

## ANÁLISIS DE VALIDEZ DE CONSTRUCTO DE UN INSTRUMENTO DE MEDICIÓN UTILIZANDO ANÁLISIS FACTORIAL



Eusebio Olivo, Carmen Batanero  
 eusebio.olivo@itesm.mx, batanero@ugr.es  
 Instituto Tecnológico de Estudios Superiores de Monterrey, México  
 Universidad de Granada, España  
 Reporte de Investigación  
 Superior

### Resumen

En este informe presentamos los resultados del análisis factorial exploratorio del conjunto de respuestas a los ítems de un cuestionario sobre los intervalos de confianza. El análisis factorial fue utilizado para determinar la validez de constructo del instrumento de medición. Se trataba de confirmar, por un lado, la existencia de un constructo subyacente que agrupase a la mayor parte de los ítems y que explicase el razonamiento sobre el intervalo de confianza. Al mismo tiempo se esperaba un constructo multidimensional en que algunos de los factores explicasen los conflictos denunciados en la investigación previa y los evaluados en el cuestionario. Los resultados muestran factores que dan cuenta de la dimensión lógico-matemática del constructo “razonamiento en intervalos de confianza” y el resto de los factores, los sesgos de razonamiento que tienen base psicológica.

**Palabras clave:** *Validez, evaluación, análisis factorial, superior*

### 1. Introducción

El proceso de medida en educación y psicología se propone vincular ciertos conceptos abstractos a indicadores empíricos. En particular en la evaluación educativa tratamos de establecer relaciones de los conocimientos de los alumnos sobre un concepto o bien relacionar sus actitudes con sus respuestas a los ítems de un cuestionario o una prueba. El análisis de los datos se hace sobre las respuestas, debido a que son observables. El interés teórico, no obstante, es el concepto subyacente (conocimientos, actitudes, necesidades especiales, etc.) que no podemos observar directamente, pero que tratamos de inferir a partir de las respuestas.

En la evaluación de los conocimientos de los estudiantes a partir de las respuestas al cuestionario se está sujeto a errores sistemáticos (sesgos) y aleatorios, como en todo proceso de inferencia (Carmines y Zeller, 1979; Thorndike, 1989). Los sesgos, de naturaleza determinista, aunque de magnitud desconocida, se desprenden de nuestros procedimientos, tanto durante la selección de la muestra, como en la construcción de los instrumentos de medición y en la toma o análisis de los datos. Por otro lado, cuando evaluamos a un alumno podemos plantearle cualquier número de posibles preguntas sobre el mismo tema, e incluso con las mismas preguntas, las respuestas del alumno pueden variar dependiendo de factores como atención, motivación, fatiga, etc. Debemos entonces reconocer el carácter aleatorio a los resultados de la evaluación. Los análisis de validez y fiabilidad, atienden a estas dos dimensiones de sesgo y aleatoriedad. En particular, en este estudio, la validez que analizamos es la de constructo.

Ya que un constructo es una conceptualización teórica sobre un aspecto medible del comportamiento, la validez de constructo trata de evaluar hasta qué punto una prueba mide los constructos sobre los que se sustenta. Se trata de comprobar si el instrumento mide el rasgo o concepto teórico o si se cumplen las hipótesis sobre la estructura del constructo (Martínez Arias, 1995).

Entre las técnicas posibles para analizar la validez de constructo la más usada es el análisis factorial o el análisis de componentes principales. En nuestro caso se usó el análisis factorial exploratorio del conjunto de respuestas a los ítems del cuestionario para examinar la estructura factorial de las puntuaciones al cuestionario e identificar las fuentes de variación en las medidas observadas. Se trataba de confirmar, por un lado, la existencia de un constructo subyacente que agrupase a la mayor parte de los ítems y que explicase el razonamiento sobre el intervalo de confianza. Al mismo tiempo se esperaba un constructo multidimensional en que algunos de los factores explicasen los conflictos denunciados en la investigación previa y los evaluados en el cuestionario.

