7

Consistencia interna

1. Descomposición de la varianza del total del test en varianzas y covarianzas de los items

La teoría clásica está pensada básicamente para el caso sencillo en que la puntuación total de un sujeto en un test se obtiene como la simple suma de sus puntuaciones en los items que forman el test.

Es decir, siendo X_T la puntuación total y X_n la puntuación del ítem n, tendremos:

$$X_T = X_1 + X_2 + X_3 + \ldots + X_n$$

Esto puede escribirse más brevemente:

$$X_T = \sum_{i=1}^{n} X_i$$

En estas condiciones es fácil descomponer la varianza del test en la varianza y covarianza de sus items componentes.

Si un test tuviese solo dos items, entonces:

$$X_T = X_1 + X_2$$

y su varianza, la del total del test, sería:

$$s_x^2 = s_{x_1}^2 + s_{x_2}^2 + 2s_{x_1 x_2}$$

En la fórmula anterior tenemos que:

s_x^2 simboliza la varianza del test total, es decir, la varianza de la variable X_T

$s_{x_1}^2$ y $s_{x_2}^2$ representan respectivamente la varianza del primer y del segundo ítem, y en adelante por simplicidad representaremos como s^2_1 y s^2_2 respectivamente.

$s_{x_1 x_2}$ representa la covarianza entre ambos items.

Es decir, la varianza del total del test de un test formado por dos items puede descomponerse en la suma de las varianzas de los items más el duplo de su covarianza.

Esto puede demostrarse fácilmente. En puntuaciones diferenciales

$$x = X_T - \bar{X}_T$$

para mayor comodidad, y representando por N el número de sujetos, tendremos:

$$s^2_x = \frac{\sum x^2}{N} = \frac{\sum (x_1 + x_2)^2}{N} = \frac{\sum (x_1^2 + x_2^2 + 2\, x_1 x_2)}{N} =$$

$$= \frac{\sum x_1^2}{N} + \frac{\sum x_2^2}{N} + 2\frac{\sum x_1 x_2}{N} = s^2_{x_1} + s^2_{x_2} + 2 s_{x_1 x_2}$$

Puede observarse que la varianza del test total es igual a la suma de la matriz de varianzas-covarianzas entre los items:

items:	X_1	X_2
X_1	$s^2_{x_1}$	$s_{x_1 x_2}$
X_2	$s_{x_1 x_2}$	$s^2_{x_2}$

En la matriz de varianzas covarianzas, como puede apreciarse, la varianza de cada ítem aparece una vez, en la diagonal principal, por lo que para n items hay n varianzas. Sin embargo la covarianza entre cada par de items aparece dos veces. Para n items hay n(n-1) covarianzas.

Del mismo modo puede demostrarse que para un test formado por n items tendremos:

$$\boxed{s^2_x = \sum s^2_i + \sum s_{ij}}$$

Es decir, que la varianza del total del test es igual a la suma de las n varianzas de los items más la suma de las n(n-1) covarianzas entre items.

Reexpresión de las covarianzas como productos de correlaciones y desviaciones típicas

Por otra parte, el coeficiente de correlación de Pearson entre dos variables puede expresarse como el cociente entre su covarianza y el producto de sus desviaciones típicas. Es decir:

$$r_{ij} = \frac{s_{ij}}{s_i s_j}$$

De donde puede deducirse que la covarianza entre dos variables es igual al producto de su correlación por sus desviaciones típicas:

$$s_{ij} = r_{ij} s_i s_j$$

Considerando esta igualdad pueden reexpresarse las fórmulas anteriores de descomposición de la varianza del total del test:

$$\boxed{s_x^2 = \sum s_i^2 + \sum r_{ij} s_i s_j}$$

Esta descomposición del test en el sumatorio de las n varianzas más el sumatorio de las n(n-1) covarianzas, expresadas a veces como productos de correlaciones y desviaciones típicas, servirá para deducir, bajo algunos supuestos, fórmulas destacadas de la teoría clásica de tests.

Descomposición de la varianza verdadera y de error del total del test en función de los items

El modelo básico de la teoría clásica de tests sostiene que una puntuación empírica X se descompone en una puntuación verdadera V y un error de medida E. Es decir:

$$X = V + E$$

Este modelo puede aplicarse a la puntuación total de un test. De ese modo la puntuación total del test puede verse como formada por esas dos partes, puntuación verdadera y error.

A su vez, cada uno de los items también puede considerarse una medida compuesta por una puntuación verdadera y una puntuación de error. Por ello, una puntuación empírica en el total del test puede verse como la suma de las puntuaciones verdaderas y de error de los items.

Si la puntuación total es igual a

$$X_T = \sum_{i=1}^{n} X_i$$

siendo

$$X_T = V_T + E_T$$

y, por otra parte,

$$X_i = V_i + E_i$$

entonces:

$$V_T + E_T = \sum_{i=1}^{n} V_i + \sum_{i=1}^{n} E_i$$

Por coherencia del modelo parece razonable suponer que:

$$V_T = \sum_{i=1}^{n} V_i$$

y que, del mismo modo:

$$E_T = \sum_{i=1}^{n} E_i$$

En este punto podemos analizar cual es la varianza de la puntuación verdadera del total del mismo modo que hemos hecho con la puntuación total del test.

Como la puntuación verdadera V_T ha resultado la suma de n variables V_i, entonces, la varianza de V_T sera igual a la suma de n varianzas V_i más n(n-1) covarianzas entre V_i.

En resumen:

$$\boxed{s_V^2 = \sum s_{V_i}^2 + \sum r_{V_i V_j} s_{V_i} s_{V_j}}$$

Del mismo modo se puede razonar con respecto a los errores. En este caso tendremos que la varianza de los errores de las puntuaciones totales del test es:

$$s_E^2 = \sum s_{E_i}^2 + \sum r_{E_i E_j} s_{E_i} s_{E_j}$$

Como la correlación entre errores es en cualquier caso 0, dada la aleatoriedad del error, la fórmula anterior se simplifica:

$$\boxed{s_E^2 = \sum s_{E_i}^2}$$

2. Descomposición de la varianza de un test supuesto que sus items son paralelos

Si los items de un test fueran mediciones paralelas entre sí, entonces las puntuaciones verdaderas de un sujeto en cada ítem serían iguales y las varianzas de error de los items serían iguales. Sabemos que, al satisfacer estas dos condiciones se cumplirían también, necesariamente, otras:

-Las medias de las puntuaciones empíricas de los items serían iguales.

-Las medias de las puntuaciones verdaderas de los items serían iguales.

-Las medias de las puntuaciones de error de los items serían iguales entre sí e iguales a 0 (este es un supuesto general que se aplica también a items no paralelos),

-Las varianzas de las puntuaciones verdaderas serían iguales entre si (ello implica, claro está, que las desviaciones típicas de las puntuaciones verdaderas serían iguales entre si).

-Las varianzas de las puntuaciones empíricas serían iguales (ello implica que las desviaciones típicas de las puntuaciones empíricas serían iguales).

-Las correlaciones entre las puntuaciones verdaderas de los items han de ser todas iguales entre si y todas iguales a 1.

-Las correlaciones entre las puntuaciones empíricas de cualesquiera dos items paralelos entre si han de ser iguales entre si. (Si un ítem j es paralelo a otro i, entonces la correlación entre ellos es el coeficiente de fiabilidad del ítem).

-Si las desviaciones típicas son iguales entre sí y las correlaciones también son iguales entre sí, entonces las covarianzas también serán iguales entre sí.

Estas igualdades entre los estadísticos de los diferentes items sugieren inmediatamente que, bajo el supuesto de items paralelos, las ecuaciones anteriores que relacionan la varianza total del test y las varianzas y covarianzas de los items se pueden simplificar.

En efecto, si las *varianzas empíricas de los items* son iguales, entonces el sumatorio de n varianzas de items es igual a n por la varianza de un ítem:

$$\sum s_i^2 = n s_i^2$$

Por otra parte, si las *covarianzas* son iguales entre sí, entonces el sumatorio de n(n-1) covarianzas empíricas ha de ser igual a n(n-1) por la covarianza de dos items cualquiera:

$$\sum s_{ij} = n(n-1) s_{ij}$$

Expresado en términos de correlaciones y desviaciones típicas:

$$\sum s_{ij} = n(n-1) r_{ij} s_i s_j$$

A su vez como las desviaciones típicas son iguales puede escribirse:

$$\sum s_{ij} = n(n-1) r_{ij} s_i^2$$

Por tanto, la descomposición de la varianza del total del test:

$$s_x^2 = \sum s_i^2 + \sum s_{ij}$$

puede escribirse, si los items son mediciones paralelas, como:

$$s_x^2 = n s_i^2 + n(n-1) r_{ij} s_i^2$$

Puede sacarse factor común, para simplificar la expresión:

$$\boxed{s_x^2 = n s_i^2 (1 + (n-1) r_{ij})}$$

Esta es la *varianza de las puntuaciones empíricas del total del test supuesto que los items son paralelos entre sí.*

En la fórmula n es el número de items, s_i^2 la varianza de un ítem cualquiera (todas son iguales entre sí), y r_{ij} la correlación entre dos items cualquiera (todas las correlaciones entre items son iguales entre sí). Obsérvese que la correlación r_{ij} entre dos items paralelos es la fiabilidad del ítem.

La media de las puntuaciones empíricas en el total del test, supuesto que los items son paralelos, como función de las medias de los items

Si los items son paralelos, entonces sus medias empíricas son iguales. Si el test total tiene n items paralelos cuyas medias son iguales entonces la media en el total del test es igual a n por la media de un ítem.

$$\bar{X}_T = n\bar{X}_i$$

3. Descomposición de la varianza verdadera del total de un test supuesto que sus items son paralelos

Hemos visto que la varianza de las puntuaciones verdaderas del total del test puede descomponerse en la suma de las n varianzas de las puntuaciones verdaderas de los items más la suma de sus n(n-1) sus covarianzas:

$$s_V^2 = \sum s_{V_i}^2 + \sum r_{V_i V_j} s_{V_i} s_{V_j}$$

Si se supone que los items son paralelos, entonces las n varianzas serán iguales entre si, y las n(n-1) covarianzas iguales entre si, lo que permite eliminar los sumatorios de la fórmula anterior:

$$s_V^2 = n s_{V_i}^2 + n(n-1)\, r_{V_i V_j} s_{V_i} s_{V_j}$$

Además, todas las correlaciones entre puntuaciones verdaderas de mediciones paralelas entre sí han de ser 1, y todas las desviaciones típicas han de ser iguales entre sí.

