



# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Facultad de Ciencias

Modelos Lineales

Act. Susana Barrera Ocampo, Mat. Erick Eduardo Aguilar Hernández



**Alumno:** Nombres ApellidoPaterno ApellidoMaterno **Cuenta:** 666

1. Una estudiante realizó un diseño de varianza modelo 1 de un solo factor. La ANOVA de este experimento es:

Fuente	GL	Suma de Cuadrados	Cuadrados Medios	F	P-Value
Factor			14.18		
Error		37.5			
Total	23	108.63			

Complete la información que falta y responda las siguientes preguntas sobre este experimento.

- ¿Explique brevemente que es el Análisis de la Varianza (ANOVA)? ¿Cuáles son las pruebas estadísticas de comparación múltiple y para que se utilizan?
  - Con un nivel de significancia es 0.05, establezca la hipótesis y mencione su conclusión.
  - ¿Cuántos niveles del factor se usaron en este experimento?
  - ¿Cuántas replicas tiene este experimento?
  - Suponga que la estudiante realmente había llevado a cabo este experimento, como un diseño de bloques al azar y la suma de cuadrados para bloques fue 12. Reconstruya la ANOVA. (ojo cambia Cuadrados medios del factor y suponer los mismos factores)
  - ¿Cuánto ha reducido el bloque la estimación del error experimental?
2. Suponiendo que se realizó una ANOVA de bloques al azar, con 4 tratamientos y 6 bloques aleatorios. Y CM Residual de 7.32575, realice la prueba de Scheffé para estos tratamientos con su hipótesis y una alfa del 5% y los intervalos de confianza al 95%.

Grupo A	Grupo B	Media A	Media B	Diferencia	na	nb	EE Diferencia	Estadística Scheffé
Trat1	Trat2	92.816	91.6833					
Trat1	Trat2	92.816	88.9167					
Trat1	Trat2	92.816	85.7667					

3. Un agrónomo desea determinar el efecto de diferentes fuentes de nitrógeno en la producción de una materia seca sobre cebada forrajera. Tenemos cinco fuentes a ser comparadas:  $(\text{NH}_4)_2\text{SO}_4$ ,  $\text{NH}_4\text{NO}_3$ ,  $\text{NaNO}_3$ ,  $\text{CO}(\text{NH}_2)_2$ ,  $\text{Ca}(\text{NO}_3)_2$  y un tratamiento control sin nitrógeno.

Se desean aplicar los resultados sobre un rango bastante amplio de condiciones; se hicieron ensayos sobre cuatro tipos de suelo.

Para el diseño experimental se eligió un diseño en bloques completamente aleatorizado con los tipos de suelo como factor de bloqueo, se localizaron seis parcelas en cada uno de los cuatro tipos de suelo, y se asignó aleatoriamente los tratamientos a las parcelas dentro de cada tipo de suelo. La variable de interés es la producción (kg/parcela) de cebada bajo varias fuentes de nitrógeno.

Tratamiento	Suelo I	Suelo II	Suelo III	Suelo IV
$(\text{NH}_4)_2\text{SO}_4$	32.1	35.6	41.9	35.4
$\text{NH}_4\text{NO}_3$	30.1	31.5	37.1	30.8
$\text{CO}(\text{NH}_2)_2$	25.4	27.1	33.8	31.1
$\text{CA}(\text{NO}_3)_2$	24.1	33.0	35.6	31.4
$\text{NaNO}_3$	26.1	31.0	33.8	31.9
Control	23.2	24.8	26.7	26.7

4. Una empresa textil teje una tela en un gran número de telares, se desea que los telares sean homogéneos para que obtenga una tela de resistencia uniforme. El ing. de procesos, sospecha de la variación en la fuerza dentro de tela del mismo telar, también puede haber y variaciones significativas en la resistencia entre telares. Se selecciona 4 telares al azar y hace 7 pruebas de resistencia en el tejido fabricado en cada telar Este experimento se ejecuta en orden aleatorio.

Telar 1	Telar 2	Telar 3	Telar 4
98	91	96	95
97	90	95	96
99	93	97	99
96	92	95	98
95	89	95	97
98	90	98	99
99	93	94	96

5. Para valorar la variabilidad del contenido de zumo de una cierta variedad de limón, se tomaron 4 árboles al azar y se midió el contenido de zumo de 3 limones de cada árbol. Esta observación se hizo durante 5 días, eligiendo fechas al azar. Los resultados fueron (en cm3):

Día	Árbol 1			Árbol 2			Árbol 3			Árbol 4		
1	24	26	26	28	20	27	28	18	21	27	24	20
2	18	25	19	21	24	23	27	29	17	25	23	22
3	16	21	15	24	20	21	22	25	24	29	27	27
4	21	24	22	23	20	26	24	24	23	20	21	27
5	23	24	28	27	21	28	26	25	27	25	27	28

Verifique si existe variabilidad entre árboles, entre días y entre las interacciones árboles x días.

6. El dataset universitarios.csv contiene registros del porcentaje de egresados mujeres por cada carrera:
- Ajustar un modelo de árbol de regresión para estimar el porcentaje de mujeres graduadas en matemáticas con respecto a las demás carreras y seleccionar las 5 variables mas importantes, (no usar el año de la generación).
  - Con las 5 variables mas importantes ajustar un conjunto de modelos de arboles de regresión realizando folios de validación cruzada, ajustando los parámetros del árbol hasta obtener al menos una R cuadrada de 0.7 de manera general en valiación cruzada.
7. El dataset severidades.csv contiene históricos de hospitalizaciones de una aseguradora y la cantidad que esta ha tenido que pagar en dolares en la columna severidad la cual se desea estimar en funcion de las demás.
- Para las variables categóricas aplicar la codificación OneHotEncoding para convertir los datos categóricos a variables de nivel como se muestra continuación.

sexo	fumador	region	sexo_femenino	sexo_masculino	fumador_no	fumador_si	region_centro	region_noreste	region_noroeste	region_sureste
femenino	si	centro	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
masculino	no	sureste	0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0
masculino	no	suroeste	0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0
masculino	no	noroeste	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0
masculino	no	noreste	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0

- b) Con las 11 variables obtenidas (numéricas mas las obtenidas en el OHE) separar los datos en una partición de 70% para entrenamiento y 30% para validación ciega.

Especificar una arquitectura neuronal y entrenar la red con los datos de entrenamiento y aplicando el modelo a los datos de validación ciega. Ajustar los parámetros hasta obtener al menos una R cuadrada del 0.8 en los datos ciegos.

sexo_femenino	sexo_masculino	fumador_no	fumador_si	region_centro	region_noreste	region_noroeste	region_sureste	edad	masacorporal	hijos
1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	19	27.900	0
0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	18	33.770	1
0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	28	33.000	3
0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	33	22.705	0
0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	32	28.880	0