

## CAPITULO III

## DISEÑO COMPLETAMENTE AL AZAR : DCA

Es un diseño en el cual los tratamientos son asignados completamente al azar a las unidades experimentales o viceversa. Este diseño es usado ampliamente.

## Características principales

1. Aplicable sólo cuando las unidades experimentales son homogéneas (verificar si existe tal homogeneidad).
2. Los tratamientos pueden tener igual o diferente número de unidades experimentales.
3. La distribución de los tratamientos es al azar en las unidades experimentales.

El número de tratamientos esta en función del número de unidades experimentales que se dispone. Es conveniente tener pocos tratamientos y más unidades experimentales que muchos tratamientos con pocas unidades experimentales.

## DCA CON UNA OBSERVACIÓN POR UNIDAD EXPERIMENTAL.

Cada tratamiento ( $i=1, \dots, t$ ), dispone de un número de unidades igual a  $r_i$ , cada unidad experimental es una repetición y el valor observado en la u.e. es  $Y_{ij}$ .

Cada observación es expresado en términos de una ecuación según el modelo estadístico:

$$Y_{ij} = \mu + \tau_i + e_{ij}$$

Corresponde al modelo del diseño experimental y es una expresión aditivo lineal del valor observado  $Y_{ij}$  como la suma de tres elementos:

$\mu$  = Factor constante ( parámetro).

$\tau_i$  = Efecto del tratamiento (parámetro) en la unidad experimental.

$e_{ij}$  = Error, valor de la variable aleatoria Error experimental.

Cada observación es expresado de la misma forma; el conjunto de ecuaciones constituye el modelo, son "r." ecuaciones.

El modelo lineal general para el diseño completamente al azar es entonces:

$$Y_{ij} = \mu + \tau_i + e_{ij} \quad \begin{array}{l} i=1,2,\dots, t \\ j=1,2,\dots,r_i \end{array}$$

La descripción de este modelo se complementa con los supuestos de los efectos de tratamiento según:

1)  $\sum r_i \tau_i = 0$ .

2) Los  $E_{ij}$  están distribuidos normal e independiente con media cero y varianza  $\sigma^2_t$ .

Cuando el Investigador fija los tratamientos en estudio y se interesa en los resultados de estos tratamientos, el modelo se denomina de EFECTOS FIJOS, conocido como modelo I. Las hipótesis son:

$$H_p : \tau_i = 0$$

$$H_a : \tau_i \neq 0$$

Si el Investigador está interesado en una población de tratamientos y para el estudio se extrae una muestra de tratamientos, entonces se denomina EFECTOS AL AZAR, conocido como modelo II. Las hipótesis son:

$$H_p: \sigma_t^2 = 0$$

$$H_a: \sigma_t^2 \neq 0$$

Las conclusiones de dan en términos del modelo, En el modelo I puede llegar a encontrar un ranking de significación entre las medias de tratamientos, en el modelo II no es posible y el análisis termina en el Análisis de variancia.

### Estimación de parámetros del modelo DCA

Para el modelo DCA, el número de parámetros a estimar es igual a (t+1); "t" tratamientos y  $\mu$  el parámetro constante en todo el modelo.

Una forma de obtener estimados a estos parámetros, es mediante el método de mínimos cuadrados del error, que consiste en determinar los valores de los  $\tau_i$  y de  $\mu$  que minimicen la suma de los cuadrados del error, es decir:

$$\min \sum \sum e_{ij}^2$$

Para ello se utiliza el modelo planteado y se despeja  $e_{ij}$

$$e_{ij} = Y_{ij} - \mu - \tau_i$$

Luego se obtiene la expresión cuadrática, elevando al cuadrado y sumando todos los elementos:

$$\sum \sum e_{ij}^2 = \sum \sum (Y_{ij} - \mu - \tau_i)^2$$

Para obtener el valor mínimo de esta función, dependiente de  $\mu$  y  $\tau_i$ , se deriva respecto a cada uno y la ecuación se iguala a cero. El resultado es un sistema de ecuaciones simultáneas con (t+1) ecuaciones y (t+1) incógnitas:

$$\mu: Y_{..} - \hat{\mu} \sum r_i - \sum r_i \hat{\tau}_i = 0$$

$$\tau_i: Y_{i.} - r_i \hat{\mu} - r_i \hat{\tau}_i = 0 ; i=1,2,\dots,t$$

La primera ecuación es dependiente de las otras ecuaciones, esta es igual a la suma de las demás ecuaciones, por lo tanto, el sistema es consistente, tiene solución. Son "t" ecuaciones linealmente independientes con (t+1) incógnitas, por lo tanto se tiene múltiples soluciones. Para lograr una única solución, se agrega la siguiente restricción :

$$\sum r_i \hat{\tau}_i = 0$$

Así, la solución del sistema de ecuaciones es:

$$\hat{\mu} = \bar{Y}_{..}$$

$$\hat{\tau}_i = \bar{Y}_{i.} - \bar{Y}_{..} \quad \text{para } i=1,2,\dots,t$$

Ejemplo: Experimento en camote (variedad costanero: Tacna) se estudio el efecto de dos virus (Spmv y Spcsv). Los tratamientos fueron los siguiente: T1 = CC (Spcsv) =Enanismo clorotico del camote, T2 = FF (Spmv) = Moteado plumoso, T3 = FC (Spmv y Spcsv) = Complejo viral y T4 =OO (testigo) plantas sanas. En cada parcela se sembro 50 plantas de camote, se utilizaron 12 parcelas, cada tratamiento con 3 repeticiones; al final del experimento se evaluo el peso total en kilos. La trasmision de virus se hizo en los esquejes y estos se sembraron en el campo.

