Unidad IV

PRUEBAS ESTADÍSTICAS PARA K MUESTRAS INDEPENDIENTES

“Es mejor tener una respuesta aproximada a la pregunta correcta que una respuesta exacta a la pregunta equivocada”

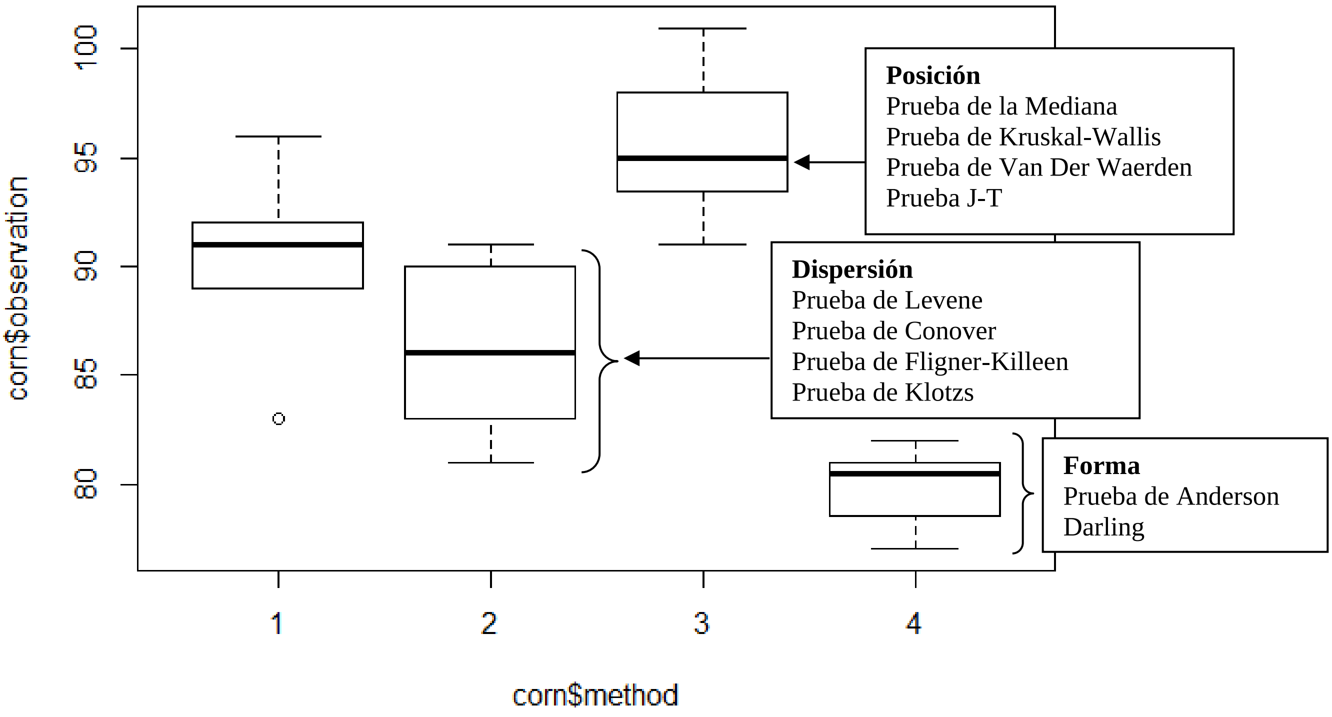
John Tukey

Introducción
Uno de los diseños experimentales más utilizados para comparar k tratamientos o niveles de un factor es el Diseño Completamente al Azar (DCA).
El DCA parte del hecho de que la fuente de variabilidad se debe a los tratamientos y al error experimental. Por tal motivo, para aplicar este diseño se deben utilizar unidades experimentales homogéneas.
Para el análisis de datos que provienen de este tipo de diseño se usan técnicas paramétricas derivadas de la descomposición de la variabilidad total, es decir, se debe realizar el Análisis de Varianza con el fin de utilizar la prueba estadística F.
La prueba F permite solo responder la pregunta si la media de al menos uno de los tratamientos presenta diferencias significativas con respecto a los demás. Para determinar que tratamiento(s) es(son) (el) los mejor(es) se deben recurrir a pruebas de comparación por pares como: DLS, Tukey, Contrastes Ortogonales, etc.
Los resultados de la prueba F y de las pruebas de comparación tendrían validez solo si se cumplen los supuestos que estas exigen como: normalidad de errores, homogeneidad de varianzas e independencia de errores.
Muchas veces, estos supuestos no se cumplen por lo que se debe recurrir a métodos alternativos (procedimientos no paramétricos) o transformación de las variables. Las transformaciones poco a poco se han ido dejando de lado debido a que las unidades de medida (cm, kg, etc.) de la variable se ven afectadas por la transformación y esto lleva a que las conclusiones se brinden en base a las unidades de medida transformadas.
Por otro lado, un error que se cometía muy a menudo era realizar el Análisis de Varianza para variables medidas en una escala ordinal o variables de tipo cuantitativa discreta esto debido a que se “asumía” erróneamente el supuesto de normalidad, a pesar de la existencia de procedimientos que permiten analizar variables de este tipo.
Para poder llevar a cabo estas pruebas se requiere de k muestras aleatorias e independientes. Si no se realizar un diseño experimental estas muestras usualmente son obtenidas siguiendo un esquema de muestreo aleatorio con k estratos.

El análisis estadístico se ha simplificado con los programas estadísticos, los cuales deben ayudarnos en la verificación de los supuestos, por lo tanto, debemos eliminar la idea de simplemente asumirlos.

El presente capítulo tiene como objetivo desarrollar algunas pruebas no paramétricas que permitan comparar k tratamientos sin ninguna restricción en la aleatorización.
También se desarrollarán las respectivas pruebas no paramétricas de comparación.

El siguiente gráfico puede ayudar a la decisión de la prueba no paramétrica adecuada de acuerdo al objetivo de la investigación



1. **Para un parámetro de Locación (Posición)**
   Para probar la hipótesis nula de que k muestras son extraídas de poblaciones con medias iguales, se utiliza la estadística F cuando se satisfacen las condiciones que esta prueba exige. Si una de las condiciones no se cumple se puede usar alguna de las pruebas que serán discutidas en esta sección.

Prueba de Hipótesis

$H_0 : Me_1 = Me_2 = \cdots = Me_k$

$H_1 : Al\ menos\ una\ Me_i\ es\ diferente\ a\ las\ demás\ i = 1,2....,k$

Si se rechaza H_0, se deben realizar las pruebas de comparaciones múltiples
Las hipótesis son:

$H_0 : Me_i = Me_j$ i,j=1,2,…, k $\forall i \neq j$
$H_1 : Me_i \neq Me_j$

A. Prueba de la Mediana

- **Aspectos Generales**

  Con frecuencia, se tiene interés en hacer inferencia basada en k muestras que no están relacionadas, es decir, que son independientes.
  Una alternativa no paramétrica la proporciona la prueba de la mediana, que puede utilizarse para probar la hipótesis nula de que k muestras independientes se han extraído de poblaciones con medianas iguales contra la hipótesis alterna que al menos una mediana es diferente a las demás.
  Esta prueba también es conocida como la prueba de la Mediana de Mood.

- **Supuesto**

  - La variable de interés debe estar medida en al menos una escala intervalo es decir la variable debe ser cuantitativa.

- **Inferencia Estadística**

  Prueba de Hipótesis
  Procedimiento para el Desarrollo de la Prueba
  - Calcular la mediana común de las k muestras (o grupos) es decir de todo el conjunto de observaciones sin importar a que grupo pertenecen.
  - Para cada grupo se determina en una columna el número de observaciones que son menores o iguales a mediana común y en otra columna aquellas observaciones que son mayores a ella.
  - Las frecuencias resultantes se disponen en una tabla k *x* 2.
  - Aplicar la prueba χ^2 para una tabla de contingencia.

$$\chi_c^2 = \sum_{i=1}^{2}\sum_{j=1}^{k}\frac{\left(o_{ij}-e_{ij}\right)^2}{e_{ij}} \sim \chi^2_{[1-\alpha,k-1]}$$

  - Comparar el valor calculado con el valor tabulado $\chi^2_{[1-\alpha,k-1]}$
  - Utilizar el siguiente criterio

  Si $\chi_c^2 \leq \chi^2_{[,1-\alpha,k-1]}$ no se rechaza la H_0.
  Si $\chi_c^2 > \chi^2_{[,1-\alpha,k-1]}$ se rechaza la H_0.

  Si, en efecto las muestras provienen de poblaciones con la misma mediana, es de esperar que aproximadamente la mitad de las calificaciones en cada muestra se encuentren por arriba de la mediana combinada y alrededor de la otra mitad por debajo.

  Para realizar las pruebas de comparación, se aplica la prueba antes discutida en cada una de las C_2^k posibles comparaciones.

- **Aplicación**

  Ejemplo
  A un grupo de consumidores compuesto por 15 individuos se le pide calificar una marca de refresco de cola que se le fue asignado aleatoriamente. Los resultados son presentados en la tabla siguiente:

Marca 1	Marca 2	Marca 3
20	16	23
24	26	20
28	18	18
24	17	21
20	20	17

Pruebe la hipótesis de que existe diferencia en el nivel mediano de calificaciones de las tres marcas de refresco de cola al nivel de significación de 5%.

Solución

$H_0 : Me_1 = Me_2 = Me_3$

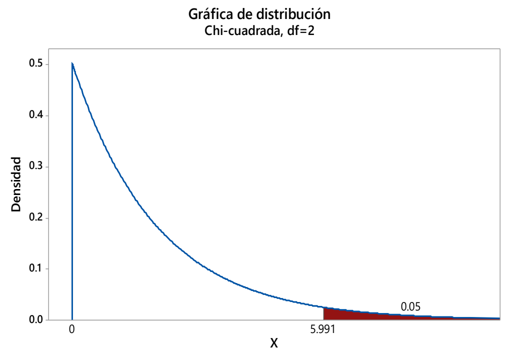
$H_1 : Al\ menos\ una\ Me_i\ es\ diferente\ a\ las\ demás\ i = 1,2,3$

α=0.05

Cálculo del Estadístico

$\chi_c^2 = 1.67$

Criterio de Decisión

<table>
<tr>
<td>
</td>
<td>Como $\chi^2_{(0.95,2)} = 5.99 > 1.67$ no se rechaza H0</td>
</tr>
</table>

Conclusión

A un nivel de significación de 0.05 existe suficiente evidencia estadística para no rechazar la hipótesis nula.

