# 5

# Estimación del coeficiente de fiabilidad

## 1. ¿Cómo estimar el coeficiente de fiabilidad?

Si pudieran realizarse N mediciones empíricas, en las mismas condiciones, de una misma dimensión de un mismo sujeto con un mismo instrumento de medida, sin que el sujeto cambiase, por sí o debido al instrumento, entonces, podríamos obtener directamente la puntuación verdadera V del sujeto como promedio de esas mediciones y la fiabilidad del instrumento expresada a través de la desviación típica de los errores de medida encontrados.

El problema reside en que, obviamente, existe una extensísima lista de variables de interés en Psicología para las que ese proceso es inviable. Esa lista incluye las aptitudes, las actitudes, buena parte de los rendimientos, dimensiones de personalidad, etc. Toda medición que exige actividad del sujeto de un modo u otro, queda excluída. Esto incluye tests, cuestionarios, procedimientos que impliquen entrevista, etc. No es posible, por ejemplo, aplicar a un sujeto N veces un test de inteligencia o de personalidad, por razones obvias de cansancio, aprendizaje, memoria, cambio del sujeto, etc. No se puede sostener que el sujeto se mantenga estable en la dimensión medida, ni que el instrumento no afecte a lo medido, ni que las circunstancias y condiciones de medición sean las mismas.

Si el proceso básico que está a la base lógica de la formulación del modelo

$$X = V + E$$

no es susceptible de realización práctica, ¿para qué sirve el modelo y todas sus deducciones formales? La ecuación anterior tiene más incógnitas que datos, dado que sólo X es observable. Si se consideran N sujetos el problema se mantiene al disponer de N puntuaciones empíricas para 2N incógnitas, una puntuación verdadera y un error de medida por sujeto.

Si se vuelven a mirar las fórmulas y deducciones de los dos capítulos anteriores buscando el modo de aplicarlas a mediciones reales se encontrará que esta situación se mantiene a través de todas ellas.

Por ejemplo, no podemos calcular el error típico de medida porque no conocemos ni los errores de medida ni el coeficiente de fiabilidad. Por ejemplo, no podemos calcular el coeficiente de fiabilidad porque aunque conocemos la varianza de las puntuaciones empíricas no conocemos la varianza de las puntuaciones verdaderas.

Hasta ahora tenemos un modelo formal no conectado con las mediciones empíricas. Hay que recordar que el interés de un modelo formal reside precisamente en su capacidad para hacer afirmaciones sobre el mundo real de modo que es fundamental encontrar ese punto de conexión para hacer operativo el modelo.

En el tema anterior hemos definido lo que serían mediciones paralelas y hemos visto que la correlación entre ellas sería igual al coeficiente de fiabilidad. *Si pudiésemos encontrar o elaborar mediciones paralelas* entonces podríamos calcular su correlación que la teoría considera el coeficiente de fiabilidad. A partir del coeficiente de fiabilidad se obtiene el error típico de medida, y a partir de ahí todos los términos desconocidos del modelo.

Sin coeficiente de fiabilidad conocemos la media, desviación típica y varianza de las puntuaciones empíricas y, además, podemos admitir, según el modelo, que conocemos la media de los errores de medida, que es cero, y la media de las puntuaciones verdaderas, que es la media de las empíricas. Nada más.

Si dispusiéramos del coeficiente de fiabilidad ya podríamos calcular el error típico de medida, dado que la fórmula del error típico de medida requiere conocer solo la desviación típica empírica y el coeficiente de fiabilidad. Conocido el error típico de medida se tiene inmediatamente la varianza del error, que es su cuadrado. Conocida la varianza de error y la varianza empírica se obtiene enseguida la varianza verdadera, dado que la varianza empírica es la suma de la verdadera más la de error. Por último, conocido el coeficiente de fiabilidad se tiene directamente el índice de fiabilidad, que es su raíz cuadrada. Aunque hay varios caminos para deducir los términos referidos a "inobservables" (varianza de las puntuaciones verdaderas y errores, error típico de medida e índice de fiabilidad) los que acabamos de sugerir pueden resultar los más sencillos si se dispone del coeficiente de fiabilidad.

Hay tres situaciones básicas en las que la teoría clásica puede encontrar mediciones paralelas: *los tests paralelos o formas paralelas de un test, la medición test-retest y la comparación entre mitades de un mismo test. Estos son los tres métodos de estimación del coeficiente de fiabilidad.* Se consideran tres aproximaciones empíricas o métodos prácticos de obtención de mediciones paralelas que

permiten, mediante el calculo de la correlación entre ellas, estimar el coeficiente de fiabilidad.

Por ello, la deducción que hemos visto en el tema anterior que establece que la correlación entre dos mediciones paralelas es el coeficiente de fiabilidad ocupa un papel central como gozne entre la teoría clásica de la fiabilidad y la medición real. Veamos ahora estos tres métodos prácticos.

## 2. Método de los tests paralelos o de las formas paralelas de un test.

Este método consiste en:

> *1. Elaborar dos formas paralelas* de un mismo test, o lo que es lo mismo, dos tests paralelos.
>
> *2. Aplicar una forma del test* a la muestra de interés, y, tras un lapso de tiempo que no sea relevante para la aparición de cambios en los sujetos, *aplicar la segunda forma del test a la misma muestra.*
>
> *3. Calcular el coeficiente de correlación* entre las puntuaciones empíricas obtenidas por los sujetos en las dos ocasiones. Si las formas son paralelas esa correlación es el coeficiente de fiabilidad del test.

Las denominaciones “tests paralelos” o “formas paralelas” son equivalentes y pueden usarse indistintamente.

Este método produce una estimación de la fiabilidad basada en la equivalencia entre formas paralelas, por ello al coeficiente de fiabilidad así estimado se le denomina *coeficiente de equivalencia* entre formas.

Veamos algunos aspectos importantes de cada uno de estos tres pasos.

**Paso 1. Elaborar formas paralelas:**

Primero, ¿Qué son y cómo se elaboran dos tests paralelos o dos formas paralelas de un test? Hay dos tipos de criterios que dos tests han de cumplir para que los consideremos paralelos:

*1. Criterio estadístico. Los tests han de mostrar que satisfacen la definición de tests paralelos que hemos dado en el tema anterior. Si no la satisfacen su correlación no podrá considerarse el coeficiente de fiabilidad.*

La cuestión de como se pone a prueba si determinados tests cumplen las condiciones estadísticas de paralelidad merece consideración aparte.