Datos:

Peso total en Kgrs. de raíces de 50 plantas de comote variedad Costanero

CC	FC	FF	OO		
28.5	14.9	41.8	38.2		
21.7	10.6	39.2	40.4		
23	13.1	28	32.1		
73.2	38.6	109	110.7	331.5	
24.4	12.86667	36.33333	36.9	27.625	

FC= SPUD = sweetpotato virus disease = SPFMV + SPCSV

FF= SPFMV = sweetpotato feathery mottle virus

CC= SPCSV = sweetpotato chlorotic stunt virus

OO=Sano

Bajo el modelo DCA, los estimadores son:

$$\hat{\mu} = 27.625$$

$$\hat{\tau}_1 = 24.4 - 27.625 = -3.225$$

$$\hat{\tau}_2 = 12.86667 - 27.625 = -14.7583$$

$$\hat{\tau}_3 = 36.33333 - 27.625 = 8.708333$$

$$\hat{\tau}_4 = 36.9 - 27.625 = 9.275$$

Así también,  $\sum r_i \hat{\tau}_i = 3(-3.225 - 14.7583 + 8.708333 + 9.275) = 0$

Cada observación del experimento puede ser expresado como:

$$Y_{12} = \hat{\mu} + \hat{\tau}_i + e_{12}$$

$$Y_{12} = 21.7 = 27.625 - 3.225 + e, \text{ Entonces } e = -2.7$$

Los errores experimentales de cada unidad experimental son:

4.1	2.0	5.5	1.3
-2.7	-2.3	2.9	3.5
-1.4	0.2	-8.3	-4.8

Ejercicio. Hallar los errores de cada unidad experimental, la variancia de los errores y la variancia común como un promedio ponderado de las cuatro variancias que es igual a la variancia del error, el valor estimado de  $\mu$  y el coeficiente de variación del experimento:

Resp. CV = 17.16%  
Estimado de  $\mu = 27.625$

## ANÁLISIS DE LA VARIANCIA

Es la descomposición de la variación total (suma de cuadrados total) en cada una de las fuentes controlables y no controlables.

Con esa descomposición se realiza el análisis de los factores mediante las pruebas de hipótesis de las fuentes de variación (controlables).

Supuestos para la validez del ANVA.

Para que el análisis de la variancia sea considerado válido, debe cumplir ciertos supuestos:

Independencia de los errores .- Los  $e_{ij}$  errores definidos por el modelo, además de ser aleatorios y distribuidos normalmente, estos deberán ser independiente estadísticamente, de sospecharse que no hay independencia, debe realizarse la prueba de CICLOS (prueba de ESTADÍSTICA NO PARAMÉTRICA)

Homogeneidad de variancia .- Las variancias de cada grupo deben ser homogéneas. Para probar el supuesto se realiza una prueba de homogeneidad de variancia, el TEST de BARTLETT. El termino de homogeneidad es sinonimo de HOMOSCEDASTICIDAD (proviene de la raíz griega, que quiere decir igualdad de dispersión)

Normalidad de los errores y la variable de estudio.- El error experimental es una variable aleatoria y se asume para las pruebas estadísticas, que esta tiene una distribución normal. En alguna situación que no satisface el requerimiento, se recomienda una transformación de la variable observada como un cambio de escala. Así por ejemplo:

Transformación logarítmica cuando las variancias son heterogéneas; valores pequeños frente a valores muy grandes.

$$L_{ij} = \text{Log}(Y_{ij})$$

Transformación mediante la raíz cuadrada, aplicada para variables discretas (contadas).

$$R_{ij} = \sqrt{y_{ij} + \frac{1}{2}} \text{ para valores enteros menos de 10}$$

$$R_{ij} = \sqrt{y_{ij}} \text{ para valores enteros variados}$$

Transformación arco seno.- Conocida como transformación angular, apropiada para porcentajes ó proporciones

$$A_{ij} = \text{ArcSin}(\sqrt{Y_{ij}})$$

En este ultimo caso, el rango debe estar comprendido entre 0 y 1 o 0 y 100% según el caso. Si los valores estan en el primer cuartil o en el tercer cuartil, es recomendable solo transformación de la raíz cuadrada, en caso de que se encuentre solo en el intervalo del primer cuartil y el tercer cuartil, no es recomendable hacer ninguna transformación.

Aditividad.- esta referida la aditividad de los parametros del modelo estadístico. La prueba para adecuada para este fin es el Test de Tukey de no aditividad.

El cumplimiento de estos supuestos es para la validez del análisis de variancia. En algunos experimentos posiblemente no satisfacen los supuestos, en esos casos se debe aplicar alguna transformación, si aún continua el problema, no insistir con transformaciones, es recomendable las pruebas no paramétricas, aunque son menos eficientes, permite conocer las tendencias en las comparaciones de tratamientos.

