

Kruskal-Wallis test. Alternativa no paramétrica al ANOVA independiente

Joaquín Amat Rodrigo j.amatrodrigo@gmail.com

Enero, 2016

Índice

Introducción.....	2
Condiciones	3
Comparaciones <i>Post-Hoc</i>	3
Tamaño del efecto.....	4
Ejemplo	4

Introducción

El test de Kruskal-Wallis, también conocido como test H, es la alternativa no paramétrica al test ANOVA de una vía para datos no pareados. Se trata de una extensión del test de Mann-Whitney para más de dos grupos. Se trata por lo tanto de un test que emplea rangos para contrastar la hipótesis de que k muestras han sido obtenidas de una misma población.

A diferencia del ANOVA en el que se comparan medias, el test de Kruskal-Wallis contrasta si las diferentes muestras están equidistribuidas y que por lo tanto pertenecen a una misma distribución (población). Bajo ciertas simplificaciones puede considerarse que el test de Kruskal-Wallis compara las medianas.

- H_0 : todas las muestras provienen de la misma población (distribución).
- H_A : Al menos una muestra proviene de una población con una distribución distinta.

El test de Kruskal-Wallis es el test adecuado cuando los datos tienen un orden natural, es decir, cuando para darles sentido tienen que estar ordenados o bien cuando no se satisfacen las condiciones para poder aplicar un ANOVA. Por ejemplo, si se quiere estudiar la diferencia entre hombres y mujeres en una carrera, se puede disponer de dos tipos de datos: los tiempos de cada participante (análisis con ANOVA) o las posiciones en las que ha terminado la carrera cada participante (análisis con Kruskal-Wallis test).

Supóngase que se dispone de k grupos cada uno con n observaciones. Si se ordenan todas las observaciones de menor a mayor y se le asigna a cada una de ellas su rango, cuando se obtenga la suma de rangos para cada uno de los grupos (R_i) es de esperar que, si se cumple la hipótesis nula, todos los grupos tengan un valor similar. Partiendo de esta idea se calcula el estadístico H como:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

Condiciones

- No es necesario que las muestras que se comparan provengan de una distribución normal.
- Homocedasticidad: dado que la hipótesis nula asume que todos los grupos pertenecen a una misma población y que por lo tanto tienen las mismas medianas, es requisito necesario que todos los grupos tengan la misma varianza. Se puede comprobar con representaciones gráficas o con los test de *Levenne* o *Barttlet*.
- Misma distribución para todos los grupos: la distribución de los grupos no tiene que ser normal pero ha de ser igual en todos (por ejemplo que todos muestren asimetría hacia la derecha).

Si estos requerimientos se cumplen, el estadístico H del test de Kruskal-Wallis se compara con:

- Si el tamaño de grupos k es igual a 3 y el número de observaciones en cada uno no es mayor que 5, se recurre a tablas tabuladas con valores teóricos de H .
- En el resto de casos se asume que el estadístico H sigue una distribución χ^2 con $k-1$ grados de libertad (siendo k el número de grupos a comparar).

En la mayoría de la bibliografía consultada consideran que el ANOVA es bastante robusto a la falta de normalidad sobre todo con muestras medianas o grandes. Solo recomiendan el uso del test de Kruskal-Wallis cuando las poblaciones a comparar sean claramente asimétricas, se cumpla que todas lo sean en la misma dirección y que la varianza sea homogénea. Si la varianza no es homogénea el test adecuado es un ANOVA con corrección de Welch. En los casos en los que se trata de datos puramente ordinales el ANOVA no es una opción.

Comparaciones *Post-Hoc*

Al igual que ocurre con un ANOVA, si el test de Kruskal-Wallis es significativo, implica que al menos dos grupos de entre los comparados son significativamente diferentes, pero no indica cuales. Para saberlo es necesario compararlos todos entre ellos. Esto implica realizar una corrección del nivel de significancia para evitar incrementar el error de tipo I. Los dos métodos de comparación post-hoc más empleados para un test de Kruskal-Wallis son:

- Test de Mann-Whitney entre cada par de grupos con corrección de significancia `pairwise.wilcox.test()`.
- Tukey's range test: en R existe la función `kruskalmc()` del paquete *pgirmess*.