Por lo tanto, no podemos afirmar que al menos una de las marcas de refresco de cola tiene diferente preferencia del consumidor.

- **Funciones en R**

En R existe la función mood.medtest del paquete RVAideMemoire

```
mood.medtest(vector de datos, grupos,exact=FALSE)
```

También, se puede construir la tabla de contingencia y posteriormente aplicar la prueba Chi Cuadrado.

Adicionalmente existe la función median_test del paquete coin, sin embargo, esta muestra una variante de la prueba de la mediana que es la de Brown-Mood.

También se puede utilizar la función Median.test del paquete agricolae que también realiza las pruebas de comparación

```
Median.test(vector de datos, grupos)
```

La ventaja de esta última función es que brinda las pruebas de comparación por pares de grupos.

- **Resultados con R**

```
x1<-c(20,24,28,24,20)
x2<-c(16,26,18,17,20)
x3<-c(23,20,18,21,17)
datos<-c(x1,x2,x3)
grupos<-as.factor(rep(1:3,c(5,5,5)))
mood.medtest(datos,grupos,exact=FALSE)
```

```
        Mood's median test
data:  datos by grupos
X-squared = 1.6667, df = 2, p-value = 0.4346
```

```
Median.test(datos,grupos)
```

```
The Median Test for datos ~ grupos

Chi Square = 1.666667   DF = 2   P.Value 0.4345982
Median = 20

  Median r Min Max Q25 Q75
1     24 5  20  28  20  24
2     18 5  16  26  17  20
3     20 5  17  23  18  21

Post Hoc Analysis

Groups according to probability of treatment differences
and alpha level.

Treatments with the same letter are not significantly
different.

  datos groups
1    24      a
3    20      a
2    18      a
```

B. Prueba de Kruskal Wallis

- **Aspectos Generales**

  Esta prueba es una alternativa no paramétrica del Diseño Completamente al Azar (DCA) de un factor de análisis de varianza.
  Esta prueba puede ser utilizada cuando no se cumplen los supuestos que exige el DCA. Es decir, cuando no se cumple que las poblaciones de las cuales se extraen las muestras están distribuidas normalmente con varianzas iguales. También se debe utilizar cuando la variable de interés es de tipo cualitativa ordinal o cuantitativa discreta.

- **Supuestos**

  - Las muestras a ser evaluadas de cada uno de los k grupos son aleatorias y mutuamente excluyentes.
  - La variable respuesta esta medida en una escala al menos intervalo.
  - Las poblaciones tienen la misma distribución, difiriendo solo en su localización.

- **Inferencia Estadística**
  Prueba de Hipótesis
  Procedimiento para el Desarrollo de la Prueba

  - Las n_1, n_2,…n_k observaciones de los k grupos se combinan en una sola serie de tamaño n y se disponen en orden de magnitud desde la más pequeña hasta la más grande. Cuando dos o más observaciones tienen el mismo valor, a cada una de ellas se le asigna la media de los rangos con los cuales está relacionado.
  - Los rangos asignados a las observaciones en cada uno de los k grupos se suman por separado para dar k sumas de rangos.

$$R_i = \sum_{j=1}^{n_i} R\left(X_{ij}\right)$$

  - Aplicar la siguiente prueba estadística:

$$H = \frac{1}{S^2}\left[\sum_{i=1}^{k}\frac{R_i^2}{n_i} - \frac{n\left(n+1\right)^2}{4}\right] \sim \chi^2_{(1-\alpha,\text{k-1})}$$

Donde:

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} R^2\left(X_{ij}\right) - \frac{n\left(n+1\right)^2}{4}\right]$$

Si no hay empates S^2 se simplifica a $\frac{n\left(n+1\right)^2}{4}$, entonces H se simplifica a:

$$H = \frac{12}{n(n+1)}\sum_{i=1}^{k}\frac{1}{n_i}\left[R_i - \frac{n_i\left(n+1\right)}{2}\right]^2 \sim \chi^2_{(1-\alpha,\text{k-1})}$$

donde:
n: Tamaño total de la muestra
R_j: Suma de los rangos de la j-ésima muestra o grupo de tratamiento.
n_j: Número de observaciones de la j-ésima muestra.
k: Número de tratamientos o grupos.

Comparaciones Múltiples
Si la hipótesis nula es rechazada, se puede usar el siguiente procedimiento para determinar cuál de los pares de tratamientos tienden a ser diferentes.

Una primera forma de realizar las pruebas de comparaciones es hacer la prueba de Mann-Whitney sobre todas las parejas. Sin embargo, esto eleva el error de tipo I bastante rápido, por eso algunos autores recomiendan seleccionar primero los pares de grupos que más nos interese comparar y de esta forma hacer menos comparaciones, posteriormente se puede hacer alguna corrección de la significación mediante Bonferroni u otra.
Estas pruebas de comparación se pueden hacer utilizando la función pairwise.

Otra forma es comparar cada pareja con Mann-Whitney, pero esta vez utilizando una desigualdad que se obtiene aplicando el test de Tukey (Tukey's range test en inglés). La fórmula propuesta por Nemenyi es:

$$\left|\frac{R_i}{n_i}-\frac{R_j}{n_j}\right| > z_{(\alpha/k(k-1))}\sqrt{\frac{n(n+1)}{12}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

Esta es la alternativa que aparece explicada en A. Field, Miles, & Field (2012) y Gibbons & Chakraborti (2003). Para realizar esta prueba se puede utilizar la función kruskalmc() del paquete pgirmess.

Una variante de la propuesta anterior para analizar si existe diferencia entre los tratamientos i y j a un nivel de significación α es dada por:

$$\left|\frac{R_i}{n_i}-\frac{R_j}{n_j}\right| \text{ con } ALS(K-W) = t_{\left(1-\frac{\alpha}{2},n-k\right)}\sqrt{\left[\frac{S^2(n-1-H)}{n-k}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)\right]}$$

Así si, $\left|\frac{R_i}{n_i}-\frac{R_j}{n_j}\right| > t_{\left(1-\frac{\alpha}{2},n-k\right)}\sqrt{\left[\frac{S^2(n-1-H)}{n-k}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)\right]}$

Existe diferencia entre los tratamientos i y j a un nivel de significación α

- **Aplicación**

  Quince alumnos son aleatoriamente asignados a tres tipos diferentes de métodos de instrucción, todos los cuales persiguen el desarrollo de un nivel específico de habilidad en diseño asistido por computadora. Para analizar la efectividad de los programas se realizó una prueba consistente en comparar la calificación obtenida al finalizar la capacitación. Los resultados se presentan a continuación:

Método A_1	**Método A_2**	**Método A_3**
86	90	82
79	76	68
81	88	63
70	82	71
84	89	61

Pruebe si al menos uno de los métodos produce un número de diseños desarrollado distinto. Use un nivel de significación de 0.05.

Los rangos correspondientes a los datos son:

Método A_1	**Método A_2**	**Método A_3**
12.0	15.0	9.5
7.0	6.0	3.0
8.0	13.0	2.0
4.0	9.5	5.0
11.0	14.0	1.0
R_1=42	R_2=57.5	R_3=20.5

$H_0: Me_1 = Me_2 = Me_3$

$H_1: \text{Al menos una } Me_i \text{ es diferente a las demás } i = 1,2,3$

α=0.05

Prueba Estadística

$$H = \frac{1}{S^2}\left[\sum_{i=1}^{k}\frac{R_i^2}{n_i} - \frac{n(n+1)^2}{4}\right] \sim \chi^2_{(1-\alpha,\text{k-1})}$$

Donde:

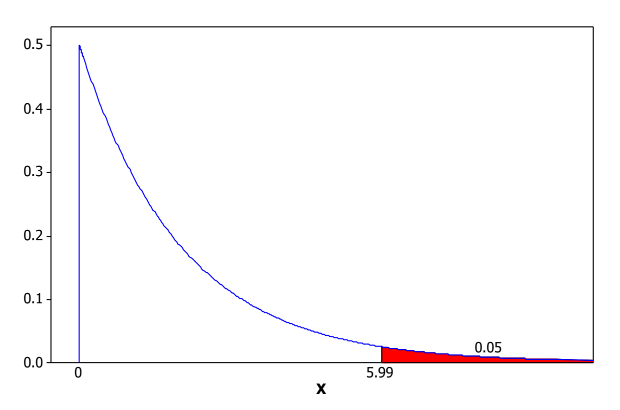
$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k}\sum_{j=1}^{n_i}R^2\left(X_{ij}\right) - \frac{n(n+1)^2}{4}\right]$$

Desarrollo de la Prueba

$$S^2 = \frac{1}{15-1}\left[\left(12^2 + \ldots + 1^2\right) - \frac{15(15+1)^2}{4}\right] = 1239.5 - 960 = 19.9643$$

$$H = \frac{1}{19.9643}\left[\frac{\left(42^2 + 57.5^2 + 20.5^2\right)}{5} - \frac{15(15+1)^2}{4}\right] = 6.917$$

Criterio de Decisión



Como $\chi^2_{(0.95,2)} = 5.99 < 6.917$ se rechaza H_0

Conclusión
A un nivel de significación de 0.05, existe suficiente evidencia estadística para afirmar que la calificación mediana con los métodos de instrucción en estudio no son las mismas.

Como la calificación no es la misma con al menos uno de los métodos de instrucción se debe proceder a realizar las pruebas de comparación.