## DESCOMPOSICIÓN DE LA SUMA DE CUADRADOS DEL TOTAL:

Del modelo estadístico, reemplazado por sus estimadores resulta:

$$Y_{ij} = \bar{Y}_{..} + (\bar{Y}_{i.} - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_{i.})$$

La Suma de cuadrados del total SC(total) es igual a:

$$\sum \sum (Y_{ij} - \bar{Y}_{..})^2$$

En el modelo equivale a:

$$\sum \sum (Y_{ij} - \bar{Y}_{..})^2 = \sum \tau_i^2 + \sum \sum e_{ij}^2 + 2 \sum \sum \tau_i e_{ij}$$

$$\sum \sum (Y_{ij} - \bar{Y}_{..})^2 = \sum r_i \hat{\tau}_i^2 + \sum \sum e_{ij}^2 + 0$$

$$\sum \sum (Y_{ij} - \bar{Y}_{..})^2 = \sum t_i \hat{\tau}_i^2 + \sum \sum e_{ij}^2$$

$$SC(\text{total}) = SC(\text{trat}) + SC(\text{error})$$

## Ejercicios

Comprobar las siguientes igualdades:

$$\sum \sum \hat{\tau}_i e_{ij} = 0$$

$$\sum \sum \hat{\tau}_i^2 = \sum r_i \hat{\tau}_i^2 = \sum \frac{Y_{i.}^2}{r_i} - \frac{Y_{..}^2}{r}$$

$$\sum \sum e_{ij}^2 = \sum \sum (Y_{ij} - \bar{Y}_{i.})^2 = \sum Y_{ij}^2 - \sum \frac{Y_{i.}^2}{r_i}$$

$$SC(\text{error}) = SC(\text{total}) - SC(\text{trat})$$

Nota: el valor  $\frac{Y_{..}^2}{r}$  es llamado termino de corrección de las sumas de cuadrados, se expresa como T.C.

## CUADRO DEL ANVA Y LA PRUEBA DE F

Las sumas de cuadrados son respuestas aleatorias y su distribución es una ji cuadrada con grados de libertad para tratamientos igual a (t-1) y para el error  $\sum (r_i - 1)$ . La razón de dos ji cuadradas resulta una F con (t-1),  $\sum (r_i - 1)$ .

El cuadro del ANVA es un resumen en el que se expresa todo o parte del análisis del experimento:

Fuentes de Variación (F.V.)	Grados de Libertad (G.L.)	Suma de Cuadrados (S.C.)	Cuadrado Medio (C.M.)	F
Tratamientos	t-1	$\sum r_i \hat{\tau}_i^2$	$\frac{\sum r_i \hat{\tau}_i^2}{t-1}$	$\frac{CM(trat)}{CM(Error)}$
Error	$\sum(r_i-1)$	$\sum \sum e_{ij}^2$	$\frac{\sum \sum e_{ij}^2}{r_i - t}$	
Total	$\sum r_i - 1$	$\sum \sum Y_{ij}^2 - \frac{Y_{..}^2}{r_i}$		

$$CV = \frac{\sqrt{CM_{Error}}}{\bar{Y}_{..}} 100\%$$

Ejemplo. Construir el cuadro del ANVA, para el experimento de camote.

Así también,  $\sum r_i \hat{\tau}_i =$

$$SC(trat) = 3(-3.225)^2 + (-4.7583)^2 + (8.708333)^2 + (9.275)^2 = 1170.2092$$

$$SC(error) = (4.1)^2 + (-2.7)^2 + (-1.4)^2 + \dots + (3.5)^2 + (-4.8)^2 = 179.9133$$

(F.V.)	(Gl.)	(SC.)	(CM.)	F
Virus	3	1170.2092	390.0697	17.34
Error	8	179.9133	22.4892	
Total	11	1350.123		

$$CV = \frac{\sqrt{22.4892}}{27.625} 100\% = 17.2$$

Conclusión : El valor 17.34 supera al valor tabular  $F(3,8, \alpha=0.01) = 7.59$  por lo tanto existen diferencias altamente significativas en el rendimiento del camote por efecto de los virus.

Ejercicio. Calcular las sumas de cuadrados del ejemplo presentado en termino de las identidades expuestas.

Esperados Cuadrados Medios.

Viene a ser la esperanza matemática (promedio) de los cuadrados medios del ANVA. Permiten definir las relaciones de los cuadrados medios para las pruebas de hipótesis. Los Esperados Cuadrados cambian de un modelo I a un modelo II, así:

	MODELO FIJO (I)	MODELO AL AZAR (II)
Tratamientos	$\sigma_{error}^2 + \frac{\sum r_i \tau_i^2}{t-1}$	$\sigma_{error}^2 + r_0 \sigma_{\tau}^2$
Error Experimental	$\sigma_{error}^2$	$\sigma_{error}^2$

Si la  $H_0$  es cierta, la relación  $CM(\text{trat}) / CM(\text{error})$  se distribuye como F con los grados de libertad correspondiente a cada fuente de variación

Comparación de tratamientos seleccionados.

Consiste en probar un grupo de tratamientos versus otro grupo. Así por ejemplo:

Testigo vs los demás.  
comparación entre grupos

Las posibles formas de comparar depende del número de tratamientos.

