Análisis de la Varianza, comparación de *2* tratamientos

1. Se estudian dos tipos de neumáticos con los resultados siguientes:

Tipo	n_i	$\overline{x}_i(Km)$	$\widehat{s}_i(Km)$
A	121	27465	2500
B	121	27572	3000

   Calcular, con $\alpha = 0.01$:

   a) Un intervalo de confianza para $\frac{\sigma_1^2}{\sigma_2^2}$.

   b) Un intervalo de confianza para $\mu_1 - \mu_2$.

2. Se dispone de rendimientos de dos máquinas. Los resultados de la máquina A son 137.5; 14.07; 106.9; 175.1; 177.3; 120.4; 77.9 y 104.2, mientras que los reultados para la B son: 103.3; 121.7; 98.4; 161.5; 167.8 y 67.3. ¿Son las máquinas iguales? (Suponer que los rendimientos de ambas máquinas siguen distribuciones normales).

3. Un fabricante de automóviles debe elegir entre un determinado tipo de piezas de acero suministradas por un proveedor A y otras suministradas por otro proveedor B. Para proceder a la elección se ha analizado la resistencia a la tracción de las piezas suministradas por ambos proveedores, tomando una muestra de tamaño 10 de las piezas del primero, y otra de tamaño 12 del segundo. La resistencia media de la muestra de A es de 54000 unidades y la de la muestra de B es de 49000 unidades, siendo las desviaciones típicas muestrales corregidas $\widehat{s}_A = 2100$ y $\widehat{s}_B = 1900$. Las resistencias de las piezas de ambos proveedores se distribuyen normalmente. Las piezas del proveedor B son más baratas que las del proveedor A, por lo que estas últimas sólo son rentables si tienen una resistencia media al menos 2000 unidades mayor que las de B, y la misma variabilidad.

   a) ¿A qué proveedor habría que comprar las piezas a la vista de los resultados muestrales?

   b) Obtener un intervalo de confianza al 90% para la diferencia de medias de la resistencia de las piezas de los proveedores A y B.

Análisis de la Varianza, comparación de *k* tratamientos

1. En una fábrica de automóviles se utiliza una misma planta para el ensamblaje de tres modelos distintos $(A, B$ y $C)$. Para determinar si los modelos reciben el mismo tratamiento, se ha realizado un control de calidad a una muestra tomada para cada modelo. El número de defectos encontrados para cinco vehículos del modelo A son $5, 4, 6, 6$ y 7; para seis vehículos del modelo B son $7, 8, 6, 7, 6$ y 5; y para ocho vehículos del modelo C: $9, 7, 8, 9, 10, 11, 10$ y 10. Contrastar si existen diferencias en el tratamiento que se da a los distintos modelos.

2. Una empresa debe elegir entre cinco procedimientos para fabricar un cierto producto químico. Se sospecha que existen diferencias entre ellos aunque pequeñas. Para detectar estas diferencias se pretende realizar un experimento a gran escala con el mismo número de observaciones en cada grupo. Para determinar este tamaño muestral se ha realizado un experimento piloto con 6 observaciones de cada método y los resultados (medias de cada grupo) han sido los siguientes:

METODO	1	2	3	4	5
Media	425.6	423.2	418.8	430.2	422.2

y la varianza residual $\hat{s}_R^2 = 198.5$.

(a) ¿ Cúal debe ser el tamaño muestral del experimento a gran escala para que el contraste de análisis de la varianza sea significativo con $\alpha = 0.01$ si el coeficiente de determinación es igual al del experimento piloto?.

(b) El método A es el procedimiento habitual y el método D es el que se sospecha proporciona mejor rendimiento. Una hipótesis que se pretende contrastar es $H_0 : \mu_D = \mu_A$, frente a la hipótesis alternativa $H_1 : \mu_D > \mu_A$. ¿ Qué condición debe cumplir la diferencia entre las medias muestrales de los dos métodos para rechazar H_0 con $\alpha = 0.01$?

3. Se ha realizado un experimento para estudiar el efecto de un único factor con I niveles en la variable respuesta y con un número diferente de observaciones en cada tratamiento: $n_1, n_2, ..., n_I$ siendo el total $n = n_1 + n_2 + \cdots + n_I$. Llamando y_{ij} a la observación j del tratamiento i, $i = 1, ..., I$, $j = 1, 2, ..., n_i$ e $\bar{y}_{i\bullet}$ la media del tratamiento i. Se desea estimar la media general ¿cuál de los dos estimadores siguientes

$$\overline{y}_{\bullet\bullet} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{n_i} y_{ij}}{n}, \qquad \tilde{y}_{\bullet\bullet} = \frac{\sum_{i=1}^{I} \bar{y}_{i\bullet}}{I}$$

tiene mínima varianza? Realiza la comprobación para el caso $I = 5$, con $n_i = 3, 2, 3, 5, 6$ el número de observaciones en cada tratamiento. Asumir que las observaciones son independientes y que se cumple la hipótesis de homocedasticidad.