$H_0 : Me_i = Me_j$ i,j=1,2,3 $\forall i \neq j$
$H_1 : Me_i \neq Me_j$

α=0.05

$$ALS(K-W) = t_{\left(1-\frac{\alpha}{2},n-k\right)} \sqrt{\left[\frac{S^2\left(n-1-H\right)}{n-k}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)\right]}$$

$$ALS(K-W) = t_{(0.975,15-3)} \sqrt{\left[\frac{19.9643(15-1-6.917)}{15-3}\left(\frac{1}{5}+\frac{1}{5}\right)\right]} = 2.18(2.17) = 4.7306$$

Comparaciones	$\left\lvert \frac{R_i}{n_i} - \frac{R_j}{n_j} \right\rvert$	**ALS(K-W)**	**Sig**
A_1 vs A_2	$\left\lvert \frac{42}{5} - \frac{57.5}{5} \right\rvert = 3.3$	4.7306	n.s.
A_1 vs A_3	$\left\lvert \frac{42}{5} - \frac{20.5}{5} \right\rvert = 3.9$	4.7306	n.s
A_2 vs A_3	$\left\lvert \frac{57.5}{5} - \frac{20.5}{5} \right\rvert = 7.2$	4.7306	*

Conclusión
A un nivel de significación de 0.05 se puede afirmar que con los métodos de instrucción A_1 y A_2 se obtiene la mayor calificación mediana.

Resumen

A_3 $\underline{A_1 \quad A_2}$

- **Funciones en R**

Se cuenta con la función kruskal.test, el problema de esta función es que no realiza la prueba de comparaciones múltiples por pares.

```
kruskal.test(vector de datos, grupos)
```

También se puede hacer uso de la función kruskal del paquete agricolae que si brinda las pruebas de comparaciones múltiples por pares. Esta función ofrece diferentes pruebas de comparación como: t, Holm, Hochberg, Bonferroni, BH, BY, fdr.

```
kruskal(datos, grupos, group=FALSE, p.adj=””)
```

El argumento group=FALSE es para que brinde los pvalores en las pruebas de comparación. El argumento p.adj es para elegir la prueba de comparación.
Esta función requiere que los resultados sean almacenados en un objeto.

La función kruskalmc del paquete pgirmess también realiza las comparaciones múltiples.

```
kruskalmc(datos, grupos)
```

La función pairw.kw del paquete asbio permite realiza las comparaciones múltiples.

```
pairw.kw (datos, grupos)
```

La función pairwise.wilcox.test permite realizar las pruebas de comparación aplicando la corrección para comparaciones múltiples

```
pairwise.wilcox.test( datos, grupos, p.adjust = "bonferroni")
```

- **Resultados con R**

```
M1<-c(86,79,81,70,84)
M2<-c(90,76,88,82,89)
M3<-c(82,68,63,71,61)
datos<-c(M1,M2,M3)
grupos<-rep(1:3,c(5,5,5))

kruskal.test(datos,grupos)
```

```
        Kruskal-Wallis rank sum test
data:  datos and grupos
Kruskal-Wallis chi-squared = 6.9174, df = 2, p-value =
0.03147
```

```
library(agricolae)
kruskal(datos,grupos, group=FALSE,console=TRUE)
```

```
$statistics
     Chisq   p.chisq
  6.917352 0.0314714
$parameters
  Df ntr  t.value
   2   3 2.178813
$means
   y      std r Min Max
```

```
1 80 6.204837 5  70  86
2 85 5.916080 5  76  90
3 69 8.276473 5  61  82
$rankMeans
  x    y r
1 1  8.4 5
2 2 11.5 5
3 3  4.1 5

$comparison
      Difference   pvalue sig.        LCL       UCL
1 - 2       -3.1 0.178826      -7.830249  1.630249
1 - 3        4.3 0.071024    . -0.430249  9.030249
2 - 3        7.4 0.005188   **  2.669751 12.130249
```

```
library(pgirmess)
kruskalmc(datos, grupos)
```

```
Multiple comparison test after Kruskal-Wallis
p.value: 0.05
Comparisons
    obs.dif critical.dif difference
1-2     3.1     6.771197      FALSE
1-3     4.3     6.771197      FALSE
2-3     7.4     6.771197       TRUE
```

```
library(asbio)
pairw.kw(eje1[,1],as.factor(eje1[,2]))
```

```
95% Confidence intervals for Kruskal-Wallis comparisons

                     Diff    Lower    Upper  Decision Adj. P-value
Avg.rank1-Avg.rank2 -3.1 -9.86515  3.66515    FTR H0      0.81793
Avg.rank1-Avg.rank3  4.3 -2.46515 11.06515    FTR H0     0.384299
Avg.rank2-Avg.rank3  7.4  0.63485 14.16515 Reject H0     0.026485
```

```
pairwise.wilcox.test(datos,grupos,p.adjust="bonferroni")
```

```
     Pairwise comparisons using Wilcoxon rank sum test
data:  datos and grupos
  1     2
2 0.667 -
3 0.286 0.083
```

- **Algunas consideraciones en R**

- Las funciones kruskal del paquete agricolae y kruskalmc del paquete pgirmess son mejores que la función kruskal.test debido a que estas brindan las pruebas de comparación por pares de tratamientos.
- La función kruskal del paquete agricolae ofrece algunos estadísticos (como promedios) a pesar que la variable sea de tipo ordinal.

C. Prueba de Van Der Waerden

- **Aspectos Generales**

La prueba no paramétrica más común para el modelo de un factor es la prueba de Kruskal-Wallis, la cual se basa en los rangos de los datos. Sin embargo, también existe la prueba de Van Der Waerden que convierte los rangos en cuantiles de la distribución normal estándar a los que llama “valores normales” y la prueba es realizada sobre dichos valores.
La ventaja de la prueba de Van Der Waerden es su eficiencia en el análisis de varianza (ANOVA) cuando se satisface el supuesto de normalidad, y es robusta al igual que la prueba Kruskal-Wallis en ausencia de normalidad.

- **Supuestos**

- La variable respuesta esta medida en una escala al menos intervalo.
- Las poblaciones tienen la misma distribución, difiriendo solo en su localización.

- **Inferencia Estadística**

Prueba de Hipótesis
Procedimiento para el Desarrollo de la Prueba
Sea n_i $(i = 1,2,\ldots,k)$ los tamaños muestrales para cada uno de los k grupos, n el tamaño de la muestra total.
Sea X_{ij} el valor i en el grupo j.

- Calcular las puntuaciones normales de la siguiente manera:

$$A_{ij} = \phi^{-1}\left(R\left(X_{ij}\right)/(n+1)\right)$$

con $R\left(X_{ij}\right)$ el rankeo de todos los datos sin considerar a que grupo pertenecen y ϕ que denota el rango de observación X_{ij} y la función de punto porcentual normal, respectivamente.
Para realizar la clasificación en rangos de las X_{ij} observaciones en orden ascendente y cada observación se reemplaza con su rango, por ejemplo $R\left(X_{ij}\right)$, asignándole a la observación menor el rango 1.
El promedio de los puntajes normales para cada muestra se puede calcular como:

$$\overline{A_\iota} = \frac{1}{n_i}\sum_{j=1}^{n_i} A_{ij} \qquad i = 1,2,\ldots,k$$

La varianza de los puntajes normales se puede calcular como:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} A_{ij}^2$$

Estadística de prueba: $T_1 = \frac{1}{s^2}\sum_{i=1}^{k} n_i\left(\overline{A_\iota}\right)^2$
Región crítica: $T_1 > \chi^2_{(1-\alpha,k-1)}$ donde χ^2es el cuantil 1-α de la distribución chi-cuadrado con k-1 graos de libertad.
Conclusión: Rechazamos la hipótesis nula si el estadístico de prueba se encuentra en la región crítica.

Prueba de Comparación

$$\left|\overline{A_\iota} - \overline{A_J}\right| > t_{(1-\alpha/2,k-1)}\ s\sqrt{\frac{n-1-T_1}{n-k}}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}} \qquad i = 1,2,\dots,k$$

- **Aplicación**

En una clínica se quiere contrastar si existe diferencia significativa entre los efectos de tres tipos de calmantes. Con este fin se suministran estos calmantes independientemente a pacientes comparables midiéndose el tiempo en segundos desde la inyección del calmante hasta la desaparición del dolor.

CALMANTE	TIEMPO EN QUE DESAPARECE EL DOLOR				
C1	30	35	60	61	
C2	41	32	70	55	77
C3	62	57	35	41	

$H_0 : Me_1 = Me_2 = Me_3$
$H_1 : Al\ menos\ una\ Me_i\ es\ diferente\ a\ las\ demás\ i = 1,2,3$
α=0.05

Prueba Estadística

CALMANTE	$R(X_{ij})$				
C1	1	3.5	9	10	
C2	5.5	2	12	7	13
C3	11	8	3.5	5.5	

Calmante	$A_{ij} = \phi^{-1}\left(R(X_{ij})/(n+1)\right)$					$\overline{A_\iota} = \frac{1}{n_i}\sum_{j=1}^{n_i} A_{ij}$
C1	-1.4652338	-0.674489	0.3661064	0.5659488		-0.3019171
C2	-0.27188	-1.067570	1.0675705	0	1.4652338	0.23867076
C3	0.7916386	0.1800124	-0.67449	-0.27188		0.0063203

Calmante	A_{ij}^2					$\sum_{j=1}^{n_i} A_{ij}^2$
C1	2.14691007	0.45493642	0.13403386	0.32029807		3.05617842
C2	0.07391874	1.13970682	1.13970682	0	2.146910	4.50024245
C3	0.62669169	0.03240445	0.45493642	0.07391874		1.1879513

$\sum_{i=1}^{k}\sum_{j=1}^{n_i} A_{ij}^2$	8.74437217
s^2	$8.74437217/(13-1) = 0.72869768$

Estadística de prueba: $T_1 = \frac{1}{s^2}\sum_{i=1}^{k} n_i(\overline{A_i})^2$= 0.8914453

Región crítica: $T_1 = 0.8914453$**<** $\chi^2_{(0.04,3-1)}$ = 6.437752,no se rechaza H_0

P-valor = 0.6403613 > $\alpha = 0.04$

Conclusión**:**
A un $\alpha = 0.04$, no se puede afirmar que existen diferencias significativas entre los tres tipos de calmante en el tiempo en segundos desde que se le aplica el calmante hasta la desaparición del dolor.

