

# Escalamiento de Thurstone en R

Realizado por: Nahuel Antonio Salem García

Dirigido por: Teresa Rivas Moya

## Introducción

El escalamiento de Thurstone (1927):

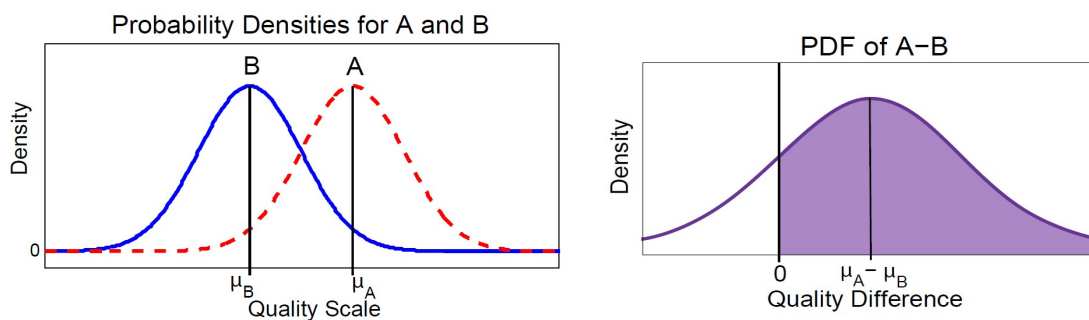
- Ordena estímulos en un continuo psicológico
- Utiliza el método de comparaciones de pares de estímulos

Cada estímulo se compara con todos los demás por pares. En cada par el sujeto indica cuál es superior en un atributo específico, por ejemplo preferencia.

Dados  $n$  estímulos hay  $n(n-1)/2$  comparaciones posibles.

El resultado es una tabla de frecuencias absolutas. El número en cada celda es la cantidad de personas que eligieron el estímulo de la columna como superior al de la fila. Si se divide por el número de personas se obtienen las frecuencias relativas (entre 0 y 1).

Aquí se estudia el **Caso V**, que considera la distancia entre estímulos como la puntuación típica de la distribución normal asociada a la proporción de veces que un estímulo es preferido a otro.



Figuras de Maya y Gupta(2011). La diferencia de calidad percibida entre dos estímulos se distribuye normalmente.

Ejemplo de tabla de frecuencias relativas. Preferencias para distintas verduras (Guilford, 1954).

##	Turn	Cab	Beet	Asp	Car	Spin	S.Beans	Peas	Corn
## Turn	0.500	0.818	0.770	0.811	0.878	0.892	0.899	0.892	0.926
## Cab	0.182	0.500	0.601	0.723	0.743	0.736	0.811	0.845	0.858
## Beet	0.230	0.399	0.500	0.561	0.736	0.676	0.845	0.797	0.818
## Asp	0.189	0.277	0.439	0.500	0.561	0.588	0.676	0.601	0.730
## Car	0.122	0.257	0.264	0.439	0.500	0.493	0.574	0.709	0.764
## Spin	0.108	0.264	0.324	0.412	0.507	0.500	0.628	0.682	0.628
## S.Beans	0.101	0.189	0.155	0.324	0.426	0.372	0.500	0.527	0.642
## Peas	0.108	0.155	0.203	0.399	0.291	0.318	0.473	0.500	0.628
## Corn	0.074	0.142	0.182	0.270	0.236	0.372	0.358	0.372	0.500

El paquete **psych** de R (Revelle, en preparación) contiene la función **thurstone**. Esta función permite hallar valores de escala en base a comparaciones múltiples entre estímulos. Pero:

- No funciona con **datos incompletos**
- No devuelve la **discrepancia media** (sí otra medida de bondad de ajuste)

En este trabajo se explora cómo hallar con R:

- Los **valores de escala** para datos completos e incompletos
- La **discrepancia media**

## Método

Para este trabajo se han utilizado varios conjuntos de datos y funciones de R. Todos los archivos se pueden encontrar en el siguiente enlace:

<https://github.com/NSalem/Thurstone/archive/master.zip>

Para replicar los resultados de los ejemplos es necesario:

- Cargar el archivo `datos_thurstone.RData`
- Cargar y ejecutar el archivo `thurstone.R`. Se puede ejecutar directamente o bien hacer lo siguiente:
  - 1- escribir `getwd()` en la consola. Aparece una ruta (ejemplo: `C:/Users/Usuario/Documentos`).
  - 2- guardar el archivo `thurstone.R` en la ruta que aparece
  - 3- escribir en la consola `source("thurstone.R")`

- Instalar el paquete `psych` si no está instalado, escribiendo en la consola `install.packages("psych")`
- Cargar el paquete `psych`. Escribiendo `library("psych")`.