4. Considere la comparación de dos tratamientos en poblaciones normales. Demuestre que el contraste t para comparar dos medias es análogo al contraste de la F en Análisis de la Varianza (suponga $n_1 = n_2$).

5. Cinco tipos (A, B, C, D y E) de material sintético se han sometido a un ensayo de desgaste. Para cada tipo de material la prueba se repitio 6 veces. El desgaste medio y la desviación típica corregida en cada caso es la siguiente:

	A	B	C	D	E
media $\overline{x}_i$	14.1	16.3	13.5	14.8	15.3
d. típica $\hat{s}_i$	1.3	1.2	1.4	1.2	1.5

(a) Contrastar ($\alpha = 0.05$) la hipótesis

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D = \mu_E$$

frente a la hipótesis alternativa,

$$H_1 : \text{ alguna media es distinta de las demás.}$$

(b) Indicar con nivel de confianza 0.95 el material con desgaste menor y qué materiales tienen desgaste medio, distinto.

(c) Obtener un intervalo de confianza con $\alpha = 0.01$ para la varianza del error experimental.

6. Se desea comprobar el efecto de un tratamiento térmico sobre la resistencia de un nuevo material. Se han tomado 15 probetas y se han asignado al azar a los tres tratamientos T_1, T_2 y T_3 obteniendo como medida de resistencia superficial los valores siguientes:

T_1	T_2	T_3
2.65	4.31	4.81
2.67	3.96	5.32
2.46	4.64	4.93
1.90	4.74	5.49
2.62	4.00	4.45

(a) Contrastar mediante el test de análisis de la varianza si existen diferencias significativas entre los tratamientos térmicos ($\alpha = 0.01$).

(b) La temperatura del tratamiento 2 es la media de las temperaturas de los otros dos tratamientos. Si la relación entre la resistencia y la temperatura es lineal, es de esperar que la media del tratamiento 2 verifique : $H_0 : \mu_2 = \frac{1}{2}(\mu_1 + \mu_3)$. Hacer el contraste bilateral de esta hipótesis con $\alpha = 0.05$. (Nota.- Usar la distribución de $\overline{y}_2 - (\overline{y}_1 + \overline{y}_3)/2$, donde $\overline{y}_i$ es la media de los datos correspondientes al tratamiento T_i).

7. Un fabricante sospecha que los lotes de materia prima recibidos de un proveedor difieren significativamente de su contenido en calcio. Elige al azar 5 lotes diferentes y un químico hace cinco determinaciones del contenido en calcio de cada lote. Los resultados obtenidos han sido

Lote 1	Lote 2	Lote 3	Lote 4	Lote 5
23.46	23.59	23.51	23.28	23.29
23.48	23.46	23.64	23.40	23.46
23.56	23.42	23.46	23.37	23.37
23.39	23.49	23.52	23.46	23.32
23.40	23.50	23.49	23.29	23.38

La tabla de análisis de la varianza se proporciona a continuación. Comparar las medias de los cinco tratamientos con nivel de significación total $\alpha_T = 0.10$.

Análisis de la varianza					
Fuente	Variabilidad	g.l.	Var. Media	F	Nivel crítico
Lote	0.096976	4	0.024244	5.54	0.0036
Residuos	0.08760	20	0.00438		
Total	0.184576	24			

Capítulo 2. Diseño de experimentos

2.1. Se pretende estudiar el efecto que produce los factores (1) *Porcentaje de algodón* (10%, 20% y 30%) (2) *Tipo de confección* (A y B) en la resistencia al desgaste de ciertos tejidos de fibra sintética. Se ha realizado el siguiente diseño con tres replicaciones (archivo *desgaste.txt*)

	10%	20%	30%
A	115	120	126
	112	135	118
	133	139	142
B	107	110	132
	114	102	114
	108	117	125

1. Construir la tabla de Análisis de la Varianza y contrastar la influencia de los dos factores y la presencia de la interacción.
2. Hacer un contraste de diferencia de medias y decidir el tratamiento más adecuado para conseguir la mayor resistencia al desgaste.

