

Manual de uso del paquete DisFac (Diseño Factorial)

Judith Angel Rivera

INTRODUCCIÓN

Un **diseño factorial** es un diseño experimental que se utiliza para evaluar el efecto producido por la combinación de dos o más factores que actúan de manera simultánea, y para el cual, cada uno de los factores del experimento se estudia a distintos niveles, formando varios tratamientos. Entonces, el número total de tratamientos a evaluar se obtiene de la multiplicación del número de niveles de los distintos factores que intervienen en el experimento. Además, comunmente, a los factores se les denomina con letras mayúsculas (A, B, C, ect), mientras que los niveles de cada factor, se representan con letras minúsculos y con subíndices que dependen del número de niveles de cada factor ($a_1, a_2, a_3, b_1, b_2, b_3$) (Castillo Marquez, 2000).

En base a un Diseño Factorial, realicé el presente paquete en R, que lleva por nombre “**DisFac**”, con este paquete se pretende analizar, el efecto de las interacciones que tienen los factores sobre la variable respuesta

Al crear este paquete, se automatizan procesos como la creación de la Tabla Anova, con lo que solo queda con hacer un análisis en base a los resultados que este arroja, y para que sea de fácil acceso el código fuente se encuentra en la plataforma GitHub, y se puede encontrar con en el siguiente enlace:

<https://github.com/JudithAnRi/DisenioFactorial.git>; y se puede instalar con el siguiente código al ejecutarlo en la consola:

```
devtools::install_github("JudithAnRi/DisenioFactorial")
```

OBJETIVOS

1. Elaborar un paquete en R para obtener una Tabla Anova con los datos obtenidos del diseño Factorial y poder analizarla con facilidad.
2. Dependiendo del caso de estudio, y si se rechaza la Hipótesis Nula, se realiza una prueba DMS (Diferencia Mínima Significativa).
3. Mostrar ejemplos en donde se pueda aplicar la paquetería “DisFac”.

MODELO ESTADÍSTICO Y PRUEBA DE HIPÓTESIS

El modelo estadístico que se maneja para este diseño experimental, es el siguiente:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

donde

$i = 1, 2, \dots, a$ $j = 1, 2, \dots, b$ $k = 1, 2, \dots, r$

- α = Número . de niveles del factor A.

- β = Número de niveles del factor B.
- r = Número de repeticiones.
- y_{ijk} = Respuesta obtenida en la k-ésima repetición del i-ésimo nivel del factor A y el j-ésimo nivel del factor B.
- μ = Efecto medio general.
- α_i = Efecto atribuido al i-ésimo nivel del factor A.
- β_j = Efecto atribuido al j-ésimo nivel del factor B.
- $(\alpha\beta)_{ij}$ = Efecto de la interacción entre el i-ésimo nivel del factor A y del j-ésimo nivel del factor B.
- ϵ_{ijk} = Error, con distribución Normal con media 0 y varianza σ^2 .

Pruebas de Hipótesis

Para un diseño factorial, se toma en cuenta los siguientes juegos de Hipótesis:

1. $H_0: a_1 = a_2$ $H_a: a_1 \neq a_2$
2. $H_0: b_1 = b_2$ $H_a: b_1 \neq b_2$
3. $H_0: a_1b_1 = a_1b_2 = a_2b_1 = a_2b_2$ H_a : Por lo menos una de las interacciones entre los niveles de los factores produce un efecto diferente al de los demás.

Y la decisión se toma en base a:

1. Se rechaza H_0 si $F_{calA} > F_{tabA}$
2. Se rechaza H_0 si $F_{calB} > F_{tabB}$
3. Se rechaza H_0 si $F_{calAB} > F_{tabAB}$

Comparación de Medias por Factor

En caso de llegar a concluir que existe una diferencia entre los niveles de un factor, se realiza una comparación de medias por niveles del factor, la cual, se basa en las siguientes fórmulas:

Para el nivel 1

$$\bar{Y}_{1..} = Y_{1..}/o$$

Para el nivel 2

$$\bar{Y}_{2..} = Y_{2..}/o$$

Prueba DMS o Diferencia Mínima Significativa

La cual va a determinar la mejor combinación de los niveles de los factores, es decir, el mejor tratamiento.