Por tanto, la fórmula anterior se simplifica:

$$s_V^2 = n s_{V_i}^2 + n(n-1)\, s_{V_i}^2$$

Sacando factor común:

$$s^2_v = n s^2_{V_i} (1 + (n - 1))$$

Por último, la expresión puede resumirse:

$$\boxed{s^2_v = n^2 s^2_{V_i}}$$

4. Descomposición de la varianza de error del total de un test supuesto que sus items son paralelos

Hemos visto que la varianza de las puntuaciones de error del total del test puede descomponerse en la suma de las n varianzas de las puntuaciones de error de los items más la suma de sus n(n-1) sus covarianzas, habiendo llegado a la expresión:

$$s^2_E = \sum s^2_{E_i}$$

Recuérdese que las n(n-1) covarianzas de error son nulas, debido a la naturaleza aleatoria del error.

Si suponemos items paralelos entonces las n varianzas de error son iguales entre si lo que permite llegar a la expresión:

$$\boxed{s^2_E = n s^2_{E_i}}$$

5. La relación entre fiabilidad y longitud: La profecía de Spearman-Brown

Hemos definido el coeficiente de fiabilidad de un test como el cociente entre la varianza verdadera y la varianza empírica,

$$r_{XX} = \frac{s^2_V}{s^2_X}$$

Los análisis anteriores nos han permitido definir la varianza verdadera y empírica del total de un test formado por items paralelos.

Ahora vamos a reescribir la fórmula anterior (fiabilidad del test total) en función de las nuevas expresiones que hemos encontrado para varianza verdadera y para varianza de error.

De acuerdo con esas fórmulas podemos decir, sustituyendo, que la fiabilidad de ese test cuyos items son paralelos entre sí será igual a:

$$r_{xx} = \frac{n^2 s^2_{V_i}}{n s^2_i (1 + (n-1) r_{ij})}$$

Esta fórmula puede reescribirse de modo más sencillo,

$$r_{xx} = n \frac{s^2_{V_i}}{s^2_i} \cdot \frac{1}{1 + (n-1) r_{ij}}$$

La expresión:

$$\frac{s^2_{V_i}}{s^2_i}$$

es el cociente entre la varianza verdadera y empírica de un ítem, es decir, el coeficiente de fiabilidad del ítem, es decir, tratándose de items paralelos, la correlación entre dos items paralelos:

$$\frac{s^2_{V_i}}{s^2_i} = r_{ij} = r_{xx}$$

Esto permite escribir la fórmula de la fiabilidad del test total de un modo mucho más simple, a partir de la fiabilidad de los items:

$$\boxed{r_{xx} = \frac{n r_{ij}}{1 + (n-1) r_{ij}}}$$

La fórmula anterior se conoce como profecía de Spearman-Brown.

Uso y significado de la fórmula de Spearman-Brown

La fórmula que acabamos de deducir permite calcular la fiabilidad de un test de longitud n a partir de la fiabilidad de los items r_{ij} .

Aunque hemos deducido la fórmula de Spearman-Brown para un test compuesto de n items paralelos, nada impide utilizarla para calcular la fiabilidad de cualquier medida compuesta por n submedidas paralelas, sean estas items, grupos de items o tests completos.

Es decir, la fórmula sería igualmente cierta si en lugar de considerar un test formado por n items paralelos estuviésemos hablando en todas las deducciones anteriores de un test formado por n componentes paralelos. Entendiendo por componente paralelo cualquier subtest paralelo tratado como una unidad.

La fórmula permite un uso general en el que se puede estimar la fiabilidad de un test compuesto por n veces una medida de fiabilidad conocida r_{ij}.

La fórmula de Spearman-Brown expresa la relación entre longitud y fiabilidad bajo el supuesto de que todos los items (o partes) son mediciones paralelas (entre sí).

La fórmula puede escribirse de un modo más sencillo, prescindiendo de los subíndices.

$$R = \frac{nr}{1 + (n - 1)r}$$

La fórmula de Spearman-Brown profetiza el coeficiente de fiabilidad R que tendrá el test de fiabilidad inicial r si se aumenta (o disminuye) su longitud n veces (añadiendo o retirando items paralelos).

La n de la fórmula de Spearman-Brown significa el número de veces que el test final contiene al test inicial:

$$n = \frac{\text{número} - \text{de} - \text{items} - \text{del} - \text{test} - \text{final}}{\text{número} - \text{de} - \text{items} - \text{del} - \text{test} - \text{inicial}}$$

Si se trata de un test compuesto por la suma de n items paralelos y estimamos la fiabilidad del test a partir de la correlación entre dos cualquiera de ellos (fiabilidad), como en nuestra deducción anterior, entonces n representa directamente el número de items, el número de unidades de medición paralelas que forman el compuesto aditivo. O, dicho de otro modo, el número de veces que la medición final (test) contiene la medición inicial (unidad).

La fórmula de Spearman-Brown es importante en teoría clásica de tests. Se puede llegar a ella por varios caminos y, a su vez, posibilita después otros desarrollos importantes.

Hemos visto con anterioridad alguna aplicación importante de la misma, al permitir estimar la fiabilidad del test total a partir de la correlación entre dos (ó tres) partes paralelas entre sí. En esos casos n vale respectivamente 2 (ó 3) y r es la correlación entre esas partes.

Otra aplicación de la fórmula consiste en calcular cuantas veces (n) ha de incrementarse la longitud inicial del test, dada su fiabilidad (r), para alcanzar una fiabilidad (R). Este uso de la fórmula permite determinar cuantos items paralelos a los del test habría que añadir para alcanzar, teóricamente, una determinada fiabilidad (R) del test.

Reducción de la fiabilidad al reducir la longitud del test (multiplicando esta por el factor n) para tests de fiabilidad inicial 0'99, 0'95, 0'90, 0'85 y 0'80

	Fiabilidad inicial:				
n	**.990**	**.950**	**.900**	**.850**	**.800**
.010	.497	.160	.083	.054	.038
.020	.664	.275	.153	.102	.074
.030	.748	.363	.213	.145	.107
.040	.798	.432	.265	.185	.138
.050	.832	.487	.310	.221	.167
.060	.856	.533	.351	.254	.194
.070	.874	.571	.387	.284	.219
.080	.888	.603	.419	.312	.242
.090	.899	.631	.448	.338	.265
.100	.908	.655	.474	.362	.286
.110	.916	.676	.497	.384	.306
.120	.922	.695	.519	.405	.324
.130	.928	.712	.539	.424	.342
.140	.933	.727	.558	.442	.359
.150	.937	.740	.574	.459	.375
.160	.941	.752	.590	.476	.390
.170	.944	.764	.605	.491	.405
.180	.947	.774	.618	.505	.419
.190	.950	.783	.631	.518	.432
.200	.952	.792	.643	.531	.444
.210	.954	.800	.654	.543	.457
.220	.956	.807	.664	.555	.468
.230	.958	.814	.674	.566	.479
.240	.960	.820	.684	.576	.490
.250	.961	.826	.692	.586	.500
.260	.963	.832	.701	.596	.510
.270	.964	.837	.708	.605	.519
.280	.965	.842	.716	.613	.528
.290	.966	.846	.723	.622	.537
.300	.967	.851	.730	.630	.545

.310	.968	.855	.736	.637	.554
.320	.969	.859	.742	.645	.561
.330	.970	.862	.748	.652	.569
.340	.971	.866	.754	.658	.576
.350	.972	.869	.759	.665	.583
.360	.973	.872	.764	.671	.590
.370	.973	.875	.769	.677	.597
.380	.974	.878	.774	.683	.603
.390	.975	.881	.778	.688	.609
.400	.975	.884	.783	.694	.615
.410	.976	.886	.787	.699	.621
.420	.977	.889	.791	.704	.627
.430	.977	.891	.795	.709	.632
.440	.978	.893	.798	.714	.638
.450	.978	.895	.802	.718	.643
.460	.979	.897	.805	.723	.648
.470	.979	.899	.809	.727	.653
.480	.979	.901	.812	.731	.658
.490	.980	.903	.815	.735	.662
.500	.980	.905	.818	.739	.667
.510	.981	.906	.821	.743	.671
.520	.981	.908	.824	.747	.675
.530	.981	.910	.827	.750	.679
.540	.982	.911	.829	.754	.684
.550	.982	.913	.832	.757	.688
.560	.982	.914	.834	.760	.691
.570	.983	.915	.837	.764	.695
.580	.983	.917	.839	.767	.699
.590	.983	.918	.842	.770	.702
.600	.983	.919	.844	.773	.706
.610	.984	.921	.846	.776	.709
.620	.984	.922	.848	.778	.713
.630	.984	.923	.850	.781	.716
.640	.984	.924	.852	.784	.719
.650	.985	.925	.854	.786	.722
.660	.985	.926	.856	.789	.725
.670	.985	.927	.858	.792	.728
.680	.985	.928	.860	.794	.731
.690	.986	.929	.861	.796	.734
.700	.986	.930	.863	.799	.737
.710	.986	.931	.865	.801	.740
.720	.986	.932	.866	.803	.742
.730	.986	.933	.868	.805	.745
.740	.987	.934	.869	.807	.747
.750	.987	.934	.871	.810	.750
.760	.987	.935	.872	.812	.752
.770	.987	.936	.874	.814	.755
.780	.987	.937	.875	.815	.757
.790	.987	.938	.877	.817	.760
.800	.988	.938	.878	.819	.762
.810	.988	.939	.879	.821	.764
.820	.988	.940	.881	.823	.766
.830	.988	.940	.882	.825	.769
.840	.988	.941	.883	.826	.771
.850	.988	.942	.884	.828	.773
.860	.988	.942	.886	.830	.775
.870	.989	.943	.887	.831	.777
.880	.989	.944	.888	.833	.779
.890	.989	.944	.889	.835	.781
.900	.989	.945	.890	.836	.783
.910	.989	.945	.891	.838	.784
.920	.989	.946	.892	.839	.786
.930	.989	.946	.893	.841	.788
.940	.989	.947	.894	.842	.790
.950	.989	.948	.895	.843	.792
.960	.990	.948	.896	.845	.793
.970	.990	.949	.897	.846	.795
.980	.990	.949	.898	.847	.797
.990	.990	.950	.899	.849	.798
1.00	**.990**	**.950**	**.900**	**.850**	**.800**

Representación gráfica del efecto sobre la fiabilidad de la reducción de la longitud

En la figura siguiente se ha representado la relación entre el valor n (en el eje de abscisas) y la fiabilidad final (en ordenadas) al reducir la longitud del test para un test de fiabilidad inicial 0'99 (linea superior). 0'95 (linea siguiente), 0'90 (tercera linea), 0'85 (cuarta linea) y 0'80 (última linea).