2.2 En una planta piloto se obtiene un nuevo producto mediante un proceso químico. Con el fin de mejorar el rendimiento se emplean dos catalizadores distintos y se trabaja con tres temperaturas diferentes. Los resultados del experimento son (archivo *rendimiento.txt*)

	Temperatura					
Catalizador	20^0		30^0		40^0	
A	115	125	130	140	110	120
B	115	105	135	145	100	110

1. Contrastar si los factores Temperatura y Catalizador tienen efectos significativos. ($\alpha = 0.05$)
2. ¿Qué tratamiento se debe utilizar para obtener el mayor rendimiento, si se desea garantizar una probabilidad de error tipo I total, $\alpha_T = 0.03$?

2.3 Un investigador quiere estudiar el efecto de sexo (hombre, mujer) y tipo de formación (ciencias, letras) en el dominio del inglés escrito en profesores universitarios. Para ello analiza el número de incorrecciones gramaticales en artículos científicos enviados a publicación. Para cada combinación de niveles de los factores se han elegido al azar tres profesores. En la tabla se proporciona el número de fallos detectados en artículos de 15 páginas (archivo *error.txt*)

	Letras	Ciencias
Hombre	8, 6, 13	22, 28, 33
Mujer	5, 10, 6	12, 14, 9

Contrastar con nivel de significación 0.05 si los efectos principales y la interacción son significativos. Tener en cuenta que $P(F_{1,8} \leq 5.32) = 0.95$, siendo $F_{1,8}$ la distribución F con grados de libertad 1 y 8. Interpretar los resultados.

2.4 Un alumno, como trabajo de la asignatura de estadística, ha comparado tres marcas distintas (A,B,C) de palomitas de maíz precocinadas. Cada marca puede prepararse friendolas en una sartén (método 1) o en el horno microondas (método 2). El alumno ha realizado un diseño factorial completo 3×2 con cinco replicaciones en cada uno de los seis tratamientos. La variable respuesta medida es el porcentaje de granos de maíz que no se han inflado adecuadamente. Los resultados del experimento se muestran en la tabla, en cada tratamiento se proporciona la media y entre paréntesis la desviación típica corregida para las cinco replicaciones. Contrastar si la interacción entre los dos factores es significativa.

	A	**B**	**C**
Sartén	5.5 (1,4)	3.6 (1,8)	7.5 (2,5)
Horno	3.8 (1,3)	3.4 (0,9)	4.3 (1,3)

2.5. La tabla muestra el tiempo de supervivencia de grupos de cuatro animales a los que se ha asignado al azar tres venenos y posteriormente cuatro tratamientos. (archivo *venenos.txt*)

	Tratamiento			
	A	B	C	D
Veneno				
I	0.31	0.82	0.43	0.45
	0.45	1.10	0.45	0.71
	0.46	0.88	0.63	0.66
	0.43	0.72	0.76	0.62
II	0.36	0.92	0.44	0.56
	0.29	0.61	0.35	1.02
	0.40	0.49	0.31	0.71
	0.23	1.24	0.40	0.38
III	0.22	0.30	0.23	0.30
	0.21	0.37	0.25	0.36
	0.18	0.38	0.24	0.31
	0.23	0.29	0.22	0.33

1. ¿Son los venenos y tratamientos significativos? ¿Existe interacción entre el veneno y el tratamiento?

2. Analice los residuos del modelo anterior. ¿Se verifican las hipótesis básicas del modelo? ¿Qué transformación de los datos hace que se verifiquen las hipótesis?

3. Calcule la tabla de análisis de la varianza con los datos transformados. ¿Tiene la transformación realizada algún efecto sobre los efectos principales y la interacción?

2.6 Se ha realizado un experimento para estudiar el efecto de la temperatura (T) y tiempo de exposición (E) sobre la cantidad absorbida de un compuesto químico por un material sumergido en él. En el estudio se han empleado tres temperaturas (T1, T2, T3) y tres tiempos de exposición (E1, E2, E3): cada tratamiento se ha replicado tres veces. La cantidad absorbida (mg) del compuesto químico en cada uno de los 27 experimentos se muestra en la tabla 1 (archivo *absorbida.txt*) y las medias en la tabla 2:

Tabla 1: Cantidad Absorbida (mg)

Tiempo de	Temperatura		
Exposición	T1	T2	T3
	35.5	91.2	70.1
E1	29.7	100.7	64.1
	31.5	82.4	70.1
	52.5	71.0	79.4
E2	53.3	77.0	77.7
	55.0	75.6	75.1
	85.9	87.0	83.0
E3	85.2	86.1	87.0
	80.2	88.1	78.5

Tabla 2: Medias de Cantidad Absorbida (mg)

Tiempo de	Temperatura			
Exposición	T1	T2	T3	Medias
E1	32.23	91.43	68.10	63.92
E2	53.60	74.53	77.40	68.51
E3	83.76	87.06	82.83	84.56
Medias	56.53	84.34	76.11	72.33

La tabla 3 corresponde al análisis de la varianza del experimento.