Para 4 tratamientos, de los cuales el cuarto es el testigo, los posibles grupos a comparar son:

- a) T4 vs T1, T2, T3  
T1, T2, T3
- b) T4 vs T1, T2, T3  
T3 vs T1, T2  
T1 vs T2

Son muchas las combinaciones posibles, pero en cada clasificación las comparaciones esta representada por contrastes que son ortogonales o la comparación dentro de un grupo. El grupo (a) esta formado por un tratamiento vs los demás y la otra comparación entre los demás, en el caso (b) son comparaciones definidos por contrastes ortogonales.

Para (b), los contrastes son C1, C2 y C3. así:

	T1	T2	T3	T4
C1	-1	-1	-1	3
C2	-1	-1	2	0
C3	-1	1	0	0

Cada uno constituye un vector, el producto escalar  $C_i$  vs  $C_k$  es igual a cero. para "i" diferente de "k".

$$C1 \cdot C2 = (-1)(2) + (-1)(-1) + (-1)(-1) + (3)(0) = 0$$

El número de contrastes es igual a (t-1), y cada contraste tiene un grado de libertad.

Los elementos de los contrastes son  $c_{ij}$ ; así el contraste

$$C1 = (c_{11}, c_{12}, c_{13}, c_{14})$$

$$C1 = (-1, -1, -1, 3)$$

La suma de cuadrados de un contraste es igual a:

$$\frac{(\sum c_{ij} Y_{i.})^2}{\sum rc_{ij}^2} ; r_1 = r_2 = \dots = r_t = r$$

Aplicación: Determinar los contrastes para el ejemplo de aplicación de virus en camote y determine las sumas de cuadrados.

Los tratamientos son CC, FC, FF y Testigo

Contraste 1 : testigo vs CC, FC y FF

Contraste 2 : FFvs CC, FC

Contraste 3 : CC vs FC

Cálculo de las sumas de cuadrados:

Contrates	CC	FC	FF	TESTIGO	$\sum c_{ij} y_{ij}$	$\sum rc_{ij}^2$	SC
C1	-1	-1	-1	3	111.3	36	344.1025
C2	-1	-1	2	0	106.2	18	626.58
C3	-1	1	0	0	-34.6	6	199.5267
Yi.	73.2	38.6	109	110.7			1170.209

Estos resultados se presentan en un cuadro del ANVA:

(F.V.)	(Gl.)	(SC.)	(CM.)	F
Virus	3	1170.209	390.0697	17.34
C1	1	344.1025	344.1025	15.30
C2	1	626.58	626.58	27.86
C3	1	199.5267	199.5267	8.87
ERROR	8	179.9133	22.4892	

El cuadro del ANVA para el caso (a) resulta sólo dos fuentes

Testigo vs (CC, FC, FF) y entre (CC,FC, FF)

Las sumas de cuadrados de estas dos fuentes, la primera fuente es calculada por contrastes y la segunda por diferencia respecto a la suma de cuadrados de tratamientos.

El ANVA para el caso de virus de camote es el siguiente:

(F.V.)	(Gl.)	(SC.)	(CM.)	F
Virus	3	1170.209	390.0697	17.34
Testigo Vs. (CC,FC,FF)	1	344.1025	344.1025	15.30
CC,FC,FF	2	826.1067	413.0533	18.36
ERROR	8	179.9133	22.4892	

Ejercicio. Realizar la prueba estadística para el siguiente contraste:

C1 = Virus "CC", "FF" vs "FC"

C2 = Virus "CC" vs. "FF".



## COMPARACION MULTIPLE ENTRE TRATAMIENTOS

Para comparar dos promedios se usa la prueba de T-student, sin embargo, en un experimento pueden haber muchos tratamientos, y si se emplea la prueba de "t", habría tantas pruebas de "t" como pares a comparar de tratamientos, así, si son "t" tratamientos, se requieren  $t(t-1)/2$  comparaciones. La comparación múltiple puede facilitarse con una DIFERENCIA MINIMA DE SIGNIFICACION (D.M.S), en Ingles se define como LSD (least significant difference)

$$DMS = T(\alpha, gl = n.-t) Sd$$

Este valor representa el valor crítico; para que una diferencia sea significativa, la diferencia absoluta de los promedios debe ser mayor al valor crítico DMS.

Sd : Desviación estandar de la diferencia de promedios, definido por:

$$S_d = \sqrt{CMerror\left(\frac{1}{r_1} + \frac{1}{r_2}\right)}$$

$T(\alpha, gl(error\ exp))$  es el valor de "t" tabular con grados de libertad del error experimental.

Aplicación: Caso "Virus en Camote"

Testigo = 36.9  
 FF = 36.3  
 CC = 24.4  
 FC = 12.8

$\alpha = 0.05$   $gl=8$   $CM(error)=22.48917$   $n=3$   $T\alpha = 2.31$   
 $Sd = 3.87$   $LSD=8,93$

Comparación	d	DLS	Significacion
Testigo vs FF	0.3	8.93	Ns
Testigo vs CC	12.5	8.93	*
Testigo vs FC	24.1	8.93	*
FF vs CC	11.9	8.93	*
FF vs FC	23.5	8.93	*
CC vs FC	11.6	8.93	*

Testigo a |  
 FF a |  
 CC b |  
 FC c |

Uso del programa R para la comparación múltiple:

Datos creados con el block de notas "notepad" de windows.