## 2. Metodología del trabajo

Un cuestionario, elaborado como parte de una investigación sobre la comprensión de los intervalos de confianza (Olivo, 2008) fue aplicado a 252 alumnos del tercer semestre de las distintas carreras de ingeniería impartidas en una Universidad Privada del Norte de México. Los alumnos, que habían estudiado el tema, tuvieron que responder nuestro cuestionario de 18 ítems; 12 ítems de opción múltiple y seis de respuesta abierta. En el anexo 1 se presentan los contenidos del cuestionario y un par de ítems.

La extracción de factores se llevó a cabo mediante el método de componentes principales; con objeto de obtener factores estadísticamente independientes y de máxima variabilidad al tiempo que no se deformara la estructura de los datos. Este método parte de una estimación inicial más alta de las comunalidades (Martínez Arias, 1995). Como método de rotación se usó la Varimax, método ortogonal, que conserva la suma de porcentajes de varianza explicados por los factores (Afifi y Clark, 1990). Está orientado a maximizar la varianza de los factores. Antes de aplicar el método, se comprobaron los supuestos de aplicación (más de 10 casos por variable, unidad experimental, factorizabilidad de la matriz de correlaciones, normalidad, linealidad y ausencia de multicolinealidad).

## 3. Resultados

Las comunalidades obtenidas oscilan entre .459 (ítem 2, efecto tamaño muestra) y .764 (ítem 1, definición) lo que indica la variabilidad de cada ítem que es explicada por el conjunto de factores retenidos (Afifi y Clark, 1990). Esto indica que cada ítem tiene una parte específica fuerte, en especial alguno de ellos. Los ítems más específicos son: el ítem 2 que evalúa la comprensión del efecto del tamaño de la muestra sobre el ancho del intervalo, debido a que su proporción común a todos los ítems es la más baja, con un valor de .459 y, el que le sigue, con un valor de .512 en la proporción común, es el ítem 9, en el que se pide al alumno construir un intervalo de confianza para estimar una proporción poblacional. De los tres parámetros: media poblacional, varianza poblacional y proporción poblacional, este último es el que menos aparece en el conjunto de ítems.

En tanto que los ítems menos específicos son: el ítem 1, con un valor de .764 en la proporción común, evalúa la comprensión de la definición del intervalo de confianza; y el ítem 15, con un valor de .751 en la proporción común con el resto de los ítems, evalúa el conocimiento de la distribución muestral necesaria para calcular un intervalo de confianza para la varianza en muestras pequeñas. La comprensión de la definición del intervalo de confianza y el uso apropiado de las distribuciones muestrales son determinantes en el proceso de construcción de intervalos de confianza, de ahí su baja especificidad.

Se usaron varios métodos para determinar cuántos factores extraer. La extracción inicial obtuvo 8 factores con autovalores mayor que 1, que explicaron el 62.7 % de la varianza total (Tabla 1). También se siguió la sugerencia de Peña (2002) de que el número máximo de factores a extraer ha de ser menor a la mitad del número inicial de variables menos 1.

Tabla 1. Varianza total explicada por cada uno de los factores extraídos

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	2.919	16.219	16.219	2.919	16.219	16.219	2.225	12.364	12.364
2	1.464	8.132	24.350	1.464	8.132	24.350	1.418	7.878	20.242
3	1.344	7.466	31.816	1.344	7.466	31.816	1.373	7.625	27.868
4	1.207	6.704	38.520	1.207	6.704	38.520	1.357	7.540	35.408
5	1.177	6.539	45.059	1.177	6.539	45.059	1.279	7.106	42.514
6	1.126	6.258	51.317	1.126	6.258	51.317	1.278	7.102	49.617
7	1.039	5.772	57.089	1.039	5.772	57.089	1.219	6.770	56.387
8	1.012	5.620	62.708	1.012	5.620	62.708	1.138	6.322	62.708
9	.937	5.204	67.912						
10	.844	4.688	72.601						
11	.774	4.301	76.902						
12	.735	4.084	80.986						
13	.690	3.833	84.820						
14	.670	3.723	88.542						
15	.620	3.443	91.986						
16	.562	3.121	95.107						
17	.512	2,842	97.949						
18	.369	2.051	100.000						

El primer factor explica un 16 % de la varianza, mientras que los siguientes explican entre el ocho y el cinco por ciento, cada uno, lo que indica la importancia relativa del primer factor. Este porcentaje proporciona una evidencia de validez del constructo, en cuanto que no hay acuerdo sobre el porcentaje de varianza mínima que debe explicar el primer factor, pero sí que este porcentaje debe ser claramente superior al explicado por los restantes (y en nuestro caso es el doble). Además, el resto de los factores explican cada uno aproximadamente la misma varianza.