Siempre que dos instrumentos satisfagan estas condiciones de paralelidad los consideraremos tests paralelos, de modo que *desde un punto de vista formal el criterio estadístico es necesario y suficiente.* Sin embargo, en la práctica existen otros criterios que conviene considerar.

*2. Criterio de formato y contenido.* En la práctica dos tests paralelos consisten en dos conjuntos distintos de ítems referidos a una misma variable o constructo psicológico, habitualmente con las mismas instrucciones y el mismo formato de prueba y de ítems. Pretenden medir la misma variable por el mismo procedimiento.

El problema práctico reside en obtener ítems distintos pero razonablemente equivalentes. Esto es relativamente fácil de cumplir en algunas variables de aptitudes y rendimientos, ya sean académicos o laborales.

> Por ejemplo, si estamos interesados en medir la capacidad relativa a resolver 'productos de números de dos cifras por números de dos cifras' en niños de cierto grado de la escuela elemental, es fácil concebir dos pruebas de 50 ítems cada una que puedan considerarse desde el punto de vista de formato y contenido como paralelas. Por ejemplo, es fácil encontrar un ítem razonablemente paralelo para "12 por 37". Es razonable admitir que el ítem "32 por 17" activará los mismos procesos y conocimientos.

Las formas paralelas pretenden muestrear el mismo contenido. Simplemente las cuestiones formuladas, los ítems, son distintos, aunque referidos a los mismos contenidos. Si se dispone de una tabla de contenidos del constructo (una clasificación lógica de las zonas o partes del constructo), dos formas paralelas habrían de mantener normalmente el mismo muestreo de contenido a lo largo de la tabla.

> Por ejemplo, supongamos una prueba de rendimiento en Física con la que se aspira a conocer

el grado de conocimiento teórico y práctico de un grupo de estudiantes de cierto grado universitario en una materia que está estructurada en 10 temas con teoría y problemas. Supongamos que teoría y problemas son de igual importancia y que están igualmente distribuidos en los 10 temas. La "tabla de contenidos" tiene dos dimensiones: temas (de 1 a 10) y teoría-problemas. Dentro de las 20 celdillas de la tabla de contenido se encuentran todas las unidades elementales de contenido bien teóricas bien prácticas que pueden preguntarse (conocimientos, definiciones, problemas, fórmulas ...). Si la prueba tiene 20 items, un muestreo razonable de la materia sería 1 ítem de teoría y otro de práctica de cada tema. Es decir, un ítem por celdilla. Una forma paralela debería mantener el mismo muestreo de contenido aproximadamente. La forma paralela estaría formada bien por alguna variación sobre los mismos items que los hiciera distintos aunque midieran exactamente el mismo contenido, bien por otros items razonablemente equivalentes que mantuvieran el mismo muestreo de contenido, es decir, manteniendo un ítem de teoría y un problema de cada tema.

Si se dispone de un contenido muy acotado, donde las unidades de información están clara y distintamente establecidas y pueden considerarse razonablemente equivalentes desde el propósito de la medición, puede aplicarse un procedimiento de muestreo aleatorio simple sin reposición sobre el total de las unidades de contenido para obtener formas paralelas. En casos distintos la expresión "razonablemente equivalentes desde el propósito de la medición" puede significar cosas distintas. Por ejemplo, en una prueba de conocimientos, podríamos considerar si la importancia del contenido, su carácter teórico o práctico y su dificultad son razonablemente equivalentes.

Imaginemos que deseamos evaluar el vocabulario adquirido por un estudiante después del estudio de 12 lecciones, cada una de las cuales con su correspondiente vocabulario (palabras y definiciones). Supongamos que en total los vocabularios de los 12 temas suman 240 palabras. Aquí las unidades de información (cada palabra y su correspondiente definición) están claramente establecidas y la materia a evaluar bien acotada (no hay ninguna duda sobre que palabras son objeto de examen). Desde el punto de vista de la medición las unidades de información pueden considerase equivalentes. Es decir, en principio, podemos considerar todas las palabras contenidos igualmente importantes, no hay razones para subdividirlas en ningún tipo de agrupación relevante (por ejemplo aquí no tiene sentido hablar de teoría- práctica), y su dificultad puede considerarse razonablemente equivalente. Supongamos que deseamos dos pruebas paralelas de 30 items cada una. Podemos extraer

simplemente al azar dos conjuntos distintos de 30 palabras. Si las 240 palabras están numeradas una tabla de números aleatorios nos puede ayudar a garantizar la aleatoriedad de la elección.

Algunas veces, cuando no es tan sencillo garantizar que dos items miden lo mismo, por ejemplo porque no está tan claro que los contenidos sean igualmente importantes, puede seguirse la estrategia de obtener nuevos items para la forma paralela como variantes de los items iniciales. La variación introducida ha de ser lo bastante importante para que podamos considerar el resultado un ítem distinto, pero sin alterar aquello que medía el ítem original.

Por ejemplo, en una prueba de conocimientos de verdadero-falso se puede alterar el enunciado del nuevo ítem, variando o no su valor de verdad (si es cierto o falso), modificando la información pero sin variar el objeto de la pregunta. Estas variaciones son relativamente fáciles de elaborar en pruebas de rendimiento académico donde se necesitan muchas formas paralelas en muchas ocasiones distintas tendiendo a garantizar la comparabilidad de las evaluaciones.

No pueden considerarse formas paralelas aquéllas en las que la diferencia consiste en que se ha variado el orden de los items o el orden de las alternativas. Por la misma razón que no pensaríamos que un test es forma paralela de otro porque se ha cambiado el tipo de letra o el color del papel. En ese caso se trata simplemente de la misma prueba bajo dos presentaciones.

En otras mediciones como cuestionarios de personalidad o cuestionarios de actitudes puede ser un poco más difícil obtener formas que puedan considerarse adecuadamente paralelas desde el punto de vista del contenido. Un procedimiento práctico que a veces se utiliza consiste en construir ítems con contenidos paralelos uno a uno. Es decir, para cada cuestión elaborar otra distinta pero estrechamente relacionada que pregunte sobre el mismo contenido.

En todo caso el propósito de la construcción de formas paralelas es obtener dos mediciones que *globalmente* sean paralelas, por tanto no es requisito necesario -aunque puede ser a veces recomendable- que los items pretendan ser paralelos. Este modelo de estimación de la fiabilidad solo requiere que las formas como un todo sean paralelas, no necesita ni suponer ni comprobar que además los items lo sean entre sí, ni dentro de la prueba ni entre pruebas. Otras aproximaciones a la estimación de la fiabilidad que se basan en comparar partes de la prueba necesitarán que estas partes sean paralelas, y otras más que después veremos operaran como si los items de la prueba, uno a uno, fueran medidas paralelas entre sí. En este sentido el método de formas paralelas es menos 'exigente' en condiciones de paralelidad.

