FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE

DEPARTAMENTO DE MATEMÁTICAS

TITULO:

DISEÑOS DE EXPERIMENTOS EN R

Fernando Ernesto Manzanares Morán

Docente:

Salvador Enrique Rodriguez Hernandez

Definición de un diseño experimental y conceptos básicos

El **diseño de un experimento** es la secuencia completa de los pasos que se deben tomar de antemano, para planear y asegurar la obtención de toda la información relevante y adecuada al problema bajo investigación, la cual será analizada estadísticamente para obtener conclusiones válidas y objetivas con respecto a los objetivos planteados.

Un **diseño experimental** es una prueba o serie de pruebas en las cuales existen cambios deliberados en las variables de entrada de un proceso o sistema, de tal manera que sea posible observar e identificar las causas de los cambios que se producen en la respuesta de salida.

Proposito de un diseño experimental

El propósito de cualquier Diseño Experimental, es proporcionar una cantidad máxima de información pertinente al problema que se está investigando. Y ajustar el diseño que sea lo mas simple y efectivo; para ahorrar dinero, tiempo, personal y material experimental que se va ha utilizar. Es de acotar, que la mayoría de los diseños estadísticos simples, no sólo son fáciles de analizar, sino también son eficientes en el sentido económico y en el estadístico. De lo anterior, se deduce que el diseño de un experimento es un proceso que explica tanto la metodología estadística como el análisis económico.

Conceptos básicos

Diseño: Consiste en planificar la forma de hacer el experimento, materiales y métodos a usar, etc.

Experimento: Conjunto de pruebas o ensayos cuyo objetivo es obtener información, que permita mejorar el producto o el proceso en estudio.

Tratamiento: - Es un conjunto particular de condiciones experimentales definidas por el investigador; son el conjunto de circunstancias creadas por el experimento, en respuesta a la hipótesis de investigación y son el centro de la misma.

Factor: Es un grupo específico de tratamientos. (Ejemplo, Temperatura, humedad, tipos de suelos, etc.).

Niveles del factor: Son diversas categorías de un factor. (Por ejemplo, los niveles de temperatura son 20°C, 30°C, etc.). Un factor Cuantitativo tiene niveles asociados con puntos ordenados en alguna escala de medición, como temperatura; mientras que los niveles de un factor cualitativo representan distintas categorías o clasificaciones, como tipo de suelo, que no se puede acomodar conforme a alguna magnitud.

Réplica: Son las repeticiones que se realizan del experimento básico.

Unidad experimental:

Es el material experimental unitario que recibe la aplicación de un tratamiento.

Es la entidad física o el sujeto expuesto al tratamiento independientemente de las otras unidades. La unidad experimental una vez expuesta al tratamiento constituye una sola réplica del tratamiento.

Es el objeto o espacio al cual se aplica el tratamiento y donde se mide y analiza la variable que se investiga.

Es el elemento que se está estudiando.

Unidad muestral: Es una fracción de la unidad experimental que se utiliza para medir el efecto de un tratamiento.

Error experimental: Es una medida de variación que existe entre dos o más unidades experimentales, que han recibido la aplicación de un mismo tratamiento de manera idéntica e independiente.

Factores controlables: Son aquellos parámetros o características del producto o proceso, para los cuales se prueban distintas variables o valores con el fin de estudiar cómo influyen sobre los resultados.

Factores incontrolables: Son aquellos parámetros o características del producto o proceso, que es imposible de controlar al momento de desarrollar el experimento.

Variabilidad natural: es la variación entre las unidades experimentales, que el experimentador no puede controlar ni eliminar.

Variable dependiente: es la variable que se desea examinar o estudiar en un experimento. (Variable Respuesta).

Hipótesis:

Es una suposición o conjetura que se plantea el investigador de una realidad desconocida.

Es el supuesto que se hace sobre el valor de un parámetro (constante que caracteriza a una población) el cual puede ser validado mediante una prueba estadística.

Diseños unifactoriales

En el análisis de los resultados de los experimentos se pueden observar diferentes aplicaciones de los Diseños Experimentales. Hay experimentos muy útiles en los cuales existe un sólo factor de interés; el cual se analiza por medio de la comparación de dos condiciones que intervienen en el Experimento (a menudo llamadas tratamientos o niveles del factor); a este tipo de experimentos se le denomina Experimentos de Comparación Simple, estudiados en los cursos de Estadística básica. El análisis de los datos de este tipo de Experimentos resulta ser sencillo, ya que se utilizan técnicas de la Inferencia Estadística, llamada Prueba de Hipótesis (o pruebas de significancia) que son las que ayudan al experimentador a comparar estas condiciones.

Si en el tipo de Diseño Experimental planteado anteriormente se requiere más de dos niveles del factor que se analiza, éstos son considerados como ”Diseños Unifactoriales”. Teniendo en cuenta que para el análisis de éstos se utiliza el Análisis de Varianza, ya que se requiere probar la igualdad de varias medias.

En los experimentos de los Diseños Unifactoriales, el número de observaciones recolectadas en cada tratamiento pueden ser iguales o diferentes. Cuando el número de observaciones sea diferente se dice que el Diseño está Desequilibrado o Desbalanceado; en caso contrario el Diseño está Equilibrado o Balanceado.

En este documento se hará uso del caso en el que el Diseño está Equilibrado o Balanceado y se utilizará un modelo de efectos fijos (en este tipo de Modelos se desea probar hipótesis en relación a las medias de los tratamientos y las conclusiones sólo se aplicarán a los niveles del factor considerados en el análisis, las conclusiones no pueden extenderse a tratamientos similares que no se consideraron.) y no se profundizara mas teóricamente ya que el fin de este documento es la aplicación de código del software estadistico R para la aplicación de dicho modelo.

Ejemplo 1.

El Ministerio de Educación esta interesado en implementar tres programas de estudio; con el objetivo de medir la habilidad de lectura en los alumnos. Para ello, se eligen alumnos del sexto grado de un Colegio de San Salvador, de los cuales fueron asignados al azar 27 alumnos, a cada uno de los tres grupos. Se utilizó un programa diferente en cada grupo, se llevó a cabo un examen al inicio y al final de la implementación de los programas, los valores obtenidos representan la diferencia que hay entre la nota del examen que se hizo al inicio y al final de la implementación del programa, obteniéndose los siguientes datos, en base 100:

Tratamiento	Datos								
Programa 1	20	18	18	23	22	17	15	13	21
Programa 2	15	20	13	12	16	17	21	15	13
Programa 3	12	15	18	20	18	17	10	24	16

Cuadro 1: Diferencias entre nota de inicio y nota final

Solución. Antes de realizar los cálculos matemáticos, se definirá la variable de estudio y las hipótesis que se desean probar.

Variable de estudio: Habilidad de Lectura.

Hipótesis:

$H_0 : \mu_1 = \mu_2 = \mu_3$ (No existe diferencia entre los grupos)

$H_1 : \mu_1 \neq \mu_2 \neq \mu_3$ (Existe diferencia entre los grupos)

El significado verbal de las hipótesis es:

H_0 : Con la implementación de los tres programas de estudio, no existe diferencia significativa en la habilidad de lectura entre los grupos de alumnos del sexto grado.