- **Funciones en R**

Se cuenta con la función waerden.test del paquete agricolae que si brinda las pruebas de comparaciones múltiples por pares. Esta función ofrece diferentes pruebas de comparación como: t, Holm, Hochberg, Bonferroni, BH, BY, fdr.

```
waerden.test(datos,grupos,group=TRUE,alpha = 0.04,console =
TRUE,p.adj=" ")
```

El argumento group=FALSE es para que brinde los p-valores en las pruebas de comparación, el argumento p.adj es para elegir la prueba de comparación y el argumento console=TRUE es para que me muestren los resultados.

- **Resultados con R**

```
calmante<-as.factor(c(1,1,1,1,2,2,2,2,2,3,3,3,3))
tiempo<-
as.numeric(c(30,35,60,61,41,32,70,55,77,62,57,35,41))
data<-cbind(calmante,tiempo)
data
```

```
library(agricolae)
waerden.test(tiempo,calmante,group=TRUE,alpha = 0.04,consol
e = T)


Study: tiempo ~ calmante
Van der Waerden (Normal Scores) test's

Value : 0.8951296
Pvalue: 0.6391828
Degrees of Freedom:  2

calmante,  means of the normal score

         tiempo       std r
1 -0.302818951 0.9483235 4
2  0.239375069 1.0271414 5
3  0.006385294 0.6294832 4
```

```
Post Hoc Analysis

Groups according to probability of treatment differences an
d alpha level( 0.04 )

Treatments with the same letter are not significantly diffe
rent.

Means of the normal score

         score groups
2  0.239375069      a
3  0.006385294      a
1 -0.302818951      a
```

D. Prueba de Jonckheere-Terpstra

➢ Aspectos Generales

Al igual que la prueba de Kruskal-Wallis, prueba la hipótesis de que las medianas de k grupos o muestras independientes son las mismas, en contra de la hipótesis alterna de que uno o más de esos grupos difieren de los otros.
La hipótesis alterna asociada al análisis de varianza unifactorial por rangos de Kruskal-Wallis es demasiado general en algunos casos. La Prueba de Jonckheere-Terpstra para niveles ordenados de la variable prueba la hipótesis de que las muestras (o grupos) se encuentran ordenadas en una secuencia específica a priori.
Entonces en la hipótesis nula las medianas son las mismas, mientras que en la hipótesis alterna las medianas se encuentran ordenadas por magnitud. Si la hipótesis alterna es verdadera, al menos una de las diferencias es estrictamente desigual.

➢ Supuestos

- Las k muestras de tamaños n_1, n_2,…,n_k que se utilizan para el análisis se deben haber extraído independientemente.
- La variable de interés debe estar medida en una escala al menos intervalo.

➢ Inferencia Estadística

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

- Para cada una las k muestras (o tratamientos) de manera independiente ordenar las observaciones de menor a mayor.
- Formar todos los diferentes pares de muestras.
- En cada par fijar una muestra, para cada una de las observaciones de esta muestra contabilizar cuantas observaciones de la otra muestra son superiores (o inferiores). Repetir este proceso para cada uno de los pares.
- Calcular la suma de las contabilizaciones de cada par.

$$U_{ij} = \sum_{h=1}^{n_i} \#\left(X_{hi}, j\right)$$

$\#(X_{ij},j)$ es el número de veces que el dato X_{hi} precede al dato de las muestras j donde i<j.

- Determinar el estadístico J que es la sumatoria de los recuentos determinados en el paso anterior.

$$J = \sum_{i<j}^{n_i} U_{ij} = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} U_{ij}$$

Si la muestra es suficientemente grande se puede hacer uso de la aproximación asintótica a la normal

$$\mu_J = \frac{n^2 - \sum_{j=1}^{k} n_j^2}{4} \qquad \sigma_J^2 = \frac{n^2(2n+3) - \sum_{j=1}^{k} n_j^2(2n_j+3)}{72} \qquad Z = \frac{J-\mu_J}{\sigma_J} \sim \mathrm{N}(0,1)$$

Unilateral izquierda	Bilateral	Unilateral derecha
H_0: $Me_1 = Me_2 = \ldots = Me_k$	H_0: $Me_1 = Me_2 = \ldots = Me_k$	H_0: $Me_1 = Me_2 = \ldots = Me_k$
H_1: $Me_1 < Me_2 < \ldots < Me_k$	H_1: *Al menos un* Me_i *es diferente* $i=1,2,..,k$	H_1: $Me_1 > Me_2 > \ldots > Me_K$

- **Aplicación**
  Se desea realizar un experimento para verificar si los juicios de salinidad se incrementarían conforme se disminuye la proporción de NaCl (cloruro de sodio) en los ensayos de prueba. Los datos se presentan a continuación:

Porcentaje de NaCl puro			
80	**50**	**17**	**10**
8.82	13.53	19.23	73.51
11.27	28.42	67.83	85.25
15.78	48.11	73.68	85.82
17.39	48.64	75.22	88.88
24.99	51.40	77.71	90.33
39.05	59.91	83.67	118.11
47.54	67.98	86.83	
48.85	79.13	93.25	
71.66	103.05		
72.77			
90.38			
103.13			

Realice la prueba respectiva use α=0.05.

Porcentaje de NaCl puro										
				80			**50**		**17**	**10**
i	**1**	**1**	**1**		**2**	**2**		**3**		
j	**2**	**3**	**4**		**3**	**4**		**4**		
	9	8	6	8.82	8	6	13.53	6	19.23	73.51
	9	8	6	11.27	7	6	28.42	6	67.83	85.25
	8	8	6	15.78	7	6	48.11	5	73.68	85.82
	8	8	6	17.39	7	6	48.64	5	75.22	88.88
	8	7	6	24.99	7	6	51.40	5	77.71	90.33
	7	7	6	39.05	7	6	59.91	5	83.67	118.11
	7	7	6	47.54	6	6	67.98	3	86.83	
	5	7	6	48.85	3	5	79.13	1	93.25	
	2	6	6	71.66	0	1	103.05			
	2	6	6	72.77						
	1	1	1	90.38						
	0	0	1	103.13						
U_{ij}	66	73	62		52	48		36		

H_0: $Me_1 = Me_2 = Me_3 = Me_4$
H_1: $Me_1 < Me_2 < Me_3 < Me_4$
$\alpha=0.05$
p-valor = 0.0003 < α se rechaza H_0

Conclusión
Existe suficiente evidencia estadística a un nivel de significación de 0.05, para rechazar H_0.
Por lo tanto, se puede afirmar que los juicios de salinidad se incrementarían conforme se disminuye la proporción de NaCl (cloruro de sodio) en los ensayos de prueba.

- **Funciones en R**

Se cuenta con la función jonckheere.test del paquete clinfun
```
jonckheere.test (vector de datos, grupos, alternativa)
```

Existe la función JonckheereTerpstraTest del paquete DescTools
```
JonckheereTerpstraTest(x, g, alternativa)
```

En el paquete kSamples se encuentra la función jt.test, esta no permite indicar la hipótesis alterna
```
jt.test (vector de datos, grupos, método)
```

Hay la función jonckheere.test del paquete PMCMR
```
jonckheere.test (vector de datos~ grupos, método)
```

También se puede hacer uso de la función JT.test del paquete SAGx pero este paquete debe ser obtenido desde el Bioconductor.
```
JT.test (vector de datos, grupos, alternativa)
```

- **Resultados con R**

Función:jonckheere.test
```
x1<-
c(8.82,11.27,15.78,17.39,24.99,39.05,47.54,48.85,71.66,72.7
7,90.38,103.13)
x2<-
c(13.53,28.42,48.11,48.64,51.40,59.91,67.98,79.13,103.05)
x3<-c(19.23,67.83,73.68,75.22,77.71,83.67,86.83,93.25)
x4<-c(73.51,85.25,85.82,88.88,90.33,118.11)
todo<-c(x1,x2,x3,x4)
grupos<-rep(1:4,c(12,9,8,6))

jonckheere.test(todo,grupos,alternative="in")
```

```
Jonckheere-Terpstra test

data:
JT = 337, p-value = 0.0003476
alternative hypothesis: increasing
```

Función:JT.test

```
JT.test(todo,grupos,alternative="in")
```

```
class was not an ordered factor.  Redefined to be one.

        Jonckheere-Terpstra

data:  todo by grupos
J = 113, p-value = 0.0004547
alternative hypothesis: increasing: 1 < 2 < 3 < 4
```

Función: JonckheereTerpstraTest

- **Algunas consideraciones en R**

- Existen dos funciones que brindan resultados distintos, en la función jonckheere.test el estadístico coincide con el del SPSS mientras que en la función JT.test el estadístico es corregido por su media y el pvalor es calculado en base a la aproximación asintótica a la Normal.

2. Para un parámetro de Escala (Dispersión)

Una prueba paramétrica muy utilizada para comparar la dispersión en k muestras independientes es la prueba de Bartlett. Sin embargo, esta prueba exige la normalidad de los datos en cada una de las k muestras.
En esta sección se presentarán algunas pruebas alternativas cuando no se cumple el supuesto de normalidad.
Sea θ_i un parámetro de escala
Las hipótesis son:

$H_0: \theta_1 = \theta_2 = \cdots = \theta_k$

$H_1: Al\ menos\ una\ \theta_i\ es\ diferente\ a\ las\ demás\ i = 1,2....,k$

Si se rechaza Ho se deben realizar las pruebas de comparación por pares

$H_0: \theta_i = \theta_j$
$H_1: \theta_i \neq \theta_j$ i,j=1,2,…, k $\forall i \neq j$

A. Prueba de Levene

➢ Aspectos Generales

Es una prueba usada para probar si k poblaciones tienen igual varianza. La prueba de Levene es una alternativa a la prueba de Bartlett, sin embargo, la prueba de Levene es menos sensible que la prueba de Bartlett.