Es decir, la figura siguiente es la representación gráfica de los datos de la tabla anterior.

El eje de abscisas puede leerse como la proporción que el test final representa de uno inicial de tamaño unidad. Así por ejemplo el valor 0'8 en abscisas se refiere a un test compuesto solo por el 80% de items del test inicial, En ordenadas podemos leer que fiabilidad se espera después de aplicar esa reducción para un test de fiabilidad inicial dada.

Por ejemplo, para un test cuya fiabilidad inicial es 0'99 al reducir el número de sus elementos multiplicándolo por 0'8 la fórmula de Spearman-Brown pronostica una fiabilidad de 0'0988.

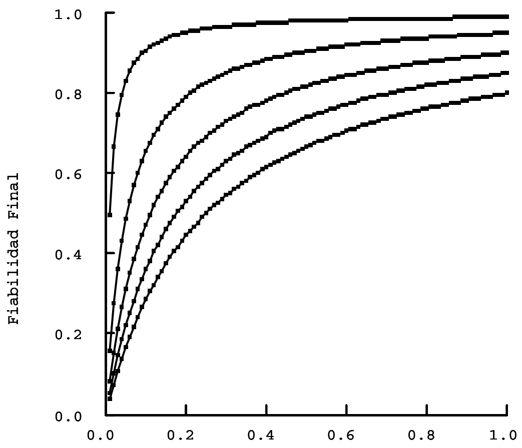
Para un test de fiabilidad inicial 0'8 al reducir sus elementos hasta dejar un 80% de los iniciales la fiabilidad pronosticada es 0'762.

Puede apreciarse que cuanto menor es la fiabilidad inicial (posición de cada curva sobre el valor 1 de abscisas, a la derecha del gráfico), mayor es el efecto sobre la fiabilidad de una determinada reducción de la longitud.

La reducción de la fiabilidad al acortar la longitud de una prueba es inversamente proporcional a n (que es el factor reductor), de modo que a más n menos reducción de la fiabilidad. Si se prefiere decirlo de otra forma, la fiabilidad final es directamente proporcional a n, cuanto mayor es n mayor es la fiabilidad final (menos se reduce está).

Obsérvese que la reducción de la fiabilidad al acortar la longitud depende por entero de n bajo este planteamiento, y es independiente del valor absoluto de la longitud inicial o final. Por ejemplo, si un test tiene una fiabilidad x y 100 items que reducimos a 50 (n=0'5) obtendremos la misma fiabilidad final que si el test tenía solo 10 items los reducimos a 5 (n=0'5) con la misma fiabilidad inicial x.

Figura. Efecto sobre la fiabilidad de la reducción de la longitud para tests de fiabilidad inicial 0'99, 0'95, 0'90, 0'85 y 0'80



6. Una nueva fórmula para el error típico de medida

Hemos visto que:

$$s_x^2 = s_V^2 + s_E^2$$

y por otro lado que:

$$s_x^2 = \sum s_i^2 + \sum s_{ij}$$

por tanto:

$$\sum s_i^2 + \sum s_{ij} = s_V^2 + s_E^2$$

De donde puede deducirse:

$$s_E^2 = \sum s_i^2 + \sum s_{ij} - s_V^2$$

$$s_E = \sqrt{\sum s_i^2 + \sum s_{ij} - s_V^2}$$

que es una nueva expresión, aunque poco operativa, del error típico de un test.

Teniendo en cuenta que:

$$r_{xx} = \frac{s_V^2}{s_X^2}$$

y que por tanto,

$$s_V^2 = r_{xx} s_X^2$$

podemos escribir esta nueva expresión del error típico de medida en función de observables:

$$\boxed{s_E = \sqrt{\sum s_i^2 + \sum s_{ij} - r_{xx} s_X^2}}$$

7. Una nueva fórmula para el error típico de medida bajo el supuesto de items paralelos

Si los items son paralelos, entonces, según hemos visto:

$$\sum s_i^2 = n s_i^2$$

$$\sum s_{ij} = n(n-1) r_{ij} s_i^2$$

$$r_{xx} = \frac{n r_{ij}}{1 + (n-1) r_{ij}}$$

$$s_X^2 = n s_i^2 (1 + (n-1) r_{ij})$$

Estas igualdades permiten reexpresar la fórmula anterior del error típico de medida del test. En efecto, el error típico de medida:

$$s_E = \sqrt{\sum s_i^2 + \sum s_{ij} - r_{xx} s_X^2}$$

puede expresarse, si los items son mediciones paralelas, como:

$$s_E = \sqrt{n s_i^2 + n(n-1) r_{ij} s_i^2 - \frac{n r_{ij}}{1 + (n-1) r_{ij}} \cdot n s_i^2 (1 + (n-1) r_{ij})}$$

Podemos operar en la fórmula anterior para encontrar una expresión más sencilla:

$$s_E = \sqrt{n s_i^2 + n(n-1) s_i^2 r_{ij} - n^2 s_i^2 r_{ij}}$$

$$s_E = \sqrt{n s_i^2 + n s_i^2 r_{ij} ((n-1) - n)}$$

$$s_E = \sqrt{ns_i^2 - ns_i^2 r_{ij}}$$

Para llegar a una expresión simple del error típico de medida del test, bajo el supuesto de items paralelos:

$$s_E = s_i \sqrt{n(1 - r_{ij})}$$

Esta es una expresión nueva para el error típico de medida de un test obtenida bajo el supuesto de items paralelos.

El supuesto puede parecer muy restrictivo, pero es el mismo bajo el que se ha desarrollado la fórmula de Spearman-Brown. La nueva fórmula que acabamos de elaborar tiene por tanto las mismas limitaciones que aquella clásica.

8. La relación entre error típico de medida y longitud del test: Una nueva fórmula de profecía

La fórmula anterior puede reescribirse de este modo:

$$s_E = \sqrt{n} \cdot s_i \sqrt{1 - r_{ij}}$$

El interés de este nuevo formato reside en que pone en relación claramente el error típico de medida del ítem y el error típico de medida del test total.

En efecto, bajo el supuesto de items paralelos r_{ij} es el coeficiente de fiabilidad del ítem (la correlación de un ítem con otro paralelo). Por tanto, el *error típico de medida del ítem* puede escribirse:

$$s_{E_i} = s_i \sqrt{1 - r_{ij}}$$

Si ponemos en relación la nueva fórmula de error típico de medida del test que acabamos de deducir y el error típico de medida del ítem observaremos que el error típico de medida del test cuando todos los items son paralelos depende exclusivamente de dos factores: del número de items y del error típico de medida de sus ítems (que, si son paralelos, es igual para todos ellos).

$$s_E = \sqrt{n} \cdot s_{E_i}$$

Este resultado resulta parcialmente paradójico con resultados clásicos dado que, según la fórmula de Spearman-Brown cuando crece un test por adición de nuevos items, ha de crecer la fiabilidad (si bien de forma negativamente acelerada). De acuerdo con la última igualdad dado un test con un error típico determinado, si crece su número de items paralelos también crecerá su error típico de medida, (si bien de modo proporcional a la raíz cuadrada del número de items).

Si la fórmula de Spearman-Brown estableció la predicción de la fiabilidad de un test al aumentar su tamaño n veces (por adición de items paralelos), nada impide considerar la fórmula que acabamos de obtener como un pronóstico del error típico de medida s_E que podemos esperar para un test que crece n veces con items paralelos a partir de uno de error típico de medida conocido s_{E_i}.

Esta fórmula no supone ningún supuesto más que la bien conocida de Spearman-Brown de modo que habrá de presentar exactamente las mismas restricciones de aplicación. Esta deducción tiene un estatus teórico "paralelo" a la de Spearman-Brown y comparte, como habrá podido verse a lo largo del capítulo, el mismo cuerpo teórico, la misma lógica deductiva y los mismos supuestos.

Esta fórmula que pronostica el error típico de medida que tendrá un test si se varía su longitud n veces puede obtenerse por diversos caminos deductivos. Así, si tenemos en cuenta que la varianza de las puntuaciones de error del test que ha variado su longitud es:

$$s_E^2 = n s_{E_i}^2$$

resulta inmediato obtener que:

$$s_E = \sqrt{n} \cdot s_{E_i}$$

Síntesis: Pronósticos para un test cuya longitud varía n veces la del test original modificando el número de elementos (items o partes) paralelos que contiene

	Test Unidad Original	Test variado n veces
Media	$\bar{X}_i$	$\bar{X}_T = n\bar{X}_i$
Varianza	s_i^2	$s_x^2 = n\,s_i^2(1 + (n - 1)r_{ij}$
Varianza de las puntuaciones verdaderas.	$s_{V_i}^2$	$s_v^2 = n^2 s_{V_i}^2$
Varianza de las puntuaciones de error.	$s_{E_i}^2$	$s_E^2 = ns_{E_i}^2$
Coeficiente de Fiabilidad	r_{ij}	$r_{XX} = \dfrac{n\,r_{ij}}{1 + (n - 1)r_{ij}}$
Error Típico de Medida.	s_{E_i}	$s_E = \sqrt{n} \cdot s_{E_i}$

En las fórmulas anteriores, n puede interpretarse como el número de items del test final partido por el número de items del test inicial.

