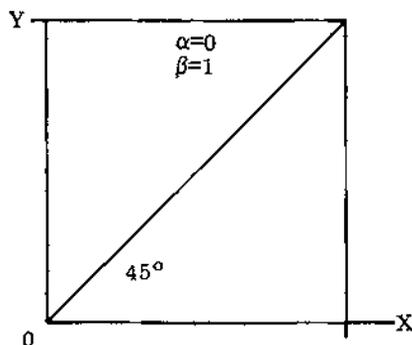
Sin embargo, al comparar las funciones lineales obtenidas en ambas situaciones hipotéticas de observación, como lo muestra la Figura 6, se observa que si bien ambas funciones son lineales, éstas difieren en el punto de intersección con el eje de las ordenadas, puesto que los datos de la Tabla 3 arrojaron un $\alpha = 0$ y $\beta = 1$, mientras que los datos de la Tabla 4 dieron un $\alpha = 5$ y $\beta = 1$.

Figura 6. Representación conjunta de las funciones lineales obtenidas en ambas situaciones hipotéticas de observación.



De lo anterior, se desprende en conclusión que si bien el concepto de confiabilidad en la evaluación conductual, al igual que en la psicometría clásica, supone la existencia de una linealidad entre el conjunto de datos de los dos observadores, esta suposición de linealidad no presupone cualquier tipo de relación lineal, sino que es una relación especial que parte del origen y su inclinación con respecto al eje de las $X'x$ es de 45° , cuyos parámetros óptimos son $\alpha = 0$ y $\beta = 1$; como se muestra en la Figura siguiente:

Figura 7. Representación de la relación lineal especial que presupone el concepto de confiabilidad en la evaluación conductual.

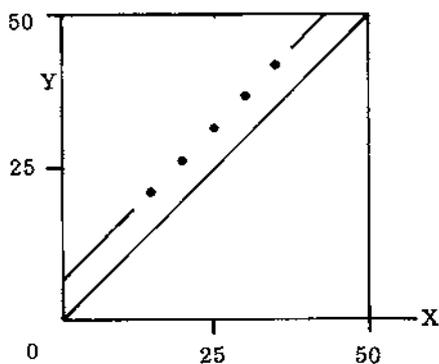


De tal modo, debido a que el coeficiente de correlación r de Pearson evalúa cualquier tipo de relación lineal, su uso con cualquier otra recta que se aleje ya sea tanto del origen como de los 45° , pondrá en tela de juicio la validez de su uso como estimador de la confiabilidad entre observadores.

Sin embargo, es posible determinar si el coeficiente de correlación r de Pearson es un índice adecuado de acuerdo entre observadores, a través de considerar la relación entre los dos conjuntos de datos, como un modelo de regresión lineal simple, en el cual, ambas variables (los puntajes del observador 1 y los puntajes del observador 2) son consideradas variables aleatorias. Conforme a esta aproximación, es factible probar la hipótesis de que $\alpha = 0$ y $\beta = 1$; en caso de comprobarse esta hipótesis, existirían evidencias que nos permitirían afirmar que el coeficiente r de Pearson es un índice adecuado para evaluar la confiabilidad entre observadores.

Esto se deriva del hecho de que como se mencionó anteriormente, la linealidad dentro del concepto de confiabilidad entre observadores presupone medidas absolutas idénticas en ambos conjuntos de datos, es decir, $Y = X$. De tal manera, que un error aditivo en uno de los observadores arrojaría medidas absolutas diferentes ($Y \neq X$), cuyo efecto estaría reflejado en cambios en el parámetro α de la función lineal, como se muestra en la siguiente Figura:

Figura 8. Comparación de la función lineal de la Figura 5 con la función lineal óptima de confiabilidad.



La función lineal de la Figura 8, fue estimada a partir de los datos hipotéticos de la Tabla 4; en esta función se observa que el único parámetro que difiere con los valores óptimos de linealidad es α , puesto que el valor estimado es igual a 5 y el valor óptimo es igual a cero; mientras que, el valor de β estimado es igual a uno, por lo que no difiere con el valor óptimo, de lo cual se obtiene el modelo lineal siguiente:

$$Y = 5 + X$$

De esta forma, se ve claramente que un error aditivo sistemático en

cualquiera de los observadores produce un efecto de desplazamiento de la función lineal a lo largo de la ordenada, mientras que la inclinación seguirá siendo de 45° , por lo que será importante saber si la magnitud del desplazamiento es lo suficientemente significativa como para invalidar el uso del coeficiente r de Pearson como índice de confiabilidad.

La forma de probar si el valor estimado de α se aleja de cero es por medio del estadístico " t_α ", que permite probar la hipótesis nula de que α es igual a cero.

El hecho de aceptar la hipótesis nula, es decir, que α sea igual a cero, no garantiza que el valor estimado de acuerdo entre observadores a partir de la r de Pearson sea un valor adecuado de confiabilidad, puesto que con esto, sólo se habrá probado que no existe un error sistemático aditivo en los observadores que afecte el valor del coeficiente de correlación, ya que un error sistemático multiplicativo en los observadores, también produce que la r de Pearson sobreestime el acuerdo entre los registradores. Pongamos por ejemplo, la observación del mismo segmento conductual de lavarse las manos en un paciente obsesivo-compulsivo, por dos observadores independientes. Los datos hipotéticos obtenidos por ambos registradores durante cinco días se muestran en la Tabla 5, y el diagrama de dispersión que generan estos datos se muestra en la Figura 9.