En cualquier caso es una cuestión empírica (criterio estadístico) decidir si efectivamente dos formas que han sido construidas desde el punto de vista de formato y contenido para ser paralelas realmente lo son.

Es decir, cumplir con el criterio de elaborar formas de formato y contenido razonablemente paralelos, según las consideraciones expuestas, no garantiza que las formas sean paralelas desde el punto de vista estadístico. Habitualmente, desde un punto de vista práctico, cumplir el criterio de forma y contenido se considera condición necesaria, -pero no suficiente- para disponer de formas paralelas.

Si se cumple el criterio de formato y contenido obtendremos formas ***nominalmente paralelas***. Unicamente si cumplen el criterio estadístico podremos hablar de formas paralelas.

Ahora bien ¿con qué propósito se elaboran las supuestas formas paralelas? Si el propósito es estimar el coeficiente de fiabilidad mediante su correlación entonces las formas han de ser realmente paralelas, es decir, han de satisfacer los supuestos de paralelidad establecidos en el tema anterior al definir mediciones paralelas (criterio estadístico).

Dado que existen otros métodos que requieren menos esfuerzo y tiempo para estimar un coeficiente de fiabilidad de un test, parece poco razonable elaborar dos formas paralelas *únicamente* para estimar el coeficiente de fiabilidad por el método de formas paralelas.

Paradójicamente, el principal propósito práctico para construir formas supuestamente paralelas *no* es estimar el coeficiente de fiabilidad, sino disponer de dos estimaciones razonablemente comparables de la posición de los sujetos en la variable de interés. Para esto, en muchas situaciones prácticas no necesitaremos exactamente formas paralelas (igual puntuación verdadera sujeto a sujeto e igual varianza de error), ni siquiera formas que cumplan condiciones más relajadas de paralelidad, sino simplemente dos pruebas razonablemente comparables en función del criterio de formato y contenido.

Dos muestras aleatorias de items en la prueba de vocabulario antes mencionada podrían no satisfacer el criterio estadístico de paralelidad, aun estando construidas con todo rigor bajo el criterio de forma y contenido siguiendo las indicaciones antes comentadas. Dos pruebas nominalmente paralelas de productos de números de dos cifras como las reseñadas antes, perfectamente elaboradas desde el criterio de forma y contenido, podrían no cumplir el criterio estadístico. Dos pruebas de rendimiento académico que preguntan exactamente los mismos contenidos obtenidas mediante variantes de los items podrían no satisfacer el criterio estadístico. En todas estas circunstancias las pruebas, adecuadas según criterio de forma y contenido, serían *nominalmente paralelas pero no paralelas.* La mayoría de los

psicólogos las considerarían razonablemente equivalentes. Es decir, dos mediciones razonablemente igual de buenas (o de malas) de la misma variable medida del mismo modo. Y las utilizarían sin excesiva preocupación. La naturaleza de este uso común tanto en trabajo profesional como en investigación podrá juzgarse mejor después de las consideraciones críticas de la teoría de la fiabilidad que se discuten en el capítulo siguiente.

**Paso 2. La aplicación de las formas del test.**

¿Bajo qué condiciones y con qué lapso de tiempo deben aplicarse ambas formas a la misma muestra? Esta es una cuestión que tiene diversas facetas y dificultades.

En primer lugar, las dos formas deben ser administradas bajo las mismas condiciones, o, al menos bajo los mínimos cambios posibles en las condiciones. Por ejemplo, si es posible deberían aplicarse en los mismos locales, por el mismo/s psicólogo/s, manteniendo iguales todas las características formales de la administración. Se trata de no introducir factores que puedan provocar cambios en los resultados.

En segundo lugar, respecto al tiempo, debe utilizarse un lapso entre ambas formas lo suficientemente corto como para que los sujetos no hayan cambiado en la variable de interés y lo suficientemente largo para que factores de memoria, fatiga, o entrenamiento tengan el mínimo efecto. Se trata, por un lado, de intentar medir a los sujetos de nuevo antes de que factores madurativos, ambientales, sociales, etc. les produzcan cambios en la variable de interés. Por otro, de evitar que la administración de la primera prueba presente efectos sobre la segunda.

Si se trata, por ejemplo, de un test de razonamiento abstracto de 60 ítems y administramos la forma B a continuación de la forma A es razonable esperar que los sujetos sufran algún tipo de fatiga que tendería a disminuir su rendimiento. Esto parece sugerir que para estas pruebas sería razonable aplicar la forma B otro día, sin dejar pasar mucho tiempo. La memoria de las respuestas dadas, que tiende a aumentar la correlación obtenida en algunas pruebas de personalidad o actitudes, y el efecto de entrenamiento y el de recuerdo de respuestas que produce la misma prueba especialmente en algunas pruebas de rendimiento, también parecen aconsejar administrar la segunda forma algunos días después.

Gulliksen (1950; 194) dice explícitamente que en la mayoría de situaciones “el mejor método para obtener la fiabilidad de un test es construir formas paralelas del test y administrarlas en días diferentes al mismo grupo de sujetos”.

Por supuesto no es razonable administrar las formas paralelas en días distintos con el fin de estimar el coeficiente de fiabilidad si los sujetos están expuestos entre

tanto a entrenamiento o tratamiento en la variable que se mide. O, al revés, si la variable que se mide, una habilidad o un rendimiento por ejemplo, sufre fuertes pérdidas por ausencia de práctica en cortos periodos. En algunos de esos casos el lapso temporal puede reducirse administrándose las dos formas en el mismo día, normalmente con un tiempo de descanso y cambio de actividad entre ellas.

Cuando se administran las dos formas en el mismo día, puede esperarse que la correlación entre ellas sea mayor que si se administran en días distintos. En muchas variables los resultados de los sujetos presentan pequeñas variaciones asistemáticas "día a día" debidas a innumerables factores que tienden desaparecer si las formas se administran el mismo día, y más todavía si se administran consecutivamente.

Si las formas resultan ser paralelas (criterio estadístico) entonces las formas A y B administradas el mismo día obtendrán un coeficiente de fiabilidad mayor que administradas en días distintos. Como lo normal, después en la práctica, será que las pruebas sean utilizadas en días distintos, el coeficiente de fiabilidad obtenido con pruebas administradas el mismo día tenderá a sobrestimar la fiabilidad que puede atribuirse a la prueba en sus condiciones más habituales de uso. Por ello, si es posible y razonable, se recomienda en general dejar pasar unos pocos días entre ambas formas.