En el gráfico de sedimentación (scree plot) (Figura 1) se observa de nuevo que la mayor varianza es debida al primer factor (que interpretaríamos como razonamiento sobre intervalo de confianza).

Podemos notar, además, que el punto de inflexión se produce en el sexto factor, lo que sería también un criterio para decidir el número de factores a retener, pues a partir de él los autovalores de los factores siete y ocho apenas cambian de magnitud (Martínez Arias, 1995).

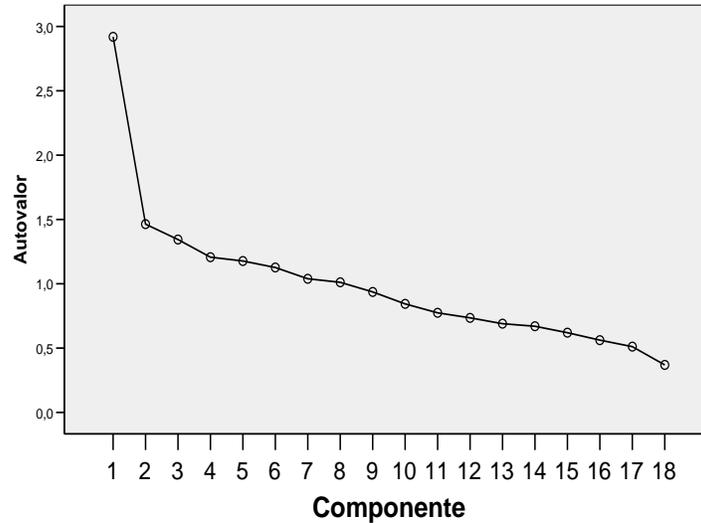


Figura 1. Gráfico de sedimentación

Por otro lado, los factores cuyo autovalor es menor que uno suelen descartarse (Díaz, 2007)

Tabla 2. Matriz no rotada de componentes

	Componente							
	1	2	3	4	5	6	7	8
i1	-.053	.346	-.069	.517	.175	-.040	.033	.580
i2	.234	-.103	.200	.144	-.568	.077	-.064	.024
i3	.140	.463	.301	-.326	-.408	.170	.129	.054
i4	.176	-.216	.413	.299	.369	-.230	.337	-.165
i5	.152	.391	.485	.173	.190	-.214	-.385	.129
i6	.546	-.399	.131	.134	.182	.222	-.090	-.078
i7	.598	-.247	-.183	-.086	-.168	-.290	.105	.130
i8	.397	-.380	-.034	.270	.000	.335	-.452	-.170
i9	.624	.116	-.093	.095	-.145	-.238	.055	.133
i10	.668	.148	-.309	.021	-.090	-.127	-.021	.072
i11	.366	.094	-.035	-.130	.322	.566	.105	.250
i12	.410	-.274	.148	-.286	.230	-.074	.419	.262
i13	.277	.422	.325	-.188	.222	.118	.129	-.438
i14	.347	.035	.127	.185	-.296	.423	.286	.068
i15	.313	.461	-.334	.432	.069	.110	.117	-.335
i16	.335	.164	-.222	-.468	.291	.083	-.390	.157
i17	.516	.127	-.263	-.099	.021	-.320	.009	-.289
i18	.404	-.036	.522	-.041	-.108	-.178	-.277	.071

Método de extracción: Análisis de componentes principales a 8 componentes extraídos

Incluimos en la Tabla 2 la matriz no rotada de componentes, donde las variables se presentan ordenadas según la importancia relativa de su contribución al primer factor.

Incluso antes de la rotación se nota que la mayoría de los ítems contribuyen con correlaciones positivas al primer factor y un número apreciable de ellos (nueve) tiene un peso importante en él, lo cual aporta una nueva evidencia de existencia del constructo subyacente (Díaz, 2007).