H_1 : Con la implementación de los tres programas de estudio, existe diferencia significativa en la habilidad de lectura entre los grupos de alumnos del sexto grado.

```
> library(xtable)
> #Leemos la hoja de datos
> datos<-read.csv("Ej1HL.csv", header = TRUE, sep = ";")
> #Analysis of Variance(aov) para un factor
> unifact<- aov(HL~Programa, data = datos)
> Resultado<-summary(unifact)
> #su equivalente
> Resultado2<- anova(unifact)
> #genarando el codigo latex para la tabla
> # de ANOVA
> Resultado3<- xtable(Resultado2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Programa	2	36.22	18.11	1.44	0.2566
Residuals	24	301.78	12.57		

Cuadro 2: Anova de un factor

Conclusión

Si se observa la columna Pr(>F) ésta muestra el P-valor generado por el ANOVA y se interpreta de la siguiente manera, como el nivel de significancia de la prueba F con 2 grados de libertad en el numerador y 24 en el denominador es de del 5 por ciento y el P-valor es mayor a este, no hay sufciente evidencia estadistica para rechazar la hipostesis nula y puede concluirse que con la implementación de los tres programas de estudio, no existe diferencia significativa en la habilidad de lectura entre los grupos de alumnos del sexto grado.

Si al efectuar el Análisis de Varianza para un Modelo de Efectos Fijos, la hipótesis nula es rechazada. Se llega a la conclusión que existe diferencia entre las medias o que hay diferencia entre los tratamientos. En muchas situaciones en la industria, este resultado es de poco interés; ya que no se especifica exactamente cuales tratamientos son diferentes y el experimentador espera hallar diferencias, y está más interesado en investigar que tratamientos difieren entre si, o dicho de otra manera, en investigar contrastes entre los tratamientos.

Cuando se da esta situación puede ser útil realizar comparaciones adicionales entre grupos de medias de los tratamientos. Las comparaciones entre medias de tratamientos se realizan en términos de los totales de tratamientos $y_{i.}$ o de los promedios de tratamientos $\bar{y_{i.}}$. Los procedimientos para efectuar esta comparación se conocen como Métodos de Comparación Múltiple o Pruebas a Posteriori.

Esto quiere decir que si no existe diferencia significativa entre la media de los tratamientos no tiene sentido aplicar pruebas posteriori, sin embargo por efectos de la utilización del código en R, se aplicará la prueba de tukey (”TukeyHSD.^{en} R) cuya hipótesis nula es: ”No esxiste difencia significativa entre las comparaciones”.

Estimacion de las medias de los tratamientos:

```
> library(ggplot2)
> library(gplots)
> boxplot(datos$HL~datos$Programa,
+          ylab = "Habilidad de lectura",
+          xlab = "Tratamientos",
+          col=rainbow(5, alpha = .3))
> plotmeans(datos$HL~datos$Programa,
+            ylab = "Habilidad de lectura",
+            xlab = "Tratamientos")
```

Evaluacion de las diferencias entre los tratamientos:

Prueba de Tukey:

```
> data.Prueba<-TukeyHSD(unifact, ordered = TRUE)
> data.Prueba.result<- data.frame(data.Prueba$Programa)
> Tabla<- xtable(data.Prueba.result)
```

	diff	lwr	upr	p.adj
Programa3-Programa2	0.89	-3.29	5.06	0.86
Programa1-Programa2	2.78	-1.40	6.95	0.24
Programa1-Programa3	1.89	-2.29	6.06	0.51

Cuadro 3: TukeyHSD

Puede observarse que el P-valor en la columna P.adj es superior al 5 porciento por lo que se concluye que no existe diferencia significativa entre las medias que se han comparado, el resultado es obvio dado que en el analisis de varianza se conluyo que no existia diferencia significativa entre las medias de los tratamientos.

Ejemplo 2.

Se supone que la cantidad de carbón usada en la producción de acero tiene un efecto en su resistencia a la tensión. En la tabla se muestran los valores de la resistencia a la tensión del acero para cada uno de los 4 diferentes porcentajes de carbón. Con estos datos efectúe el análisis apropiado e interprete sus resultados.

% de Carbon	Observaciones			
0.10	23	28	28	30
0.20	31	29	36	38
0.30	36	40	42	44
0.40	48	45	40	40

Cuadro 4: Resistencia a la Tensión

Variable de estudio: Resistencia a la tensión.

Hipótesis:

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ (No existe diferencia entre los grupos)

$H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$ (Existe diferencia entre los grupos)

El significado verbal de las hipótesis es:

H_0 : La cantidad de carbón usada en la producción de acero no tiene efecto significativo en la resistencia a la tensión

H_1 : La cantidad de carbón usada en la producción de acero tiene efecto significativo en la resistencia a la tensión.

```
> FCarbon<-c("0.10%","0.10%","0.10%",
+             "0.10%","0.20%","0.20%",
+             "0.20%","0.20%","0.30%",
+             "0.30%","0.30%","0.30%",
+             "0.40%","0.40%","0.40%",
+             "0.40%")
> Nivel<-factor(FCarbon)
> Datos<-c(23,28,28,30,31,29,36,38,36,40,42,44,48,45,40,40)
> Unifact<-aov(Datos~Nivel)
> Result<-summary(Unifact)
> Tabla <- xtable(Result)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Nivel	3	622.25	207.42	15.41	0.0002
Residuals	12	161.50	13.46		

Cuadro 5: Anova de un factor

Conclusión.

Se concluye que la cantidad de carbón usada en la producción de acero tiene efectos significativos en la resistencia a la tensión. Como se ha rechazado H_0, existe diferencia entre las medias de los tratamientos, pero no se especifica entre que medias de tratamientos existen las diferencias. Se podría estar interesado en querer saber entre que medias de tratamientos existe diferencia, para ello se utilizará el método de Tukey y LSD para contestar esta inquietud.

Estimacion de las medias de los tratamientos:

```
> library(ggplot2)
> library(gplots)
> boxplot(Datos~FCarbon,
+         ylab = "Resistencia a la tensión",
+         xlab = "Porcentaje de Carbón",
+         col=rainbow(5, alpha = .3))
> plotmeans(Datos~FCarbon,
+           ylab = "Resistencia a la tensión",
+           xlab = "Porcentaje de Carbón")
```

Evaluacion de las diferencias entre los tratamientos:

Prueba LSD:

```
> library(agricolae)
> data.Prueba1<- LSD.test(Unifact, "Nivel" )
> data.Prueba1.result1<- data.frame(data.Prueba1$statistics)
> data.Prueba1.result2<- data.frame(data.Prueba1$means)
> data.Prueba1.result3<- data.frame(data.Prueba1$groups)
> tabla1 <- xtable(data.Prueba1.result1)
```

```
> tabla2 <- xtable(data.Prueba1.result2)
> tabla3 <- xtable(data.Prueba1.result3)
```

Mean	CV	MSerror	LSD
36.12	10.16	13.46	5.65

Cuadro 6: Calculo del LSD

Porcentaje de carbón	Means	std	r	LCL	UCL	Min	Max
0.10 %	27.25	2.99	4	23.25	31.25	23.00	30.00
0.20 %	33.50	4.20	4	29.50	37.50	29.00	38.00
0.30 %	40.50	3.42	4	36.50	44.50	36.00	44.00
0.40 %	43.25	3.95	4	39.25	47.25	40.00	48.00

Cuadro 7: Estadisticos de los Tratamientos

	trt	means	M
1	0.40 %	43.25	a
2	0.30 %	40.50	a
3	0.20 %	33.50	b
4	0.10 %	27.25	c

Cuadro 8: Comparaciones

Las conclusiones principales se toman apartir de el cuadro 8, de manera que si se observa la columna ”M”se deben hacer las combinaciones de las letras y éstas diferiran en media significativamente si en las combinaciones las letras no se repiten, es decir, la media en la resistencia a la tensión según en el porcentaje de carbón para los grupos 1 y 2 (30 % y 40 %) no difiere significativamente, sin embargo para las demas combinaciones 1 y 3, 1 y 4, 2 y 3, 2 y 4, 3 y 4, si difieren significativamente en la media de la resistencia, Para reforzar esta conlclusión se realizara la prueba de tukey.