Todas las deducciones suponen que el total del test se obtiene por adición y que todos los items (o, en su caso, partes) del test son paralelos.

Una lectura interesante pero que raras veces se subraya de los análisis anteriores es que *si los items del test son paralelos,* basta conocer la media, la desviación típica y la fiabilidad de un ítem para conocer los estadísticos de interés del test sea cual sea su longitud (media, desviación típica, fiabilidad,...). Dicho de otro modo, si los items son paralelos los estadísticos de un ítem “tipo” y la longitud determinan completamente el test.

9. Descomposición de la varianza de un test en proporciones de varianza

Hemos visto que la varianza empírica de la puntuación total de un test puede descomponerse en la suma de las n varianzas de los items más la suma de las n(n-1) covarianzas de los items:

$$s_x^2 = \sum s_i^2 + \sum s_{ij}$$

Vamos a reexpresar en términos de proporciones de la varianza total esta fórmula. Para ello dividimos a ambos lados por la varianza empírica del test:

$$\frac{s_x^2}{s_x^2} = \frac{\sum s_i^2}{s_x^2} + \frac{\sum s_{ij}}{s_x^2}$$

Podemos apreciar así que la proporción total (es decir, 1) de la varianza total se descompone en dos partes.

El primer término,

$$\frac{\sum s_i^2}{s_x^2}$$

representa la proporción de la varianza total del test debida a las varianzas de los items.

El segundo término,

$$\frac{\sum s_{ij}}{s_x^2}$$

representa la proporción de la varianza total del test debida a la covarianza entre los items.

Por tanto, puede decirse que una proporción de la varianza total del test se debe a las covariaciones entre los items y que la proporción restante se debe a las varianzas de los items.

La varianza total se descompone en estas dos partes, de modo que el incremento proporcional de una parte implicará la reducción proporcional de la otra.

La proporción de la varianza total debida a covariación entre los items puede reexpresarse de otro modo. Dado que, como acabamos de ver,

$$1 = \frac{\sum s_i^2}{s_x^2} + \frac{\sum s_{ij}}{s_x^2}$$

entonces, puede deducirse que:

$$1 - \frac{\sum s_i^2}{s_x^2} = \frac{\sum s_{ij}}{s_x^2}$$

Es decir, la proporción de varianza total debida a covariación puede expresarse como uno menos la proporción debida a la varianza de los items.

$$1 - \frac{\sum s_i^2}{s_x^2}$$

Este término juega un papel instrumental para obtener o interpretar a continuación otras fórmulas de interés práctico.

La proporción de la varianza del test total debida a covariación entre los items puede considerarse un indicador del efecto sobre el resultado del test total del grado de relación que mantienen los items entre sí.

La descomposición anterior de la varianza del test total en debida a la covariación y debida a la varianza no ha introducido ningún supuesto, excepto que la puntuación total del test es igual a la suma de las puntuaciones de los items.

10. Fiabilidad del test como función de sus componentes: el coeficiente alfa

Como hemos analizado, la varianza del test total puede descomponerse en:

$$s_x^2 = \sum s_i^2 + \sum s_{ij}$$

Si las varianzas entre items son iguales, por un lado, y las correlaciones entre items son también iguales, entonces:

$$\sum s_i^2 = ns_i^2$$

Por otra parte:

$$\sum s_{ij} = n(n-1)s_{ij} = n(n-1)r_{ij}s_i s_j = n(n-1)r_{ij}s_i^2 = (n-1)r_{ij} \cdot ns_i^2$$

De donde:

$$\sum s_{ij} = (n-1)r_{ij} \cdot \sum s_i^2$$

Podemos reescribir la descomposición de la varianza del test total teniendo en cuenta este último resultado:

$$s_X^2 = \sum s_i^2 + (n-1)r_{ij} \cdot \sum s_i^2$$

De esta última expresión podemos despejar el término r_{ij} lo que permitirá tener una fórmula de la fiabilidad de un ítem:

$$r_{ij} = \frac{s_X^2 - \sum s_i^2}{(n-1)\sum s_i^2}$$

Si ésta es la fiabilidad de un ítem, entonces a partir de la fórmula de Spearman-Brown

$$r_{XX} = \frac{nr_{ij}}{1+(n-1)r_{ij}}$$

podremos obtener una fiabilidad para la fiabilidad del test. Para ello sustituiremos r_{ij} por su fórmula anterior:

$$r_{XX} = \frac{n \cdot \dfrac{s_X^2 - \sum s_i^2}{(n-1)\sum s_i^2}}{1 + (n-1) \cdot \dfrac{s_X^2 - \sum s_i^2}{(n-1)\sum s_i^2}}$$

El resultado es fácilmente simplificable en diversos pasos:

$$r_{XX} = \frac{\dfrac{n}{(n-1)} \cdot \dfrac{s_X^2 - \sum s_i^2}{\sum s_i^2}}{1 + \dfrac{s_X^2 - \sum s_i^2}{\sum s_i^2}}$$

$$r_{XX} = \frac{\dfrac{n}{(n-1)} \cdot \dfrac{s_X^2 - \sum s_i^2}{\sum s_i^2}}{\dfrac{s_X^2}{\sum s_i^2}}$$

$$r_{XX} = \left(\frac{n}{(n-1)} \cdot \frac{s_X^2 - \sum s_i^2}{\sum s_i^2}\right) : \frac{s_X^2}{\sum s_i^2}$$

$$r_{XX} = \frac{n}{(n-1)} \cdot \frac{s_X^2 - \sum s_i^2}{s_X^2}$$

Por último:

$$\boxed{r_{XX} = \frac{n}{n-1} \cdot \left(1 - \frac{\sum s_i^2}{s_X^2}\right)}$$

Esta fórmula puede leerse así: El coeficiente de fiabilidad del test es igual al [número de items partido por el (número de items menos uno)] por [uno menos (la suma de las varianzas de los items partida por la varianza del test total)].

La fórmula anterior se conoce como ***coeficiente Alfa***. Es una fórmula muy importante en teoría clásica de tests.

En su concepto esencial fue deducida originalmente por Kuder y Richardson (1937) bajo diversos formatos, aunque inicialmente restringida a items con valoración dicotómica. Después ha sido presentada de diversas formas y deducida bajo diversos supuestos por numerosos autores.

11. Interpretación y supuestos del Coeficiente alfa

El coeficiente alfa expresa la fiabilidad total en función de dos términos: el número de items o longitud de la prueba y la proporción de la varianza total del test debida a covariación entre los items.

El primer término resalta que la fiabilidad depende de la longitud, tal y como vimos al tratar de la fórmula de Spearman-Brown. En principio cuantos más items tiene un test más fiable tiende a ser.

En realidad, en la forma de alfa de la fórmula anterior el primer término es una corrección entre 2 y 1. El primer término vale 2 cuando n=2, es decir, cuando analizamos la consistencia entre dos items o partes. El primer término vale 1 cuando $N = \infty$. Después veremos otro modo de expresar alfa donde el efecto de la longitud incrementando la consistencia es patente.

El segundo término ya fue deducido y estudiado en un apartado anterior. Su presencia en la fórmula alfa expresa que la fiabilidad del test depende del grado de covariación entre los items. Cuanto más covaríen los items entre sí mayor tiende a ser la fiabilidad del test.

Según hemos visto anteriormente:

$$1 - \frac{\sum s_i^2}{s_x^2} = \frac{\sum s_{ij}}{s_x^2}$$

Por tanto alfa puede escribirse también:

$$\boxed{r_{xx} = \frac{n}{n-1} \cdot \frac{\sum s_{ij}}{s_x^2}}$$

Lo que pone más de manifiesto las consideraciones sobre el segundo término que acabamos de hacer.

Este tipo de estimación de la fiabilidad únicamente tiene en cuenta lo consistentes que son entre sí los items y el número de los mismos, por ello este tipo de estimación de la fiabilidad se clasifica como *consistencia interna del test.*

Desde el punto de vista de la interpretación de la fórmula alfa resulta interesante considerar los diversos supuestos bajo los que puede ser deducida. Aquí hemos operado bajo supuestos implicados en la definición de mediciones paralelas. Pero, como señalábamos, hay varios caminos alternativos para llegar a alfa. Jackson y Ferguson (1941) y luego Gulliksen (1950) mostraron que basta suponer que la covarianza media entre los items paralelos del test es igual a la covarianza media entre los items no paralelos para poder deducir alfa. Estos son supuestos necesarios y suficientes.

Guttman (1945) orientó su trabajo a encontrar diversos límites inferiores del coeficiente de fiabilidad (el límite superior admisible es 1). En ese contexto, demostró que *alfa es un límite inferior del coeficiente de fiabilidad.*

$$\boxed{\alpha \leq r_{XX}}$$

Que alfa es un límite inferior del coeficiente de fiabilidad significa que alfa es menor o igual que el coeficiente de fiabilidad, es decir:

$$\frac{n}{n-1} \cdot \left(1 - \frac{\sum s_i^2}{s_X^2}\right) \leq \frac{s_V^2}{s_X^2}$$

o, de otro modo, que:

$$\frac{n}{n-1} \cdot \frac{\sum s_{ij}}{s_x^2} \leq \frac{s_V^2}{s_X^2}$$

Expresada la desigualdad de un modo más sencillo:

$$\frac{n\sum s_{ij}}{n-1} \leq s_V^2$$

Para una demostración de que alfa es un límite inferior del coeficiente de fiabilidad, ver, por ejemplo, Yela (1984).

El lector puede estar perplejo debido a que hemos deducido alfa como el coeficiente de fiabilidad para terminar diciendo que es igual o menor que el coeficiente de fiabilidad. La paradoja se desvanece si se considera que alfa es igual al coeficiente de fiabilidad *si* los componentes son *al menos* mediciones esencialmente tau-equivalentes.