Archivo: camote.txt

```

virus    rdto
CC       28.5
CC       21.7
CC       23
FC       14.9
FC       10.6
FC       13.1
FF       41.8
FF       39.2
FF       28
OO       38.2
OO       40.4
OO       32.1

```

Escriba las siguientes instrucciones en la consola de R.

Asegurese que se encuentra en el directorio de trabajo donde grabo el archivo de datos.

Procedimiento para leer los datos y hallar el análisis de variancia.

```

> datos<-read.table("camote.txt", header=TRUE)
> modelo<-aov(rdto ~ virus, data=datos)
> anova(modelo)

```

Analysis of Variance Table

Response: rdto

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
virus	3	1170.21	390.07	17.345	0.0007334 ***
Residuals	8	179.91	22.49		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Llamar a los paquetes para otras pruebas.

```

> library(agricolae)
> cv.model(modelo)

```

```
[1] 17.16660
```

```

> gl <- df.residual(modelo)
> cm <- deviance(modelo)/gl

```

```

> attach(datos)
> comparacion <- LSD.test(rdto, virus, gl, cm, alpha=0.05,group=FALSE)

```

Study:

LSD t Test for rdto

```

          ..*..*..*
Alpha                0.050000
Error Degrees of Freedom 8.000000
Error Mean Square      22.489167
Critical Value of t    2.306004

```

Treatment Means

	virus	rdto	std.err	replication
1	cc	24.40000	2.084067	3
2	fc	12.86667	1.246774	3

```
3    ff 36.33333 4.233727      3
4    oo 36.90000 2.482606      3
```

Comparison between treatments means

```
tr.i tr.j      diff pvalue
1    1    2 11.5333333 0.0176
2    1    3 11.9333333 0.0150
3    1    4 12.5000000 0.0120
4    2    3 23.4666667 0.0004
5    2    4 24.0333333 0.0002
6    3    4  0.5666667 0.8872
```

Agrupamiento de tratamientos

```
> comparacion <- LSD.test(rdto, virus, gl, cm, alpha=0.05, group=TRUE)
```

Study:

LSD t Test for rdto

```
Alpha          0.050000
Error Degrees of Freedom 8.000000
Error Mean Square 22.489167
Critical Value of t 2.306004
```

Treatment Means

```
virus      rdto  std.err replication
1    cc 24.40000 2.084067      3
2    fc 12.86667 1.246774      3
3    ff 36.33333 4.233727      3
4    oo 36.90000 2.482606      3
```

Least Significant Difference 8.928965

Means with the same letter are not significantly different.

Groups, Treatments and means

```
a    oo    36.9
a    ff    36.33333
b    cc    24.4
c    fc    12.86667
```

El tratamiento testigo es "00"

Compara sus resultados.

## METODO DE TUKEY

Tambien conocido como Tukey's HSD o metodo de Tukey-Kramer.  
Las siglas HSD se derivan del ingles Honestly Significant Difference

No requiere la prueba de F. Los valores son más altos que LSD de T-student, la prueba es más rigurosa, porque requiere que la diferencia sea más grande para alcanzar una significación.

Se utiliza un solo valor para todas las comparaciones. La conclusión esta referida al experimento en si y no a la comparación como es el caso de DLS. Si el nivel de significación  $\alpha$  es 0.05, significa que de 100 experimentos similares en 95 de ellos se espera tener las mismas conclusiones y en 5 experimentos conclusiones diferentes; es por esto que en TUKEY necesita un solo valor y esto hace suponer igual

repeticiones por tratamiento, aunque es posible con diferente repetición, para este caso se tendrá diferentes valores de la desviación estándar del promedio, pero el valor de la tabla seguirá siendo un solo valor.

Este método requiere el cálculo del error estándar de promedios  $S_{\bar{X}}$

$$S_{\bar{X}} = \sqrt{\frac{CMerror}{r}};$$

r=Nro de observaciones por cada promedio.

Con la tabla de TUKEY "Amplitudes Estudiantizadas Significativas prueba de Student-Newman-Keul" AES. y el error estándar de promedios, se halla el valor de la diferencia significativa, es el valor crítico para la prueba (HSD).

$$HSD = AES S_{\bar{X}}$$

El valor AES depende del nivel  $\alpha$ , los grados de libertad del error y el número de tratamientos del experimento.

Para el caso "Virus en camote ", con  $\alpha=0.05$ , la prueba de Tuckey da el siguiente resultado:

Testigo	= 36.9
FF	= 36.3
CC	= 24.4
FC	= 12.8

$$S_{\bar{X}} = 2.7379536$$

$$AES = 4.529$$

$$HSD = 4.529(2.7379536)=12.4$$

Testigo	FF	CC	FC
-----			
	-----		
		-----	

Tambien,

Testigo	a			
FF	a b			
CC	b c			
FC	c			

Mediante el programa "R" se puede obtener estos resultados.