➢ Supuestos

- Las muestras son independientes.
- Las muestras son tomadas al azar.
- La variable en estudio debe ser cuantitativa continua.

➢ Inferencia Estadística

- Las *observaciones iniciales son corregidas brindando valores Z_{ij}. Las correcciones pueden ser:*

$Z_{ij} = \left|Y_{ij} - \bar{Y}_{i\bullet}\right|$ donde $\bar{Y}_{i\bullet}$ es la media de la i-esima muestra.

$Z_{ij} = \left|Y_{ij} - \hat{Y}_{i\bullet}\right|$ donde $\hat{Y}_{i\bullet}$ es la mediana de la i-esima muestra.

$Z_{ij} = \left|Y_{ij} - \tilde{Y}'_{i\bullet}\right|$ donde $\tilde{Y}'_{i\bullet}$ es la media podada al 10% de la i-esima muestra.

$\bar{Z}_{i\bullet}$ son los grupos de medias de Z_{ij} y $\bar{Z}_{\bullet\bullet}$ es la media de todos los Z_{ij}.

Las tres opciones para definir Z_{ij} determinan la robustez y poder de prueba de Levene. La robustez, indica la capacidad de la prueba de detectar no falsas variancias distintas cuando los datos no se distribuyen normalmente. Poder de prueba significa la capacidad de la prueba de detectar variancias distintas cuando las variancias son realmente distintas.
El paper original de Levene solo propuso usar la media. Brown y Forsythe (1974) extendieron la prueba de Levene para usarlo con la mediana o la media podada.

Ellos realizaron estudios de simulación mediante Monte Carlo que indicaban que el uso de la media podada era mejor cuando los datos provienen de una distribución de Cauchy y la mediana es mejor cuando los datos siguen una distribución muy sesgada. La media provee el mejor poder para distribuciones simétricas, o con distribuciones con colas moderadas.

La prueba de Levene ofrece una alternativa más robusta que el procedimiento de Bartlett, ya que es poco sensible a la desviación de la normalidad. Esto significa que será menos probable que rechace una verdadera hipótesis de igualdad de varianzas solo porque las distribuciones de las poblaciones muestreadas no son normales.

- Aplicar el estadístico de prueba que tiene una distribución F con k-1 y n-k grados de libertad

$$W = \frac{(n-k)\sum_{i=1}^{k} n_i \left(\overline{Z}_{i.} - \overline{Z}_{..}\right)^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Z_{ij} - \overline{Z}_{i.}\right)^2} \sim F_{(1-\alpha,k-1,n-k)}$$

- Se rechaza la hipótesis nula con un nivel de significación α si

$$W > F_{(1-\alpha,k-1,n-k)}.$$

El p-valor se calcula de la siguiente manera:

$$P\left[F_{(k-1,n-k)} > F_{cal}\right] = 1 - P\left[F_{(k-1,n-k)} \leq F_{cal}\right]$$

Una forma sencilla de entender la prueba de Levene es:
Calcular $Z_{ij} = \left|Y_{ij} - \overline{Y}_{i.}\right|$
Realizar el Análisis de varianza sobre los datos Z_{ij}
Por lo tanto, si la prueba F sobre la variable Z_{ij} da resultado significativo, se rechazará la igualdad de varianzas.

- **Aplicación**

Utilice los datos de la aplicación de la prueba de Kruskal Wallis para verificar si la dispersión del puntaje obtenido es diferente en al menos uno de los métodos utilizados

$H_0: \theta_1 = \theta_2 = \theta_3$

$H_1: Al\ menos\ una\ \theta_i\ es\ diferente\ a\ las\ demás\ i = 1,2,3$

$\alpha = 0.05$

pvalor= 0.834
Conclusión
A un nivel de significación de 0.05, no se puede afirmar que la variabilidad de la calificación es diferente al utilizar al menos uno de los métodos de instrucción considerados en el estudio.

- **Funciones con R**

Existe la función levene.test del paquete lawstat o leveneTest del paquete car

```
levene.test(datos, grupos)
leveneTest(datos, grupos)
```

- **Resultados con R**

```
M1<-c(86,79,81,70,84)
M2<-c(90,76,88,82,89)
M3<-c(82,68,63,71,61)
datos<-c(M1,M2,M3)
grupos<-rep(1:3,c(5,5,5))
library(lawstat)
levene.test(datos,grupos)
```

```
modified robust Brown-Forsythe Levene-type test based on the
absolute
     deviations from the median
data:  datos
Test Statistic = 0.18391, p-value = 0.8343
```

```
library(car)
leveneTest(datos,as.factor(grupos))

Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.1839 0.8343
      12
```

B. Prueba de Fligner-Killeen

- **Aspectos Generales**

Cuando se desea comparar un parámetro de dispersión como la varianza, en k muestras independientes, en el caso paramétrico existe la prueba de Bartlett; sin embargo, esta exige que cada una de las muestras provengan de una distribución normal; supuesto que muchas veces no se cumple.
Existen muchas pruebas no paramétricas que permiten comparar un parámetro de dispersión en k muestras independientes una de ellas es la prueba de Fligner-Killeen.

- **Supuestos**

- Las k muestras de tamaños $n_1, n_2, \ldots, n_k$ que se utilizan para el análisis se deben haber extraído independientemente.
- La variable de interés debe estar medida en una escala al menos intervalo.

- **Inferencia Estadística**

- Ordenar las observaciones $\left(x_{ij} - me_i\right)$ de menor a mayor, donde me_i es la mediana del grupo i.
- Calcular $a_{n,i} = \Phi^{-1}\left(\frac{1}{2} + \frac{i}{2(n+1)}\right)$ para i=1,2,…,n. donde $\Phi^{-1}(p)$ es el percentil 100p de la distribución N(0,1)
- Se definen

$$\bar{a}_i = \sum_{j \in Gi} \frac{a_{n,j}}{n_i} \qquad \qquad \bar{a} = \sum_{j=1}^{n} \frac{a_{n,j}}{n}$$

Donde G_i denota la muestra de la población i, i=1,2,..,k

- Entonces el estadístico de prueba es

$$F - K = \frac{\sum_{i=1}^{k} n_i\left(a_i - \bar{a}\right)^2}{\sum_{j=1}^{n}\left(a_{n,j} - a\right)^2 \Big/ (n-1)} \sim \chi^2_{(1-\alpha,k-1)}$$

Las hipótesis son:

$H_0: \theta_1 = \theta_2 = \cdots = \theta_k$

$H_1: Al\ menos\ una\ \theta_i\ es\ diferente\ a\ las\ demás\ i = 1, 2...., k$

- **Aplicación**

Utilice los datos de la aplicación de la prueba de Kruskal Wallis para verificar si la dispersión del puntaje obtenido es diferente en al menos uno de los métodos utilizados

$H_0: \theta_1 = \theta_2 = \theta_3$

$H_1: Al\ menos\ una\ \theta_i\ es\ diferente\ a\ las\ demás\ i = 1, 2, 3$

α=0.05

pvalor= 0.7032
Conclusión
A un nivel de significación de 0.05, no se puede afirmar que la variabilidad de la calificación es diferente al utilizar al menos uno de los métodos de instrucción considerados en el estudio.

- **Funciones con R**
  Existe la función fligner.test

```
fligner.test(datos, grupos)
```

- **Resultados con R**

```
M1<-c(86,79,81,70,84)
M2<-c(90,76,88,82,89)
M3<-c(82,68,63,71,61)
datos<-c(M1,M2,M3)
grupos<-rep(1:3,c(5,5,5))
fligner.test(datos,grupos)
```

```
     Fligner-Killeen test of homogeneity of variances

data:  datos and grupos
Fligner-Killeen:med chi-squared = 0.70427, df = 2,
p-value = 0.7032
```

C. Prueba de Klotz

➢ Aspectos Generales

Jerome H. Klotz desarrolló una prueba que está en parte basado en el trabajo de Siegel-Tukey, en el cual se evalúa un parámetro de escala, en su trabajo llamado "Nonparametric Test For Scale" publicado en 1962 en la Universidad de California.

A pesar de que existen diferentes pruebas para comprobar la homogeneidad de varianzas en el campo de la estadística no paramétrica, como son el Test de Mood, Test de Conover, Test de Siegel-Tukey, entre otras. Klotz tiene ciertas ventajas particulares que son descritas por Hollander and Wolfe (1999):

- Es altamente recomendable particularmente en distribuciones simétricas, por su potencia y robustez.
- No necesita el supuesto de igualdad de medianas a comparación de pruebas similares como la de Siegel-Tukey.
- Recomendable para muestras grandes.

Se debe resaltar que inicialmente Klotz desarrolló la prueba para dos muestras únicamente. Más adelante el estadístico Puri, en 1965, lo generalizó para k muestras independientes.

➢ Supuesto

- La variable de interés debe estar en una escala por lo menos intervalo.

➢ Inferencia Estadística

La prueba de Klotz se basa en los cuadrados de los scores normales. Bajo normalidad su eficiencia respecto a la prueba F es 1.

La ventaja de muchas pruebas basadas en puntajes normales es que se desempeñan bien cuando se cumplen los supuestos de las pruebas paramétricas estándar.

Procedimiento para el desarrollo de la prueba

- Obtener los rangos r_{ij} de las n observaciones sin importar a que grupo pertenecen, no se efectuara ningún ajuste si hay empates
- Calcular la función de puntuación.

$$m_{ij} = \Phi^{-1}\left(\frac{r_{ij}}{n+1}\right)^2$$

Donde:
r_{ij}: denota el rango en el total de la muestra de i-esimo dato en el j-esimo grupo
n: Es el tamaño de la muestra total ($n=n_1+\ldots+n_k$).
Φ^{-1}: Es la función de puntuación de la distribución normal estándar.