El valor de DMS, se calcula de la siguiente manera:

$$DMS = [t_{\alpha/2}(GLE)]\sqrt{2(CME)/r}$$

Después de obtener el valor de DMS, posteriormente se calculan los Totales por Interacción o Tratamiento

$$(Y_{11\cdot}), (Y_{12\cdot}), (Y_{21\cdot}), (Y_{22\cdot})$$

Después se calculan las Medias cada interacción, es decir, de los tratamientos

$$(\bar{Y}_{11\cdot}), (\bar{Y}_{12\cdot}), (\bar{Y}_{21\cdot}), (\bar{Y}_{22\cdot}),$$

Y luego se obtienen las diferencias entre cada tratamiento como se muestra a continuación:

$$D1 = |\bar{Y}_{11\cdot} - \bar{Y}_{12\cdot}| \quad D2 = |\bar{Y}_{11\cdot} - \bar{Y}_{21\cdot}| \quad D3 = |\bar{Y}_{11\cdot} - \bar{Y}_{22\cdot}| \quad D4 = |\bar{Y}_{12\cdot} - \bar{Y}_{21\cdot}| \quad D5 = |\bar{Y}_{12\cdot} - \bar{Y}_{22\cdot}| \quad D6 = |\bar{Y}_{21\cdot} - \bar{Y}_{22\cdot}|$$

Estas diferencias se ordenan de mayor a menor, se localizan las diferencias que sean mayor al valor DMS, y posteriormente, de estas se busca la Media de la interacción o tratamiento mayor, siendo esta, la mejor interacción o el mejor tratamiento.

EJEMPLOS DE APLICACIÓN

Para ilustrar como se utiliza el paquete DisFac, a continuación se muestran dos ejemplos.

Ejemplo 1

Se quiere probar distintas dietas en cerdos de 100 kg de Peso Vivo durante 15 días, para lo cual, se tienen dos factores, el *factor A* es el **% de Proteína** de la dieta, con dos niveles **a1:16% de proteína** y **a2:17.2% de proteína**; y el *factor B* el cual es el **% de lisina** en la dieta, con dos niveles **b1:0.60% de lisina** y **b2:0.70% de lisina**. Por lo tanto, teniendo dos factores, cada uno con dos niveles, obtenemos un diseño Factorial 2x2, con 4 tratamientos, los cuales son:

- 16% Proteína 0.60% Lisina
- 16% Proteína 0.70% Lisina
- 17.2% Proteína 0.60% Lisina
- 17.2% Proteína 0.70 Lisina

Y la variable respuesta será el promedio de la ganancia diaria de peso que tuvieron los cerdos durante los 15 días.

Y las hipótesis a probar son las siguientes:

1. $H_0: a_1:16\% \text{ de proteína} = a_2:17.2\% \text{ de proteína}$ $H_a: a_1:16\% \text{ de proteína} \neq a_2:17.2\% \text{ de proteína}$
2. $H_0: b_1:0.60\% \text{ de lisina} = b_2:0.70 \text{ de lisina}$ $H_a: b_1:0.60\% \text{ de lisina} \neq b_2:0.70 \text{ de lisina}$
3. $H_0: a_1b_1 = a_1b_2 = a_2b_1 = a_2b_2$ $H_a: \text{Por lo menos una de las interacciones entre los niveles de los factores produce un efecto diferente al de los demás.}$

y queremos saber si ¿existen diferencias entre los niveles de cada factor?, ¿existen diferencias entre las combinaciones de % de proteína y % de lisina en la dieta?, y ¿cuál es la mejor combinación de % de proteína y % de lisina para obtener una mayor ganancia diaria de peso en los cerdos?.