Prueba de Tukey:

```
> data.Prueba<-TukeyHSD(Unifact, ordered = TRUE)
> data.Prueba.result<- data.frame(data.Prueba$Nivel)
> Tabla<- xtable(data.Prueba.result)
```

Diferencias	diff	lwr	upr	p.adj
0.20 %-0.10 %	6.25	-1.45	13.95	0.13
0.30 %-0.10 %	13.25	5.55	20.95	0.00
0.40 %-0.10 %	16.00	8.30	23.70	0.00
0.30 %-0.20 %	7.00	-0.70	14.70	0.08
0.40 %-0.20 %	9.75	2.05	17.45	0.01
0.40 %-0.30 %	2.75	-4.95	10.45	0.72

Cuadro 9: TukeyHSD

Y efectivamente puede observarse que en la columna P.adj la comparación entre los grupos con 40 % y 30 % de carbón no difieren significativamente en media (igual al LSD) sin embargo en los grupos 20 % y 10 % y 30 % y 20 % de carbón dado que el P.adj es mayor a 0.05 no existe diferencia significativa en las medias de estas comparaciones y si existe diferencia en las demas combinaciones, en definitiva puede decirse que el la prueba de tukey es mas precisa que el la prueba LSD dado que esta reporto un grupo más que no difieren significativamente en media.

Ejemplo 3.

Una empresa tiene cinco vendedores y proyecta reapartir bonificaciones entre ellos segun su capacidad de venta. Para ello mide su volumen de ventas en cinco instantes distintos de tiempo obteniendo los siguientes resultados:

Vendedor 1	Vendedor 2	Vendedor 3	Vendedor 4	Vendedor 5
9.34	6.46	5.79	8.37	4.94
8.53	4.83	5.13	7.57	4.11
9.43	5.89	6.17	8.69	5.45
8.37	5.30	4.72	8.06	5.21
9.64	6.33	5.60	7.23	5.00

Cuadro 10: Volumen de ventas

Contrastar al 95 % si la capacidad de venta difiere de unos vendedores a otros ordenando dichas capacidades en caso de que existan diferencias.

Solución.

Variable de estudio: Capacidad de Venta.

Hipótesis:

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (No existe diferencia entre los grupos o niveles)

$H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5$ (Existe diferencia entre los grupos)

El significado verbal de las hipótesis es:

H_0 : Los vendedores tienen igual capacidad de venta.

H_1 : Los vendedores no tienen igual capacidad de venta.

```
> library(xtable)
> #Ejemplo del libro de Cesar Perez
```

```
> #pagina 130, Dise?o unifactorial
> datos<-read.csv("BonificacionesCP.csv", header = TRUE, sep = ",")
> UnifactCP<- aov(VLV~Vendedor, data = datos)
> anovaU<-anova(UnifactCP)
> Res<- xtable(anovaU)
> #Agregar la prueba tukey y LSD
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Vendedor	4	63.34	15.83	45.75	0.0000
Residuals	20	6.92	0.35		

Cuadro 11: Análisis de Varianza

Conclusión.

Se concluye que los 5 vendedores no tienen igual capacidad de venta dado que se ha rechazado la hipotesis nula, debido a que el valor de P.adj es menor a 0.05. Como se ha rechazado H_0, existe diferencia entre las medias de los tratamientos, pero no se especifica entre que medias de tratamientos existen las diferencias, para ello se utilizará el método de Tukey y LSD para contestar esta inquietud.

Estimacion de las medias de los tratamientos:

```
> library(ggplot2)
> library(gplots)
> boxplot(datos$VLV~datos$Vendedor,
+         ylab = "Volumen de venta",
+         xlab = "Vendedor",
+         col=rainbow(5, alpha = .3))
> plotmeans(datos$VLV~datos$Vendedor,
+           ylab = "Volumen de venta",
+           xlab = "Vendedor")
```

Evaluacion de las diferencias entre los tratamientos:

Prueba LSD:

```
> library(agricolae)
> data.Prueba1<- LSD.test(UnifactCP, "Vendedor"  )
> data.Prueba1.result1<- data.frame(data.Prueba1$statistics)
> data.Prueba1.result2<- data.frame(data.Prueba1$means)
> data.Prueba1.result3<- data.frame(data.Prueba1$groups)
> tabla1 <- xtable(data.Prueba1.result1)
> tabla2 <- xtable(data.Prueba1.result2)
> tabla3 <- xtable(data.Prueba1.result3)
```

Mean	CV	MSerror	LSD
6.65	8.85	0.35	0.78

Cuadro 12: Calculo del LSD

	VLV	std	r	LCL	UCL	Min	Max
Vendedor1	9.06	0.57	5	8.51	9.61	8.37	9.64
Vendedor2	5.76	0.69	5	5.21	6.31	4.83	6.46
Vendedor3	5.48	0.57	5	4.93	6.03	4.72	6.17
Vendedor4	7.98	0.59	5	7.44	8.53	7.23	8.69
Vendedor5	4.94	0.51	5	4.39	5.49	4.11	5.45

Cuadro 13: Estadísticos

	trt	means	M
1	Vendedor1	9.06	a
2	Vendedor4	7.98	b
3	Vendedor2	5.76	c
4	Vendedor3	5.48	cd
5	Vendedor5	4.94	d

Cuadro 14: Comparaciones

Conclusión. Pude observarse que el vendedor con mayor capacidad de venta es el vendedor 1, y que las combinaciones (según la tabla de Comparaciones) Vendedor 1 con los vendedores 2,3,4,5 y vendedor 4 con 2,3,5 son significativamente diferentes en capacidad de venta (media) y para las combinaciones de vendedores 2, 3, 5 su capacidad de venta no es significativamente diferente.

Prueba de Tukey:

```
> data.Prueba<-TukeyHSD(UnifactCP, ordered = TRUE)
> data.Prueba.result<- data.frame(data.Prueba$Vendedor)
> Tabla<- xtable(data.Prueba.result)
```

	diff	lwr	upr	p.adj
Vendedor3-Vendedor5	0.54	-0.57	1.65	0.60
Vendedor2-Vendedor5	0.82	-0.29	1.93	0.22
Vendedor4-Vendedor5	3.04	1.93	4.16	0.00
Vendedor1-Vendedor5	4.12	3.01	5.23	0.00
Vendedor2-Vendedor3	0.28	-0.83	1.39	0.94
Vendedor4-Vendedor3	2.50	1.39	3.62	0.00
Vendedor1-Vendedor3	3.58	2.47	4.69	0.00
Vendedor4-Vendedor2	2.22	1.11	3.34	0.00
Vendedor1-Vendedor2	3.30	2.19	4.41	0.00
Vendedor1-Vendedor4	1.08	-0.04	2.19	0.06

Cuadro 15: Prueba de Tukey

Conclución.

Puede conluirse que el vendedor 1 con los vendedores 2,3,4,5 y vendedor 4 con 2,3,5 son significativamente diferentes en capacidad de venta (media) y para las combinaciones de vendedores 2, 3, 5 su capacidad de venta no es significativamente diferente además esta prueba muestra que el vendedor 1 con el vendedor 4 no poseen diferencia significativa en capacidad de venta.

En definitiva el mejor vendedor es el 1 seguido del 4 y 2,3,5.

Diseños por Bloques

En todo Diseño de Experimento se desea que el Error Experimental sea lo más pequeño posible, ya que refleja tanto el Error aleatorio como la variabilidad entre las unidades experimentales o grupos de Unidades Experimentales; es decir, se trata de sustraer del Error Experimental la variabilidad producida por las unidades experimentales o grupos de Unidades Experimentales.

Para lograr lo anterior, las unidades experimentales se deben distribuir aleatoriamente en grupos o bloques, de tal manera que dentro de un bloque sean relativamente homogéneas y que el número de ellas dentro de un bloque sea igual al número de tratamientos por investigar. A esta forma de organizar y distribuir las unidades Experimentales se le denomina **Diseño por Bloques.**

El principio de bloqueo se basa en que las unidades experimentales dentro de cada bloque o grupo deben ser parecidas entre si (que exista homogeneidad dentro del bloque) y que los bloques debieran ser diferentes entre si (que exista heterogeneidad entre los bloques); y además que las unidades experimentales se deben distribuir aleatoriamente en los grupos o bloques de tal manera que dentro de un bloque sean relativamente homogéneas; lo que significa que el bloqueo o agrupamiento del material experimental debe hacerse de tal forma que las unidades experimentales dentro de un bloque sean tan homogéneas como sea posible y que las diferencias entre las unidades experimentales sean explicadas, en mayor proporción, por las diferencias entre los bloques. Esta técnica es una generalización de la prueba de comparación de medias de muestras apareadas, donde cada pareja es un bloque. Para la validez del análisis, se supone que no existe interacción entre los bloques y los tratamientos. En experimentos de campo, cuando no se tenga información segura sobre la uniformidad del terreno, siempre es preferible introducir bloques, que servirán para separar la variabilidad debido a dicha heterogeneidad.

Los Diseños que cumplen estas características y que son objeto de estudio, se clasifican en:

1. Diseños Aleatorizados por Bloques Completos.
2. Diseños Aleatorizados por Bloques Incompletos Balanceados.

En este documento se hara referencia unicamente a los Diseños Aleatorizados por Bloques Completos este Diseño se da cuando cada bloque tiene tantas unidades Experimentales como Tratamientos, y todos los tratamientos son asignados al azar dentro de cada bloque. En este tipo de Diseño la palabra "Completo" significa que todos los tratamientos son probados en cada bloque, y los bloques forman una unidad experimental más homogénea con la cual se comparan los tratamientos, y "Aleatorizado" porque dentro de cada bloque los tratamientos son asignados de forma aleatoria a las unidades Experimentales; es decir, que todas las Unidades Experimentales de un mismo bloque tienen la misma probabilidad de recibir cualquiera de los tratamientos. Esta estrategia de Diseño mejora efectivamente la precisión en las comparaciones al eliminar la variabilidad en los tratamientos.

Ejemplo 1.

Se probaran 5 raciones respecto a sus diferencias en el engorde de novillos. Se dispone de 20 novillos para el experimento, que se distribuyen en 4 bloques (5 novillos por bloque) con base a sus pesos, al iniciar la prueba de engorde. Los 5 tratamientos (raciones) se asignaron al azar dentro de cada bloque. Los novillos más pesados se agruparon en un bloque, en otro se agruparon los 5 siguientes más pesados y así sucesivamente. Se obtuvieron los siguientes datos:

Raciones	Bloques			
	1	2	3	4
Ración 1	0.9	1.4	1.4	2.3
Ración 2	3.6	3.2	4.5	4.1
Ración 3	0.5	0.9	0.5	0.9
Ración 4	3.6	3.6	3.2	3.6
Ración 5	1.8	1.8	0.9	1.4

Cuadro 16: Peso de los novillos

Solución

Antes de realizar los cálculos matemáticos, se definirán las hipótesis que se desean probar.

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (No existe diferencia entre las Raciones) $H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5$ (Existe diferencia entre las Raciones)

Variable Respuesta: Peso de los Novillos.

El significado verbal es:

H_0 : La cantidad de ración no influye en el engorde de los novillos. H_1 : La cantidad de ración influye en el engorde de los novillos

```
> library(xtable)
> #Leemos la hoja de datos
```

```
> datos<-read.csv("Ej1_B.csv", header = TRUE, sep = ";")
> #Analysis of Variance(aov) para un factor
> DBloques<- aov(datos$Observaciones ~ datos$Raciones + datos$Bloques)
> Resultado<-summary(DBloques)
> #su equivalente
> Resultado2<- anova(DBloques)
> #genarando el codigo latex para la tabla
> # de ANOVA
> Resultado3<- xtable(Resultado2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Raciones	4	30.71	7.68	39.11	0.0000
Bloques(novillos)	3	0.46	0.15	0.78	0.5257
Residuals	12	2.36	0.20		

Cuadro 17: Análisis de Varianza

Conclusión

Las cinco raciones no son igualmente efectivas en el engorde de novillos o la cantidad de ración influye significativamente en el engorde de los novillos, otro punto importante que puede observarse es que los bloques (novillos) no son significativos, es decir no hay un efecto significativo de los bloques hacia el factor, o dicho de otra manera separar a los novillos en bloques no reduce significativamente la variabilidad, en palabras sencillas, dado que los bloques no son significativos, da igual utilizar un diseño unifactorial.

Como la hipotesis nula fue rechazada es necesario determinar que medias son significativamente diferentes; para observar esas diferencias entre las medias se utilizará la prueba LSD y tukey.

Estimacion de las medias de los tratamientos:

```
> library(ggplot2)
> library(gplots)
```

```
> boxplot(datos$Observaciones~datos$Raciones,
+         ylab = "Pesos",
+         xlab = "Raciones",
+         col=rainbow(5, alpha = .3))
> plotmeans(datos$Observaciones~datos$Raciones,
+           ylab = "Pesos",
+           xlab = "Raciones")
```

Prueba LSD:

```
> library(agricolae)
> data.Prueba1<- LSD.test(DBloques, "datos$Raciones")
> data.Prueba1.result1<- data.frame(data.Prueba1$statistics)
> data.Prueba1.result2<- data.frame(data.Prueba1$means)
> data.Prueba1.result3<- data.frame(data.Prueba1$groups)
> tabla1 <- xtable(data.Prueba1.result1)
> tabla2 <- xtable(data.Prueba1.result2)
> tabla3 <- xtable(data.Prueba1.result3)
```

Mean	CV	MSerror	LSD
2.21	20.10	0.20	0.68

Cuadro 18: Calculo del LSD

Raciones	Means	std	r	LCL	UCL	Min	Max
R1	1.50	0.58	4	1.02	1.98	0.90	2.30
R2	3.85	0.57	4	3.37	4.33	3.20	4.50
R3	0.70	0.23	4	0.22	1.18	0.50	0.90
R4	3.50	0.20	4	3.02	3.98	3.20	3.60
R5	1.48	0.43	4	0.99	1.96	0.90	1.80

Cuadro 19: Estadisticos

	trt	means	M
1	R2	3.85	a
2	R4	3.50	a
3	R1	1.50	b
4	R5	1.48	b
5	R3	0.70	c

Cuadro 20: Comparaciones

Conclusión

Puede observarse que las combinaciones 1 y 2, 3 y 4 no difieren significativamente en media y las combinaciones 1 y 3, 1 y 4, 1 y 5, 2 y 3, 2 y 4, 2 y 5, 3 y 5, 4 y 5, difieren significativamente en media, en el contexto las raciones que se le den al novillo si afectan en su peso.