Figura 9. Comparación de la función lineal estimada a partir de los datos hipotéticos de la Tabla 5 con la función lineal óptima de confiabilidad.

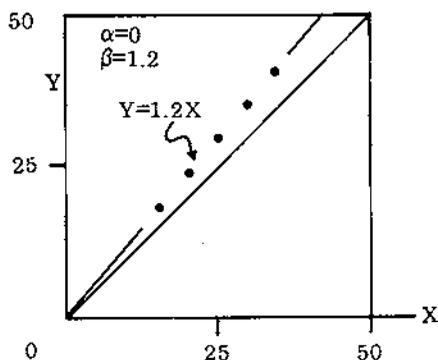


Tabla 5. Datos hipotéticos de la ocurrencia de lavarse las manos de un paciente obsesivo-compulsivo durante cinco días.

Día	OBSERVADOR 1 (X)	OBSERVADOR 2 (X)
1	30	36
2	15	18
3	25	30
4	35	42
5	20	24

A simple vista se observa en la Tabla 5, que el supuesto básico de acuerdo de $Y = X$ no fue obtenido, puesto que en todos los días de registro se ve que $Y \neq X$. Sin embargo, al representar gráficamente los pares de valores para cada día, éstos dan como resultado que todos los puntos estén sobre una línea recta. Por consiguiente, el valor que se obtiene en este caso particular es $r = 1.00$.

Los parámetros estimados de la relación lineal, generados por estos datos hipotéticos, dan como resultado un $\alpha = 0$ y una $\beta = 1.20$, de lo que se obtiene el modelo lineal siguiente:

$$Y = 1.20X$$

De esta forma, se ve claramente que un error multiplicativo sistemático en cualquiera de los observadores produce un efecto de desplazamiento de la función lineal, como se observa en la Figura 9, puesto que esta recta aumenta su ángulo de inclinación, mientras que la ordenada al origen, sigue teniendo el parámetro óptimo, es decir, α es igual a cero.

De aquí que también es importante saber, si la magnitud del aumento del ángulo de inclinación de la recta es lo suficientemente grande para invalidar el uso de la r de Pearson como índice de confiabilidad entre observadores.

La forma de probar esta suposición es por medio del estadístico " t_β ", que permite probar la hipótesis nula de que $\beta = 1$.

En conclusión, sólo hasta que se demuestre que la función lineal generada por los datos de los observadores, no se aleja significativamente de los parámetros óptimos del supuesto de linealidad especial que presupone el concepto de confiabilidad, el coeficiente de correlación r de Pearson no puede ser tomado como estimador adecuado de la confiabilidad en la evaluación conductual.

CONCLUSIONES

Como se habrá notado, todas las anteriores consideraciones sobre el uso de la r de Pearson como índice de acuerdo, se han abocado a través del modelo de regresión lineal simple a proponer un procedimiento cuantitativo que permita tomar una decisión acerca de lo adecuado de dicho coeficiente de correlación, para estimar la confiabilidad entre observadores; pero de ninguna manera este procedimiento estadístico da elementos de análisis sobre la magnitud aceptable de la r de Pearson para afirmar la existencia de confiabilidad en los datos de los observadores.

De aquí que si el concepto de confiabilidad en la evaluación conductual supone una relación lineal positiva específica en donde $Y = X$, con parámetros $\alpha = 0$ y $\beta = 1$, y la r de Pearson evalúa la dispersión de los datos con respecto a cualquier relación lineal y , si por otro lado, el modelo de regresión simple sólo permite saber si dicho coeficiente es adecuado como índice de

confiabilidad, ¿no sería mejor desecharlo y buscar otras alternativas para cuantificar la confiabilidad en la evaluación conductual?

Sin embargo, el hecho de que se desprecie el coeficiente de correlación r de Pearson como índice de confiabilidad, no invalida el uso del modelo de regresión lineal simple dentro de la evaluación conductual, puesto que al conceptualizar el acuerdo entre observadores como un modelo de regresión, permitiría detectar errores aditivos y/o multiplicativos en las medidas de los observadores, independientemente de la confiabilidad que existiera entre ellos. De esta forma, la detección de estos tipos de errores en los puntajes de los observadores tiene implicaciones metodológicas y de procedimiento más importantes, que el demostrar sólo si la r de Pearson es un estimador adecuado de la confiabilidad.

Para finalizar, es importante mencionar que este tipo de consideraciones acerca del uso del modelo de regresión lineal simple para detectar errores aditivos y/o multiplicativos entre los datos de los observadores, se abordarán en un trabajo posterior.