Tabla 3: Tabla de análisis de la varianza

Fuente Variabilidad	Suma de Cuadrados	Grados de Libertad	Varianzas	F	p-valor
Temperatura	3673.61	2	1836.80	110.58	0.0000
T. Exposición	2112.65	2	1056.32	63.59	0.0000
Interacción	2704.44	4	676.11	40.70	0.0000
Residual	299.00	18	16.61		
Total	8789.7	26			

1. (a) Interpreta los resultados del análisis de la varianza.

2. Realiza las comparaciones dos a dos de los nueve tratamientos y elige aquél o aquellos que proporcionan una absorción mayor (95%).

3. Comprueba gráficamente la hipótesis de homocedasticidad e interpreta los resultados.

2.7. Se ha realizado un diseño experimental para determinar la influencia de dos factores combinación de hidrocarburos y cantidad de hidrógeno en el rendimiento de un proceso químico complejo. Se estudiaron cuatro combinaciones de hidrocarburos (A,B, C y D) y tres niveles en el contenido de hidrógeno (1,2 y 3). En cada tratamiento se realizaron cuatro réplicas. En la tabla 1 se presentan los resultados: mejora en tanto por mil respecto a procedimiento estándar (archivo *hidrocarburos.txt*). Los números entre paréntesis de la tabla se corresponden con las medias de cada tratamiento, de los cuatro niveles del factor hidrocarburos y de los tres niveles de hidrógeno. En la tabla 2 se muestra la tabla de análisis de la varianza del experimento.

Tabla 1. Datos y medias entre paréntesis

	A	**B**	**C**	**D**	Medias	Etapa
	10.3	10.5	7.2	13.0		1
	11.1	8.2	5.3	12.9		1
1	15.3	9.7	12.5	5.3		2
	2.1	8.9	19.1	12.0		2
Medias	(9.7)	(9.325)	(11.025)	(10.8)	(10.213)	
	25.8	20.6	29.7	17.6		1
	25.7	17.1	26.3	12.0		1
2	28.9	21.4	22.4	24.6		2
	27.8	17.3	25.9	23.1		2
Medias	(27.05)	(19.1)	(26.075)	(19.325)	(22.888)	
	28.5	21.0	30.4	20.5		1
	31.2	26.8	26.6	26.2		1
3	24.8	19.4	34.4	27.8		2
	26.5	22.2	27.5	21.9		2
Medias	(27.75)	(22.35)	(29.975)	(24.1)	(25.981)	
Medias	(21.5)	(16.925)	(22.275)	(18.075)		

Tabla 2. *ANOVA -*

Fuentes	Suma Cuadrados	Grados Libertad	Var.	F	p-valor
Hidrocarburos	242.5	3	80.85	5.55	.0031
Hidrógeno	2234	2	1117	76.7	.0000
Interacción	119.3	6	19.88	1.36	.2546
Residual	523.7	36	14.55		
Total	3120	47			

1. Comparar las medias de los cuatro niveles del factor *Hidrocarburo* y las de los tres niveles del factor *Hidrógeno*. Indica si existen diferencias significativas con nivel de significación 0.05.

2. Elige el tratamiento que proporciona el rendimiento óptimo, justificando la respuesta. Da un intervalo de confianza para el valor medio en dichas condiciones con nivel de confianza del 95%.

3. El experimento se realizó en dos etapas, en una primera etapa se recogieron las 24 observaciones que se indican en la tabla 1 como etapa 1 y las otras 24 como etapa 2. Los resultados del análisis de la varianza correspondientes a cada etapa se muestran en las tablas 3 y 4.

Tabla 3. *ANOVA - Etapa 1*

Fuentes	Suma Cuadrados	Grados Libert.	Var.	F	p-valor
Hidrocarburos	115.9	3	38.63	6.07	.0093
Hidrógeno	1175.0	2	587.7	92.4	.0000
Interacción	218.4	6	36.39	5.72	.0051
Residual	76.3	12	6.358		
Total	1586.0	23			

Tabla 4. *ANOVA - Etapa 2*

Fuentes	Suma Cuadrados	Grados Libert.	Var.	F	p-valor
Hidrocarburos	162.9	3	54.31	3.35	.0555
Hidrógeno	1076	2	537.9	33.19	.0000
Interacción	94.94	6	15.82	0.976	.9762
Residual	194.5	12	16.21		
Total	1528	23			

¿Se puede concluir que en las dos etapas la varianza del error experimental es la misma? (Realiza el contraste con $\alpha = 0.05$)

2.8 Se ha estudiado el efecto de tres hornos diferentes y dos temperaturas (290 ^oC y 320 ^oC) en la duración de cierto componente. Para cada combinación de horno y temperatura se ha replicado el experimento 3 veces. En la tabla siguiente se proporcionan las medias y desviaciones típicas (corregidas) de los datos de cada tratamiento.