A continuación, explico como se obtendría respuestas a estas preguntas, utilizando la paquetería “DisFac”:

```
##Se fija la ruta de trabajo y se importan los datos
ruta <- "C:/Users/judit/Desktop/Paq R/Manual_DF/Datos"
setwd(ruta)
datos <- read.csv(file = "Datos.csv")
##Los datos se almacenan en un dataframe llamado "df"
df <- data.frame(datos)
print(df)
```

	Factor.A	Factor.B	Tratamiento	Replica	Respuesta
1	16 % Proteína	0.60% lisina	16%P 0.60%l	1	803
2	16 % Proteína	0.60% lisina	16%P 0.60%l	2	795
3	16 % Proteína	0.60% lisina	16%P 0.60%l	3	752
4	16 % Proteína	0.60% lisina	16%P 0.60%l	4	896
5	16 % Proteína	0.60% lisina	16%P 0.60%l	5	856
6	16 % Proteína	0.60% lisina	16%P 0.60%l	6	819
7	16 % Proteína	0.70% lisina	16%P 0.70%l	1	869
8	16 % Proteína	0.70% lisina	16%P 0.70%l	2	906
9	16 % Proteína	0.70% lisina	16%P 0.70%l	3	913
10	16 % Proteína	0.70% lisina	16%P 0.70%l	4	926
11	16 % Proteína	0.70% lisina	16%P 0.70%l	5	902
12	16 % Proteína	0.70% lisina	16%P 0.70%l	6	893
13	17.2 % Proteína	0.60% lisina	17.2%P 0.60%l	1	859
14	17.2 % Proteína	0.60% lisina	17.2%P 0.60%l	2	803
15	17.2 % Proteína	0.60% lisina	17.2%P 0.60%l	3	856
16	17.2 % Proteína	0.60% lisina	17.2%P 0.60%l	4	869
17	17.2 % Proteína	0.60% lisina	17.2%P 0.60%l	5	884
18	17.2 % Proteína	0.60% lisina	17.2%P 0.60%l	6	819
19	17.2 % Proteína	0.70% lisina	17.2%P 0.70%l	1	823
20	17.2 % Proteína	0.70% lisina	17.2%P 0.70%l	2	916
21	17.2 % Proteína	0.70% lisina	17.2%P 0.70%l	3	846
22	17.2 % Proteína	0.70% lisina	17.2%P 0.70%l	4	867
23	17.2 % Proteína	0.70% lisina	17.2%P 0.70%l	5	808
24	17.2 % Proteína	0.70% lisina	17.2%P 0.70%l	6	794

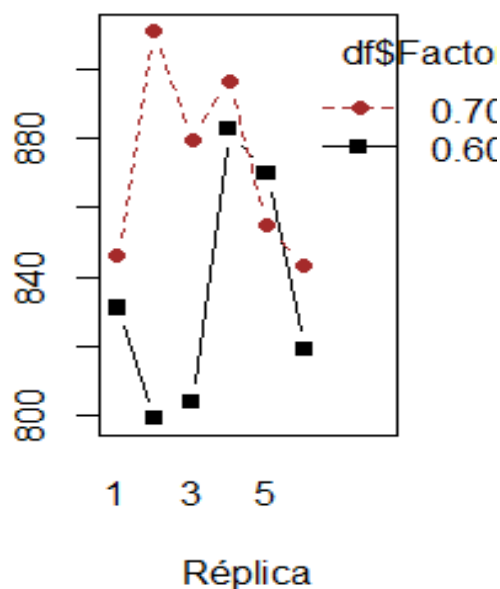
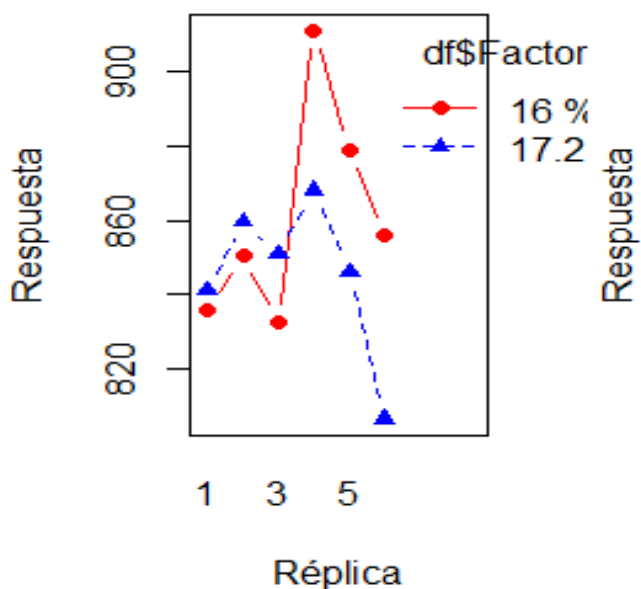
Una vez que se fija la ruta y se importan los datos, se instala la librería del paquete “DisFac”

```
library(DisFac)
```