Prueba de Tukey:

```
> data.Prueba<-TukeyHSD(DBloques, ordered = TRUE)
> data.Prueba.result<- data.frame(data.Prueba$`datos$Raciones`)
> Tabla<- xtable(data.Prueba.result)
```

		diff	lwr	upr	p.adj
1	R5-R3	0.78	-0.22	1.77	0.16
2	R1-R3	0.80	-0.20	1.80	0.14
3	R4-R3	2.80	1.80	3.80	0.00
4	R2-R3	3.15	2.15	4.15	0.00
5	R1-R5	0.02	-0.97	1.02	1.00
6	R4-R5	2.03	1.03	3.02	0.00
7	R2-R5	2.38	1.38	3.37	0.00
8	R4-R1	2.00	1.00	3.00	0.00
9	R2-R1	2.35	1.35	3.35	0.00
10	R2-R4	0.35	-0.65	1.35	0.79

Cuadro 21: TukeyHSD

Conclusión

Puede observarse que en las combinaciones 1,2,5,10 no existe diferencia significativa en la media del peso de los novillos.

Ejemplo 2.

Un ingeniero de control de calidad para manufacturas de componentes electrónicos, intuitivamente siente que hay mucha variabilidad entre los 36 hornos usados por la compañía manufacturera para la cual trabaja, en las pruebas de duración de los diversos componentes. Para determinar si tiene razón, escoge un sólo tipo de componente y obtiene los siguientes datos para las 2 temperaturas (T) normalmente usadas en las pruebas de duración de estas unidades; T1 = 550ºF y T2 = 600ºF. El componente se pone a funcionar en un horno hasta que falla. En este experimento se utilizaron 3 hornos (H) seleccionados aleatoriamente y se midió los minutos de duración del componente electrónico. Con un 5 % de significancia, pruebe si el ingeniero tiene razón en su creencia de la diferencia entre los 36 hornos. Los datos obtenidos son los siguiente:

Hornos	Temperatura	
	550ºF	600ºF
H_1	246	180
H_2	191	144
H_3	187	134

Cuadro 22: Temperaturas

Solución

Antes de realizar los cálculos matemáticos, se definirán las hipótesis que se desean probar; ya que hay dieciséis hornos y solo se toman tres aleatoriamente es un Modelo de Efectos aleatorios.

$H_0 : \sigma_t^2 = 0$ (No existe variabilidad entre los Hornos)

$H_1 : \sigma_t^2 > 0$ (Existe variabilidad entre los Hornos)

Variable Respuesta: Duración del componente electrónico.

El significado verbal es:

H_0 : La variabilidad entre los Hornos no influyen significativamente en la duración del componente electrónico probado.

H_1 : La variabilidad entre los Hornos influyen significativamente en la duración del componente electrónico probado.

```
> library(xtable)
> #Leemos la hoja de datos
> datos<-read.csv("Ej2_B.csv", header = TRUE, sep = ";")
> #Analysis of Variance(aov) para un factor
> DBloques<- aov(datos$Obs ~ datos$Hornos + datos$Temperaturas)
> Resultado<-summary(DBloques)
> #su equivalente
> Resultado2<- anova(DBloques)
> #genarando el codigo latex para la tabla
> # de ANOVA
> Resultado3<- xtable(Resultado2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Hornos	2	3250.33	1625.17	34.46	0.0282
Temperaturas	1	4592.67	4592.67	97.37	0.0101
Residuals	2	94.33	47.17		

Cuadro 23: Análisis de Varianza

Conclusión

Puede observarse de la tabla de análsis e varianza que tanto bloques (temperaturas) y hornos son significativos por lo tanto, los tipos de hornos influyen significativamente en la duración del componente electrónico probado y separar los hornos en bloques por temperaturas ayuda a reducir la variabilidad del factor, el cual es el proposito del bloque.

Como en el ejercicio 2, fue rechazada la hipótesis nula. Supongamos que se desea saber cuales son las parejas de medias que son diferentes; para ello se utilizará el Método de la Mínima Diferencia Significativa (LSD) y tukey.

Estimacion de las medias de los tratamientos:

```
> library(ggplot2)
> library(gplots)
> boxplot(datos$Obs~datos$Hornos,
+         ylab = "Temperaturas",
+         xlab = "Hornos",
+         col=rainbow(5, alpha = .3))
> plotmeans(datos$Obs~datos$Hornos,
+           ylab = "Temperaturas",
+           xlab = "Hornos")
```

Prueba LSD:

```
> library(agricolae)
> data.Prueba1<- LSD.test(DBloques, "datos$Hornos")
> data.Prueba1.result1<- data.frame(data.Prueba1$statistics)
> data.Prueba1.result2<- data.frame(data.Prueba1$means)
> data.Prueba1.result3<- data.frame(data.Prueba1$groups)
> tabla1 <- xtable(data.Prueba1.result1)
> tabla2 <- xtable(data.Prueba1.result2)
> tabla3 <- xtable(data.Prueba1.result3)
```

Mean	CV	MSerror	LSD
180.33	3.81	47.17	29.55

Cuadro 24: Calculo del LSD

	datos.Obs	std	r	LCL	UCL	Min	Max
H1	213.00	46.67	2	192.11	233.89	180.00	246.00
H2	167.50	33.23	2	146.61	188.39	144.00	191.00
H3	160.50	37.48	2	139.61	181.39	134.00	187.00

Cuadro 25: Estadísticos

	trt	means	M
1	H1	213.00	a
2	H2	167.50	b
3	H3	160.50	b

Cuadro 26: Comparaciones

Conclusión

Se observa que la pareja de medias que no difieren significativamente son la media dos y la media tres; por lo tanto, no existe diferencia significativa entre el Horno dos y tres, las parejas de medias uno y dos, uno y tres al hacer las combinaciones las letras se no se repiten; por lo tanto, las dos parejas de medias difieren significtivamente; es decir que existe diferencia significativa entre el Horno uno con los Hornos dos y tres.

Prueba de Tukey:

```
> data.Prueba<-TukeyHSD(DBloques, ordered = TRUE)
> data.Prueba.result<- data.frame(data.Prueba$`datos$Hornos`)
> Tabla<- xtable(data.Prueba.result)
```

	diff	lwr	upr	p.adj	
1	H2-H3	7.00	-33.46	47.46	0.64
2	H1-H3	52.50	12.04	92.96	0.03
3	H1-H2	45.50	5.04	85.96	0.04

Cuadro 27: TukeyHSD

Conclusión

Puede notarse que el P.adj para la comparacion 1 es mayor a 0.05 por lo que no existe diferencia significativa entre las medias del horno 2 y 3.

Con respecto a las Temperaturas:

Pudo observarse que adiferencia del ejemplo 1, en el ejemplo dos los bloques si son significativos por lo tanto puede decirse que éstos tiene un efecto significativo en los componenetes electronicos, para saber cual es su efecto se realiza el siguiente procedimiento:

```
> model.tables(DBloques)

Tables of effects

 datos$Hornos
datos$Hornos
    H1     H2     H3
 32.67 -12.83 -19.83

 datos$Temperaturas
datos$Temperaturas
     T1      T2
 27.667 -27.667
```

Para los hornos en H1 aumentara la temperatura si se colocan los componentes en el horno 1. y disminuirá en -12.83ºF y -19.83ºF si se coloca el componente en el Horno 2 y 3 respectivamente.

El 27.667 para T1 significa que la tempreratura aumentara en 27.667ºF si se colocan los componentes en el Bloque 1 y disminuira -27.667ºF si se coloca en el Bloque 2.

Ejemplo 3. Como ejemplo se consideran cuatro máquinas diferentes para el emsamblaje de un producto, las cuatro máquinas son utilizadas por seis operadores diferentes (Variable Bloque) en un diseño de bloques aleatorizados para comprobar el rendimiento de las máquinas, las máquinas se asignan en orden aleatorio a cada operador y se anotan los tiempos para ensamblar el producto obteniéndose los siguientes resultados:

Máquina (Factor)	Operador (Bloque)					
	1	2	3	4	5	6
1	42.5	39.3	39.6	39.9	42.9	43.6
2	39.8	40.1	40.5	42.3	42.5	43.1
3	40.2	40.5	41.3	43.4	44.9	45.1
4	41.3	42.2	43.5	44.2	45.9	42.3

Cuadro 28: Temperaturas

Se trata de probar la hipótesis de que las máquinas funcionan a la misma tasa de velocidad al nivel $\alpha = 0{,}05$, es decir comparar si existen diferencias en los operadores en la rapidez con la que operan las maquinas.

Solución

Antes de realizar los cálculos matemáticos, se definirán las hipótesis que se desean probar.

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$ (No existe diferencia en la velocidad de la máquinas)
$H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5 \neq \mu_6$ (Existe diferencia entre la velocidad de la máquinas)

Variable Respuesta: Velocidad de las maquinas.

El significado verbal es:

H_0 : Los operadores no influyen en la velocidad de la máquinas. H_1 : Los operadores influyen en la velocidad de la máquinas.

```
> library(xtable)
> #Ejemplo del libro de Cesar Perez
```

```
> #Diseño Por Bloques, el ejemplo original esta
> #en la pagina 186 y sigue en la 255
> datosB<-read.csv("BLOQUESCP.csv", header = TRUE, sep = ",")
> #cuando los bloques o factores no son cadenas
> #es necesario generar los niveles con
> #el comando factor()
> BLOCCP<- aov(TIEMPO~factor(MAQUINA) + factor(OPERADOR), data = datosB)
> anovaB<-anova(BLOCCP)
> res<-xtable(anovaB)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
MAQUINA (Factor)	3	15.92	5.31	3.34	0.0479
OPERADOR (Bloque)	5	42.09	8.42	5.29	0.0053
Residuals	15	23.85	1.59		

Cuadro 29: Análisis de Varianza

Conclusión.

Puede observarse que existe diferencia entre la velocidad de la máquinas y que los operadores influyen en la velocidad de la máquinas ya que el P.Adj es menor a 0.05 por lo tanto las máquinas no funcionan a la misma tasa media de velocidad.

Diseños Factoriales

En muchas situaciones experimentales resulta de interés estudiar los efectos producidos por dos o más factores simultáneamente; esto se logra con la ayuda de los Diseños Factoriales. En general los Diseños Factoriales producen experimentos más eficientes, ya que cada observación proporciona información sobre todos los factores, y es posible ver las respuestas de un factor en diferentes niveles de otro factor en el mismo experimento. Por lo tanto, se entiende por Diseño Factorial a aquel diseño en el cual se pueden estudiar los efectos de dos o más factores de variación a la vez; es decir, que se puede investigar todas las posibles combinaciones de los niveles de los factores en cada ensayo completo o réplica del experimento. Cada uno de los factores en estudio varían en su aplicación, a esta variación se le llama Niveles del Factor. Las combinaciones de los niveles de cada factor, forman los respectivos tratamientos. En un diseño factorial, los factores en estudio se representan por letras mayúsculas (A,B,C,. . . . . .) y los niveles de cada uno por sus respectivas letras minúsculas (a,b,c,. . . . . .). Los cuales pueden tomar valores de 2,3,4, ... Existen experimentos factoriales Balanceados y Desbalanceados; diremos que es balanceado cuando el número de réplicas es igual para cada uno de los tratamientos usados en el experimento; en caso contrario es Desbalanceado; también se puede dar el caso en que sólo exista una sola réplica para cada tratamiento. Los Diseños factoriales se pueden combinar con los Diseños Completamente al Azar (Unifactoriales), o con el Diseño de Bloques Aleatorios, etc., dependiendo de la naturaleza del experimento.

Entre las Ventajas de usar un diseño Factorial, se pueden mencionar las siguientes:

1. Ahorro y economía del recurso experimental; ya que cada unidad experimental provee información acerca de dos o más factores, lo que no sucede cuando se realiza con una serie de experimentos simples.

2. Da información respecto a las interacciones entre los diversos factores en estudio.

3. Permite realizar estimaciones de las interacciones de los factores, además de los efectos simples.

4. Permite estimar los efectos de un factor en diversos niveles de los otros factores, produciendo conclusiones que son válidas sobre toda la extensión de las condiciones experimentales.

La única desventaja es que si el número de niveles de algunos de los factores o el número de factores es demasiado grande, entonces el número de todas las combinaciones posibles de tratamientos de factores llega a ser un número grande, en consecuencia la variabilidad en el experimento podría ser grande. Estas dos situaciones, pueden hacer difícil detectar los efectos significativos en el experimento. Se entiende por efecto de un factor al cambio en la respuesta media ocasionada por un cambio en el nivel de ese factor.

En los diseños factoriales existen tres efectos, los cuales son:

1. **Efecto Simple:** son comparaciones entre los niveles de un factor a un sólo nivel del otro factor.

2. **Efecto Principal:** son comparaciones entre los niveles de un factor promediados para todos los niveles del otro factor.

3. **Efecto de Interacción:** Miden las diferencias entre los efectos simples de un factor a diferentes niveles de otro factor; es decir, la diferencia en la respuesta entre los niveles de un factor no es la misma en todos los niveles de los otros factores.

Diseño Bifactorial

El diseño factorial más simple o sencillo es aquel que involucra en su estudio sólo dos factores o conjunto de tratamientos; es decir, que sólo se está interesado en los efectos que producen estos dos factores. A este tipo de diseño se le llama Bifactorial. Si A y B son los factores que se van ha estudiar en un diseño factorial, el factor A tendrá .ª nivelesz el factor B tendrá "b niveles", entonces cada repetición o réplica del experimento contiene todas las .ªbçombinaciones de los tratamientos y en general hay "nrepeticiones, es necesario tener al menos dos réplicas (n=2), para poder obtener la suma de cuadrados del error.

Ejemplo 1.

Se llevó a cabo un estudio del efecto de la temperatura sobre el porcentaje de encogimiento de telas teñidas, con dos réplicas para cada uno de cuatro tipos de tela en un diseño totalmente aleatorizado. Los datos son el porcentaje de encogimiento de dos réplicas de tela secadas a cuatro temperaturas; los cuales se muestran a continuación.