Es decir, *alfa es igual al coeficiente de fiabilidad si:*

-Los items (o componentes aditivos del test) son *mediciones esencialmente tau-equivalentes.* Es decir, las puntuaciones verdaderas de una medición son iguales a las verdaderas de otra medición más una constante. Este es el supuesto mínimo, relativo al tipo de puntuaciones de los items, que proporciona alfa igual al coeficiente de fiabilidad. Es decir, podría deducirse alfa utilizando únicamente este supuesto.

-Los items son *mediciones tau-equivalentes.* Es decir, son mediciones que producen las mismas puntuaciones verdaderas.

-Los items son *mediciones paralelas.* Es decir, son mediciones que producen las mismas puntuaciones verdaderas y, además, las mismas varianzas de error.

-Los items son *mediciones replicadas.* Es decir, mediciones que producen las mismas puntuaciones verdaderas y de error. En este caso la información de los items es totalmente redundante. (Las varianzas y las covarianzas son iguales entre sí e iguales a la varianza del test total partido por n al cuadrado. El coeficiente alfa es igual a 1) El caso carece de interés práctico.

Alfa cuando los items son paralelos

Si los items son paralelos (supuesto de igual puntuación verdadera e igual varianza de error) alfa expresa la fiabilidad del test.

En ese caso, dado que todos los items tienen igual varianza y que ya deducimos a que es igual la varianza del test total para este supuesto, la fórmula alfa puede escribirse así:

$$r_{xx} = \frac{n}{n-1} \cdot \left(1 - \frac{ns_i^2}{ns_i^2(1+(n-1)r_{ij})}\right)$$

Esta fórmula es fácilmente simplificable en pasos sucesivos:

$$r_{xx} = \frac{n}{n-1} \cdot \left(1 - \frac{1}{1+(n-1)r_{ij}}\right)$$

$$r_{xx} = \frac{n}{n-1} \cdot \frac{1+(n-1)r_{ij} - 1}{1+(n-1)r_{ij}}$$

$$r_{xx} = \frac{n r_{ij}}{1 + (n - 1) r_{ij}}$$

No sorprendentemente esta última expresión es de nuevo la fórmula de Spearman-Brown. Es decir, cuando los items del test son paralelos el coeficiente alfa se reduce a la fórmula de Spearman-Brown. Es decir, cuando los componentes del test son paralelos basta conocer su fiabilidad (la correlación entre dos de ellos) y el número de componentes para conocer la fiabilidad del test completo.

Por cierto que si se siguen los cuatro últimos pasos en dirección inversa puede verse otro modo de deducir alfa a partir de Spearman-Brown bajo el supuesto de componentes paralelos.

Aspectos de interpretación práctica de alfa

Desde un punto de vista de la interpretación práctica es esencial recordar que *alfa representa la consistencia interna del test,* el grado que todos los items del test covarían entre sí. Es sin duda el estadístico de fiabilidad más utilizado.

Salvo que tengamos un interés expreso en conocer la consistencia entre dos o más partes de un test (por ejemplo, primera parte frente a segunda parte; items pares frente a impares; sección primera de un test, frente a segunda, frente a tercera, etc.) será preferible calcular alfa a aplicar métodos de dos mitades. Esto es así debido a que alfa es un indicador global de la consistencia, único para una aplicación de un test y que tiene en cuenta la covariación entre cualquier par de items. Sin embargo los métodos de mitades, o más en general de partes del test, únicamente ofrecen información sobre la consistencia entre esas dos (o más) partes fruto de una división específica entre otras muchas posibles.

El coeficiente alfa ha de presentar valores positivos entre 0 y 1. Cuanto más próximo a 1 los items son más consistentes entre sí. Generalmente los instrumentos muy consistentes producen valores por encima de 0’75 ó 0’80, pero esto es solo indicativo. La interpretación de alfa ha de tener en cuenta la longitud del test (cuantos más items mayor tiende a ser alfa) y el contexto de investigación en que se trabaja. Una buena referencia es considerar los coeficientes alfa que han obtenido otros investigadores para ese test o cuestionario u otros semejantes en condiciones semejantes. Los tests de aptitudes de composición homogénea obtienen frecuentemente coeficientes alfa por encima de 0´90 y cuestionarios de actitudes, personalidad etc. homogéneos y con un número considerable de items pueden alcanzar valores por encima de 0’80. Pero estos valores son solo orientaciones indicativas, basadas en algunas experiencias concretas, sin ningún valor como criterios.

El coeficiente alfa puede aplicarse, como se desprende del párrafo anterior, a tests, pruebas y cuestionarios virtualmente en cualquier área de la medición psicológica donde se utilicen items o partes que se espera razonablemente que midan lo mismo. Por supuesto en la medida en que se trate de un test o cuestionario cuyo contenido es heterogéneo es razonable esperar un alfa menor. Si los items (o partes en su caso) fueran completamente linealmente independientes entre sí (es decir correlacionaran 0) alfa valdría 0.

Respecto a los test de potencia y de velocidad, parece claro que alfa, como un método que evalúa la consistencia interna, no resulta adecuado para tests de velocidad dado que éstos, básicamente, contienen items acertados hasta el límite de tiempo e items omitidos desde el límite de tiempo.

Alfa puede producir resultados aberrantes, tales como números negativos, a veces incluso varios enteros por debajo de 0. Este fenómeno se produce si el cuestionario incluye items que correlacionan negativamente con otros o con el total de la prueba. En este tema hay que tener especial cuidado con las escalas de actitudes donde se recomienda la presencia de items invertidos (que miden en sentido inverso al de la escala) pero que deben ser desinvertidos en sus puntuaciones *antes* de sumarlos en el total y, antes, también, de estimar la fiabilidad mediante alfa (o cualquier otro método).

12. Relación de alfa con otros procedimientos de fiabilidad

Alfa indica la consistencia interna entre los items. No indica la estabilidad temporal ni tampoco la relación entre formas paralelas. Estos tres métodos, aunque teóricamente pueden verse como tres aproximaciones a la obtención de mediciones paralelas, son prácticamente de naturaleza y significado bien distinto.

Estimar la fiabilidad por un procedimiento de consistencia interna, como el coeficiente alfa, evalúa un aspecto del comportamiento de las puntuaciones del test completamente distinto al método de las formas paralelas o al método test-retest. Este último evaluaba la estabilidad temporal de los resultados del test cuando no habían variaciones en éstas debidas a los sujetos o a las circunstancias.

Dado que alfa es un límite inferior del coeficiente de fiabilidad cabría esperar que para un mismo test alfa fuera inferior a los coeficientes de fiabilidad obtenidos por formas paralelas o test-retest. Pero esto, en la práctica, puede no ser exactamente cierto en algunos casos debido a los aspectos diferenciales y ajenos a la fiabilidad misma que introducen estos métodos.

Sin embargo, el coeficiente alfa está muy relacionado con los métodos de las dos mitades. De hecho, las fórmulas de Rulon y la fórmula L_4 de Guttman son casos particulares del coeficiente alfa, para un test formado por dos componentes o partes.

En efecto, si tenemos un test con dos mitades y queremos obtener una estimación de la fiabilidad del test a partir de su covariación, tendremos, sustituyendo en alfa:

$$r_{xx} = 2 \cdot \left(1 - \frac{s_{X_1}^2 + s_{X_2}^2}{s_X^2} \right)$$

que es precisamente la fórmula L_4 de Guttman. Por otra parte la fórmula de Rulon es simplemente otro modo de expresión de esta fórmula.

La fórmula anterior puede, además, reexpresarse de otras formas más o menos sintéticas. Por ejemplo, dado que alfa es igual a:

$$r_{xx} = \frac{n}{n-1} \cdot \frac{\sum s_{ij}}{s_x^2}$$

aplicando esta igualdad a este caso, supuesto que los componentes son paralelos, tenemos:

$$r_{xx} = \frac{2}{2-1} \cdot \frac{2 s_{ij}}{s_x^2}$$

$$r_{xx} = \frac{4 s_{ij}}{s_x^2}$$

13. Alfa para items valorados dicotómicamente

Un caso particular de aplicación de alfa extraordinariamente común es aquel en que el test está formado por items con respuesta verdadera que se valoran, usualmente, 1 para el acierto y 0 para el error. Este es un caso muy frecuente con pruebas de inteligencia, aptitudes, rendimientos etc.

Un ítem i valorado dicotómicamente (1=Acierto; 0=Error) para una muestra de N sujetos puede describirse por los siguientes estadísticos:

1. Indice de dificultad, p ó media del ítem i:

$$p_i = I.D._i = \bar{X}_i = \frac{A_i}{N}$$

El índice de dificultad (ID), p ó media del ítem es la proporción de sujetos que aciertan el ítem. Es decir, el número de sujetos que aciertan el ítem (A) partido por el número total de sujetos (N).

Cuanto mayor es el ID más fácil es el ítem (mayor proporción de gente lo acierta).

2. q:

$$q_i = 1 - p_i = \frac{N - A_i}{N}$$

El estadístico q es simplemente la proporción de sujetos que no aciertan el ítem.

3. La varianza del ítem, s_i^2 *:*

$$s_i^2 = p_i \cdot q_i$$

Para un ítem dicotómico la varianza máxima se alcanza cuando:

$$p_i = q_i = 0'5$$

en cuyo caso la varianza es 0’25. La varianza mínima posible es 0, caso que se produce si todas las respuestas son aciertos o todas no aciertos (es decir, p ó q vale 1, con lo que, respectivamente q ó p vale 0 y la varianza 0).Por tanto puede afirmarse que la varianza de un ítem dicotómico va entre 0 y 0’25. La desviación típica por tanto va entre 0 y 0’5.