Una vez cargada la librería del paquete “DisFac”, se llama a la función **AnovaDF**, para poder obtener la tabla Anova, y poder hacer el análisis y concluir si se rechazan las hipótesis Nulas o no. Esta función, recibe cinco argumentos, el primero es *datos*, es

decir, debemos ingresar el nombre del dataframe donde estan almacenados nuestros datos a analizar, que en este caso se llama "df"; el segundo es *a* que nos pide los niveles del Factor A, en este caso, el factor A tiene dos niveles; el tercer argumento es *b*, el cual nos pide los niveles del factor B, que en nuestro ejercicio tiene dos niveles; el cuarto es *r*, es decir, el número de repeticiones, que para el ejercicio es de 6; y finalmente, nos pide el *alfa*, el cual nos pide el valor de significancia que utilizaremos en el ejercicio, y nosotros podemos definir, para nuestro caso, utilizaremos un alfa de 0.05.

```
AnovaDF(df, 2, 2, 6, 0.05)
```



	FuenteVar	GradosLibertad	SumaCuadrados	CuadradosMedios	FCalculada
1	Factor A	1	1441.500	1441.500	0.9867543
2	Factor B	1	8512.667	8512.667	5.8272011
3	Interacción AB	1	11440.667	11440.667	7.8315136
4	Error	20	29217.000	1460.850	
5	Total	23	50611.833		
FTablas					
1	4.351244				
2	4.351244				
3	4.351244				
4					
5					

Con esta tabla Anova, vemos que 1. No se rechaza H_0 ya que $F_{calA} = 0.9867 < F_{tabA} = 4.3512$, por lo tanto, los dos niveles del factor A, es decir, los dos niveles de % de

proteína no tienen diferencias en la ganancia diaria de peso de los cerdos. 2. Si se rechaza H_0 debido a que $F_{calB} = 5.8272 > F_{tabB} = 4.3512$, es decir, los dos niveles del factor B: % de Lisina si tienen diferencias para la ganancia diaria de peso de los cerdos. 3. Se rechaza H_0 puesto que $F_{calAB} = 7.8315 > F_{tabAB} = 4.3512$, por lo tanto, podemos concluir que por lo menos una de las interacciones de los niveles de % de proteína y % de Lisina produce un efecto diferente a los demás en la ganancia diaria de peso de los cerdos.

Una vez concluido que si hay diferencias entre los niveles del Factor B, es decir, hay diferencias entre b_1 : 0.60% de lisina y b_2 : 0.70 de lisina, se calcula comparando las Medias por nivel de Factor, y para esto, se utiliza la función **MediasFactor**, la cual recibe dos argumentos, de los cuales, el primero f , es decir, el factor del que se quieren calcular las Medias y que solo recibe 1 para indicar que se quiere calcular para el Factor A o 2 para Factor B, y en este ejemplo, se utilizará el número 2; y el argumento o , que nos pide el número de observaciones por nivel, es decir, para nuestro ejemplo, se tienen 12 observaciones.

```
MediasFactor(2,12)
```

```
0.60% lisina 0.70% lisina
834.2500    871.9167
```

Al obtener estas dos medias, vemos que la media del nivel 2: 0.70% lisina es mayor, por lo tanto, la dieta con un 0.70% de lisina produce una mayor ganancia diaria de peso.