Factor A (Telas)	Factor B (Temperatura)			
	210ºF	215ºF	220ºF	225ºF
1	1.8	2.0	4.6	7.5
	2.1	2.1	5.0	7.9
2	2.2	4.2	5.4	9.8
	2.4	4.0	5.6	9.2
3	2.8	4.4	8.7	13.2
	3.2	4.8	8.4	13.0
4	3.2	3.3	5.7	10.9
	3.6	3.5	5.8	11.1

Cuadro 30: Porcentaje de encogimiento de la tela

Solución

En este ejemplo el análisis se hará como un Modelo de Efectos Fijos; ya que el investigador define con su propio criterio las temperaturas y los tipos de telas que va ha utilizar para llevar a cabo este experimento.

Variable Respuesta: Porcentaje de encogimiento de la tela teñida.

Planteamiento de las Hipótesis a probar:

Con el objetivo de ejemplificar el Análisis de Varianza de este tipo de Modelo, se plantearan las tres hipótesis en forma Estadística que se desean probar.

1. $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$

   H_1 : Cuando menos un $\tau_i \neq 0$

2. $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

   H_1 : Cuando menos un $\beta_j \neq 0$

3. $H_0 : (\tau\beta)_{ij} = 0$, para todo ij.

   H_1 : Cuando menos un $(\tau\beta)_{ij} \neq 0$

Forma verbal de las Hipótesis

1. H_0 : El tipo de tela no influye en el porcentaje de encogimiento de la tela teñida.

   H_1 : El tipo de tela influye en el porcentaje de encogimiento de la tela teñida .

2. H_0 : Los niveles de temperatura no influyen en el porcentaje de encogimiento de la tela teñida.

   H_1 : Los niveles de temperatura influyen en el porcentaje de encogimiento de la tela teñida.

3. H_0 : La combinación del tipo de tela teñida y la temperatura no influye significativamente en el porcentaje de encogimiento de la tela.

H_1 : La combinación del tipo de tela teñida y la temperatura influye significativamente en el porcentaje de encogimiento de la tela.

```
> library(xtable)
> #Leemos la hoja de datos
> datos<-read.csv("Ej1_F.csv", header = TRUE, sep = ";")
> #Analysis of Variance(aov) para 2 factores
> D2Fact<- aov(datos$obs~ datos$Tela*datos$Temperatura)
> Resultado<-summary(D2Fact)
> #su equivalente
> Resultado2<- anova(D2Fact)
> #genarando el codigo latex para la tabla
> # de ANOVA
> Resultado3<- xtable(Resultado2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Tela	3	41.88	13.96	279.18	0.0000
Temperatura	3	283.94	94.65	1892.91	0.0000
Tela:Temperatura	9	15.86	1.76	35.24	0.0000
Residuals	16	0.80	0.05		

Cuadro 31: Análisis de Varianza

Conlusiones

1. Respecto a la hipótesis 1. (Factor A(Tipo de Tela))

   Se observa de la tabla de Análisis de Varianza que el P.adj <0.05; por lo tanto, se rechaza H_0; es decir, el tipo de tela teñida influye significativamente en el porcentaje de encogimiento de ella.

2. Respecto a la hipótesis 2 (Factor B(Temperatura))

Se observa de la tabla de Análisis de Varianza que el P.adj <0.05; por lo tanto, se rechaza H_0; es decir, los niveles de temperatura influyen significativamente en el porcentaje de encogimiento de la tela teñida.

3. Respecto a la hipótesis 3 (Interacción(Tipo de tela y Temperatura))

   Se observa de la tabla de Análisis de Varianza que el P.adj <0.05; por lo tanto, se rechaza H_0; es decir, la combinación de el tipo de tela teñida y los niveles de temperatura influyen en el porcentaje de encogimiento de la tela teñida.

Ejemplo 2.

Como ejemplo consideramos las diferencias que aparecen al pesar un reactivo en un laboratorio debido a las balanzas utilizadas y a la habilidad del personal que realiza las pesadas, se usa una muestra de tres balanzas yd e cuatro personas del laboratorio para realizar las pesadas. A fin de contrastar la hipótesis de igualdad en las balanzas y de similaridad en la habilidad de personal, cada una de las cuatro personas efectúa tres pesadas con cada balanza, se obtienen los siguientes resultados:

Factor A (Balanzas)	Factor B (Personal)			
	1	2	3	4
1	1.81	2.04	2.03	2.05
	1.91	1.97	1.98	1.96
	1.91	1.99	2.07	2.07
2	1.94	2.08	2.03	2.23
	1.90	2.14	1.98	2.34
	1.99	2.08	2.00	2.32
3	1.83	1.98	1.91	2.19
	1.92	2.05	2.06	2.24
	1.96	2.03	2.04	2.21

Cuadro 32: Peso del Reactivo

¿Puede asegurarse a un nivel de significación del 10 % que hay habilidad homogénea de todos los obreros? ¿Puede asegurarse que las tres balanzas pesan igual? Estime las variabilidades originadas por als personas, las balanzas y la interacción entre las mismas y por el error asi como por la variabilidad total, ¿Qué componente tiene más influencia en la variación total?

Solución.

Variable Respuesta: Pesada calculada.

Planteamiento de las Hipótesis a probar:

Con el objetivo de ejemplificar el Análisis de Varianza de este tipo de Modelo, se plantearan las tres hipótesis en forma Estadística que se desean probar.

1. $H_0 : \sigma^2_\tau = 0$

   $H_1 : \sigma^2_\tau > 0$

2. $H_0 : \sigma^2_\beta = 0$

   $H_1 : \sigma^2_\beta > 0$

3. $H_0 : (\sigma^2_{\tau\beta}) = 0$, para todo ij.

   $H_1 : (\sigma^2_{\tau\beta}) > 0$

Forma verbal de las Hipótesis

1. H_0 : La variabilidad entre las balanzas no influyen significativamente en el peso calculado.

   H_1 : La variabilidad entre las balanzas influyen significativamente en el peso calculado.

2. H_0 : La variabilidad entre las habilidadades del personal para efectuar las pesadas es homogenea.

   H_1 : La variabilidad entre las habilidadades del personal para efectuar las pesadas no es homogenea.

3. H_0 : La combinación de la habilidad del personal y las balanzas no influyen significativamente al efectuar las pesadas.

   H_1 : La combinación de la habilidad del personal y las balanzas influyen significativamente al efectuar las pesadas.

```
> library(xtable)
> #Ejemplo del libro de Cesar Perez
```

```
> #pagina 189, Diseño Bifactorial
> datos2F<-read.csv("BIFACTORIALCP.csv", header = TRUE, sep = ",")
> #cuando los bloques o factores no son cadenas
> #es necesariso generar los niveles con
> #el comando factor()
> fact2CP<- aov(PESO~BALANZAS*PERSONAL , data = datos2F  )
> anovaBi<-summary(fact2CP)
> RES<- xtable(anovaBi)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
BALANZAS	2	0.13	0.07	2.40	0.1121
PERSONAL	3	0.43	0.14	5.24	0.0063
BALANZAS:PERSONAL	6	0.15	0.03	0.93	0.4924
Residuals	24	0.65	0.03		

Cuadro 33: Análisis de Varianza

Conclusión.

Se concluye que:

La variabilidad entre las balanzas no influyen significativamente en el peso calculado, la variabilidad entre las habilidadades del personal para efectuar las pesadas no es homogenea, la combinación de la habilidad del personal y las balanzas no influyen significativamente al efectuar las pesadas.