## Datos

Los datos que se utilizan para los ejemplos son:

- **veg** : Comparaciones de preferencias por pares de 9 verduras (Guilford, 1954). Disponible en el paquete `psych`.
- **veg2** : `veg` con valores perdidos (NA) en las celdas (2,1) y (1,2)
- **veg3** : `veg` con valores de 0.99 y 0.01 en las celdas (2,1) y (1,2) respectivamente.
- **edwards** : Comparaciones por pares de 9 ítems, que son juzgados respecto a cuál es más favorable. (Edwards, 1957; citado en Rivas, 2014). 72 participantes.
- **golfo** : Comparación por pares de 9 ítems acerca del grado en el que se está a favor de la participación de EEUU en la guerra del golfo (Edwards, 1957, citado en Rivas, 2014). 94 participantes.

En los cuatro primeros casos, cada celda de la tabla representa la **frecuencia relativa** o proporción de participantes que eligieron como preferida la opción de la fila a la de la columna. En el último conjunto de datos, se usan las **frecuencias absolutas**.

## Funciones y paquetes

Las funciones de R que se han desarrollado son:

**thurstone\_dm** : devuelve los valores de escala de una matriz de preferencias y la bondad de ajuste medida como discrepancia media. La matriz elegida puede ser de frecuencias relativas o absolutas. Si alguna proporción es superior a 0.99 o inferior a 0.01 da una advertencia.

**thurstone\_incompletos** : devuelve los valores de escala de una matriz de preferencias con datos incompletos. Permite establecer umbrales para considerar valores perdidos las proporciones extremas. Si no se establecen umbrales, se eliminarán automáticamente los valores fuera del intervalo (0.02, 0.98), o fuera del intervalo (0.01, 0.99) si hay más de 200 juicios.

Para utilizar estas funciones (una vez ejecutado el archivo que las contiene) sólo hace falta escribir en la consola el nombre de la función seguido del nombre del conjunto de datos que se quiere usar entre paréntesis. Ejemplo: `thurstone_dm(veg)`.

Aparte de estas funciones propias se ha usado la función **thurstone** incluida en el paquete **psych**. Esta función devuelve los valores de escala y la bondad de ajuste de los ítems de una matriz de preferencia dada. Deben usarse frecuencias relativas.

## Ejemplos

Veamos cómo funciona la función `thurstone` en los datos de preferencia de verduras.

```
library("psych")
thurstone(veg)

## Thurstonian scale (case 5) scale values
## Call: thurstone(x = veg)
##      Turn      Cab      Beet      Asp      Car      Spin S.Beans      Peas
##      0.00      0.52      0.65      0.98      1.12      1.14      1.40      1.44
##      1.63
##
## Goodness of fit of model  0.99
```

Los valores de escala indican que el **maíz** (Corn) sería la verdura más preferida de las 9, mientras que el **nabo** (Turn) sería la menos preferida. La medida de bondad de ajuste (Goodness of fit of model) que utiliza esta función puede estar entre 0 y 1, siendo mejor cuando más cerca esté de 1.

Para hallar la discrepancia media se hace lo mismo usando la función `thurstone_dm` (una vez se haya cargado y ejecutado el archivo `thurstone.R`).

```
thurstone_dm(veg)

## $val_escalas
##      Turn      Cab      Beet      Asp      Car      Spin
##      0.0000000 0.5220458 0.6544394 0.9795439 1.1170831 1.1437223
##      1.4001155
##      Peas      Corn
##      1.4438341 1.6294420
##
## $DM
## [1] 0.03508756
```

Ahora veamos cómo funciona la función `thurstone` en datos incompletos. Para ello, tenemos los datos `veg2`, que son los mismos que los de `veg` sólo que con dos celdas con valores perdidos.

```
thurstone(veg2)
```

```
## Thurstonian scale (case 5) scale values
## Call: thurstone(x = veg2)
##      Turn      Cab      Beet      Asp      Car      Spin S.Beans      Peas
Corn
##      NA       NA       NA       NA       NA       NA       NA       NA
NA
##
## Goodness of fit of model  NA
```

La función no puede hallar los valores de escala.