	Temperatura ^{o}C			
	290 ^{o}C		**320 ^{o}C**	
	Media	Desv. T.	Media	Desv. T.
Horno 1	24.56	0.850	18.00	0.265
Horno 2	19.10	1.539	14.40	0.265
Horno 3	18.70	0.458	17.43	0.862

Contrasta si existe interacción entre los factores horno y temperatura ($\alpha = 0.05$).

2.9. Cierto Organismo Público (O.P.) encargado de certificar la composición de aleaciones de metales preciosos, debe seleccionar entre dos Laboratorios al más capacitado para la realización de futuros análisis de gran precisión. Para tomar la decisión les somete a la siguiente prueba: Prepara tres aleaciones A, B y C que contienen proporciones distintas de oro. De cada una de ellas envía cuatro muestras a cada uno de los dos laboratorios. Así pues, cada laboratorio recibe un lote de 12 muestras (codificadas) ordenadas aleatoriamente sin conocer como han sido obtenidas. Los resultados recibidos por el O.P. son (entre paréntesis las medias de las casillas) (archivo *laboratorios.txt*):

	Aleac. A		*Aleac. B*		*Aleac. C*	
Lab. I	10.96	11.03	10.95	11.00	11.07	11.01
	11.08	11.01	11.04	10.97	10.97	11.03
	(11.02)		*(10.99)*		*(11.02)*	
Lab. II	10.97	10.96	10.97	10.96	11.02	11.00
	10.94	10.95	10.97	10.98	11.01	11.01
	(10.955)		*(10.97)*		*(11.01)*	

1. Determinar si existen diferencias entre los resultados de los laboratorios y si éstos han encontrado diferencias entre las aleaciones.

2. Aceptando que los datos cumplen la hipótesis de normalidad, indicar si podemos aceptar que verifican el resto de las hipótesis del modelo y en caso negativo que medidas se deben adoptar para analizar los datos.

3. Realizar un test de razón de varianzas para contrastar que las varianzas de los dos laboratorios son iguales, sabiendo que las tres aleaciones tienen composición distinta. Interpretar el resultado.

4. El O.P. conoce exáctamente el porcentaje en oro de la aleación A (11 %), de la B (11.02 %) y de la C (11.04 %). Con esta información comparar los resultados de los laboratorios.

2.10 Un laboratorio de Análisis Clínicos ha adquirido un nuevo equipo (B) para medir el colesterol en la sangre de los enfermos. Para evaluar si el nuevo equipo está ajustado se decide analizar muestras de 5 enfermos que previamente han sido analizadas con otro equipo (A), dando como resultado

Enfermo	1	2	3	4	5	Media
Equipo A	215	305	247	221	286	254.8
Equipo B	224	312	251	232	295	262.8

Contrastar con $\alpha = 0.05$ existen diferencias entre los dos equipos. (archivo *colesterol.txt*)

2.11. El análisis de la varianza de un diseño en bloques aleatorizados proporciona los siguientes resultados: $VT = 232$, VE(factor) $= 156$, VE(bloque) $= 15$ y $VNE = 61$. El número de niveles del factor es 5 y el número de bloques 8. Construir la tabla ADEVA. ¿ Cuál sería el resultado del análisis si no se tiene en cuenta el efecto de los bloques ? Indicar en qué circunstancias es preferible cada uno de los modelos.

2.12. Se realiza un experimento para estudiar si la presencia de fluorita reduce el coste de fabricación de clinker de cemento en tres tipos diferentes de mezcla. Los resultados del mismo (en miles de pesetas por Tm) se muestran en la siguiente tabla (archivo $fluorita2.txt$):

FLUORITA	MI	MII	MIII	$\bar{y}_{i\bullet}$
0%	15.4	10.6	17.8	14.6
1%	10.3	5.5	10.9	8.9
2%	7.4	1.2	8.1	5.5
3%	10.7	6.5	9.6	8.9
4%	13.5	11.6	15.5	13.5
$\bar{y}$	11.4	7.1	12.4	

$$\sum_{i=1}^{5}\sum_{j=1}^{3} e_{ij}^2 = 10.2 \quad \bar{y}_{\bullet\bullet} = 10.3$$

1. (a) Determinar si el tipo de mezcla y el nivel de fluorita añadido influyen significativamente en el coste de fabricación. Se supone que no existe interacción entre los dos factores.
   (b) Contrastar que porcentaje de fluorita produce el menor coste del clinker.