Con objeto de obtener una estructura más simple se realizó una rotación Varimax, que maximiza la varianza de los coeficientes que definen los efectos de cada factor sobre las variables observadas (Peña, 2002). Tras la rotación (ver Tabla 3), la estructura se clarifica, ya que las variables con correlaciones negativas prácticamente desaparecen o los valores de la correlación son muy pequeños (por debajo de 0.3), mientras que las correlaciones positivas fuertes en cada factor se mantienen o incluso crecen.

Tabla 3. Matriz de componentes rotados

	Componente							
	1	2	3	4	5	6	7	8
i1. Definición	-.005	-.125	.085	.133	-.013	.127	.024	.840
i2. Efecto tamaño muestra	.143	.165	-.208	.109	.580	-.104	-.084	-.037
i3. Efecto nivel confianza	.031	-.427	.197	.269	.506	.052	-.216	-.214
i4. Efecto varianza	.019	.095	-.031	.151	-.059	.051	.793	.000
i5. Variación en diferentes muestras	-.019	-.013	-.007	.794	-.071	.104	.043	.196
i6. Estimar media $\sigma$ conocida	.211	.607	.264	.075	.110	-.037	.276	-.105
i7. Estimar media $\sigma$ desconocida	.719	.124	.014	-.069	.101	-.203	.098	-.043
i8. Estimar media $\sigma$ desconocida, $n$ grande	.079	.825	.048	.065	.116	.065	-.085	-.062
i9. Estimar proporción	.661	.032	.045	.155	.177	.084	.052	.118
i10. Estimar varianza	.707	.105	.150	.045	.080	.187	-.113	.067
i11. Comparar medias, varianzas conocidas	.027	.151	.780	-.034	.101	.086	-.012	.093
i12. Comparar medias, v. desc. $n$ pequeña	.347	-.081	.437	-.059	.023	-.356	.442	-.081
i13. Comparar medias, v. desc. $n$ grande	-.009	-.166	.297	.344	.043	.457	.199	-.426
i14. Comparar varianzas	.083	.106	.245	-.087	.637	.140	.100	.064
i15. Elegir distribución muestral	.242	.064	.025	-.077	.057	.811	.006	.143
i16. Determinar valor crítico	.282	.086	.486	.212	-.347	-.056	-.410	-.128
i17. Interpretar salidas ordenador	.623	.014	-.034	.044	-.147	.278	.017	-.241
i18. Interpretar gráficos intervalos	.210	.163	-.018	.615	.203	-.214	.112	-.122

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser. La rotación ha convergido en 9 iteraciones.

Para facilitar la interpretación, se repiten los resultados en la Tabla 4, suprimiendo las correlaciones menores a 0.3, con lo que se nota más fácilmente la estructura.

Los factores separan claramente diferentes elementos de significado del intervalo de confianza, puesto que los primeros factores parecen ligados a la construcción de intervalos (campos de problemas y procedimientos) mientras que los últimos se ligan a las definiciones y propiedades.

*Factor Primero:* Agrupa con altas puntuaciones los ítems abiertos que piden construir intervalos para la media con desviación típica desconocida ( $r = 0.719$ ), proporción ( $r = 0.661$ ), varianza desconocida ( $r = 0.707$ ), comparar medias con varianzas desconocidas,  $n$  pequeñas ( $r = .347$ ); todos ellos requieren recordar y discriminar diferentes distribuciones muestrales:  $T$  de Student,  $F$ , y normal. Asimismo, incide en este factor la interpretación de salidas con ordenador ( $r = 0.623$ ), posiblemente porque el problema propuesto se refiere también a la comparación de dos medias.

Tabla 4. Matriz de componentes rotados simplificada

	Componente							
	1	2	3	4	5	6	7	8
i1. Definición								.840
i2. Efecto tamaño muestra					.580			
i3. Efecto nivel confianza		-.427			.506			
i4. Efecto varianza							.793	
i5. Variación en diferentes muestras				.794				
i6. Estimar media $\sigma$ conocida		.607						
i7. Estimar media $\sigma$ desconocida	.719							
i8. Estimar media $\sigma$ desconocida, $n$ grande		.825						
i9. Estimar proporción	.661							
i10. Estimar varianza	.707							
i11. Comparar medias, varianzas conocidas			.780					
i12. Comparar medias, v. desc. $n$ pequeña	.347		.437			-.356	.442	
i13. Comparar medias, v. desc. $n$ grande				.344		.457		-.426
i14. Comparar varianzas					.637			
i15. Elegir distribución muestral						.811		
i16. Determinar valor crítico			.486		-.347			-.410
i17. Interpretar salidas ordenador	.623							
i18. Interpretar gráficos intervalos				.615				