Si en un ítem dicotómicamente valorado la varianza se expresa como el producto de p y q, entonces alfa para items dicotómicos puede escribirse:

$$r_{XX} = \frac{n}{n-1} \cdot \left(1 - \frac{\sum p_i q_i}{s_X^2}\right)$$

Esta fórmula se conoce como *KR-20* debido a que era la fórmula número 20 del artículo de Kuder-Richardson. Es simplemente otra forma de alfa, donde el sumatorio de la varianza de los items ha sido sustituido por el modo más sencillo de calcular la varianza de un ítem dicotómico.

Si, además todos los items tienen la misma dificultad, entonces:

$$\sum p_i q_i = np_i q_i = np_i(1 - p_i) = np_i - np_i^2 = np_i - \frac{(np_i)^2}{n}$$

Ya vimos que si una variable (total del test) es la suma de otras (items), entonces la media de la variable suma (la media del total) es la suma de las medias de las variables componentes (las medias de los items):

$$\bar{X}_T = n\bar{X}_i$$

Para items dicotómicos las medias de los items se representan por p_i:

$$\bar{X}_T = np_i$$

Lo que permite que la fórmula anterior se puede reescribir en función de la media en el total del test:

$$\sum p_i q_i = \bar{X}_T - \frac{\bar{X}_T^2}{n}$$

Por tanto:

$$\boxed{r_{xx} = \frac{n}{n-1} \cdot \left(1 - \frac{\bar{X}_T - \frac{\bar{X}_T^2}{n}}{s_X^2}\right)}$$

Esta última fórmula se denomina KR-21. Se aplica para items valorados dicotómicamente con igual valor p_i.

Su resultado es igual al de KR-20 (e igual a alfa por tanto) si todos los ítem tienen igual dificultad (que en este caso equivale a decir igual media, o igual varianza). Si los items no tienen igual índice de dificultad entonces el resultado de KR-21 es menor que el de KR-20.

14. Fórmula de Hoyt

Es otra expresión general de alfa.

Hemos visto que alfa puede escribirse:

$$r_{xx} = \frac{n}{n-1} \cdot \frac{\sum s_{ij}}{s_x^2}$$

La suma de la covarianza de los items puede escribirse como el producto de la covarianza media entre items por las n(n-1) covarianzas que hay en n items:

$$r_{xx} = \frac{n}{n-1} \cdot \frac{n(n-1)\bar{s}_{ij}}{s_x^2}$$

y por tanto:

$$\boxed{r_{xx} = \frac{n^2 \bar{s}_{ij}}{s_x^2}}$$

que es una expresión de la denominada fórmula de Hoyt.

Esta expresión pone claramente de manifiesto que alfa depende de tres factores:

Primero del número de items.

Segundo de las covarianzas entre ellos (expresadas en la fórmula por la covarianza media entre items).

Tercero de la varianza del total del test.

Los dos primeros factores multiplican y el último divide. Los dos primeros incrementan la fiabilidad y el último la disminuye.

Manteniendo los otros términos constantes alfa será mayor cuanto mayor sea el número de items. Además la influencia del número de items es importante dado que el término entra al cuadrado, 1 ítem sólo multiplica por 1 el numerador, pero 10 ya multiplica por 100.

Manteniendo los otros factores constantes alfa será mayor cuanto más covaríen los items entre sí. Esta covariación entre items es la que refleja propiamente el significado de consistencia interna.

Manteniendo los otros factores constantes alfa será menor cuanto mayor sea la varianza total del test.

Resumiendo pues alfa depende claramente de la longitud del test y de la proporción de la varianza total debida a covariación entre los items (razón entre los últimos dos términos). Cuando más proporción de la varianza total pueda atribuirse a covariación mayor será alfa y más consistente la prueba.

A su vez, dado que:

$$r_{ij} = \frac{s_{ij}}{s_i s_j}$$

la covarianza es igual a:

$$s_{ij} = r_{ij} s_i s_j$$

y la covarianza media a:

$$\bar{s}_{ij} = \overline{r_{ij} s_i s_j}$$

Por lo que la fórmula de Hoyt puede escribirse así:

$$\boxed{r_{XX} = \frac{n^2 \overline{r_{ij} s_i s_j}}{s_X^2}}$$

Esta última expresión es la forma clásica de la fórmula de Hoyt.

15. Alfa estandarizado

Si los items se estandarizan, es decir, cada una de sus puntuaciones se convierten en puntuaciones típicas:

$$z_{X_i} = \frac{X_i - \bar{X}_i}{s_i}$$

entonces la fórmula de alfa se simplifica dado que todas las varianzas y todas las desviaciones típicas de los items expresados en típicas valen 1.

Hemos visto que alfa puede expresarse:

$$r_{XX} = \frac{n}{n-1} \cdot \left(1 - \frac{ns_i^2}{ns_i^2(1 + (n-1)r_{ij})}\right)$$

Eliminando las varianzas de los productos y considerando la correlación promedio $\bar{r}_{ij}$ de todas las n(n-1) correlaciones entre los items (dado que no suponemos igual correlación entre todos los items), la fórmula anterior puede simplificarse en pasos sucesivos hasta la siguiente expresión:

$$\boxed{r_{XX} = \frac{n\bar{r}_{ij}}{1 + (n-1)\bar{r}_{ij}}}$$

Esta expresión representa el *coeficiente alfa estandarizado.* Es decir, el coeficiente alfa que se obtendría sobre los items pasados a puntuaciones típicas.

Esta fórmula permite obtener una estimación del coeficiente de fiabilidad del test a partir de la correlación media entre los items.

De nuevo la fórmula presenta forma de ecuación de Spearman-Brown. Cuando todos los items tienen la misma varianza esta no afecta el valor de alfa, que, en ese caso, depende exclusivamente de la longitud del test n y de las correlaciones entre los items.

En el llamado *modelo de espacio muestral de items* (que considera los items del test como una muestra del espacio muestral de contenidos posibles del test) esta fórmula se interpreta como una indicación del grado en que el test representa su espacio muestral y sus resultados son generalizables a dicho espacio muestral. Esta es otra vía de interpretación teórica de alfa

Obsérvese que dada la fórmula anterior no es necesario transformar cada ítem a sus puntuaciones típicas para calcular alfa estandarizado. Basta con calcular las correlaciones entre los items y obtener la media de las n(n-1) correlaciones.

Dado que la matriz de correlaciones es simétrica, para calcular $\bar{r}_{ij}$ bastará con sumar las

$$\frac{n^2 - n}{2}$$

correlaciones entre pares de items y dividir por el número de correlaciones, es decir, por

$$\frac{n^2 - n}{2}$$

16. Alfa a partir del total del test obtenido como promedio de los items

Es frecuente, especialmente en escalas de actitudes, que el total del test en lugar de construirse como la suma de puntos de los items se obtenga como la media de las puntuaciones obtenidas por el sujeto en los items de la prueba.

El total como suma se expresa:

$$X_T = \sum_{i=1}^{n} X_i$$

mientras que el total como promedio de los items supone, para cada sujeto:

$$X_M = \frac{\sum_{i=1}^{n} X_i}{n}$$

La relación entre ambos es la siguiente:

$$X_T = n \cdot X_M$$

Es decir, el total como suma es igual al total como promedio multiplicado por n (siendo n el número de items del test).

Puede demostrarse fácilmente, a partir de la fórmula convencional de la varianza, que si una variable X_T es igual a otra X_M por una constante n, entonces, por lo que respecta a las varianzas:

$$s^2_{X_T} = n^2 s^2_{X_M}$$

Alfa está pensada para un total obtenido como suma, si en lugar de esto disponemos de un total obtenido como promedio, entonces debe aplicarse esta relación a la fórmula de alfa.

La fórmula usual de alfa:

$$r_{xx} = \frac{n}{n-1} \cdot \left(1 - \frac{\sum s_i^2}{s^2_{X_T}}\right)$$

se interpretará, utilizando la varianza del total como promedio de items, como:

$$r_{xx} = \frac{n}{n-1} \cdot \left(1 - \frac{\sum s_i^2}{n^2 s^2_{X_M}}\right)$$

Esta es la fórmula de alfa a partir del total construido como promedio. Esta fórmula sólo es otra presentación de alfa.

17. Síntesis de las relaciones entre estimaciones de la fiabilidad

Las principales relaciones pueden sintetizarse en el siguiente cuadro de equivalencias y desigualdades:

$$\boxed{KR-21 \;\leq\; KR-20 \;\equiv\; \alpha \;\leq\; r_{xx}}$$

Tanto alfa como sus variantes KR-20 y KR-21 indican la consistencia interna de una prueba y pueden considerarse límites inferiores del coeficiente de fiabilidad.

KR-20 y KR-21 solo se aplican para tests formados por items dicotómicamente valorados. Alfa puede aplicarse a cualquier tipo de items valorados cuantitativamente (incluidos los dicotómicamente valorados).

KR-21 se aplica para items dicotómicamente valorados que tienen todos la misma dificultad. Solo si todos los items tienen la misma dificultad entonces KR-21 = KR-20.

KR-20 equivale a alfa. Es simplemente una adaptación de la fórmula que facilitaba el trabajo antes de disponer de ordenadores. La equivalencia no puede interpretarse en el sentido de que donde pueda aplicarse alfa pueda aplicarse KR-20. KR-20 es un caso particular de alfa, y por tanto donde puede aplicarse KR-20 puede aplicarse alfa.

Por último, alfa es menor o igual que el coeficiente de fiabilidad. Alfa es igual al coeficiente de fiabilidad si los items del test son *al menos* mediciones esencialmente tau-equivalentes (las puntuaciones verdaderas de una medición son iguales a las verdaderas de otra medición más una constante).

Si los items son mediciones paralelas entonces alfa es igual a la fórmula de Spearman-Brown.

Al ser alfa un límite inferior de la fiabilidad teóricamente su valor para un mismo test habría de ser inferior al obtenido por formas paralelas o por test retest.

La fórmula L_4 de Guttman, (también conocida como fórmula de Flanagan y Guttman) es equivalente a la fórmula de Rulon. Ambas equivalen a alfa aplicado sobre las mitades del test definidas. Si las varianzas de las partes son iguales las tres equivalen a aplicar Spearman-Brown sobre la correlación entre las partes. Si las varianzas de las partes no son iguales las tres darán un valor inferior al resultado de aplicar Spearman-Brown sobre la correlación entre las partes.