Pero como el objetivo del Diseño Factorial, es saber, el efecto de las interacciones entre los niveles de los factores, se realiza la Prueba de Diferencia Mínima Significativa, para lo cual, ocuparemos la función **pDms** que recibe cuatro argumentos, el primero α , que en este caso será de 0.05; el segundo r , que es el número de réplicas, y para este ejemplo es de 6; el tercero es glE que nos solicita los grados de libertad del error, que podremos obtener de la tabla anova, y que en este caso es de 20; y el último, CME que nos pide el valor de Cuadrado Medio del Error, que igual obtenemos de la tabla anova, y para este caso es 1460.850.

```
pDms(0.05,6,20, 1460.850)
```

```
[1] 46.03084
```

		Tratamiento	TotalTratamiento	MediaTratamiento	
16%P	0.60%1	16%P	0.60%1	4921	820.1667
16%P	0.70%1	16%P	0.70%1	5409	901.5000
17.2%P	0.60%1	17.2%P	0.60%1	5090	848.3333
17.2%P	0.70%1	17.2%P	0.70%1	5054	842.3333
TRATAMIENTOS					
1	16%P	0.60%1	81.3333333333334***	28.1666666666667	22.1666666666667
2	17.2%P	0.70%1	59.1666666666666***	6	0
3	17.2%P	0.60%1	53.1666666666666***	0	<NA>
4	16%P	0.70%1	0	<NA>	<NA>
	16%P	0.60%1			

1	0
2	<NA>
3	<NA>
4	<NA>

Con estas tablas observamos que el DMS tiene un valor de 46.03084; y en la segunda tabla, observamos que las Diferencias mayores al valor de DMS se señalan con ***, por lo tanto, las interacciones 17.2%P 0.70%l, 17.2%P 0.60%l y 16%P 0.70%l tienen una mayor efecto en la ganancia diaria de peso, y de estos tres tratamientos, observamos en la primera tabla, que el tratamiento con la Media más grande es 16%P 0.70%l con una media de 901.5000; por lo tanto, podemos concluir que esta interacción tiene un mejor efecto en la ganancia diaria de peso de los cerdos.

Ejemplo 2

Se quiere probar el efecto de dos factores en el rendimiento de un terreno de Cebada; los dos factores son *Factor A* con dos niveles **a₁:44% Nitrógeno** y **a₂:47% Nitrógeno** y el *Factor B*, con dos niveles **b₁:17.5% Fósforo**, y **b₂:19.2% Fósforo**;M entonces teniendo dos factores cada uno con dos niveles tenemos un Diseño Factorial 2x2 y con cuatro tratamientos que son los siguientes:

- 44% Nitrógeno 17.5% Fósforo
- 44% Nitrógeno 19.2% Fósforo
- 47% Nitrógeno 17.5% Fósforo
- 47% Nitrógeno 19.2% Fósforo

La variable respuesta será el rendimiento medido en Toneladas por hectárea (ton/ha).

Y las hipótesis a probar son las siguientes:

1. Ho: a₁:44% Nitrógeno = a₂:47% Nitrógeno Ha: a₁: a₁:44% Nitrógeno ≠ a₂:47% Nitrógeno
2. Ho: b₁:17.5% Fósforo = b₂:19.2% Fósforo Ha: b₁: 17.5% Fósforo ≠ b₂:19.2% Fósforo
3. Ho: a₁b₁ = a₁b₂ = a₂b₁ = a₂b₂ Ha: Por lo menos una de las interacciones entre los niveles de los factores produce un efecto diferente al de los demás.