*Factor segundo: Construcción de intervalos que implican a la distribución normal.* Agrupa los problemas de estimación de la media con varianza conocida y varianza desconocida pero muestras grandes; los dos requieren de la distribución normal. Aunque se trate de construcción de intervalos, son problemas de opción múltiple, por lo que sólo podemos tener en cuenta la respuesta totalmente correcta o incorrecta, mientras que en los ítems agrupados en el primer factor tenemos en cuenta las parcialmente correctas. También el estudio de la relación del coeficiente con el ancho del intervalo, que aparece con signo contrario, lo que nos indica que algunos estudiantes que construyen bien los intervalos pedidos no comprenden la definición de coeficiente de confianza.

*Factor tercero: Construcción de intervalos que implican la distribución  $T$  de Student.* Agrupa los problemas de comparación de medias con varianza desconocida,  $n$  pequeña ( $r = 0.437$ ), y determinación de valor crítico para un intervalo con desviación estándar desconocida ( $r = 0.486$ ); los dos problemas requieren de la distribución  $T$  de Student. También se incluye el estudio de la comparación de medias, varianzas conocidas ( $r = 0.780$ ).

*Factor cuarto: Significado del nivel de confianza.* Agrupa los problemas de significado del nivel de confianza ( $r = 0.794$ ) e interpretación de gráficos de intervalos ( $r = 0.615$ ); en ambos ítems la comprensión del significado del nivel de confianza es primordial. Además se incluye en este factor la comparación de medias con varianza desconocida,  $n$  grande ( $r = 0.344$ ), e incluye también la pregunta de interpretar el resultado. El ítem es de opción múltiple y el alumno, para contestar correctamente el significado de nivel de confianza, debe comprenderlo.

*Factor quinto: Habilidad analítica e interpretativa.* Agrupa los problemas que exhiben una relación entre efecto tamaño muestra ( $r = 0.580$ ), efecto del nivel de confianza ( $r = 0.506$ ) y la comparación de varianzas ( $r = .637$ ). Podemos reconocer a este factor como la habilidad para

analizar e interpretar el efecto que se produce en el ancho del intervalo cuando cambia el tamaño de la muestra, disminuye o aumenta el nivel de confianza, o bien analizar el intervalo obtenido en el cociente de varianzas pasando o no por el uno, para finalmente hacer la interpretación correcta. Aparece en este factor el problema de determinar valor crítico con un valor negativo ( $r = -.347$ ) lo que indica que algunos estudiantes que pudieran realizar un buen análisis e interpretación no determinan correctamente el valor crítico en la construcción del intervalo.

*Factor sexto:* Agrupa los problemas de comparar medias con varianzas desconocidas,  $n$  pequeñas ( $r = -.356$ ), comparar medias con varianzas desconocidas,  $n$  grande ( $r = .457$ ); y elegir distribución muestral ( $r = .811$ ). Igual que en el factor primero, todos ellos requieren recordar y discriminar diferentes distribuciones muestrales:  $T$  de Student, normal y Chi-cuadrado. El valor negativo ( $r = -.356$ ) en el problema de comparar medias con varianzas desconocidas,  $n$  pequeñas, nos indica que la distribución que más conflictos causa en algunos estudiantes es la  $T$  de Student. Este factor, aunque de naturaleza similar al primero, solo explica el 6.25% de la varianza, en tanto que el factor primero explica el 16.22% de la varianza.

El resto de los factores sugieren que los diferentes sesgos que afectan el razonamiento de los intervalos de confianza aparecen sólo parcialmente relacionados con la ejecución matemática; algunos de los sesgos en estos factores están ligados a las definiciones y propiedades.

En consecuencia, se confirma la hipótesis previa de existencia de un constructo subyacente (definido por el peso comparativamente alto del primer factor antes de la rotación) que se subdivide en diferentes componentes, todos ellos relacionados con el razonamiento sobre intervalos de confianza.