Si un test se somete a estudio por el método de mitades, de formas paralelas y test-retest, en general el método de mitades produce la estimación de fiabilidad más alta, seguido por el método de test-retest, si bien esto depende fuertemente de las condiciones concretas de aplicación de estos métodos.

La fórmula de Hoyt equivale a alfa. Alfa a partir del total como promedio de items equivale a alfa.

18. Ejemplos de aplicación de las fórmulas de estimación

1. Pronóstico de la fiabilidad de un test, aumentando su longitud, mediante la fórmula de Spearman-Brown

Tenemos un test cuya longitud inicial es 20 y cuyo coeficiente de fiabilidad es 0´81.

¿Cuál será su fiabilidad si doblamos su longitud.(añadiendo items paralelos)?

Datos:

Longitud inicial del test (nº de items inicial): $i = 20$

Longitud final del test (nº de items final): $f = 40$

Fiabilidad inicial: $r_{xx} = 0'81$

Solución:

Relación entre longitud final y longitud inicial:

$$n = \frac{f}{i} = \frac{40}{20} = 2$$

Fiabilidad final pronosticada:

$$R_{xx} = \frac{nr_{xx}}{1 + (n-1)r_{xx}} = \frac{2 \cdot 0'81}{1 + (2-1)0'81} = 0'895$$

2. Pronóstico de la fiabilidad de un test, disminuyendo su longitud, mediante la fórmula de Spearman-Brown

Tenemos un test cuya longitud inicial es de 100 items y cuyo coeficiente de fiabilidad es 0´99.

¿Cuál será su fiabilidad si obtenemos una forma abreviada de 75 items?

¿Cuál será su fiabilidad si obtenemos una forma abreviada de 50 items?

¿Cuál será su fiabilidad si obtenemos una forma abreviada de 25 items?

¿Cuál será su fiabilidad si obtenemos una forma abreviada de 5 items?

Datos:

Longitud inicial(nº de items inicial): $i = 100$

Fiabilidad inicial: $r_{xx} = 0'99$

Primera longitud final del test (nº de items final): $f = 75$

Segunda longitud final del test (nº de items final): $f = 50$

Tercera longitud final del test (nº de items final): $f = 25$

Cuarta longitud final del test (nº de items final): $f = 5$

Solución:

Primera longitud final:

Relación entre longitud final y longitud inicial:

$$n = \frac{f}{i} = \frac{75}{100} = 0'75$$

Fiabilidad final pronosticada para 75 items:

$$R_{xx} = \frac{nr_{xx}}{1 + (n - 1)r_{xx}} = \frac{0'75 \cdot 0'99}{1 + (0'75 - 1)0'99} = 0'987$$

Segunda longitud final:

Relación entre longitud final y longitud inicial:

$$n = \frac{f}{i} = \frac{50}{100} = 0'5$$

Fiabilidad final pronosticada para 50 items:

$$R_{xx} = \frac{nr_{xx}}{1 + (n - 1)r_{xx}} = \frac{0'5 \cdot 0'99}{1 + (0'5 - 1)0'99} = 0'98$$

Tercera longitud final:

Relación entre longitud final y longitud inicial:

$$n = \frac{f}{i} = \frac{25}{100} = 0'25$$

Fiabilidad final pronosticada para 25 items:

$$R_{xx} = \frac{nr_{xx}}{1 + (n - 1)r_{xx}} = \frac{0'25 \cdot 0'99}{1 + (0'25 - 1)0'99} = 0'961$$

Cuarta longitud final:

Relación entre longitud final y longitud inicial:

$$n = \frac{f}{i} = \frac{5}{100} = 0'05$$

Fiabilidad final pronosticada para 5 items:

$$R_{xx} = \frac{nr_{xx}}{1 + (n - 1)r_{xx}} = \frac{0'05 \cdot 0'99}{1 + (0'05 - 1)0'99} = 0'832$$

Puede apreciarse que al tratarse de un test con una fiabilidad inicial muy elevada, aun drásticas reducciones en el número de items permiten mantener el pronóstico de coeficientes de fiabilidad elevados. Los primeros 75 items eliminados solo logran reducir la fiabilidad unas 3 centésimas (una variación despreciable en términos prácticos). Todavía para una longitud final de 1 ítem se pronosticaría un coeficiente de fiabilidad de 0'5 aproximadamente.

3. Estimación del número de items a añadir para mejorar una determinada fiabilidad inicial

Tenemos un test con 40 items cuyo coeficiente de fiabilidad es 0'6. ¿Cuántos items habrá que añadir para alcanzar una fiabilidad de 0'8?

Solución: [1]

[1] Nota: En algunos problemas los resultados finales se han redondeado, por simplicidad,;no así los valores de los pasos intermedios que afectarían al resultado. Las discrepancias en el resultado final por efectos de redondeo no deberían superar el intervalo $\pm 0'02$ respecto al valor obtenido.

El problema puede resolverse planteando la fórmula de Spearman-Brown y despejando n en ella:

$$R_{xx} = \frac{nr_{xx}}{1 + (n - 1)r_{xx}} \quad \rightarrow \quad 0'8 = \frac{n \cdot 0'6}{1 + (n - 1)0'6} \quad \rightarrow \quad n = 2'6667$$

Después, puede despejarse la longitud final en la fórmula de n:

$$n = 2'6667 \quad n = \frac{f}{i} \quad \rightarrow \quad 2'6667 = \frac{f}{40} \quad \rightarrow \quad f = 106'7$$

Es decir, la forma final debería tener unos 107 items.

(Dado que el número de items es variable discreta redondeamos al entero superior).

De aquí es inmediato que, según el pronóstico de Spearman-Brown, habrá que añadir 67 items a los 40 iniciales para alcanzar esa fiabilidad final

Para resolver esta clase de problemas puede resultar práctico despejar n en la fórmula anterior, para facilitar los cálculos:

$$R_{xx} = \frac{nr_{xx}}{1 + (n - 1)r_{xx}}$$

$$R + (n - 1)Rr = nr$$

$$R - Rr = nr - nRr$$

$$n = \frac{R(1-r)}{r(1-R)}$$

Dado que se persigue generalmente obtener la longitud final todavía es posible dar un paso más:

$$\frac{f}{i} = \frac{R(1-r)}{r(1-R)}$$

de donde:

$$f = \frac{iR(1-r)}{r(1-R)}$$

Aplicándolo al problema anterior tenemos:

$$f = \frac{40 \cdot 0'8(1-0'6)}{0'6(1-0'8)}$$

$$f = 106'7$$

Obteniéndose, obviamente, las mismas conclusiones.

4. Estimación del número de items que se pueden eliminar para obtener una determinada fiabilidad final

Tenemos un test con 167 items cuyo coeficiente de fiabilidad es 0'95. ¿Cuántos items podríamos suprimir para mantener una fiabilidad de 0'9?

Solución:

El problema puede resolverse planteando la fórmula de Spearman-Brown y despejando o aplicando las fórmulas con n o f despejados que hemos deducido en el problema anterior.

$$R_{xx} = \frac{nr_{xx}}{1+(n-1)r_{xx}} \quad \rightarrow \quad 0'9 = \frac{n \cdot 0'95}{1+(n-1)0'95} \quad \rightarrow \quad n = 0'474$$

$$n = \frac{f}{i} \quad \rightarrow \quad 0'474 = \frac{f}{167} \quad \rightarrow \quad f = 79'11$$

La forma final tendría unos 80 items (redondeamos los items finales al entero superior), y, por tanto, podríamos eliminar unos 87 items.

Aplicando la fórmula de f llegamos a los mismos resultados:

$$f = \frac{iR(1-r)}{r(1-R)} \quad \rightarrow \quad f = \frac{167 \cdot 0'9(1-0'95)}{0'95(1-0'9)} \quad \rightarrow \quad f = 79'11$$

5. Estimación del coeficiente alfa

Hemos aplicado un test compuesto de 8 items a 10 personas observando los siguientes resultados. Obtener el coeficiente alfa.

	i1	i2	i3	i4	i5	i6	i7	i8	T
s1	1	1	1	2	3	1	1	1	11
s2	5	4	5	4	4	3	2	1	28
s3	3	3	4	4	3	4	3	2	26
s4	5	5	5	4	4	4	5	3	35
s5	2	1	2	2	3	1	1	1	13
s6	4	3	4	4	3	5	5	3	31
s7	1	2	1	2	1	2	2	1	12
s8	5	5	5	4	5	5	4	2	35
s9	3	3	1	2	3	3	1	1	17
s10	5	5	5	5	4	4	3	1	32
Suma	34	32	33	33	33	32	27	16	240
Media	3,4	3,2	3,3	3,3	3,3	3,2	2,7	1,6	24
Des Tip	1,562	1,46969	1,735	1,1	1,005	1,4	1,487	0,8	9,263
Varianza	2,44	2,16	3,01	1,21	1,01	1,96	2,21	0,64	85,8

Sum. Var. Items =	14,64
Var. Total =	85,8
Nº de items =	8

Una vez calculadas las varianzas de los items y del total del test (columna T), simplemente se aplica la fórmula del coeficiente alfa:

$$\propto = \frac{n}{n-1} \cdot \left(1 - \frac{\sum s_i^2}{s_T^2}\right) \quad \rightarrow \quad \propto = \frac{8}{8-1} \cdot \left(1 - \frac{14'64}{85'8}\right) = 0'94$$

El coeficiente alfa vale 0'94.

Obsérvese que en la fórmula de alfa n significa el número de items, 8 en el ejemplo.

6. Estimación del coeficiente KR-20

Hemos aplicado un test compuesto de 10 items valorados dicotómicamente a 12 personas que han obtenido los siguientes resultados. Calculad el coeficiente KR-20.