Si realizo la prueba anterior, Ud. Puede continuar con las siguientes instrucciones para obtener los mismos resultados.

```
> attach(datos)
> comparacion <- HSD.test(rdto, virus, gl, cm, alpha=0.05, group=FALSE)
```

Study:

HSD Test for rdto

```

          .....
Alpha                0.05000
Error Degrees of Freedom  8.00000
Error Mean Square      22.48917
Critical Value of Studentized Range  4.52881

```

Treatment Means

	virus	rdto	std.err	replication
1	cc	24.40000	2.084067	3
2	fc	12.86667	1.246774	3
3	ff	36.33333	4.233727	3
4	oo	36.90000	2.482606	3

Comparison between treatments means

	tr.i	tr.j	diff	pvalue
1	1	2	11.5333333	0.0686
2	1	3	11.9333333	0.0592
3	1	4	12.5000000	0.0482
4	2	3	23.4666667	0.0014
5	2	4	24.0333333	0.0012
6	3	4	0.5666667	0.9988

Agrupamiento de tratamientos

```
> comparacion <- HSD.test(rdto, virus, gl, cm, alpha=0.05, group=TRUE)
```

Study:

HSD Test for rdto

```

          .....
Alpha                0.05000
Error Degrees of Freedom  8.00000
Error Mean Square      22.48917
Critical Value of Studentized Range  4.52881

```

Treatment Means

	virus	rdto	std.err	replication
1	cc	24.40000	2.084067	3
2	fc	12.86667	1.246774	3
3	ff	36.33333	4.233727	3
4	oo	36.90000	2.482606	3

Honestly Significant Difference 12.39967

Means with the same letter are not significantly different.

Groups, Treatments and means

a	oo	36.9
ab	ff	36.33333
bc	cc	24.4
c	fc	12.86667

Compara con los resultados manuales.

**METODO DE DUNNETT.**

Se utiliza cuando se tiene un tratamiento control (testigo). El método consiste en realizar comparaciones de cada tratamiento con el control.

Requiere un solo valor de la tabla para la prueba: Tabla de Dunnett

El valor de la tabla se busca con grados de libertad del error y el numero de tratamientos sin incluir el testigo.

Se calcula la desviación estándar de la diferencia ( Usado en la prueba DMS de t-student)

Se determina el valor critico de comparación de Dunnett.

$$D' = t(\text{Dunnett}) \times Sd$$

Las diferencias de promedios entre el control y el tratamiento si superan el valor  $D'$ , serán significativamente diferentes.

Nota, se puede utilizar la tabla de una o dos colas, según la hipótesis que formule el investigador.

Ejemplo:

Aplicación: Caso "Virus en Camote"

Testigo = 36.9  
 FF = 36.3  
 CC = 24.4  
 FC = 12.8

$\alpha = 0.05$   $gl=8$   $CM(\text{error})= 22.48917$   $p=3$  tratamientos  $T_{\alpha}$  (2 colas) = 2.88  
 $Sd = 3.87$   $D'=\text{Diferencia Mínima de Significación} = 11.15$

Comparación	d	D'	Significancia
Testigo vs FF	-0.567	11.15	Ns
Testigo vs CC	-12.500	11.15	*
Testigo vs FC	-24.033	11.15	*

Para la comparacion se observa el valor absoluto del valor de "d". En este caso 12.5 y 24.03 son valores mayores que 11.15; por lo tanto la diferencia es significativa.

La prueba de Dunnett puede realizar con el programa "R", necesita el paquete "multcomp"

Si realizo la lectura, puede continuar con:

```
> library(multcomp)
> library(mvtnorm)
> contraste <- rbind(c(1,0,0,-1),c(0,1,0,-1),c(0,0,1,-1))
> filas<-c("CC vs OO","FC vs OO","FF vs OO")
> columnas<-c("CC","FC","FF","OO")
> dimnames(contraste)<-list(filas,columnas)
> compara<-simint(rdto~virus, data=datos, cmatrix=contraste, type="Dunnett")
> summary(compara)
```

Simultaneous 95% confidence intervals: user-defined contrasts

Call:

```
simint.formula(formula = rdto ~ virus, data = datos, cmatrix = contraste,
  type = "Dunnett")
```

user-defined contrasts for factor virus

Contrast matrix:

```

      CC FC FF OO
CC vs OO 0  1  0  0 -1
FC vs OO 0  0  1  0 -1
FF vs OO 0  0  0  1 -1

```

Absolute Error Tolerance: 0.001

95 % quantile: 2.88

Coefficients:

	Estimate	low CI,	upp CI	t value	Std.Err.	p raw	p Bonf	p adj
CC vs OO	-12.500	-23.650	-1.350	-3.228	3.872	0.012	0.036	0.030
FC vs OO	-24.033	-35.184	-12.883	-6.207	3.872	0.000	0.001	0.001
FF vs OO	-0.567	-11.717	10.584	-0.146	3.872	0.887	1.000	0.998

Interpretación de resultados:

El valor t-tabular a un riesgo de 0.05 es 2.88  
La desviación estandar de la diferencia es: 3.872

El valor crítico sería:  $2.88 \times 3.872 = 11.15$

Los valores de la diferencia esta dado por el estimado de la comparación.

En este analisis, ademas, se reporta los limites de confianza de la diferencia, el t-calculado de la diferencia, la probabilidad de rechazo por "t" y por el metodo de "Bonferroni" y el nivel de riesgo ajustado.

Los valores de p-value por la prueba de "t" estan expresado como "p.raw" y puede ser calculado como:

```

> 2*pt(-3.228, 8)
[1] 0.0120932
> 2*pt(-6.207, 8)
[1] 0.0002574577
> 2*pt(-0.146, 8)
[1] 0.8875332

```

como se aprecia en el cuadro.