#### 4. Conclusiones

Del análisis factorial, con el que se muestra que los seis primeros factores dan cuenta de la dimensión lógico-matemática del constructo “razonamiento en intervalos de confianza” y el resto de los factores de los sesgos de razonamiento que tienen base psicológica, según se planteó hipotéticamente en la definición semántica de la variable, podemos concluir que estos resultados aportan evidencias de validez de constructo.

Frecuentemente en la investigación en educación necesitamos aplicar o construir cuestionarios para evaluar diversas variables no observables, tales como conocimientos, necesidades o actitudes. Una multitud de factores puede influir en la variabilidad de las respuestas a la prueba; además, cualquier número de posibilidades de variación que tenemos al construir una misma prueba de evaluación dificulta hacer inferencias correctas. Al revelar la existencia de componentes diferenciados, el análisis factorial contribuye a disminuir esas dificultades a los investigadores o a profesores-investigadores interesados en el tema.

#### 5. Referencias

- Afifi, A. y Clark, V. (1990). *Computer- aided multivariate analysis*, New York: Van Nostrand Reinhold.
- Carmines, E. G. y Zeller, R. A. (1979). *Reliability and validity assessment*. London: Sage.
- Díaz, C. (2007). *Viabilidad de la enseñanza de la inferencia bayesiana en el análisis de datos en psicología*. Tesis Doctoral. Universidad de Granada.
- Ghiglione, R. y Matalón, B. (1991). *Les enquêtes sociologiques. Théorie et pratique*. París: Armand Colin.

- Ghiselli, E., Campbell, J. y Zedeck, S (1981). *Measurement theory for the behavioral Sciences*. USA: W.H. Freeman and Company.
- Martínez Arias, R. (1995). *Psicometría: teoría de los tests psicológicos y educativos*. Madrid: Síntesis.
- Olivo, E. (2008). *Significados de los intervalos de confianza para los estudiantes de ingeniería en México*. Tesis doctoral. Universidad de Granada.
- Peña, D. (2002). *Análisis de datos multivariantes*. Madrid: McGraw-Hill.
- Thorndike, R. L. (1989). *Psicometría aplicada*. Mexico: Limusa.

## ANEXO 1

Contenido del cuestionario	Ítem
Definición del intervalo de confianza	1
El ancho del intervalo de confianza disminuye cuando aumenta el tamaño de la muestra	2
El ancho del intervalo de confianza aumenta cuando el nivel de confianza aumenta	3
El ancho del intervalo de confianza aumenta cuando aumenta la varianza	4
Significado del nivel de confianza (variación del intervalo en diferentes muestras)	5
Estimar la media de una población normal o en una muestra grande con $\sigma$ conocida	6
Estimar la media de una población aproximadamente normal cuando $\sigma$ es desconocida	7
Estimar la media de una población a partir de datos experimentales $\sigma$ desconocida, muestra grande	8
Estimar una proporción	9
Estimar una varianza	10
Comparar las medias en dos poblaciones conociendo $\sigma^2_1$ y $\sigma^2_2$ , muestras independientes	11
Comparar las medias en dos poblaciones $\sigma^2_1 = \sigma^2_2$ pero desconocidas, muestras independientes pequeñas	12
Comparar las medias en dos poblaciones $\sigma^2_1 = \sigma^2_2$ , desconocidas, muestras independientes grandes	13
Comparar dos varianzas poblacionales	14
Elegir un modelo de distribución muestral del estadístico	15
Determinar valores críticos en la distribución del estadístico	16
Interpretar intervalos de confianza obtenidos de un programa de ordenador	17
Interpretar gráficos de intervalos de confianza	18

Par de reactivos del cuestionario
Ítem 1. Explica con tus propias palabras la diferencia entre una estimación puntual y un intervalo de confianza
Ítem 2. Comparado a los intervalos de confianza calculados en muestras de tamaño $n=4$ , el ancho de los intervalos de confianza de la media de la población calculado en muestras de tamaño $n = 50$ :
<ol style="list-style-type: none"> <li>Variará más que los anchos de los intervalos para muestras de tamaño <math>n = 4</math>.</li> <li>Variará un poco, pero no tanto como lo hicieron los anchos de los intervalos para muestras de tamaño <math>n=4</math>.</li> <li>Tomarán valores parecidos.</li> </ol>