Tamaño del efecto

No hay una manera general de calcular el tamaño del efecto de una prueba Kruskal-Wallis. Lo que se suele hacer es calcular el tamaño del efecto sobre las parejas que hayan resultado significativamente distintas en el análisis *post-hoc*. Si las comparaciones *post-hoc* se hacen mediante Mann-Whitney, se tendrá que calcular el efecto asociado a dicho test.

Ejemplo

Un estudio compara el número de huevos que pone un determinado insecto bajo 3 condiciones distintas. ¿Existen diferencias significativas dependiendo de las condiciones?

```
datos <- data.frame(condicion = c(rep("condicion1", 18), rep("condicion2", 18),
  rep("condicion3", 18)), n_huevos = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 16, 27,
  28, 29, 30, 51, 52, 53, 342, 40, 41, 42, 43, 44, 45, 46, 47, 48, 67, 88,
  89, 90, 91, 92, 93, 94, 293, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 25,
  36, 37, 58, 59, 60, 71, 72))
head(datos)
```

```
##   condicion n_huevos
## 1 condicion1      1
## 2 condicion1      2
## 3 condicion1      3
## 4 condicion1      4
## 5 condicion1      5
## 6 condicion1      6
```

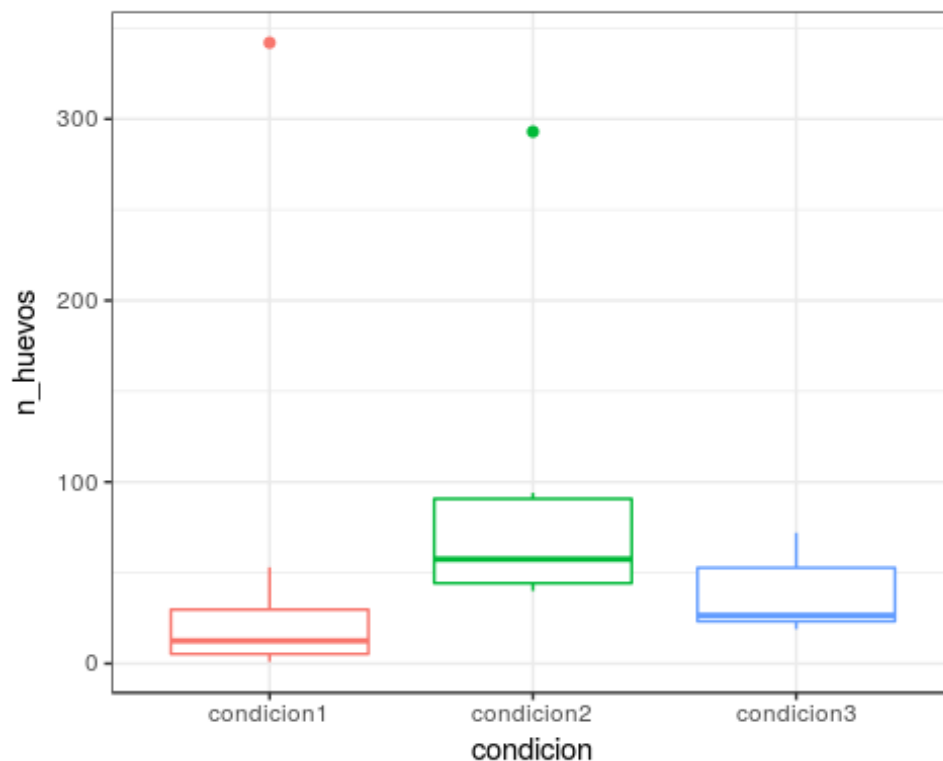
```
aggregate(n_huevos ~ condicion, data = datos, FUN = median)
```

```
##   condicion n_huevos
## 1 condicion1    12.5
## 2 condicion2    57.5
## 3 condicion3    26.5
```