El valor de p-value se utiliza para el analisis comparativo

p.raw = 0.012 es significativo  
p.raw = 0.00 es altamente significativo  
p.raw = 0.88 es no significativo.

Nota. Para formar la matriz de contraste, es necesario saber como ordena los tratamientos el programa "R", esto lo puede averiguar con la siguiente instrucción:

```

> unique(datos$virus)

[1] CC FC FF OO
Levels: CC FC FF OO
>

```

entonces, el testigo esta en el cuarto lugar, la matriz de contraste sera:

```

> contraste
      CC FC FF OO

```

CC vs OO 1 0 0 -1  
 FC vs OO 0 1 0 -1  
 FF vs OO 0 0 1 -1

### Prueba de medias con número desigual de repeticiones.

Para la prueba de DLS(t-student), el valor de  $S_d$  (desviación estándar de la diferencia) debe ser calculado por:

$$S_d = \sqrt{CMerror\left(\frac{1}{r_1} + \frac{1}{r_2}\right)}$$

Para las pruebas que requieren la desviación estándar de promedio y no se tiene igual repetición, calcular un valor que aproxime al número de repeticiones, utilizando el promedio armónico, por ejemplo:

$r_1=2, r_2=4, r_3=3, r_4 = 3$ , el promedio armónico es:

$r' = 4/(1/2+1/4+1/3+1/3) = 2.8$ , con este valor hallar la desviación estándar de promedio.

$$S_{\bar{X}} = \sqrt{\frac{CMerror}{r'}}$$

En el caso de contrastes para igual o diferente repetición, puede utilizar el siguiente procedimiento:

Si  $Y_i$  es el total del tratamiento con  $r_i$  repeticiones, un contraste utilizando los totales sería:

$$Q = \sum c_i Y_i = \sum r_i c_i \bar{y}_i \quad \text{Con } \sum r_i c_i = 0$$

Entonces :

$$E[Q] = \sum r_i c_i \mu_i \quad \text{y la variancia } V[Q] = \sigma^2 \sum r_i c_i^2$$

Así,

Para probar la hipótesis  $H_p : \sum r_i c_i \mu_i = 0$

Se utiliza la distribución de t-student, con grados de libertad del error experimental, y el valor :

$$T - \text{calculado} = \frac{Q}{\sqrt{CMerror(\sum r_i c_i^2)}}$$

si el valor calculado es superior al tabular, entonces se rechaza la Hipótesis planteada  $H_p$ . Es decir que el contraste es diferente de cero.

Nota. Debe tener presente que los coeficientes de los contrastes deben ser de tal forma que  $\sum r_i c_i = 0$

Por ejemplo:

Comparación	A	B	C	D
Repeticiones	2	4	3	3
A Vs B, C, D	-5	1	1	1



B vs C,D	0	3	-2	-2
C vs D	0	0	1	-1

Satisface  $\sum r_i c_i = 0$ . La hipótesis planteada: A diferente de (B, C, D)

Por ejemplo: Considere los siguiente datos de un DCA con desigual numero de repeticiones

	A	B	C	D	
Datos	3.7	4	4.7	6.4	
	3.2	4.4	4.3	7.5	
		5.7	5.9	8.4	
		4.7			
Totales	6.9	18.8	14.9	22.3	62.9
Repeticiones	2	4	3	3	

El ANVA para estos datos es el siguiente:

	GL	SC	CM	Fc
Tratamientos	3	22.23083	7.410278	11.62777
SC(T vs Otros)	1	7.704167	7.704167	12.08892
Error	8	5.098333	0.637292	
Total	11	27.32917		

La SC(Testigo vs otros) es igual a:

$$\frac{6.9^2}{2} + \frac{(18.8+14.9+22.3)^2}{10} - \frac{62.9^2}{12} = 7.704167$$

Utilizando contrastes:

	Testigo	A	B	C
A vs B,C,D	-5	1	1	1
B vc C, D	0	3	-2	-2
C vs D	0	0	1	-1
Totales	6.9	18.8	14.9	22.3
Repeticiones	2	4	3	3

Para el primer contraste A vs B, C, D

$$Q = \sum c_i y_i = 21.5, \text{ tambien, } SC(\text{contraste}) = \frac{Q^2}{\sum r_i c_i^2} = \frac{426.25}{60} = 7.704167$$

Tcalculado = 3.476912;  $T_{0.05}(8) = 2.3$ . Resultado significativo

La respuesta del contraste en el ANVA y esta prueba, dan los mismos resultados. Este resultado se da porque se cumple la siguiente identidad:  $F = T^2$ , para un grado de libertad.

Efectivamente  $12.08892 = (3.476912)^2$

De igual forma para los valores tabulares, para 0.05:

$$F_{0.05}(1, 8) = T^2_{0.05}(8)$$

$5.317645 = (2.306006)^2$  Significa que las pruebas son equivalentes.

Para los otros contrastes:

$Q^2 = 324$ , el denominador es 60 y la  $SC(\text{contraste}) = 5.4$

$Q^2 = 54.76$ , el denominador es 6 y la  $SC(\text{contraste}) = 9.1266$

La suma de cuadrados de los contrastes es:  $7.7041 + 5.4 + 9.1266 = 22.23$ , que es equivalente a la suma de cuadrados de tratamientos.