Para dos muestras el estadístico es:

$$KL = (n-1)\frac{\sum_{j=1}^{k}\sum_{i=1}^{n1} m_{ij}{}^{2} - \left(\frac{n_1}{n_1+n_2}\right)\sum_{j=1}^{k}\sum_{i=1}^{n_1+n_2} m_{ij}{}^{2}}{\left(\frac{n_1 * n_2}{(n_1+n_2)((n_1+n_2-1)}\right)\sum_{j=1}^{k}\sum_{i=1}^{n_j} m_{ij}{}^{2} - \frac{1}{n_1+n_2}\left(\sum_{i=1}^{k}\sum_{i=1}^{n_j} m_{ij}\right)^{2}} \sim \chi^2_{(1-\alpha,1)}$$

Para k muestras independientes (aportación de Puri):

$$KL = (n-1)\frac{\sum_{i=1}^{k} n_j(m_{.j} - m_{..})^2}{\sum_{j=1}^{k}\sum_{i=1}^{n_j} m_{ij}{}^{2} - \frac{1}{n}(\sum_{i=1}^{k}\sum_{i=1}^{n_j} m_{ij})^2} \sim \chi^2_{(1-\alpha,k-1)}$$

Donde:
$m_{.j}$: es la media de los m_{ij} del j-esimo grupo
$m_{..}$: es la media de todos los m_{ij}

- **Aplicación**

Del total de trabajadores del área de mantenimiento mecánico de la minera Volcán del Sur S.A fueron seleccionados cincuenta de ellos de manera aleatoria e independiente, y estos fueron asignados a cinco programas diferentes de capacitación para la reparación de maquinarias utilizadas para la extracción de minerales, todos los cuales persiguen el objetivo de aumentar los conocimientos técnicos de dichos trabajadores y por ende la eficiencia al realizar sus labores. Para analizar la efectividad de estos programas de capacitación se realizó una prueba al final de cada uno, dicha prueba evaluaba los conocimientos adquiridos durante las capacitaciones con calificaciones del 1 al 20. Los resultados se presentan a continuación:

Programa 1	Programa 2	Programa 3	Programa 4	Programa 5
6	11	8	16	5
9	15	19	17	6
15	9	14	12	12
14	4	18	10	6
5	16	4	16	3
8	18	13	18	5
8	11	20	5	12
14	6	3	7	5
11	10	7	7	10
15	19	17	5	13

Pruebe si la dispersión de los puntajes obtenidos es diferente en al menos uno de los programas de capacitación desarrollados.

Los rangos correspondientes a los datos son:

Programa 1	Programa 2	Programa 3	Programa 4	Programa 5
12.5	27	19	41	7.5
21.5	38	48.5	43.5	12.5

38	21.5	35	30	30
35	3.5	46	24	12.5
7.5	41	3.5	41	1.5
19	46	32.5	46	7.5
19	27	50	7.5	30
35	12.5	1.5	16	7.5
27	24	16	16	24
38	48.5	43.5	7.5	32.5

$H_0 : \theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_5$

H_1 : Al menos un θ_i es diferente a las demás para i $= 1, 2, 3, 4, 5$

$\alpha = 0.05$

Prueba estadística

$$KL = (n-1)\frac{\sum_{i=1}^{k} n_j (m_{.j} - m_{..})^2}{\sum_{j=1}^{k}\sum_{i=1}^{n_j} m_{ij}{}^2 - \frac{1}{n}(\sum_{i=1}^{k}\sum_{i=1}^{n_j} m_{ij})^2} \sim \chi^2_{(1-\alpha, k-1)}$$

Donde:

$$Z_{ij} = \Phi^{-1}\left(\frac{R_{ij}}{N+1}\right) = \left(\Phi^{-1}\frac{12.5}{50+1}\right), \ldots, \left(\Phi^{-1}\frac{32.5}{50+1}\right) = -0.689, \ldots, 0.351$$

Desarrollo de la Prueba:

$$m_{ij} = Z_{ij}^2 = \Phi^{-1}\left(\frac{R_{ij}}{N+1}\right)^2 = \left(\Phi^{-1}\frac{12.5}{50+1}\right)^2 \ldots \left(\Phi^{-1}\frac{32.5}{50+1}\right)^2 = (-0.689)^2, \ldots, (0.351)^2 = 0.476, \ldots, 0.123$$

$$m_{.j} = \frac{\left(\Phi^{-1}\frac{12.5}{51}\right)^2 + \ldots + \left(\Phi^{-1}\frac{38}{51}\right)^2}{50}, \frac{\left(\Phi^{-1}\frac{27}{51}\right)^2 + \ldots + \left(\Phi^{-1}\frac{48.5}{51}\right)^2}{50}, \ldots, \frac{\left(\Phi^{-1}\frac{7.5}{51}\right)^2 + \ldots + \left(\Phi^{-1}\frac{32.5}{51}\right)^2}{50}$$

m.j= $0.3173, 0.8315, 1.6239, 0.6964, 0.8052$

$$m_{..} = \frac{1}{N}\sum_{1}^{N} Z_{ij}^2 = \frac{\left(\Phi^{-1}\frac{12.5}{50+1}\right)^2 + \ldots + \left(\Phi^{-1}\frac{32.5}{50+1}\right)^2}{50} = 0.8549$$

$$KL = (50-1)\frac{10(0.3173-0.8549)^2 + 10(0.8315-0.8549)^2 + \ldots + 10(0.8053-0.8549)^2}{\left(0.4761^2 + 0.03916^2 + \ldots + 0.1233^2\right) - \frac{(0.4761 + 0.03916 + \ldots + 0.1233)^2}{50}} = 8.3488$$

Criterio de decisión:
Como $\chi^2_{(0.95,4)} = 9.487729 > 8.3488$, No se rechaza H_0

Conclusión:
A un nivel de significancia de 0.05 y con un p.valor= 0.07961, no se puede afirmar que la variabilidad de los puntajes obtenidos es diferente al utilizar al menos uno de los programas de capacitación considerados en el estudio.

- **Funciones con R**
  Existe la función klotz_test del paquete coin

```
klotzs_test(datos~grupos)
```

- **Resultados con R**

```
M1<-c(6,9,15,14,5,8,8,14,11,15)
M2<-c(11,15,9,4,16,18,11,6,10,19)
M3<-c(8,19,14,18,4,13,20,3,7,17)
M4<-c(16,17,12,10,16,18,5,7,7,5)
M5<-c(5,6,12,6,3,5,12,5,10,13)
datos=c(M1,M2,M3,M4,M5)
grupos<-rep(1:5,c(10,10,10,10,10))
library(coin)
klotz_test(datos~as.factor(grupos))
```

```
Asymptotic K-Sample Klotz Test

data:  datos by as.factor(grupos) (1, 2, 3, 4, 5)
chi-squared = 8.3487, df = 4, p-value = 0.07961
```

D. Prueba de Conover

- **Aspectos Generales**

En el capítulo anterior se presentó la prueba de Conover para comparar una medida de dispersión de dos poblaciones, ahora se presentará una extensión de esta prueba.
La notación que se utilizará es:
$n=n_1+\ldots+n_k$
$\bar{x}_i$ = es la i-ésima media muestral
$R(x_{ij})$= es el rango $\left(x_{ij}-\bar{x}_i\right)^2$ de todas las n observaciones.

- **Supuesto**

- La variable de interés debe estar en una escala por lo menos intervalo.

- **Inferencia Estadística**

- Calcular la media $\bar{x}_i$ para cada uno de los grupos
- Obtener las desviaciones al cuadrado de cada observación con respecto a su media $\left(x_{ij}-\bar{x}_i\right)^2$
- Obtener los rangos $R(x_{ij})$ de las n observaciones sin importar el grupo al que pertenecen
- Calcular

$$T_i=\sum_{j=1}^{n_i}R\left(x_{ij}\right)^2$$

$$\bar{T}=\frac{1}{n}\sum_{j=1}^{k}T_j$$

$$V_T=(n-1)^{-1}\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}R\left(x_{ij}\right)^4-n\bar{T}^2\right)$$

Prueba de Hipótesis
La prueba estadística es

$$T=\frac{\sum_{j=1}^{k}\left(T_j^2/n_j\right)-n\bar{T}^2}{V_T}\sim\chi^2_{(1-\alpha,k-1)}$$

Si no hay empates

$$\bar{T}=\frac{(n+1)(2n+1)}{6}\qquad V_T=\frac{n(n+1)(2n+1)(8n+11)}{180}$$

Comparaciones Múltiples

Para eso se puede hacer uso del siguiente criterio de comparaciones por pares:

$$\left|\frac{T_i}{n_i}-\frac{T_j}{n_j}\right|>t_{(n-k,1-\alpha/2)}\sqrt{\left(\frac{1}{n_i}+\frac{1}{n_j}\right)V_T\left(\frac{n-1-T}{n-k}\right)}$$

- **Aplicación**

Utilice los datos de la aplicación de la prueba de Kruskal Wallis para verificar si la dispersión del puntaje obtenido es diferente en al menos uno de los métodos utilizados

$H_0 : \theta_1 = \theta_2 = \theta_3$

$H_1 : Al\ menos\ una\ \theta_i\ es\ diferente\ a\ las\ demás\ i = 1,2,3$

$\alpha = 0.05$

Prueba Estadística

$$T = \frac{\sum_{j=1}^{k}\left(T_j^2/n_j\right) - n\overline{T}^2}{V_T} \sim \chi^2_{(1-\alpha,k-1)}$$

Donde:

$$T_i = \sum_{j=1}^{n_i} R(x_{ij})^2$$

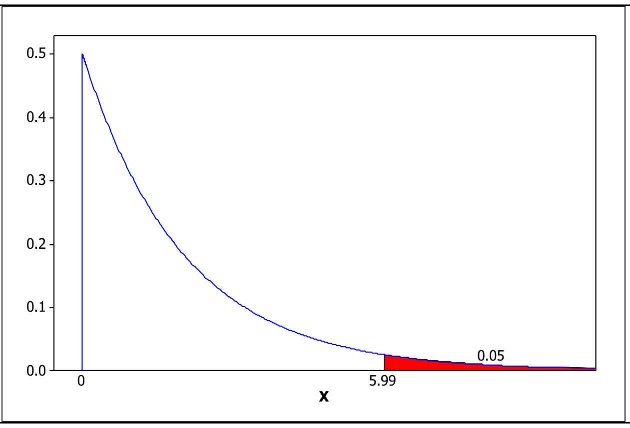
$$\overline{T} = \frac{1}{n}\sum_{j=1}^{k} T_j$$

$$V_T = (n-1)^{-1}\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} R(x_{ij})^4 - n\overline{T}^2\right)$$

Desarrollo de la Prueba

$$T = 0.421$$

Criterio de Decisión



Como $\chi^2_{(0.95,2)} = 5.99 > 0.421$ no se rechaza H_0

Conclusión

A un nivel de significación de 0.05, no se puede afirmar que existe que la variabilidad de la calificación es diferente en al menos uno de los métodos de instrucción en estudio.