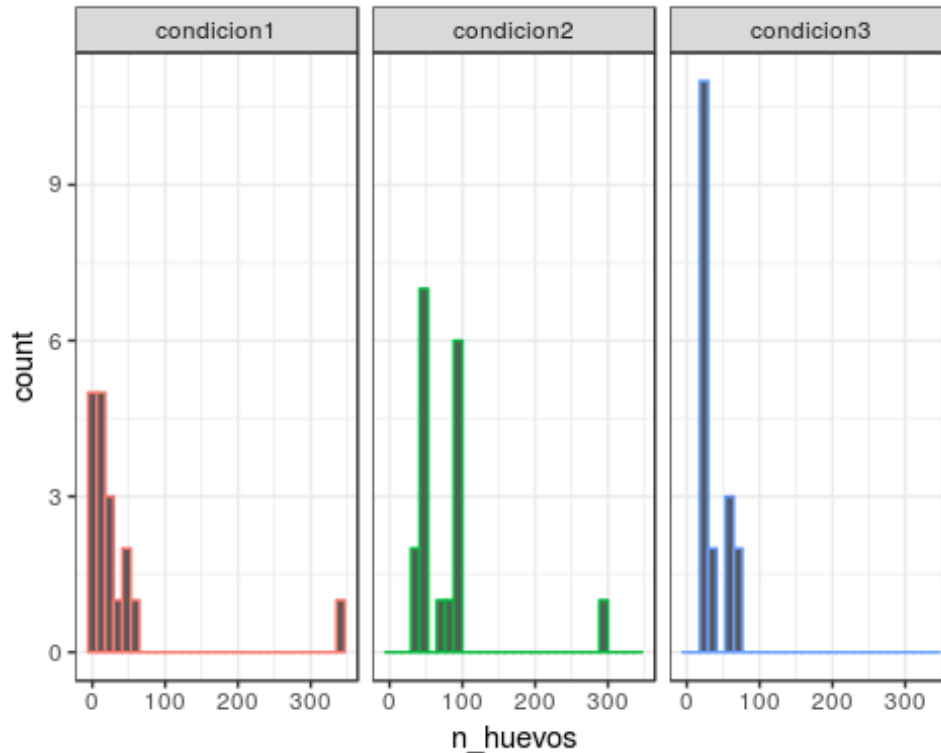
```
aggregate(n_huevos ~ condicion, data = datos, FUN = sd)
```

```
##      condicion n_huevos  
## 1 condicion1 78.10637  
## 2 condicion2 58.41750  
## 3 condicion3 18.59097
```

```
require(ggplot2)  
ggplot(data = datos, mapping = aes(x = condicion, y = n_huevos, colour =  
  condicion)) +  
  geom_boxplot() + theme_bw() + theme(legend.position = "none")
```



```
ggplot(data = datos, mapping = aes(x = n_huevos, colour = condicion)) +  
  geom_histogram() +  
  theme_bw() + facet_grid(. ~ condicion) + theme(legend.position = "none")
```



La representación gráfica de los datos muestra que las muestras no se distribuyen de forma normal, lo que supone una limitación para emplear un test ANOVA. Las tres muestras presentan el mismo tipo de distribución, asimetría hacia la derecha, a falta de comprobar la homogeneidad de varianza el test de Kruskal-Wallis es la opción más adecuada (otra alternativa serían las técnicas de *resampling*).

Condiciones

Homocedasticidad: la varianza debe de ser constante entre todos los grupos.

```
library(car)
leveneTest(n_huevos ~ condicion, data = datos, center = "median")
```

```
## Levene's Test for Homogeneity of Variance (center = "median")
##      Df F value Pr(>F)
## group 2  0.7929  0.458
##      51
```

No hay evidencias en contra de la homogeneidad de varianzas.

Test de Kruskal-Wallis

```
kruskal.test(n_huevos ~ condicion, data = datos)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  n_huevos by condicion
## Kruskal-Wallis chi-squared = 19.964, df = 2, p-value = 4.623e-05
```

El test encuentra significancia en la diferencia de al menos dos grupos.

Comparaciones *post-hoc* para saber que dos grupos difieren

Existen diferentes métodos de corrección del nivel de significancia, entre ellos destacan el de *Bonferroni* que es muy estricto y el de *holm*, este último parece ser más recomendado.

```
pairwise.wilcox.test(x = datos$n_huevos, g = datos$condicion, p.adjust.method =
"holm")
```

```
##
##  Pairwise comparisons using Wilcoxon rank sum test
##
## data:  datos$n_huevos and datos$condicion
##
##           condicion1 condicion2
## condicion2 0.00029      -
## condicion3 0.04795      0.00058
##
## P value adjustment method: holm
```

Las comparaciones por pares encuentran diferencias significativas entre todas las condiciones.

Cálculo de tamaño del efecto

Finalmente, para aquellas comparaciones que se consideren significativas se tiene que reportar el tamaño del efecto (no existe forma de calcular el tamaño del efecto para el test de Kruskal-Wallis en conjunto). En este caso se han realizado comparaciones con Mann-Whitney-wilcox (ver en capítulo dedicado a este test para saber cómo se calcula).