*¿Para qué tipos de tests es adecuado este método?* El método de las formas paralelas es adecuado para tests de potencia y para tests de velocidad en todas las áreas de medición psicológica con instrumentos de lápiz y papel y también con ciertos tests manipulativos.

La distinción y denominación de tests de potencia y tests de rapidez o velocidad se aplica a mediciones donde los items tienen respuesta verdadera (aptitudes y rendimientos básicamente).

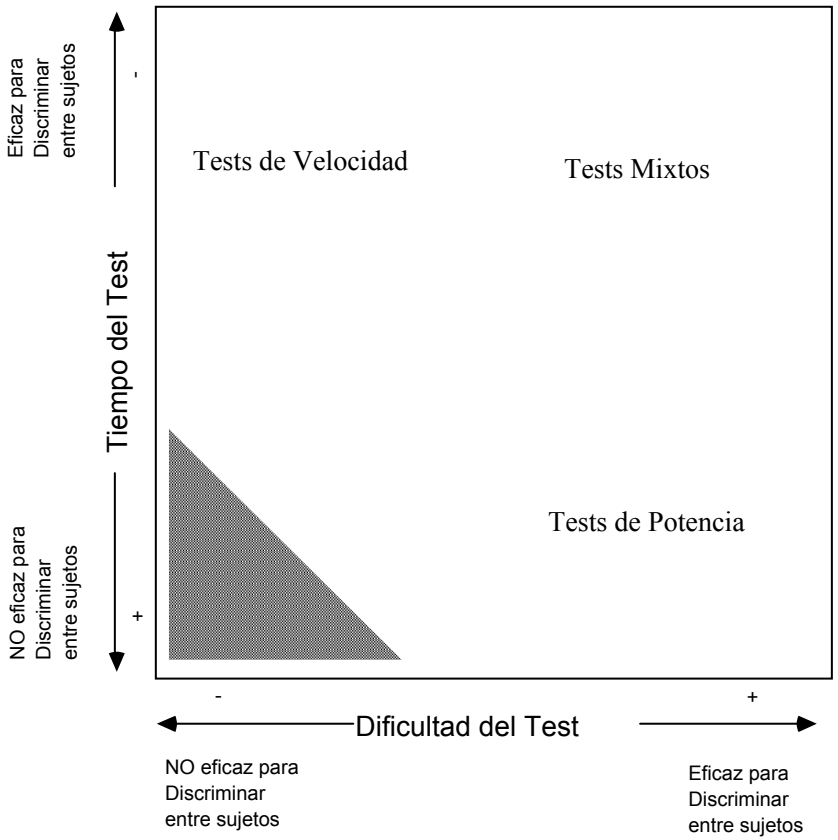
Se denominan *tests de velocidad* aquellos en los que el límite de tiempo concedido a los sujetos para realizar la prueba está calculado de modo que una proporción determinada de sujetos, generalmente mayoritaria, no terminará la prueba. Es decir, no tendrá tiempo para hacer cierto número de items al final de la prueba. Se caracterizan, generalmente, por estar formados por items muy fáciles. Items que los sujetos de la población a los que va destinada la prueba resolverían todos bien si tuviesen tiempo para ello.

Se conocen por *tests de potencia* aquellos en los que el tiempo concedido permite a los sujetos llegar a considerar todos los items de la prueba. Esto puede expresarse de otro modo diciendo que son tests sin límite de tiempo. Generalmente son tests formados por ítems de diversos grados de dificultad. Items que

un sujeto con capacidad inferior a la dificultad de los mismos sólo podrá acertar por azar. Lo usual es que la dificultad de los items sea creciente, con una meseta en la zona central del test.

Se denominan *tests mixtos* a aquellos tests de potencia que tienen establecido un tiempo límite que impide terminar a una porción no sustancial de sujetos. Por razones prácticas los tests de potencia -incluso cuando no quieren introducir ningún elemento de velocidad- han de tener fijado un límite definitivo de tiempo en el que cerrar la situación de examen. Si ese tiempo límite es eficaz para impedir concluir algunos items a los sujetos más lentos hablamos de tests mixtos.

Tests de velocidad, tests de potencia y tests mixtos. (En general no t
desarrollar tests en la zona sombreada por su escasa capacidad par
entre los sujetos examinados).



El concepto de *test mixto* permite apreciar que los conceptos de test de potencia y de velocidad son dos zonas de la retícula formada por las dimensiones dificultad y tiempo del test, pudiendo darse muchas situaciones graduales intermedias.

El método de formas paralelas no presenta especiales restricciones respecto a la materia o tipo de variable que mide el test. Puede utilizarse con tests de inteligencia y de capacidades, con tests de habilidades y destrezas, con tests de rendimientos, con cuestionarios de personalidad, con inventarios de intereses, con cuestionarios de actitudes, con encuestas de opinión, etc.

Como la paralelidad debe darse entre las dos formas tomadas como un todo, este método tiene la ventaja de poder utilizarse con tests internamente heterogéneos, con tal de que ambas formas mantengan el mismo grado y tipo de heterogeneidad.

**Paso 3. Calcular el coeficiente de correlación.**

Después de administrar las dos formas paralelas del test a una misma muestra de sujetos se dispondrá de una tabla de datos con N sujetos por 2 variables, la puntuación en la forma A y en la forma B para cada sujeto. Se procede entonces a calcular el coeficiente de correlación de Pearson:

$$r_{AB} = \frac{s_{AB}}{s_A s_B}$$

En caso de no efectuarse los cálculos mediante programa un paquete estadístico, que sería lo habitual, puede utilizarse cualquier fórmula del coeficiente de correlación de Pearson que resulte adecuada.

Para una calculadora corriente que ofrece directamente medias y desviaciones típicas he deducido la siguiente fórmula que resultará especialmente cómoda:

$$r_{AB} = \frac{\overline{X_A X_B} - \bar{X}_A \bar{X}_B}{s_A s_B}$$

Se obtiene el producto de las puntuaciones sujeto a sujeto y se calcula la media del producto. De la media del producto se sustrae el producto de las medias y el resultado se divide por el producto de las desviaciones típicas.

Como cualquier correlación el resultado obtenido puede estar entre -1 (máxima relación lineal negativa) y +1 (máxima relación lineal positiva), pasando por 0 (ausencia de relación lineal). En realidad como se trata de formas paralelas no tiene sentido esperar correlaciones negativas debiendo estar el resultado entre 0 y +1. Es más, como se trata de la correlación del test consigo mismo (en dos versiones) cabría esperar valores positivos alejados de 0.