2.13 Se ha realizado un experimento con dos factores cada uno de ellos con 3 niveles. El 20% de la variabilidad total está explicada por la interacción de los dos factores y el 40% de la variabilidad total es debida a la variabilidad residual. Determinar el número de replicaciones necesarias en cada tratamiento para que la interacción sea significativa con $\alpha = 0.01$. (Explicar el procedimiento de cálculo, dejando el resultado indicado en función de las tablas).

2.14 Sea un diseño factorial con 4 factores a 3, 4, 2 y 5 niveles. Calcular el número de parámetros totales correspondientes a efectos principales e interacciones de orden 2, 3 y 4.

2.15 Un centro ha realizado un experimento para mejorar la resistencia a la tensión de ciertos muelles de acero. En una etapa del proceso el muelle caliente se sumerge en aceite templado. Se han estudiado tres factores, A (temperatura del acero antes de la inmersión, con tres niveles), B (temperatura del baño de aceite, dos niveles) y C (concentración de carbono en el acero, dos niveles). El experimento se ha replicado tres veces. En la tabla se muestra la media y la varianza (corregida) para los tres datos de cada tratamiento.

A	B	C	$\overline{y}_i$	$\hat{s}_i^2$
1	1	1	40.2	0.25
1	1	2	61.1	2.68
1	2	1	35.9	2.43
1	2	2	57.1	4.44
2	1	1	49.0	3.49
2	1	2	70.3	7.77
2	2	1	46.7	5.08
2	2	2	67.6	1.03
3	1	1	41.9	4.27
3	1	2	62.7	11.41
3	2	1	37.1	1.33
3	2	2	60.3	6.13

1. (a) Dar un intervalo del 95 % de confianza para la varianza del error experimental, σ^2.

2. Indicar si los efectos principales de A, B y C son significativamente distintos de cero.

3. Dado σ^2, construir un intervalo que cumpla que la probabilidad de que $\hat{s}_i^2$ (la varianza muestral corregida de un tratamiento) esté contenido en él sea igual a 0.95. Sustituir σ^2 por su estimador y con ayuda de este intervalo, discutir si se puede rechazar la hipótesis de homocedasticidad de las observaciones.

2.16 Un estudio bioquímico ha valorado la cantidad de tres ácidos (a, b, c) en muestras extraídas a cuatro terneras (1, 2 ,3 y 4) de la misma raza. El análisis es bastante complejo y la determinación incluye un error de medida. ¿Se puede aceptar la hipótesis de que los tres ácidos se encuentran en la misma proporción en cada animal? Realiza el contraste con nivel de significación 0.05. (La variabilidad total es 41.90). (archivo *ultrasonidos.txt*)

1.

	a	**b**	**c**	Medias
1	11.0	11.4	12.7	11.7
2	9.8	10.8	13.7	11.43
3	7.5	10.6	11.5	9.87
4	7.9	7.6	10.1	8.53
Medias	9.05	10.1	12.0	**10.38**

OTROS EJEMPLOS

2.17. Treinta y seis adultos (18 hombres y 18 mujeres) son utilizados en un estudio para comparar los tensiómetros de tres fabricantes. Los sujetos de cada sexo son asignados de forma aleatoria en seis grupos de tres cada uno. A tres grupos de cada sexo se les mide la presión de la sangre nada más comenzar el experimento; a los otros tres grupos se les mide la presión después de diez minutos de descanso.

Los resultados son los siguientes:

	I		II		III	
	H	M	H	M	H	M
	147	122	156	131	127	110
1	124	142	127	133	122	115
	113	136	155	146	153	105
	140	108	100	141	114	103
2	130	151	140	125	139	135
	112	138	105	139	126	114

Conteste a las siguientes preguntas:

- ¿Existen diferencias entre los fabricantes en la medida de presión de la sangre?
- ¿Hay diferencia entre el descanso y el no descanso en la presión en la sangre?
- ¿Hay diferencia entre hombres y mujeres?
- Comprobar si hay interacción entre descanso y sexo.
- Comprobar las hipótesis de normalidad, homocedasticidad y homogeneidad.