Deseamos saber si, ¿existen diferencias entre los niveles de cada factor?, ¿cuál es la mejor combinación de % de Nitrógeno y % de para Fósforo a utilizar en el terreno para obtener un mayor rendimiento de cebada?.

Y para obtener respuesta a estas preguntas utilizaremos la paquetería “DisFac”:

```
##Se fija la ruta de trabajo y se importan los datos
ruta <- "C:/Users/judit/Desktop/Paq R/Manual_DF/Datos"
setwd(ruta)
datos <- read.csv(file = "Datos 2.csv")
##Los datos se almacenan en un dataframe llamado "df"
```

```
df <- data.frame(datos)
print(df)
```

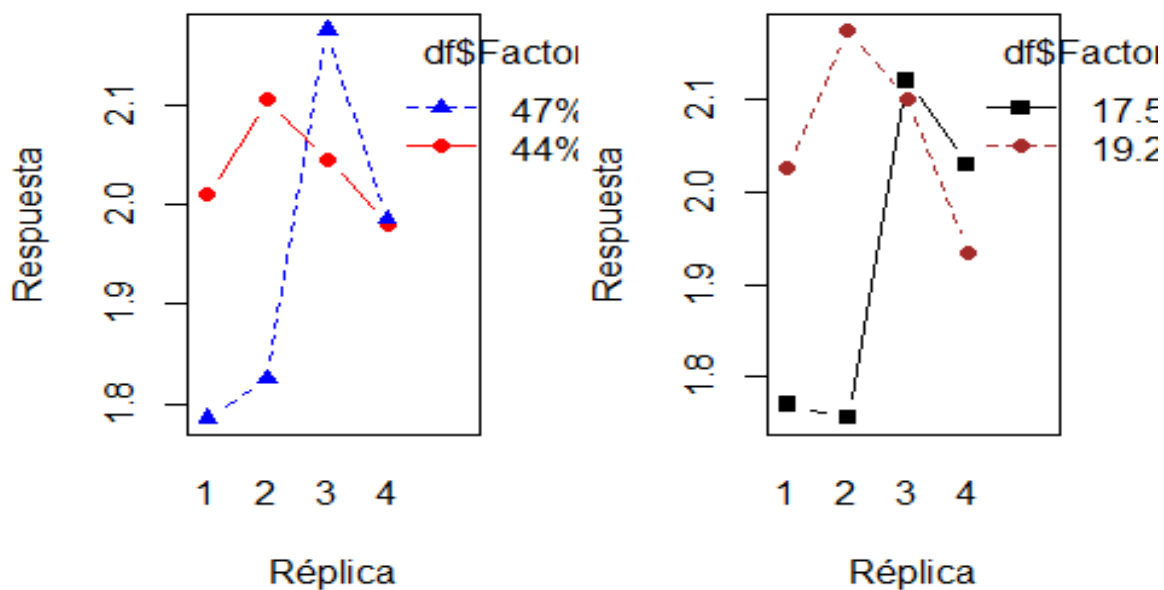
	Factor.A	Factor.B	Tratamiento	Replica	Respuesta
1	44% Nitrógeno	17.5% Fósforo	44%N 17.5%P	1	1.89
2	44% Nitrógeno	17.5% Fósforo	44%N 17.5%P	2	1.67
3	44% Nitrógeno	17.5% Fósforo	44%N 17.5%P	3	2.06
4	44% Nitrógeno	17.5% Fósforo	44%N 17.5%P	4	1.97
5	44% Nitrógeno	19.2% Fósforo	44%N 19.2%P	1	2.13
6	44% Nitrógeno	19.2% Fósforo	44%N 19.2%P	2	2.54
7	44% Nitrógeno	19.2% Fósforo	44%N 19.2%P	3	2.03
8	44% Nitrógeno	19.2% Fósforo	44%N 19.2%P	4	1.99
9	47% Nitrógeno	17.5% Fósforo	47%N 17.5%P	1	1.65
10	47% Nitrógeno	17.5% Fósforo	47%N 17.5%P	2	1.84
11	47% Nitrógeno	17.5% Fósforo	47%N 17.5%P	3	2.18
12	47% Nitrógeno	17.5% Fósforo	47%N 17.5%P	4	2.09
13	47% Nitrógeno	19.2% Fósforo	47%N 19.2%P	1	1.92
14	47% Nitrógeno	19.2% Fósforo	47%N 19.2%P	2	1.81
15	47% Nitrógeno	19.2% Fósforo	47%N 19.2%P	3	2.17
16	47% Nitrógeno	19.2% Fósforo	47%N 19.2%P	4	1.88