Tanto si A y B son formas paralelas como si no, el coeficiente de correlación calculado expresa el grado de relación lineal entre ambas formas. Si A y B son formas paralelas entonces esta correlación es el coeficiente de fiabilidad y como tal puede interpretarse en términos de proporción de varianza verdadera en la varianza empírica. Para considerar al test fiable el coeficiente de correlación obtenido debe ser alto, de modo que una gran proporción de la varianza de las puntuaciones se deba a varianza verdadera.

Así por ejemplo, si obtenemos un coeficiente de fiabilidad de 0,75 diremos que tres cuartas partes de la varianza empírica del test se deben a varianza verdadera (a la varianza de las puntuaciones verdaderas), o bien al revés, que un 25% de la varianza empírica es varianza de error.

En muchas ocasiones se construye un test sin elaborar formas paralelas del mismo, entonces hay que recurrir a los procedimientos de estimación de la fiabilidad que se describen a continuación.

## 3. Método test-retest.

El método test-retest está indicado para estimar la fiabilidad de un test del que sólo disponemos una forma. Consiste en:

1. *Administrar el mismo test en dos ocasiones* diferentes separadas por cierto lapso temporal a una misma muestra de sujetos.

2. *Calcular el coeficiente de correlación* entre las puntuaciones obtenidas por los sujetos en las dos ocasiones.

El método evalúa la estabilidad de los resultados a través de cierto tiempo. Por ello al coeficiente de fiabilidad que obtiene se le denomina *coeficiente de estabilidad* temporal.

Respecto al tiempo que debe transcurrir entre aplicaciones pueden aplicarse los comentarios generales al respecto que hemos efectuado al tratar de las formas paralelas.

En general, a menor tiempo mayor efecto de la memoria de las respuestas dadas, del aprendizaje debido al propio test y de la fatiga producida por el

propio test, esta última si el retest (segunda medición) sucede al test de un modo más o menos inmediato.

En general, a mayor tiempo mayor posibilidad de que los sujetos hayan cambiado realmente en la variable de interés debido a múltiples factores permanentes o circunstanciales: aprendizaje, cambios evolutivos, experiencias emocionales, enfermedad, condiciones ambientales y sociales varias, etc.

Gulliksen (1950) teme más los primeros factores que los segundos, y, en conjunto, no es muy partidario de este método. En sus propias palabras “el mayor riesgo en esta técnica es que la fiabilidad será demasiado alta debido a la tendencia del sujeto a duplicar su desempeño inicial.[...] los resultados en la repetición de un test es probable que sean mucho más cercanos a la puntuación original que los resultados en una forma paralela del mismo test. Este método de repetición del mismo test en un tiempo diferente debería en general no ser usado, dado que da un coeficiente alto espuriamente, y la cantidad de error no es fácil de determinar” (Gulliksen, 1950: 197-198).

Las dos mediciones deben probar su paralelidad, es decir, probar que satisfacen la definición de mediciones paralelas para que su correlación pueda considerarse estimación del coeficiente de fiabilidad.

## 4. Métodos de las dos mitades.

Cuando se dispone de una sola forma de test resulta sencillo aplicar un método de “dos mitades”.

Hay tres métodos clásicos de "dos mitades": el método basado en la fórmula de Spearman-Brown, el método basado en la fórmula de Rulon y el método basado en la fórmula de Guttman. Los tres comparten la necesidad de partir el test en dos partes con igual número de items que cumplan las condiciones de paralelidad para después aplicar las fórmulas.

A continuación se presentan los tres métodos, después se presentarán algunas consideraciones comunes.

### 4.1. Método de las dos mitades mediante la corrección de Spearman-Brown.

1. Administrar el test a una muestra de sujetos una sola vez.

2. Descomponer el test en dos partes de modo que tengan el mismo número de items y que puedan ser consideradas paralelas. Calcular la puntuación total en cada una de esas dos mitades.

3. Obtener la correlación entre los totales de las partes. Esa correlación, si las mitades son paralelas, podría considerarse la fiabilidad de un test con la mitad de items que el test completo.

4. Aplicar sobre esa correlación la ***corrección de Spearman-Brown para el caso de longitud doble***:

$$r_{XX} = \frac{2r}{1+r}$$

Donde r es la correlación entre las partes o mitades, y, el resultado de la fórmula, $r_{XX}$, es la fiabilidad estimada para el test total.

Esta corrección estima la correlación que se hubiera obtenido entre las partes si hubiesen tenido el mismo número de items que el test completo. (Es decir, si hubiesen actuado como dos auténticos tests paralelos de igual longitud al original). Se denomina ***longitud del test*** a su número de items.

La corrección del paso 4 se basa en la relación entre fiabilidad y longitud del test que trataremos después.

## 4.2. Método de las dos mitades mediante la fórmula de Rulon.

1. Administrar el test a una muestra de sujetos una sola vez.

2. Descomponer el test en dos partes de modo que tengan el mismo número de items y que puedan ser consideradas paralelas. Calcular la puntuación total en cada una de esas partes.

3. Calcular para cada sujeto la *diferencia entre las puntuaciones* totales que ha obtenido en las partes:

$$d = X_1 - X_2$$

4. Obtener la varianza del total del test y la varianza de la nueva variable d. Aplicar la *fórmula de Rulon* (1939):

$$r_{XX} = 1 - \frac{s_d^2}{s_X^2}$$

## 4.3. Método de las dos mitades mediante la fórmula $L_4$ de Guttman.

1. Administrar el test a una muestra de sujetos una sola vez.

2. Descomponer el test en dos partes de modo que tengan el mismo número de items y que puedan ser consideradas paralelas. Calcular la puntuación total en cada una de estas partes.

3. Calcular la varianza del total de cada una de las partes así como la varianza del total del test.

4. Aplicar la fórmula $L_4$ de Guttman (1939):

$$r_{XX} = 2\left[1 - \frac{s^2_{X_1} + s^2_{X_2}}{s^2_X}\right]$$

El numerador de la fórmula expresa la suma de las varianzas de los totales de las dos partes y el denominador la varianza del total del test.

## 4.4. Consideraciones sobre los métodos de las dos mitades.

La discusión sobre los modos de partir el test en dos partes, los usos, las ventajas y las desventajas es común a los tres métodos.

*¿Qué relación existe entre los métodos de las dos mitades?* La fórmula de Guttman puede considerarse una reexpresión de la fórmula de Rulon, por ello ambas darán el mismo resultado bajo cualquier situación. A su vez, ambas equivalen a Spearman-Brown cuando la varianza de las puntuaciones en ambas partes es igual. Si las varianzas de las partes no son iguales entonces las fórmulas de Rulon y de Guttman darán un valor inferior a la fórmula de Spearman-Brown.

