

Prueba de Kruskal Wallis

Fernanda Sarmiento

```
library(car)
```

```
Cargando paquete requerido: carData
```

```
library(kableExtra)
```

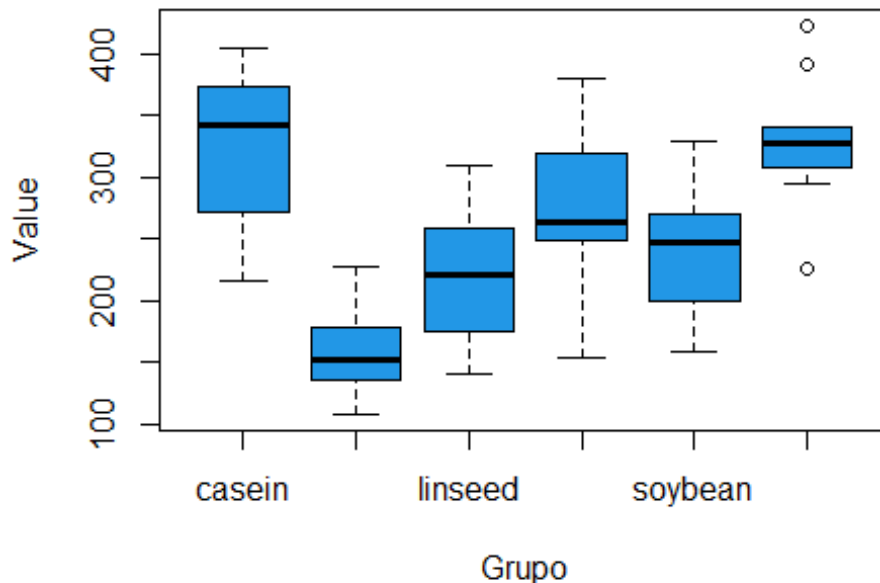
Introducción

Esta prueba es una prueba no paramétrica basada en el banco que puede utilizarse para corroborar si existen diferencias relevantes a nivel estadístico entre dos o más grupos de variables dependientes ordinales o continuas.

Para esta aplicación se va a emplear la base de datos chickwts, la cual se encuentra en la base de datos de R.

```
df <- chickwts
```

```
boxplot(df$weight ~ df$feed, col = 4, xlab="Grupo", ylab = "Value")
```



```
tapply(df$weight, df$feed, summary)
```

```
$casein
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
```

	216.0	277.2	342.0	323.6	370.8	404.0
\$horsebean	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	108.0	137.0	151.5	160.2	176.2	227.0
\$linseed	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	141.0	178.0	221.0	218.8	257.8	309.0
\$meatmeal	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	153.0	249.5	263.0	276.9	320.0	380.0
\$soybean	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	158.0	206.8	248.0	246.4	270.0	329.0
\$sunflower	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	226.0	312.8	328.0	328.9	340.2	423.0

Por una parte, podemos evidenciar las diferencias de grupos que se presentan.

Además, con los oxplots generados podemos evidenciar los cambios entre las representaciones de las gráficas sin contar que uno de los gráficos presenta datos atípicos en comparación de los demás.

En el caso de la fórmula, `tapply` nos permite comparar los valores más importantes de cada grupo, siendo estos: **Valor mínimo, Primer cuartil, Mediana, Media, Tercer cuartil y Valor máximo**. Esto nos permite poder comparar los valores de las diferentes variables que presentemos.

Como siguiente paso se debe controlar la igualdad de varianzas entre los grupos, para esto se pueden emplear dos funciones. `Bartlett` o `Levene`; para poder emplear estas funciones se debe tener activada la librería `car`.

```
leveneTest(df$weight~df$feed)
```

```
Levene's Test for Homogeneity of Variance (center = median)
  Df F value Pr(>F)
group 5  0.7493 0.5896
    65
```

```
bartlett.test(df$weight~df$feed)
```

```
Bartlett test of homogeneity of variances
```

```
data: df$weight by df$feed
Bartlett's K-squared = 3.2597, df = 5, p-value = 0.66
```

Al dar un p-valor mayor a lo habitual, en este caso no se rechaza la hipótesis nula de homogeneidad de varianzas.

Obtenido este resultado, podemos continuar aplicando así la prueba `kruskal.test` como se muestra a continuación.

```
kruskal.test(df$weight, df$feed)
```

```
Kruskal-Wallis rank sum test
```

```
data: df$weight and df$feed
Kruskal-Wallis chi-squared = 37.343, df = 5, p-value = 5.113e-07
```

Los resultados obtenidos a través de la función `kruskal.test` nos indican que las medias no son iguales, pero no nos indica cuál de las medias es la que presenta estas diferencias. Para poder establecer cuál es la variable que nos genera este problema, podemos emplear la función `pairwise.wilcox.test`. Esta realiza la prueba de Wilcoxon para cada grupo.

```
pairwise.wilcox.test(x= df$weight, g= df$feed)
```

```
Pairwise comparisons using Wilcoxon rank sum exact test
```

```
data: df$weight and df$feed
```

	casein	horsebean	linseed	meatmeal	soybean
horsebean	0.00030	-	-	-	-
linseed	0.01221	0.05001	-	-	-
meatmeal	0.36337	0.00306	0.21806	-	-
soybean	0.04736	0.00834	0.70996	0.70996	-
sunflower	1.00000	9.3e-05	0.00064	0.34409	0.01402

```
P value adjustment method: holm
```

Esta función nos va a devolver una matriz de los p-valores de la comparación de las variables. Si el p-valor de cada pareja es menor al nivel de significancia con el que se esté trabajando, se puede decir que las dos medias no presentan igualdad de medias.

En el caso de nuestro ejemplo, muestra que la mayor parte de las medias no están relacionadas.