En el archivo **tension.sf3** están la variable respuesta *presión* y las variables factores *descanso, fabricante* y *sexo.*

2.18 Se desea investigar el comportamiento de dos tipos de semilla y de tres tipos diferentes de fertilizante. Los resultados serán los diferentes rendimientos para las combinaciones de semillas y fertilizantes.

Se pide contestar a las siguientes preguntas:

- ¿Existen diferencias entre los fertilizantes?
- ¿Existen diferentes entre las semillas?
- Estudiar si la interacción entre las semillas y fertilizantes es significativa.
- Comprobar las hipótesis de normalidad, homocedasticidad e independencia e homocedasticidad.

En el archivo **rend.sf3** están la variable respuesta *rendimiento* y los factores *semilla* y *fertilizante.*

	A	B	C
1	14.3	18.1	17.6
	14.5	17.6	18.2
	11.5	17.1	18.9
	13.6	17.6	18.2
2	12.6	10.5	15.7
	11.2	12.8	17.5
	11.0	8.3	16.7
	12.1	9.1	16.6

2.19. Se ha realizado un experimento para estudiar la influencia de dos factores en el rendimiento de un proceso. Estos factores son la temperatura, que puede estar a tres niveles (alta, media y baja), y el catalizador, que puede ser el catalizador 1 o el catalizador 2. En el archivo **rend2.sf3** se presentan los resultados que se muestran en la siguiente tabla.

	Temperatura		
	Alta	**Media**	**Baja**
Catalizador 1	279	174	397
	172	277	348
	176	130	434
Catalizador 2	253	252	417
	238	367	427
	387	323	423

- ¿De qué modelo se trata?
- ¿Qué efectos son significativos?
- ¿Cuál es el tratamiento adecuado para obtener el mayor rendimiento?

2.20. Se ha realizado un experimento para estudiar las fuentes de variabilidad de la resistencia a la compresión de cemento tipo Portland. El cemento ha sido mezclado con agua por tres obreros diferentes (mezcladores) durante un tiempo fijo. Después, la resistencia de las probetas generadas ha sido medida por otros tres obreros diferentes (medidores). Cada mezclador ha generado doce probetas, que se han dividido en tres grupos de cuatro; cada uno de esos grupos de cuatro ha sido asignado a un medidor. Los datos obtenidos para la resistencia a la compresión de cada probeta, dados en libras por pulgada cuadrada, se proporcionan en la tabla siguiente y se encuentran en el archivo **portland.sf3**.

	Medidor 1	Medidor 2	Medidor 3
Mezclador 1	5280	4340	4160
	5520	4400	5180
	4760	5020	5320
	5800	6200	4600
Mezclador 2	4420	5340	4180
	5280	4880	4800
	5580	4960	4600
	4900	6200	4480
Mezclador 3	5360	5720	4460
	6160	4760	4930
	5680	5620	4680
	5500	5560	5600

- ¿Existen diferencias entre las resistencias dadas por los diferentes medidores? ¿y entre las probetas generadas por cada mezclador?
- ¿Es significativa, con nivel de significación del 5%, la interacción entre medidores y mezcladores?
- ¿Se cumplen las hipótesis del modelo?

2.21. Se está estudiando el rendimiento de un proceso químico. Se piensa que las dos variables más importantes pueden ser la presión y la temperatura. Se seleccionan tres niveles de cada factor. Los resultados del experimento son los siguientes:

	Presión		
Temperatura	200	215	230
Baja	90.4	90.7	90.2
Baja	90.2	90.6	90.4
Media	90.1	90.5	89.9
Media	90.3	90.6	90.1
Alta	90.5	90.8	90.4
Alta	90.7	90.9	90.1

Utilizando el archivo **proceso. sf3** conteste a las siguientes preguntas:

- ¿Qué conclusiones se pueden sacar de los datos?
- ¿Bajo qué condiciones podría operar este proceso?
- ¿Existe interacción entre temperatura y presión?
- Compruebe las hipótesis del modelo.

2.22. Se realiza un experimento para estudiar la influencia de la temperatura de operación y de tres tipos de cristal en la salida de luz de un osciloscopio medidas en lux. En el archivo lux.sf3 se encuentran los resultados obtenidos que se presentan a continuación:

	Temperatura		
Cristal	100	125	150
1	580	1090	1392
	568	1087	1380
	570	1085	1386
2	550	1070	1328
	530	1035	1312
	579	1000	1299
3	546	1045	867
	575	1053	904
	599	1066	889

- ¿Hay diferencia entre las temperaturas?
- ¿Hay diferencia en el cristal? ¿Cúal es el mejor?
- Estudie si existe interacción entre la temperatura y el cristal.