Recuérdese que cuando dos mediciones son paralelas ambas tienen la misma varianza de puntuaciones empíricas, por tanto los tres métodos habrán de dar el mismo resultado para la misma partición *paralela* del test en los mismos datos.

Las partes del test deben probar su paralelidad, es decir, probar que satisfacen las condiciones para que podamos hablar de mediciones paralelas. Solo de ese modo el valor que se estimará a partir de su correlación podrá ser tomado a su vez como una estimación del coeficiente de fiabilidad.

Guttman demostró que este método produce una estimación de un límite inferior del coeficiente de fiabilidad, por lo que produciría infraestimaciones de la fiabilidad.

*¿Cómo partir un test en dos mitades nominalmente paralelas?* Hay muchos modos de partir un test en dos.

La forma más sencilla es precisamente dividir el test en *dos mitades*, es decir, considerar su primera mitad y su segunda mitad. Este método de división en partes no suele ser muy recomendable por diversas razones. Primero, si el test produce algún cansancio es obvio que la fatiga tenderá a ser mayor en la segunda parte, lo que hará menos comparables las partes. Segundo, si el test actúa como entrenamiento para los sujetos, la segunda parte contará con el entrenamiento fruto de la primera y no viceversa. Tercero, muchos tests tienen una dificultad que se va incrementando lo que hace que ambas partes no sean comparables. Cuarto, en muchos tests relativamente heterogéneos ambas partes pueden no ser comparables por su contenido. La presencia de alguno o de varios de estos motivos es suficiente en la mayoría de los casos para desaconsejar este método de división.

Un segundo método muy popular consiste en separar los items *pares e impares.* Es decir, se toman los pares como una parte y los impares como la segunda parte. Las desventajas anteriores desaparecen o se reducen fuertemente, si bien hay que cuidar posibles problemas de heterogeneidad relativa en el contenido.

Un tercer método consiste en formar las *partes ad hoc* buscando pares de items que sean semejantes y asignando un ítem de cada par a cada parte por algún procedimiento de azar. Por supuesto esto tenderá a producir un coeficiente de correlación más alto que el que encontraríamos entre dos partes formadas por items menos comparables, lo que puede resultar en una sobreestimación de la fiabilidad de la prueba.

Los criterios para emparejar dos items pueden ser estadísticos o de forma y contenido. Entre los criterios estadísticos pueden mencionarse diversos acercamientos:

1) Items con semejante media y con semejante correlación con el total del test,

2) Items con semejante saturación factorial[1],

3) Items con semejante correlación con alguna o algunas variables ajenas al test que resulta de interés predecir.

[1] La saturación factorial es un concepto de análisis factorial que expresa, esencialmente, la correlación del ítem con un factor. Esa correlación puede interpretarse como el grado en que ese ítem mide ese factor. Por ejemplo, la saturación factorial de un ítem i de un test de inteligencia general en un factor de comprensión verbal podría interpretarse como el grado en que ese ítem mide comprensión verbal.

El primero es el acercamiento clásico y probablemente, de estos tres, el que más favorece encontrar una estimación de la fiabilidad más alta. Estos criterios pueden combinarse entre sí y con análisis del contenido para obtener pares de items lo más semejantes posible.

> El proceso puede realizarse en pruebas relativamente heterogéneas con criterios de contenido. Si, por ejemplo, disponemos de una prueba de conocimientos de 20 items, dos de ellos de cada uno de 10 temas, cuya posición en la prueba ha sido aleatorizada, es razonable pensar en dividir la prueba en dos mitades formando la primera con un ítem de cada tema y la segunda con el otro ítem de cada tema, si estos son comparables en dificultad y tipo. Por cierto que en este caso también podría ser razonable pares-impares e incluso dos mitades si todos los sujetos han terminado a tiempo.

*¿A qué tipos de tests se pueden aplicar estos métodos?* La respuesta general es que pueden aplicarse a cualquier prueba a la que podamos dividir en dos mitades paralelas por alguno de los procedimientos antes mencionados.

Hay una salvedad específica: No tiene mucho sentido aplicar métodos de dos mitades a tests de velocidad. Primero, porque en un test de velocidad normalmente la segunda parte tendrá un número importante de omisiones (items no contestados), lo que hará las partes poco comparables. Segundo, si se opta por pares-impares o por partes ad hoc, como los items suelen ser muy fáciles habrá el mismo número de aciertos en ambas partes, lo que puede dar lugar a una alta fiabilidad bien poco informativa.

A parte de esta salvedad no hay limitaciones en cuanto a tipo de variable medida: capacidades, personalidad, actitudes, etc.

*¿Pueden utilizarse tres o más partes en lugar de dos?* Efectivamente. Puede dividirse el test en tres o más partes, especialmente por el método de partes ad hoc, aunque también por partes determinadas al azar o por partes sucesivas, si ello es adecuado a las características de la prueba. Generalmente esto exige un número mayor de items para que las partes no sean de un tamaño irrelevante. Disponer de tres o más mediciones paralelas es condición para efectuar algunos contrastes de paralelidad.

Si hay tres partes paralelas podrán calcularse tres correlaciones entre partes que, como consecuencia de los supuestos de paralelidad, habrán de ser iguales entre sí. En ese caso sobre la correlación obtenida r, para estimar la correlación que se hubiera obtenido de un test con la longitud del test global, se aplica la fórmula de corrección de Spearman-Brown oportuna:

$$r_{xx} = \frac{3r}{1+2r}$$

Este es otro caso particular de la fórmula general de Spearman-Brown que relaciona fiabilidad y longitud del test.

*¿Qué relación hay entre la fiabilidad obtenida por un método de mitades y la obtenida por los otros métodos?* La fiabilidad obtenida por un método de mitades es un *coeficiente de homogeneidad entre mitades.* Su valor concreto depende de que mitades del test se obtengan: Dos particiones distintas del mismo test, aun satisfaciendo la definición de paralelidad, pueden dar dos estimaciones de fiabilidad distintas.

En la práctica si un mismo test es sometido a los tres tipos de estudios de fiabilidad vistos hasta aquí, la estimación de la fiabilidad por el método de mitades suele ser mayor que la producida por el método test-retest y ésta mayor que la producida por el método de las formas paralelas.