Para hallar los valores de escala para datos incompletos se ha definido la función `thurstone_incompletos`. Así, al usarla con los datos mostrados antes se consiguen los valores de escala:

```
thurstone_incompletos(veg2)
```

```
##      Turn      Cab      Beet      Asp      Car      Spin S.Beans
## Valores    0 0.411839 0.5818972 0.9070017 1.044541 1.07118 1.327573
##      Peas      Corn
## Valores 1.371292 1.5569
```

La función `thurstone_incompletos` también sirve para datos con valores fuera de un intervalo establecido. Así, se puede escribir `thurstone_incompletos(datos, umbral1, umbral2)`, donde `datos` es el conjunto de datos del que queremos obtener los valores de escala y `umbral1` y `umbral2` son valores entre 0 y 1. Los valores de la tabla fuera del intervalo (`umbral1, umbral2`) serán omitidos.

Ejemplo: Hallar valores de escala omitiendo proporciones menores que 0.02 y mayores que 0.98. Para ello, tenemos los datos `veg3`, que son los mismos que los de `veg` sólo que con dos celdas tienen valores de 0.99 y 0.01 respectivamente.

```
thurstone_incompletos(veg3, 0.02, 0.98)
```

```
##      Turn      Cab      Beet      Asp      Car      Spin S.Beans
## Valores    0 0.411839 0.5818972 0.9070017 1.044541 1.07118 1.327573
##      Peas      Corn
## Valores 1.371292 1.5569
```

Otro ejemplo con datos incompletos (Edwards, 1957; citado en Rivas, 2014). Si no se especifican valores para umbral1 y umbral2, estos se establecen automáticamente como se menciona previamente.

```
thurstone_incompletos(edwards)

##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
##           [,8]
## Valores      0 1.169323 1.200044 1.463857 2.15508 2.53036 2.703987
##           2.753748
##           [,9]
## Valores 3.01531
```

Los valores obtenidos en ambos casos son muy próximos a los de Rivas (2014).

Las funciones creadas funcionan también con las frecuencias absolutas, como se puede ver con los datos de actitudes para la guerra del golfo:

```
thurstone_dm(golfo)

## $val_escalas
##           i1           i2           i3           i4           i5           i6
##           i7
## 0.0000000 0.6024647 0.8387961 0.9044108 1.0580879 1.2428534
##           1.4574784
##
## $DM
## [1] 0.0319754
```

## Conclusiones

Las funciones diseñadas aportan:

- La posibilidad de hallar discrepancia media
- Valores de escala para datos incompletos, tanto si son introducidos como NA como si se trata de proporciones extremas.
- Posibilidad de usar como input tanto las frecuencias relativas como absolutas

Algo útil de cara al futuro sería definir una medida de bondad de ajuste para datos incompletos, ya que no parece que exista.

## Referencias

Guilford, J. (1954). *Psychometric Methods*. McGraw-Hill, New York.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Revelle, W. (in preparation) An Introduction to Psychometric Theory with applications in R. Springer. at <http://personality-project.org/r/book/>

Rivas, T. Escalamiento Centrado en el Estímulo: Thurstone – Teoría. <http://psicologia.cv.uma.es/mod/resource/view.php?id=87354> (Técnicas de Investigación en Psicología Clínica y de la Salud. Campus Virtual UMA. 14-01-2014)

Thurstone, L. (1927). A law of comparative judgments. *Psychological Review*, 34, 273-286.

Tsukida, K. y Gupta, M.R. (2011). How to Analyze Paired Comparison Data. *UWEE Technical Report*.

<http://mayagupta.org/publications/PairedComparisonTutorialTsukidaGupta.pdf>