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	T
s1	1	1	1	1	1	1	1	1	1	1	10
s2	1	1	1	1	1	1	1	1	0	1	9
s3	1	1	1	1	1	1	1	1	1	1	10
s4	1	1	1	1	1	0	1	1	0	1	8
s5	1	1	1	1	0	0	1	1	0	1	7
s6	1	1	1	1	0	1	1	1	0	1	8
s7	0	1	1	1	1	1	1	1	0	1	8
s8	0	1	0	1	1	1	1	1	0	1	7
s9	0	1	0	1	1	1	1	1	0	1	7
s10	0	1	0	0	1	0	0	1	0	1	4
s11	0	0	0	0	1	0	0	1	0	0	2
s12	0	0	0	1	0	0	0	0	0	0	1
p	0,5	0,833	0,583	0,833	0,75	0,583	0,75	0,917	0,167	0,833	6,75
q	0,5	0,167	0,417	0,167	0,25	0,417	0,25	0,083	0,833	0,167	
var	0,25	0,139	0,243	0,139	0,188	0,243	0,188	0,076	0,139	0,139	7,854
D.T.	0,5	0,373	0,493	0,373	0,433	0,493	0,433	0,276	0,373	0,373	2,803

La fórmula KR-20 tan solo es una variante de alfa especialmente orientada a items dicotómicamente valorados (específicamente, valorados con los valores 0 y 1). Allí donde puede aplicarse KR-20 puede aplicarse alfa con el mismo resultado (pero lo inverso no es cierto, porque KR-20 es un caso particular y no sirve donde los items no están valorados dicotómicamente).

Una vez obtenido p (=A/N) y q (=1- p) para cada ítem, se procede a obtener la varianza de cada ítem como producto de p por q. También necesitamos la varianza del total y la suma de la varianza de los items para aplicar finalmente la fórmula:

$$KR_{20} = \frac{n}{n-1} \cdot \left(1 - \frac{\sum pq}{s_T^2}\right) = \frac{10}{9} \cdot \left(1 - \frac{1'734}{7'854}\right) = 0'86$$

7. Estimación del coeficiente KR-20 y KR-21

El coeficiente KR-21 es una variante de KR-20 que iguala a este cuando todos los items presentan igual p. Si alguno de los items presenta un p distinto a los demás entonces KR-21 infraestima KR-20 y carece de sentido aplicarlo. Si todos los items tienen exactamente igual p entonces da igual aplicar KR-20 ó KR-21. Además, la situación en que todos los items tengan igual p es en términos prácticos inverosímil. Por estas razones KR-21 carece de interés práctico, sin que aporte nada teóricamente. El problema siguiente se introduce únicamente a efecto de ilustrar estos extremos de una fórmula clásicamente presente en los manuales de psicometría.

Caso 1. Todos los items tienen igual p. Hemos aplicado un test compuesto de 10 items valorados dicotómicamente a 12 personas que han obtenido los siguientes resultados. Calculad el coeficiente KR-20 y KR-21.

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	T	
s1	1	1	1	1	1	1	1	1	1	1	10	
s2	1	1	1	1	1	1	1	1	1	1	10	
s3	1	1	1	1	0	1	1	1	1	1	9	
s4	1	1	1	0	0	0	1	1	1	1	7	
s5	1	1	1	0	0	0	1	1	1	1	7	
s6	1	1	0	0	0	1	1	0	1	1	6	
s7	0	0	0	0	0	1	0	0	0	0	1	
s8	0	0	1	1	1	0	0	0	0	0	3	
s9	0	0	0	1	1	1	0	0	0	0	3	
s10	0	0	0	0	1	0	0	0	0	0	1	
s11	0	0	0	0	1	0	0	1	0	0	2	
s12	0	0	0	1	0	0	0	0	0	0	1	
p	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	5	Media
q	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5		
var	0,25	0,25	0,25	0,25	0,25	0,25	0,25	0,25	0,25	0,25	11,67	
des tip	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	3,416	

$$KR_{20} = \frac{n}{n-1} \cdot (1 - \frac{\sum pq}{s_T^2}) = \frac{10}{9} \cdot (1 - \frac{2'5}{11'67}) = 0'873$$

$$KR_{21} = 0'873$$

KR-21 utiliza otra fórmula (ver página siguiente) para expresar la suma de la varianza de los items, basada en la igualdad de las p entre estos siendo todos los demás términos iguales. Dado que en el problema anterior todos los items tienen igual p (es irrelevante que además en el ejemplo p=q) entonces se obtiene que KR-20=KR-21.

Por cierto que dado que si todos los items tienen igual p también tienen necesariamente igual varianza, entonces, en mi opinión, la fórmula clásica de KR-21 es innecesariamente complicada bastaría aplicar esta otra nueva fórmula para KR-21:

$$KR'_{21} = \frac{n}{n-1} \cdot (1 - \frac{n \cdot s_i^2}{s_T^2})$$

dado que, obviamente, en este caso:

$$\sum s_i^2 = n \cdot s_i^2$$

Aplicada la nueva fórmula de KR-21 al problema anterior tenemos:

$$KR'_{21} = \frac{n}{n-1} \cdot (1 - \frac{n \cdot s_i^2}{s_T^2}) \rightarrow KR'_{21} = \frac{10}{9} \cdot (1 - \frac{10 \cdot 0'25}{11'67}) = 0'873$$

La nueva fórmula de KR-21 que he propuesto no sirve si todos los items no tienen igual varianza, pero en ese caso tampoco sirve ni tiene sentido calcular KR-21.

Caso 2. Para ilustrar el efecto de la alteración de un solo valor p sobre KR-21 hemos variado en la tabla anterior el valor p del ítem 10.

Como ,puede apreciarse entonces KR-21 infraestima KR-20 (y no procede utilizarlo):

i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	T

s1	1	1	1	1	1	1	1	1	1	1	10
s2	1	1	1	1	1	1	1	1	1	1	10
s3	1	1	1	1	0	1	1	1	1	1	9
s4	1	1	1	0	0	0	1	1	1	1	7
s5	1	1	1	0	0	0	1	1	1	0	6
s6	1	1	0	0	0	1	1	0	1	0	5
s7	0	0	0	0	0	1	0	0	0	0	1
s8	0	0	1	1	1	0	0	0	0	0	3
s9	0	0	0	1	1	1	0	0	0	0	3
s10	0	0	0	0	1	0	0	0	0	0	1
s11	0	0	0	0	1	0	0	1	0	0	2
s12	0	0	0	1	0	0	0	0	0	0	1
p	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,333	4,833
q	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,667	
var	0,25	0,25	0,25	0,25	0,25	0,25	0,25	0,25	0,25	0,222	11,31
des tip	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,471	3,362

$$KR_{20} = \frac{n}{n-1} \cdot \left(1 - \frac{\sum pq}{s_T^2}\right) = \frac{10}{9} \cdot \left(1 - \frac{2'472}{11'31}\right) = 0'868$$

Fórmula clásica de KR_{21}:

$$KR_{21} = \frac{n}{n-1} \cdot \left(1 - \frac{\bar{X}_T - \frac{\bar{X}_T^2}{n}}{s_T^2}\right) \rightarrow KR_{21} = \frac{10}{9} \cdot \left(1 - \frac{4'833 - \frac{4'833^2}{10}}{11'31}\right) = 0'866$$

Caso 2. Segundo ejemplo.

En general la presencia de mayores diferencias entre los valores p de los items (como es usual y requerido en la mayoría de los tests con respuesta verdadera) producen una discrepancia mayor entre KR-21 y KR-20.

Así por ejemplo, en los siguientes datos, donde, como es usual cada ítem presenta su propio valor p, KR-20= 0'568 y KR-21=0'401.

Por supuesto, en este caso el valor que habría que considerar es el de KR-20.

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	T
s1	1	0	1	1	1	0	1	1	1	1	8
s2	1	0	1	1	1	0	1	1	1	1	8
s3	1	0	1	1	0	0	1	1	1	1	7
s4	1	0	1	0	0	0	1	1	1	1	6
s5	1	0	1	0	0	0	1	1	1	0	5
s6	1	0	1	0	0	1	1	1	1	0	6
s7	1	0	0	0	0	1	0	1	1	0	4
s8	1	0	1	1	0	0	0	1	1	0	5
s9	1	0	0	1	1	1	0	1	0	0	5
s10	1	0	0	0	1	0	0	0	0	0	2
s11	1	0	0	0	1	0	0	1	0	0	3
s12	0	1	0	1	0	0	0	0	0	0	2
p	0,917	0,0833	0,583	0,5	0,417	0,25	0,5	0,833	0,667	0,333	5,083
q	0,083	0,9167	0,417	0,5	0,583	0,75	0,5	0,167	0,333	0,667	
var	0,076	0,0764	0,243	0,25	0,243	0,188	0,25	0,139	0,222	0,222	3,91
des tip	0,276	0,2764	0,493	0,5	0,493	0,433	0,5	0,373	0,471	0,471	1,977

Problema planteado:

Sean los datos de 10 sujetos en cuatro items, y su total X

	X1	X2	X3	X4	X
S_1	1	2	3	4	10
S_2	3	5	7	6	21
S_3	2	5	7	8	22
S_4	1	2	3	2	8
S_5	1	2	2	1	6
S_6	5	7	8	9	29
S_7	7	8	9	9	33
S_8	4	4	6	6	20
S_9	2	4	2	4	12
S_{10}	2	6	7	9	24

Puntuaciones típicas.

	Z_1	Z_2	Z_3	Z_4
S_1	-.905	-1.179	-.911	-.605
S_2	.101	.236	.608	.067
S_3	-.402	.236	.608	.740
S_4	-.905	-1.179	-.911	-1.278
S_5	-.905	-1.179	-1.291	-1.614
S_6	1.106	1.179	.987	1.076
S_7	2.112	1.650	1.367	1.076
S_8	.603	-.236	.228	.067
S_9	-.402	-.236	-1.291	-.605
S_{10}	-.402	.707	.608	1.076

Matriz de correlaciones:

	X1	X2	X3	X4
X1	1.000			
X2	0.843	1.000		
X3	0.781	0.895	1.000	
X4	0.688	0.916	0.919	1.000

El propósito es estimar el coeficiente alfa desde diversas fórmulas.