Cada método en realidad evalúa cosas distintas: la equivalencia entre formas, la estabilidad temporal y la homogeneidad entre partes, de ahí las diferencias que, generalmente, pueden aparecer. La interpretación del coeficiente de fiabilidad como proporción de la varianza empírica debida a la varianza verdadera es igual de legítima para los diversos métodos si, en cada caso, se cumple la definición de mediciones paralelas.

## 5. Ejemplos de estimación del coeficiente de fiabilidad

En los ejemplos siguientes se utilizan solo unos pocos casos para ilustrar el proceso de estimación, obviamente este proceso debe realizarse con muestras representativas y de tamaño suficiente de la población bajo estudio.

### 1. Fiabilidad Test-Retest.

Aplicamos el test X en el tiempo 1 a una muestra de 10 sujetos, y el retest en tiempo 2. Calculad el coeficiente de fiabilidad por el método test-retest.

**Datos:**

| *Caso:* | I | II |
|---|---|---|
| 1 | 17 | 17 |
| 2 | 18 | 19 |
| 3 | 12 | 12 |
| 4 | 11 | 10 |
| 5 | 7 | 7 |
| 6 | 4 | 3 |
| 7 | 19 | 18 |
| 8 | 10 | 11 |
| 9 | 3 | 4 |
| 10 | 27 | 28 |

**Resultados:**

$\bar{X}_1 = 12'8 \qquad s_{x_1} = 7'096478$

$\bar{X}_2 = 12'9 \qquad s_{x_2} = 7'3$

Coeficiente de fiabilidad:

$$r_{xx} = \frac{216'6 - 12'8 \cdot 12'9}{7'096478 \cdot 7'3} = 0'9937$$

**2. Fiabilidad Formas "Paralelas".**

Aplicamos las formas A y B del test X a una muestra de 8 sujetos. Calculad el coeficiente de fiabilidad por el método de formas paralelas.

**Datos:**

| Caso: | I | II |
|---|---|---|
| 1 | 13 | 15 |
| 2 | 38 | 39 |
| 3 | 12 | 12 |
| 4 | 11 | 10 |
| 5 | 3 | 4 |
| 6 | 4 | 3 |
| 7 | 40 | 38 |
| 8 | 15 | 13 |

**Resultados:**

$\bar{X}_A = 17 \qquad s_{x_A} = 13'304135$

$\bar{X}_B = 16'75 \qquad s_{x_B} = 13'15057$

Coeficiente de fiabilidad:

$$r_{xx} = \frac{458'75 - 17 \cdot 16'75}{13'304135 \cdot 13'15057} = 0'99453$$

**3. Método de Dos Partes aplicando la Corrección de Spearman-Brown.**

Aplicamos un test que tiene 10 items a una muestra de 10 sujetos. Obtened el coeficiente de fiabilidad por el método de dos mitades aplicando la fórmula de Spearman-Brown.

**Datos:**

| | | | | | Items: | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ***Caso*** | **I1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **I10** |
| ***S1*** | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 3 | 1 | 1 |
| ***2*** | 2 | 2 | 3 | 3 | 3 | 3 | 4 | 2 | 2 | 2 |
| ***3*** | 3 | 3 | 4 | 4 | 3 | 3 | 3 | 4 | 4 | 3 |
| ***4*** | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 |
| ***5*** | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 4 |
| ***6*** | 4 | 3 | 3 | 4 | 3 | 4 | 3 | 4 | 3 | 4 |
| ***7*** | 1 | 1 | 2 | 3 | 1 | 1 | 1 | 3 | 2 | 1 |
| ***8*** | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 |
| ***9*** | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 2 |
| ***S10*** | 5 | 5 | 5 | 4 | 5 | 5 | 5 | 4 | 5 | 5 |

**Solución:**

Necesitamos elaborar en la tabla una columna P1 que refleje el total de la primera parte (suma de puntos de los cinco primero items) y otra P2 que refleje el total de la segunda parte (suma de puntos de los items 6 a 10) y obtener la correlación entre ambas. (Para obtener la correlación puede ayudar utilizar una columna "P1.P2" de productos, si es necesario).

| | **I1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **I10** | **P1** | **P2** | **P1.P2** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **S1** | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 3 | 1 | 1 | 7 | 8 | 56 |
| **2** | 2 | 2 | 3 | 3 | 3 | 3 | 4 | 2 | 2 | 2 | 13 | 13 | 169 |
| **3** | 3 | 3 | 4 | 4 | 3 | 3 | 3 | 4 | 4 | 3 | 17 | 17 | 289 |
| **4** | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 5 | 7 | 35 |
| **5** | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 4 | 22 | 22 | 484 |
| **6** | 4 | 3 | 3 | 4 | 3 | 4 | 3 | 4 | 3 | 4 | 17 | 18 | 306 |
| **7** | 1 | 1 | 2 | 3 | 1 | 1 | 1 | 3 | 2 | 1 | 8 | 8 | 64 |
| **8** | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 10 | 9 | 90 |
| **9** | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 6 | 6 | 36 |
| **S10** | 5 | 5 | 5 | 4 | 5 | 5 | 5 | 4 | 5 | 5 | 24 | 24 | 576 |
| **Sum** | 25 | 25 | 26 | 28 | 25 | 26 | 26 | 29 | 26 | 25 | 129 | 132 | 2105 |
| **Med** | 2,5 | 2,5 | 2,6 | 2,8 | 2,5 | 2,6 | 2,6 | 2,9 | 2,6 | 2,5 | 12,9 | 13,2 | 210,5 |
| **DT** | 1,565 | 1,432 | 1,356 | 1,166 | 1,285 | 1,356 | 1,356 | 1,3 | 1,497 | 1,36 | 6,457 | 6,274 | 186,1 |

Correlación de Pearson r entre los totales P1 y P2 de ambas partes:

$$r = \frac{210'5 - 170'28}{6'456779 \cdot 6'273755} = 0'992884$$

Estimación del coeficiente de fiabilidad del test aplicando la corrección de Spearman-Brown caso de longitud doble a la correlación anterior:

$$r_{xx} = \frac{2r}{1+r} = \frac{2 \cdot 0'992884}{1+0'992884} = 0'996429$$

**Observación:**

Nótese que hay tres métodos de cálculo para obtener el coeficiente de fiabilidad por un método de partes (Spearman-Brown, Rulon y Guttman). *Cualquiera de ellos puede aplicarse sobre cualquiera de los tres métodos de partición en dos partes de los items de un test (dos mitades, pares-impares, o "ad hoc")*.

*Distintas particiones dan, en general, distintas estimaciones del coeficiente de fiabilidad del mismo test.*

**4. Método de Dos partes aplicando la fórmula de Rulon.**

Aplicamos un test que tiene 10 items a una muestra de 10 sujetos. Obtener el coeficiente de fiabilidad por el método de las dos mitades aplicando la fórmula de Rulon.