Una vez importados los datos, se carga la librería del paquete

```
library(DisFac)
```

Ya que se tiene cargada la librería, se ejecuta la función para poder obtener la Tabla Anova, donde los argumento son: *Datos* que es *df*, *a* que es 2 porque el factor A tiene dos niveles, *b* que es 2 porque el factor B tiene dos niveles, *r* que es 4, porque el ejemplo cuenta con 4 réplicas y finalmente, *alfa* que será de 0.05.

```
AnovaDF(df, 2, 2, 4, 0.05)
```



	FuenteVar	GradosLibertad	SumaCuadrados	CuadradosMedios	FCalculada
1	Factor A	1	0.034225	0.0342250	0.7870078
2	Factor B	1	0.078400	0.0784000	1.8028169
3	Interacción AB	1	0.072900	0.0729000	1.6763438
4	Error	12	0.521850	0.0434875	
5	Total	15	0.707375		
FTablas					
1	4.747225				
2	4.747225				
3	4.747225				
4					
5					

Como podemos observar, en la tabla Anova que obtenemos, observamos los resultados de F Calculada y F de Tablas, para poder concluir que: no se rechaza H_0 ya que $F_{calA} = 0.7870 < F_{tabA} = 4.7472$, por lo tanto, no hay diferencia entre los dos niveles del Factor a, es decir, los dos niveles de Nitrógeno tienen el mismo efecto. También observamos que no se rechaza H_0 para el Factor B puesto que $F_{calB} = 1.8028 < F_{tabB} = 4.7472$, por lo tanto, no hay diferencia entre los dos niveles de Fósforo; y finalmente, no se rechaza H_0 puesto $F_{calAB} = 1.6763 < F_{tabAB} = 4.7472$, lo que quiere decir, que todas las interacciones entre los dos niveles de los dos factores, tienen el mismo efecto en el rendimiento de Cebada.

Una vez que concluimos que tanto los niveles del Factor A: % de Nitrógeno, como los niveles del Factor B: % de Fósforo no tienen diferencias y tienen el mismo efecto en el rendimiento de cebada, ya no se realiza la *comparación de medias por Factor*; y ya que también observamos que los cuatro tratamientos, es decir, las cuatro interacciones formadas por los dos niveles de los dos factores tienen el mismo efecto en el rendimiento de la cebada, ya no se realiza la *Prueba de Diferencia Mínima Significativa*, por lo tanto, para este ejemplo, solo utilizamos una de las tres funciones con las que cuenta el paquete "DisFac".

CONCLUSIONES

Con la elaboración del presente paquete, se observa la importancia de los diseños experimentales en la Zootecnia, estos son muy útiles para poner a prueba nuevas estrategias que se desean implementar por ejemplo, en la alimentación, reproducción, manejo, etc., y saber si los resultados que arrojan son favorables o no para los sistemas de producción en los que se desean implementar. Además, es muy importante la automatización de estos diseños, para que el usuario solo pueda ingresar sus datos, utilizar la paquetería correspondiente, como la presente, y en solo un momento obtener resultados, lo que le facilitaría la toma de decisiones.

LITERATURA CITADA

Castillo Marquez, L. E. (2000). *INTRODUCCIÓN a LA ESTADÍSTICA EXPERIMENTAL* (Vol. 1). Universidad Autónoma Chapingo.