DISCUSIÓN

En este trabajo, se mostró que el concepto de confiabilidad en ambos modelos difiere marcadamente, puesto que mientras en la evaluación tradicional se supone la existencia de confiabilidad en los instrumentos de medición cuando dicho instrumento produce posiciones relativas iguales en las distribuciones de los puntajes en cada ocasión de medida, en la evaluación conductual, la suposición en la mayoría de los estudios observacionales, es que la confiabilidad puede ser evaluada a partir de la estimación del acuerdo entre observadores, para lo cual es necesario que los puntajes absolutos en ambos observadores en cada momento de registro sean iguales, esto es, $Y = X$.

Tal vez uno de los aspectos más polémicos dentro de la evaluación conductual, sea el hecho de la extrapolación de los índices estadísticos utilizados en la psicometría clásica, derivados del modelo tradicional de evaluación, para estimar la confiabilidad, ya que la psicometría clásica, no sólo desarrolló principios teóricos que la fundamentan, sino que a la par, también creó técnicas estadísticas que le dieran un soporte formal más estructurado; de aquí que la evaluación conductual no puede extrapolar así como así las técnicas estadísticas desarrolladas a partir de principios teóricos diferentes, sino que debe por un lado analizar críticamente dicha extrapolación y, por otro lado, crear sus propias técnicas estadísticas que tomen en cuenta los principios teóricos en los cuales se sustenta.

RESUMEN

Dentro del ámbito de la evaluación conductual y, en concreto, en situaciones de observación directa, se analiza mediante ejemplos con datos hipotéticos el funcionamiento del coeficiente r de Pearson, que por otra parte es el estimador por excelencia de la confiabilidad de los instrumentos en la psicometría clásica. Se revela su inadecuación al diferir claramente el concepto de confiabilidad en los modelos psicométrico y conductual, a pesar de no invalidarse el uso de la regresión lineal simple dentro de la evaluación, puesto que al conceptualizar el acuerdo entre observadores como un modelo de regresión permitiría detectar errores aditivos y/o multiplicativos en las medidas de los observadores, independientemente de la confiabilidad que existiera entre ellos.

Las consecuencias prácticas que de aquí se derivan son importantes, favoreciéndose la creación de técnicas estadísticas que sean sensibles a los principios teóricos correspondientes en cada caso y buscándose otras alternativas para cuantificar la confiabilidad en la evaluación conductual.

SUMMARY

In the ambit of behavioural evaluation and specifically in situations of direct observation, the Pearson r coefficient functioning is analysed by means of hypothetic data. Pearson being, on the other and, the appraiser par excellence of instrument reliability in classic psychometry. Its inadequacy is made known by the differing reliability concept of the psychometric and behavioural models, despite the fact that the use of uncombined linear regression in the evaluation is not invalidated, as by considering the agreement between observers as a regression model may enable to detect additive and/or multiplicative errors in the observer's measures, independently of the reliability that may exist between them.

The practical consequences stemming from this are important and favour the creation of statistics techniques perceptive to the corresponding theoretic principles of each case and other alternatives to quantify reliability in behavioural evaluation are being sought.

RÉSUMÉ

Dans le cadre de l'évaluation conductuelle et, plus précisément dans des situations d'observation directe, on analyse à l'aide d'exemples avec des éléments d'hypothèse le fonctionnement du coefficient r de Pearson, qui par ailleurs évalue au mieux la fiabilité des instruments de la psychométrie

classique. Il se révèle inadéquat car les concepts de fiabilité pour les modèles psychométrique ou conductuel sont bien différents, et ceci sans mettre en question l'usage de la régression linéaire simple dans l'évaluation, d'autant plus que si on conceptualise l'accord entre observateurs comme un modèle de régression, il permet de détecter des erreurs additives et/ou multiplicatives proportionnelles au nombre des observateurs et indépendamment de la confiance qui existe entre eux.

Les conséquences pratiques qui en découlent sont importantes, elles favorisent la création de techniques statistiques qui soient sensibles aux principes théoriques correspondants à chaque cas et la recherche d'autres possibilités pour quantifier la fiabilité dans l'évaluation conductuelle.

REFERENCIAS BIBLIOGRÁFICAS

- Anguera, M.T. (1983). *Manual de prácticas de observación*. México: Trillas.
- Cone, J.D. (1981). Psychometric considerations. En M. Hersen y A.S. Bellack (Eds.), *Behavioral assessment: A practical handbook*. New York: Pergamon Press.
- Haber, A. y Runyon, R.P. (1973). *Estadística general*. México: Fondo Educativo Interamericano.
- Hartmann, D.P. (1977). Considerations in the choice of interobserver agreement. *Journal of Applied Behavior Analysis*, 10, 103-116.
- Kent, R.N. y Foster, S.L. (1977). Direct observational procedure: Methodological issues in naturalistic setting. En A.R. Ciminero, K.S. Calhoun y H.E. Adams (Eds.), *Handbook of behavioral assessment*. New York: Wiley and Sons.
- Magnussons, D. (1975). *Teoría de los tests*. México: Trillas.
- Nelson, R.O. y Hayes, S.C. (1981). Nature of behavioral assessment. En M. Hersen y A.S. Bellack (Eds.), *Behavioral assessment: A practical handbook*. New York: Pergamon Press.