**Datos:**

(Nota: Son los mismos datos del problema anterior).

| | | | | | Items: | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ***Caso*** | **I1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **I10** |
| ***S1*** | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 3 | 1 | 1 |
| ***2*** | 2 | 2 | 3 | 3 | 3 | 3 | 4 | 2 | 2 | 2 |
| ***3*** | 3 | 3 | 4 | 4 | 3 | 3 | 3 | 4 | 4 | 3 |
| ***4*** | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 |
| ***5*** | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 4 |
| ***6*** | 4 | 3 | 3 | 4 | 3 | 4 | 3 | 4 | 3 | 4 |
| ***7*** | 1 | 1 | 2 | 3 | 1 | 1 | 1 | 3 | 2 | 1 |
| ***8*** | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 |
| ***9*** | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 2 |
| ***S10*** | 5 | 5 | 5 | 4 | 5 | 5 | 5 | 4 | 5 | 5 |

**Solución:**

Necesitamos una columna P1 que refleje el total de la primera parte (suma de puntos de los cinco primero items) y otra P2 que refleje el total de la segunda parte (suma de puntos de los items 6 a 10) y, además, una columna "d" que refleje la diferencia entre ellas (d= P1-P2), y otra X que refleje el total (X=P1+P2) para obtener sus varianzas y aplicar la fórmula de Rulon.

| | I1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | I10 | P1 | P2 | X | d |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **S1** | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 3 | 1 | 1 | 7 | 8 | 15 | -1 |
| **2** | 2 | 2 | 3 | 3 | 3 | 3 | 4 | 2 | 2 | 2 | 13 | 13 | 26 | 0 |
| **3** | 3 | 3 | 4 | 4 | 3 | 3 | 3 | 4 | 4 | 3 | 17 | 17 | 34 | 0 |
| **4** | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 5 | 7 | 12 | -2 |
| **5** | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 4 | 22 | 22 | 44 | 0 |
| **6** | 4 | 3 | 3 | 4 | 3 | 4 | 3 | 4 | 3 | 4 | 17 | 18 | 35 | -1 |
| **7** | 1 | 1 | 2 | 3 | 1 | 1 | 1 | 3 | 2 | 1 | 8 | 8 | 16 | 0 |
| **8** | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 10 | 9 | 19 | 1 |
| **9** | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 6 | 6 | 12 | 0 |
| **S10** | 5 | 5 | 5 | 4 | 5 | 5 | 5 | 4 | 5 | 5 | 24 | 24 | 48 | 0 |
| **Sum** | 25 | 25 | 26 | 28 | 25 | 26 | 26 | 29 | 26 | 25 | 129 | 132 | 261 | -3 |
| **Med** | 2,5 | 2,5 | 2,6 | 2,8 | 2,5 | 2,6 | 2,6 | 2,9 | 2,6 | 2,5 | 12,9 | 13,2 | 26,1 | -0,3 |
| **DT** | 1,565 | 1,432 | 1,356 | 1,166 | 1,285 | 1,356 | 1,356 | 1,3 | 1,497 | 1,36 | 6,457 | 6,274 | 12,71 | 0,781 |
| **Var** | 2,45 | 2,05 | 1,84 | 1,36 | 1,65 | 1,84 | 1,84 | 1,69 | 2,24 | 1,85 | 41,69 | 39,36 | 161,5 | 0,61 |

Aplicando la fórmula de Rulon obtenemos el coeficiente de fiabilidad del test:

$$r_{xx} = 1 - \frac{s_d^2}{s_x^2} = 1 - \frac{0'61}{161'5} = 0'9962$$

**Observación:**

Los métodos de Rulon y Guttman siempre dan el mismo resultado para una misma partición del test. El método de Spearman-Brown aplicado sobre la misma partición dará igual que estos otros si las varianzas son iguales. Si las varianzas de las partes no son iguales, el procedimiento de Spearman-Brown dará una estimación del coeficiente de fiabilidad superior a la de Rulon y Guttman; (en general, solo ligeramente superior).

## 5. Método de Dos partes aplicando la fórmula $L_4$ de Guttman.

Aplicamos un test que tiene 10 items a una muestra de 10 sujetos. Obtener el coeficiente de fiabilidad por el método de las dos mitades aplicando la fórmula de Guttman.

**Datos:**

Los mismos datos del problema anterior.

**Solución:**

Necesitamos la misma tabla del problema anterior, pero podemos prescindir de la columna de diferencias d.

| | It. 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | P1 | P2 | X |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Su.1 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 3 | 1 | 1 | 7 | 8 | 15 |
| 2 | 2 | 2 | 3 | 3 | 3 | 3 | 4 | 2 | 2 | 2 | 13 | 13 | 26 |
| 3 | 3 | 3 | 4 | 4 | 3 | 3 | 3 | 4 | 4 | 3 | 17 | 17 | 34 |
| 4 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 5 | 7 | 12 |
| 5 | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 4 | 22 | 22 | 44 |
| 6 | 4 | 3 | 3 | 4 | 3 | 4 | 3 | 4 | 3 | 4 | 17 | 18 | 35 |
| 7 | 1 | 1 | 2 | 3 | 1 | 1 | 1 | 3 | 2 | 1 | 8 | 8 | 16 |
| 8 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 10 | 9 | 19 |
| 9 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 6 | 6 | 12 |
| 10 | 5 | 5 | 5 | 4 | 5 | 5 | 5 | 4 | 5 | 5 | 24 | 24 | 48 |
| Sum | 25 | 25 | 26 | 28 | 25 | 26 | 26 | 29 | 26 | 25 | 129 | 132 | 261 |
| Med | 2,5 | 2,5 | 2,6 | 2,8 | 2,5 | 2,6 | 2,6 | 2,9 | 2,6 | 2,5 | 12,9 | 13,2 | 26,1 |
| DT | 1,56 | 1,43 | 1,36 | 1,16 | 1,28 | 1,36 | 1,36 | 1,3 | 1,5 | 1,36 | 6,46 | 6,27 | 12,7 |
| Var | 2,45 | 2,05 | 1,84 | 1,36 | 1,65 | 1,84 | 1,84 | 1,69 | 2,24 | 1,85 | 41,7 | 39,4 | 161,5 |

Aplicando la fórmula de Guttman:

$$r_{xx} = 2(1 - \frac{s^2_{x_1} + s^2_{x_2}}{s^2_x}) = 2(1 - \frac{41'69 + 39'36}{161'5}) = 0'9962$$