Solucion por el programa "R"

1. Crear el archivo "estudio.txt" con el block de notas y almacenar en su folder de trabajo.
2. direccionar su programa "R" al folder de trabajo.

Archivo "estudio.txt"

```
Trat      Y
A         3.7
A         3.2
B         4.0
B         4.4
B         5.7
B         4.7
C         4.7
C         4.3
C         5.9
D         6.4
D         7.5
D         8.4
```

Lectura de los datos:

```
> datos<-read.table("estudio.txt",header=TRUE)
> modelo<-aov(y ~ trat, data=datos)
> anova(modelo)
```

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
trat	3	22.2308	7.4103	11.628	0.002749 **
Residuals	8	5.0983	0.6373		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Aplicar las funciones de agricolae para realizar las comparaciones multiples, compare con sus resultados.

### Prueba de Kruskal-Wallis

Es una prueba noprametrica, apropiada para el diseño completamente al azar. Se utiliza cuando la variable no cumple requisitos para una ANVA normal; por ejemplo normalidad, el CV demasiado grande, la variable es discreta o categorica.

La prueba esta basada en rangos, es decir todos los datos son ordenados de acuerdo a su valor y luego se asigna un valor siguiendo un orden de prioridad y con los datos del rango se realiza la prueba.

Por ejemplo se tiene la siguiente secuencia de numeros: 13, 12, 17, 12, 4, 8, 9, 5 .

Estos numeros se ordenan de menor a mayor:

4, 5, 8, 9, 12, 12, 13, 17

Se estable un numero de orden:

1, 2, 3, 4, 5, 6, 7,8 son 8 datos

Los valores del rango seran:

1, 2, 3, 4, 5.5,5.5, 7,8. El valor 5.5 es el promedio de la posicion de los datos iguales (5+6)/2.

Entonces en el orden de los datos, sus equivalentes son:

13	12	17	12	4	8	9	5
7	5.5	8	5.5	1	3	4	2

Criterio de la prueba cuando no hay rangos empatados o se considere muy pocos que no pueda significar una variacion en la respuesta, se utiliza:

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

N : total de observaciones

ni: repeticiones en el tratamiento i

Ri: es la suma de los rangos del tratamiento i

H se distribuye como una chi-cuadrada con (t-1) grados de libertad.

En el caso de Empates significativos en los rangos, se utiliza el siguiente proceso:

$$H = \frac{1}{S^2} \left( \sum \frac{R_i^2}{n_i} - N \frac{(N+1)^2}{4} \right); S^2 = \frac{1}{N-1} \left( \sum r_{ij}^2 - N \frac{(N+1)^2}{4} \right)$$

r<sub>ij</sub> : Es el rango de un valor observado.

La variable H sigue con la misma distribucion definida.

### Aplicación a los datos de virus en camote.

Con los datos observados, se calcula el rango entre todos los datos (son 12)

	CC	FC	FF	OO	
	7	3	12	9	
	4	1	10	11	
	5	2	6	8	
Ri	16	6	28	28	
ni	3	3	3	3	
Ri <sup>2</sup> / ni	85.3	12.0	261.3	261.3	620.0

$$H = \frac{12}{12 \times 13} 620 - 3 \times 13 = 8.69$$

El valor critico con 3 grados de libertad es 7.81

Se rechaza la hipótesis planteada, es decir hay diferencia en la respuesta observada por efecto del virus, es decir los virus presentan un efecto significativo entre las parcelas tratadas.

Para realizar una múltiple comparación, se utiliza los promedios de los rangos y se compara la diferencia absoluta con el valor crítico dado por:

$$t(0.05, N-t) \sqrt{\left( S^2 \frac{N-1-H}{N-t} \right) \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}$$

$n_i$  y  $n_j$  son las repeticiones de los tratamientos a comparar.

Aplicación al ejemplo propuesto.

	Promedio
CC	5.33
FC	2.00
FF	9.33
OO	9.33

CC vs FC, la diferencia absoluta es: 3.33

H =	8.69
S <sup>2</sup> =	13.0
t <sub>0.05,8</sub>	2.31
Valor.Critico	3.65

Como el valor calculado (3.33) es menor que el crítico (3.65), entonces no hay diferencia significativa entre CC y FC.

En la comparación de los otros tratamientos, CC es diferente de FF y de OO, FC también es diferente de FF y OO, pero no existe diferencia entre FF y OO.

Cuando se tiene igual repetición, el valor crítico no cambia para todas las comparaciones.

```
> library(SuppDists)
> library(agricolae)
> comparacion<-kruskal(rdto,virus,group=TRUE, main="camote")
```

```
Study: camote
Kruskal-Wallis test's
```

```
Value: 8.692308
degrees of freedom: 3
Pvalue chisq : 0.03367430
pKruskalWallis: 0.006523891
```

```
Mean of the ranks
  virus      rdto replication
1  cc 5.333333          3
2  fc 2.000000          3
3  ff 9.333333          3
4  oo 9.333333          3
```

t-Student: 2.306004  
Alpha : 0.05  
LSD : 3.646113

Means with the same letter are not significantly different

Groups, Treatments and mean of the ranks

a	ff	9.333333
a	oo	9.333333
b	cc	5.333333
b	fc	2

>