Con fines prácticos se procederá a realizar las pruebas de comparación.

$H_0 : \theta_i = \theta_j$
$H_1 : \theta_i \neq \theta_j$ i,j=1,2,3 $\forall$i≠j

α=0.05

$$\left|\frac{T_i}{n_i}-\frac{T_j}{n_j}\right| > t_{(n-k,1-\alpha/2)}\sqrt{\left(\frac{1}{n_i}+\frac{1}{n_j}\right)V_T\left(\frac{n-1-T}{n-k}\right)}$$

Comparaciones	$\left\lvert\frac{T_i}{n_i}-\frac{T_j}{n_j}\right\rvert$	**Valor Crítico**	**Sig**
A_1 vs A_2	$\left\lvert\frac{370.5}{5}-\frac{366.75}{5}\right\rvert=0.75$	107.82	n.s.
A_1 vs A_3	$\left\lvert\frac{370.5}{5}-\frac{499.25}{5}\right\rvert=25.75$	107.82	n.s
A_2 vs A_3	$\left\lvert\frac{366.75}{5}-\frac{499.25}{5}\right\rvert=26.5$	107.82	n.s.

Conclusión
A un nivel de significación de 0.05 no se puede afirmar que exista diferencia en la variabilidad de las calificaciones con los métodos en estudio.

Resumen

$\underline{A_3 \quad A_1 \quad A_2}$

- **Funciones con R**
  Existe la función conover_test del paquete coin

```
conover_test(datos~grupos)
```

- **Resultados con R**

```
M1<-c(86,79,81,70,84)
M2<-c(90,76,88,82,89)
M3<-c(82,68,63,71,61)
datos<-c(M1,M2,M3)
grupos<-rep(1:3,c(5,5,5))
conover_test(datos~as.factor(grupos))
```

```
Asymptotic K-Sample Conover-Iman Test

data:  datos by as.factor(grupos) (1, 2, 3)
chi-squared = 0.4208, df = 2, p-value = 0.8103
```

También se puede utilizar comandos en R:

```
n1<-length(M1)
n2<-length(M2)
n3<-length(M3)
n<-n1+n2+n3
xbar1<-mean(M1)
xbar2<-mean(M2)
xbar3<-mean(M3)
```

```
M1c<-(M1-xbar1)^2 ; M2c<-(M2-xbar2)^2 ; M3c<-(M3-xbar3)^2
rdc<-rank(c(M1c,M2c,M3c))
rdc2<-rdc^2
T1<-sum(rdc2[1:n1])
T2<-sum(rdc2[(n1+1):(n1+n2)])
T3<-sum(rdc2[(n1+n2+1):n])
Tbar<-(T1+T2+T3)/n
Vt<-(sum(rdc^4)-n*Tbar^2)/(n-1)
Tcon<-((T1^2/n1+T2^2/n2+T3^2/n3)-n*Tbar^2)/Vt
1-pchisq(Tcon,3-1)

#Comparaciones múltiple: "M1vsM2"
dif<-abs(T1/n1-T2/n2)
vc<-qt(0.975,n-3)*sqrt((1/n1+1/n2)*Vt*((n-1-Tcon)/(n-3)))
```

3. Prueba de Permutación para k muestras independientes

- **Aspectos Generales**

Se puede extender la noción de prueba de permutación para dos tratamientos a estudios que involucran k tratamientos. Esto es las unidades experimentales son asignadas a k tratamientos (o grupos) en un diseño completamente aleatorio o las observaciones han sido aleatoriamente seleccionadas de k poblaciones. Si hay diferencias entre los tratamientos, esto es las observaciones de al menos un tratamiento tiende a brindar observaciones distintas a la de los otros tratamientos.

Si las observaciones son seleccionadas aleatoriamente de poblaciones normales con varianzas iguales, entonces el estadístico F puede ser útil para encontrar si hay diferencias en las medias de al menos un tratamiento. Sin embargo, si no se cumplen los supuestos, entonces se puede llevar a cabo la prueba de permutación en lugar de la prueba F.

- **Supuestos**

- Las muestras a ser evaluadas son aleatorias y mutuamente excluyentes.
- La variable respuesta debe estar medida en una escala al menos ordinal.

- **Inferencia Estadística**

Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

Cada grupo está compuesto por n_1, n_2,…n_k observaciones de tal manera que $n_1+n_2+…+n_k=n$

- Obtener el estadístico F para los datos originales, denotado por F_{obs}.
- Obtener todas las posibles permutaciones de las n observaciones de los k tratamientos hay:

$$\frac{n!}{n_1!n_2!...n_k!}$$

Posibles permutaciones. Si no es posible generar todas las permutaciones, entonces se debe seleccionar una muestra de R permutaciones.

- Para cada permutación, calcular el estadístico F.
- Obtener el pvalor como la fracción de las F´s del paso anterior que son mayores o iguales a F_{obs}. El pvalor es aproximado si se calcula en base a una muestra aleatoria de permutaciones. Note que siempre es una prueba cola derecha.

$$P\,\text{valor} = \frac{\#\,F\acute{}s \geq F_{obs}}{R}$$

Las hipótesis son:

$H_0: \mu_1 = \mu_2 = ... = \mu_k$

$H_1: Al\ menos\ una\ \mu_i\ es\ diferente\ a\ las\ demás\ i = 1,2,...,k$

Donde μ_i es una medida de posición.

- **Aplicación**

  Simule datos para tres tratamientos con cinco repeticiones cada uno. Pruebe si existen diferencias significativas

- **Secuencia o funciones con programas estadísticos**

  En R

```
#Generar los datos
set.seed(10)
T1<-round(rexp(5,1/10),1)
T2<-round(rexp(5,1/13),1)
T3<-round(rexp(5,1/18),1)
tiempo<-c(T1,T2,T3)
factor<-as.factor(rep(1:3,c(5,5,5)))
res<-summary(lm(tiempo~factor))
Fobs<-as.numeric(res$fsta[1])
Fobs
R<-100
Fper<-rep(0,R)
for (i in 1:R){
  res<-summary(lm(sample(tiempo,15,F)~factor))
  Fper[i]<-as.numeric(res$fsta[1])
  }
 pvalor<-mean(ifelse(Fper>=Fobs,1,0))
pvalor
```

En R existen varias funciones que permiten realizar la prueba de permutación para k muestras independientes, algunas son:

- La función aovp el paquete lmPerm modificado para utilizar pruebas de permutación en lugar de las pruebas teóricas normales.

```
M1<-c(86,79,81,70,84)
M2<-c(90,76,88,82,89)
M3<-c(82,68,63,71,61)
datos<-c(M1,M2,M3)
grupos<-rep(1:3,c(5,5,5))
library(lmPerm)
summary(aovp(datos~as.factor(grupos)))
```

- La función perm.f.test del paquete jmuOutlier realiza una prueba de permutación en la estadística F.

```
library(jmuOutlier)
perm.f.test(datos , grupos, num.sim = 20000 )
```

4. Prueba para un experimento factorial
Prueba de Scheirer-Ray-Hare

- **Aspectos Generales**

La prueba de Scheirer-Ray-Hare es el equivalente no paramétrico de la prueba paramétrica F para dos factores con replicación balanceada.
La prueba fue publicada en 1976 por James Scheirer, William S.Ray y Nathan Hare, el método consistió en extender la prueba de rangos de Kruskal-Wallis que permite el cálculo de los efectos de interacción y los contrastes lineales.
Al igual que la prueba de Kruskall Wallis en esta prueba está presente la suposición acerca de la distribución de los datos.
Esta prueba no es muy considerada ya que tiene una potencia mucho menor que la ANOVA paramétrica, por este motivo se recomienda que cada interacción de los factores en el modelo debe contener por lo menos 5 observaciones a más para un mejor equilibrio de la potencia de esta prueba.
Groggel y Skillings en 1986 en su libro *"Pruebas sin distribución para efectos principales"* presentaron un método para realizar estas pruebas no paramétricas para más de 2 factores.

- **Supuesto**

- Las observaciones son independientes, es decir, no son pares o medidas repetidas.
- Las r repeticiones a ser evaluadas en cada uno de las k tratamientos o grupos de observaciones de la interacción de los tratamientos son aleatorias y mutuamente excluyentes.
- La variable respuesta esta medida en una escala al menos intervalo.
- Las observaciones de los k tratamientos tienen la misma distribución, difiriendo solo en su localización
- Los grupos de interacción deben ser del mismo tamaño y contener al menos r=5 repeticiones.