Diseño Trifactorial

Estos tipos de Diseños experimentales son aquellos en los cuales se involucran en su estudio tres factores; es decir, que se esta interesado en los efectos que producen los tres factores en la variable respuesta en forma individual y conjunta (interacción). A los cuales se les llama Diseños Trifactoriales.

Sean, A,B y C los factores que se van a estudiar en un experimento; el factor A tiene “a” niveles, el factor B tiene “b” niveles y el factor C tiene “c” niveles; por lo tanto, cada repetición del experimento tiene todas la “abc” combinaciones de tratamiento y en general hay “n” repeticiones ($n = 2$). El orden en que se toman las “abcn” observaciones en el experimento debe ser aleatorio, de modo que este es un Diseño completamente aleatorizado. Existen tres efectos principales (A,B y C), tres efectos dobles (AB,AC y BC) y un efecto triple (ABC).

Los niveles de cada uno de los factores pueden ser elegidos de forma aleatoria, si es así los factores son aleatorios o ser elegidos específicamente por el experimentador; es decir, los que a él le interesan estudiar, entonces los factores son fijos.

Por lo tanto, de acuerdo a la forma en que son elegidos los niveles de los factores, así es el Modelo que resulta y el Análisis de Varianza que se lleva a cabo.

Ejemplo 3.

En un experimento para investigar las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra, se utilizaron dos períodos (Edad A) diferentes de curado en combinación con dos Temperaturas(B) diferentes de curado y dos tierras(C) diferentes. Se hicieron dos réplicas para cada combinación de niveles de los tres factores, resultando los siguientes datos:

<table>
<tr><th rowspan="4">Factor A (Edad)</th><th colspan="4">Factor B (Temperatura)</th></tr>
<tr><th colspan="2">1</th><th colspan="2">2</th></tr>
<tr><th colspan="2">Factor C (tierra)</th><th colspan="2">Factor C (tierra)</th></tr>
<tr><th>1</th><th>2</th><th>1</th><th>2</th></tr>
<tr><td rowspan="2">1</td><td>471</td><td>385</td><td>485</td><td>530</td></tr>
<tr><td>413</td><td>434</td><td>552</td><td>593</td></tr>
<tr><td rowspan="2">2</td><td>712</td><td>770</td><td>712</td><td>741</td></tr>
<tr><td>637</td><td>705</td><td>789</td><td>806</td></tr>
</table>

Cuadro 34: Resistencia a la compresión de la mezcla de Cemento y Tierra

Solución

Planteamiento de las hipótesis a probar

Forma Estadística:

1. $H_0 : \tau_1 = \tau_2 = 0$

   H_1:cuando menos un $\tau_i \neq 0$

2. $H_0 : \beta_1 = \beta_2 = 0$

   H_1:cuando menos un $\beta_j \neq 0$

3. $H_0 : \gamma_1 = \gamma_2 = 0$

   H_1:cuando menos un $\gamma_k \neq 0$

4. $H_0 : (\tau\beta)_{ij} = 0$, para todo ij

   H_1: cuando menos un $(\tau\beta)_{ij} \neq 0$

5. $H_0 : (\tau\gamma)_{ik} = 0$, para todo ik

   H_1: cuando menos un $(\tau\gamma)_{ik} \neq 0$

6. $H_0 : (\beta\gamma)_{jk} = 0$, para todo jk

   H_1: cuando menos un $(\beta\gamma)_{jk} \neq 0$

7. $H_0 : (\tau\beta\gamma)_{ijk} = 0$, para todo ijk

   H_1:cuando menos un $(\tau\beta\gamma)_{ijk} \neq 0$

Variable Respuesta: Resistencia a la compresión de la mezcla de Cemento y Tierra.

Forma Verbal

1. H_0: La edad o períodos no influye significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

H_1: La edad o períodos influye significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

2. H_0: La Temperatura no influye significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

   H_1: La Temperatura influye significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

3. H_0: Los tipos de tierra no influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

   H_1: Los tipos de tierra influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

4. H_0: La edad y la temperatura no influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

   H_1: La edad y la temperatura influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra

5. H_0: La edad y los tipos de tierra no influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

   H_1: La edad y los tipos de tierra influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

6. H_0: La temperatura y los tipos de tierra no influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

   H_1: La temperatura y los tipos de tierra influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

7. H_0: La edad, Temperatura y tipos de tierra no influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

   H_1: La edad, Temperatura y tipos de tierra influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra

```
> library(xtable)
> #Leemos la hoja de datos
> datos<-read.csv("Ej2_F.csv", header = TRUE, sep = ";")
> #Analysis of Variance(aov) para 2 factores
> D3Fact<- aov(datos$OBS~ datos$Edad*datos$Temperatura*datos$Tierra)
> Resultado<-summary(D3Fact)
> #su equivalente
> Resultado2<- anova(D3Fact)
> #genarando el codigo latex para la tabla
> # de ANOVA
> Resultado3<- xtable(Resultado2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Edad	1	252255.06	252255.06	117.92	0.0000
Temperatura	1	28985.06	28985.06	13.55	0.0062
Tierra	1	2328.06	2328.06	1.09	0.3273
Edad:Temperatura	1	3393.06	3393.06	1.59	0.2434
Edad:Tierra	1	1425.06	1425.06	0.67	0.4380
Temperatura:Tierra	1	315.06	315.06	0.15	0.7111
Edad:Temperatura:datos$Tierra	1	3335.06	3335.06	1.56	0.2471
Residuals	8	17113.50	2139.19		

Cuadro 35: Análisis de Varianza

Conclusiones

Se tiene:

1. **Respecto a la Hipótesis 1 (Factor A(Edad))** Se observa de la tabla de Análisis de Varianza que el P.adj <0.05; por lo tanto, se rechaza H_0; es decir, la edad o períodos influye significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

2. **Respecto a la Hipótesis 2 (Factor B (Temperatura))** Se observa de la tabla de Análisis de Varianza que el P.adj <0.05; por lo tanto, se rechaza H_0; es decir, la Temperatura influye significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

3. **Respecto a la Hipótesis 3 (Factor C (Tierra))** Se observa de la tabla de Análisis de Varianza que el P.adj >0.05; por lo tanto, se acepta H_0; es decir, los tipos de tierra no influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

4. **Respecto a la Hipótesis 4 (Interacción(Edad y Temperatura))** Se observa de la tabla de Análisis de Varianza que el P.adj >0.05; por lo tanto, se acepta H_0; es decir, la edad y la temperatura no influye significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

5. **Respecto a la Hipótesis 5 (Interacción(Edad y Tierra))** Se observa de la tabla de Análisis de Varianza que el P.adj >0.05; por lo tanto, se acepta H_0; es decir, la edad y los tipos de tierra no influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

6. **Respecto a la Hipótesis 6 (Interacción(Temperatura y Tierra))** Se observa de la tabla de Análisis de Varianza que el P.adj >0.05; por lo tanto, se acepta H_0; es decir, la temperatura y los tipos de tierra no influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.

7. **Respecto a la Hipótesis 7 (Interacción(Edad,Temperatura y Tierra))** Se observa de la tabla de Análisis de Varianza que el P.adj >0.05; por lo tanto, se acepta H_0; es decir, la edad, Temperatura y tipos de tierra no influyen significativamente en las propiedades de resistencia a la compresión de mezclas de Cemento y Tierra.