- **Inferencia Estadística**

Procedimiento para el desarrollo de la prueba

- Se debe considerar las n observaciones totales n=rk.
- Se rankea las observaciones conjuntamente (de manera ascendente o en caso de empates se asigna el rango promedio), luego en cada grupo o tratamiento se suman los rangos correspondientes, obteniendo por lo tanto k sumas de rangos.

$$R_i = \sum_{i=1}^{n} R(X_{ij})$$

- Se aplicar la siguiente prueba estadística:

$$H = SS(R_{ij})\frac{12}{\mathrm{n(n+1}} \sim \chi^2_{(1-\alpha,k-1)}$$

Siendo SS(R_{ij}) la suma de cuadrados de los rangos y se obtiene de la siguiente manera:
Para las filas:

$$SS(R_i) = \sum_{i=1}^{p} (\overline{R_i} - \overline{R_{ij}})^2 * p$$

Donde: p es el número de niveles del factor A

Para las columnas:

$$SS(R_j) = \sum_{j=1}^{l} (\overline{R_j} - \overline{R_{ij}})^2 * l$$

Donde: l es el número de niveles del factor B

Para la interacción:

$$SS(R_{ij}) = \sum_{j=1}^{l} \left[- \sum_{i=1}^{p} (\bar{R}_{i.} + \bar{R}_{.j} \right]^2 r$$

Donde: r es la cantidad de repeticiones

Término de corrección: $c_{mi} = \frac{(n-n_{ij})}{n}$

Donde:

n : Tamaño total de las observaciones.

R_i : Suma de los rangos de la i-ésima muestra o grupo de tratamiento.

n_{ij} : Número de observaciones de la j-ésima muestra del i-ésimo factor

r : Número de repeticiones.

Si hay empates se corrige H dividiéndolo por $1 - \frac{\sum_{i=1}^{G}(t_i^3 - t_i)}{n^3 - n}$
donde t_i es el número de empates en el g-ésimo grupo.
Si el $R(X_{ij})$ se trata como la variable dependiente en un ANOVA, el álgebra de rangos muestra que H es equivalente a la relación de la suma de cuadrados a tratar, calculada en rangos (RSS), dividida por un "cuadrado medio" para la variabilidad total (RMS), también calculado por rangos en la forma del análisis de varianza (ANOVA). Eso es

$$H = \frac{RSS_{treatment}}{RSS_{total}}$$

El numerador, $RSS_{tratamiento}$, es el componente de variabilidad que se divide en componentes mediante el uso de contrastes lineales o en los principales efectos e interacciones en un ANOVA. Uno podría sospechar un resultado análogo para $RSS_{tratamiento}$, de modo que los resultados de una partición, RSS_m, podrían ser sustituidos por RSS tratamiento para su uso en una estadística H, Hm. Por lo tanto, $\sum_{m-1}^{k-i} H_m$ sería igual a H de una manera análoga a la partición de $SS_{tratamiento}$ en un ANOVA. A continuación, se presenta una demostración de estas ideas.

Prueba de hipótesis

En primer lugar, se debe demostrar que existe significancia en la interacción de los factores.

$H_0 : (\alpha\beta)_{ij} = 0$

$H_A : (\alpha\beta)_{ij} \neq 0$

Si es el caso de no rechazarse la hipótesis nula entonces se deberá probar las siguientes hipótesis:

$H_0^{(\alpha)}: \alpha_1 = ... = \alpha_i$

H_A : Al menos un α es diferente

$H_0^{(\beta)}: \beta_1 = ... = \beta_j$

H_A : Al menos un β es diferente

- **Aplicación**

Un biólogo está investigando el efecto de la luz en la ingesta de alimentos en estorninos (Sturnus vulgaris). Establece un experimento donde las aves se colocan en aviarios individuales de tamaño idéntico con iluminación controlada y se les da un exceso de comida. Las aves son sexadas y cuatro machos y cuatro hembras están expuestos a 16 h u 8 h de luz (aquí denominados días "largos" o "cortos"). Cada ave se controla durante 7 días y se registra su ingesta total de alimentos en gramos. Cada ave solo se usa una vez.

	Sexo							
Tiempo	Femenino				Masculino			
largo (16h)	78.1	75.5	76.3	81.2	69.5	72.1	73.2	71.1
corto (8h)	82.4	80.9	83.0	88.2	72.3	73.3	70.0	72.9

Probar la hipótesis de la interacción entre el sexo y el tiempo que quedan expuestas las aves, además probar la hipótesis de que existe diferencia significativa mediana del sexo y del tiempo que quedarán expuestas las aves si

es que corresponde, usar un α =0.05
Factores
A: Sexo (dos niveles: masculino y femenino)
B: Tiempo de exposición (dos niveles: 16 y 8 horas de luz)
r: 4 (número de replicaciones)

Los rangos correspondientes a los datos son:

	Sexo							
tiempo	Femenino				Masculino			
Largo (16h)	11	9	10	13	1	4	7	3
corto (8h)	14	12	15	16	5	8	2	6

$H_0: (\alpha\beta)_{ij} = 0$

$H_A: (\alpha\beta)_{ij} \neq 0$ $\forall\ i, j = 1, 2$. Siendo $i \neq j$

Calculo del estadístico:

$$H = \frac{12 \times 4}{16(16+1)} SS(R_{ij}) = 0.1765$$

Criterio de decisión:

Como $\chi^2_{(0.95,1)} = 3.841459 > 0.1765$ No se rechaza H_0

Conclusión: Para un α=0.05, se puede afirmar que no existe interacción significativa entre el sexo y el tiempo en que las aves son expuestas a la luz, por tanto, se procede a evaluar los efectos principales.

Evaluando para el sexo

$H_0^{(\alpha)} : \alpha_1 = \alpha_2$

H_A : Al menos un α_i es diferente para $i = 1, 2$

Cálculo del estadístico:

$$H = \frac{12 \times 4}{16(16+1)} SS(R_i) = 11.29$$

Criterio de decisión:

Como $\chi^2_{(0.95,1)} = 3.841459 < 11.29$ Se rechaza H_0

Conclusión: Para un α =0.05, se puede afirmar que la ingesta de alimento en las aves por sexo es diferente.

Evaluando para el tiempo de exposición a la luz

$H_0^{(\beta)} : \beta_1 = \beta_2$

H_A : Al menos un β_j es diferente, para j=1,2

Cálculo del estadístico:

$$H = \frac{12 \times 4}{16(16+1)} SS(R_j) = 1.1029$$

Criterio de decisión:

Como $\chi^2_{(0.95,1)} = 3.841459 > 1.1029$ No se rechaza H_0

Conclusión: Para un α =0.05, no se puede afirmar que la ingesta de alimento en las aves por los diferentes tiempos de exposición a la luz sean diferentes.

- **Funciones con R**

Se cuenta con la función *scheirerRayHare* del paquete *rcompanion* la cual solo funciona con un máximo de dos factores, esta prueba contiene los siguientes argumentos:

fórmula *:* Una fórmula indicando la variable respuesta y las dos variables independientes, ejemplo y ~ x1 + x2.
data **:** La data frame usada.
y **:** Si no se expresa una fórmula, esta es la variable respuesta.
x1 **:** La primera variable independiente.
x2 **:** La segunda variable independiente.
tie.correct **:** Si es “TRUE”, se aplicará la corrección para empates en la variable respuesta.
ss **:** Si "TRUE", incluye las sumas de cuadrados en la salida
verbose **:** Si "TRUE", las estadísticas de salida utilizadas en el análisis por impresión directa.

- **Resultados con R**

```
library(rcompanion)
scheirerRayHare(intake~sex+day,data = nopara)

DV:  intake
Observations:  16
D:  1
MS total:  22.66667

                   Df     Sum Sq        H      p.value
sex                1      256      11.2941    0.00078
day                1      25        1.1029    0.29362
sex:day            1      4         0.1765    0.67442
Residuals          12     55
```

5. Para comparar las funciones de distribución F(X)

En algunas pruebas mencionadas anteriormente es necesario verificar previamente como supuesto que las k muestras independientes provienen de la misma distribución continua. Esto sucede por ejemplo con la prueba de Kruskal Wallis; muchas veces esto no se verifica por lo que sus resultados nos podrían llevar a conclusiones erróneas.

Las hipótesis son:

$$H_0 : F(X_1) = F(X_2) = \cdots = F(X_k)$$

$$H_1 : Al\ menos\ una\ F(X_i)\ es\ diferente\ a\ las\ demás\ i = 1, 2...., k$$

El paquete kSamples tiene la función ad.test para comparar si las k muestras independientes provienen o no de la misma distribución.
La función ad.test utiliza el criterio de Anderson-Darling y permite realizar las hipótesis antes propuestas.

ad.test(datos~as.factor(grupos))

```
Anderson-Darling k-sample test.
Number of samples:  3
Sample sizes:  5, 5, 5
Number of ties: 1

Mean of  Anderson-Darling  Criterion: 2
Standard deviation of  Anderson-Darling  Criterion: 0.91893

T.AD = ( Anderson-Darling  Criterion - mean)/sigma
Null Hypothesis: All samples come from a common population.
              AD  T.AD  asympt. P-value
version 1: 4.451 2.667          0.02295
version 2: 4.600 2.831          0.01925
```

Se debe tener cuidado con esta prueba dado que a pesar de que los datos provengan de la misma distribución si el parámetro de locación es diferente en al menos una de ellas, se rechazará también la hipótesis de que las muestras provienen de distribuciones con igual forma. Para superar este problema lo que se puede hacer es corregir cada uno de los grupos o tratamientos por su media si la distribución de los datos es simétrica o por su mediana si los datos tienen una distribución asimétrica.

```
set.seed(10)
data1<-round(rexp(15,1/10),1)
set.seed(11)
data2<-round(rexp(15,1/16),1)
set.seed(12)
data3<-round(rexp(15,1/16),1)
tiempo1<-c(data1,data2,data3)
tiempo2<-c(data1-median(data1),data2-median(data2),data3-
median(data3))
```

```
marca<-rep(1:3,c(15,15,15))
library(kSamples)
ad.test(tiempo1~as.factor(marca))
ad.test(tiempo2~as.factor(marca))
```

```
 Anderson-Darling k-sample test.
 Number of samples:  3
 Sample sizes:  15, 15, 15
 Number of ties: 4

 Mean of  Anderson-Darling  Criterion: 2
 Standard deviation of  Anderson-Darling  Criterion: 1.02165

 T.AD = ( Anderson-Darling  Criterion - mean)/sigma

 Null Hypothesis: All samples come from a common population.

               AD   T.AD  asympt. P-value
 version 1: 2.146 0.1432           0.3402
 version 2: 2.180 0.1751           0.3292
```