2.22 Para comprobar la diferencia de rendimientos entre las distintas variedades de avena se diseño un experimento con ocho variedades distintas. Como el terreno donde fueron plantadas las distintas variedades estaba en pendiente se pensó que podría afectar la situación de la planta en su rendimiento. Los resultados obtenidos en gramos fueron los siguientes:

	I	II	III	IV	V
1	296	357	340	331	348
2	402	390	431	340	320
3	437	334	426	320	296
4	303	319	310	260	242
5	469	405	442	487	394
6	345	342	358	300	308
7	324	339	357	352	220
8	488	374	401	338	320

Si no se tiene en cuenta el efecto de las diferentes condiciones del terreno, conteste a las siguientes preguntas:

- ¿Existen diferencias entre las variedades?
- ¿Cúal es la mejor y la peor?
- La variedad ocho es autóctona y la más empleada. La cinco es la más cara. Si tuvierá que elegir ¿cuál elegiría?
- Haga un contraste de las hipótesis del modelo: normalidad, homocedasticidad, homogeneidad e independencia.

Conteste todas las preguntas anteriores si se introduce la variable que tiene en cuenta el efecto del terreno.

2.23. Se desea comparar cuatro procedimientos de obtención de la penicilina (A, B, C y D); siendo la variable respuesta producción en kg.

Una materia prima, licor de maíz, se tiene en cuenta en el experimento. Se dispone de cinco muestras de licor de maíz. A continuación se presenta la tabla de los datos.

	A	B	C	D
1	89	88	97	94
2	84	77	92	79
3	81	87	87	85
4	87	92	89	84
5	79	81	80	88

- ¿Cómo afectan los procedimientos y la materia prima?
- ¿Cuál es el mejor procedimiento y materia prima?
- Realice la diagnosis del modelo

En el archivo **penicili.sf3** se encuentra la variable respuesta *cantidad*, el factor tratamiento y el bloque *mezcla*.

2.24. En 1986 IBM realizó una serie de experimentos en varios de sus sistemas para investigar el comportamiento de nuevos algoritmos para incorporar en la librería de funciones matemáticas de su compilador FORTRAN. En el archivo fortran.sf3 se encuentran el tiempo empleado por llamada para la ejecución (dado en μs) de cinco funciones escalares, que se proporcionan en la siguiente tabla. El tiempo se ha promediado en 10000 argumentos seleccionados aleatoriamente en los intervalos de interés ([-π,π],...). Las ejecuciones se llevaron a cabo en tres sistemas IBM diferentes (4331, 4361 y 4341). Se proporcionan también los nombres de las funciones escalares consideradas.

Función	**Sistema IBM**		
	4331	**4361**	**4341**
EDUM	9,90	3,07	4,88
ACOS CIRC $[-\pi, \pi]$	179,62	33,28	33,23
SEN LINEAL $[-\pi, \pi]$	105,72	24,13	27,08
EXP LINEAL $[-16, 16]$	254,82	39,14	37,46
D2DUM	13,47	4,63	5,72

- El interés principal del experimento era el estudio de la eficacia de los tres sistemas ¿ha resultado adecuada la estrategia?
- Realice la diagnosis del modelo y proponga posibles soluciones si detecta algún problema.

2.25 Unos alumnos de la universidad de Tuffs (Massachussets, E.U.A.), preocupados por el estado de corrosión de las tuberías de su universidad, decidieron realizar el siguiente experimento. Tomaron muestras de agua corriente haciendo variar los factores Campus, Tipo de edificio y antigüedad del edificio.

Se midió la concentración de hierro en el agua corriente (mg/dm^3) y para cada posible combinación de factores se tomaron dos observaciones. En el archivo **corrosio.sf3** se muestran los resultados que se presentan en la siguiente tabla.

	Factor		**Concentración de Fe**	
Antigüedad	**Tipo**	**Campus**		
Viejo	Académico	Medford	0,23	0,28
Nuevo	Académico	Medford	0,36	0,29
Viejo	Residencial	Medford	0,03	0,06
Nuevo	Residencial	Medford	0,05	0,02
Viejo	Académico	Somerville	0,08	0,05
Nuevo	Académico	Somerville	0,03	0,08
Viejo	Residencial	Somerville	0,04	0,07
Nuevo	Residencial	Somerville	0,02	0,06

- Identifique el modelo de que se trata, estime sus parámetros y realice la diagnosis.
- Si no se cumplieren las hipótesis del modelo indique qué podría hacerse para remediarlo.
- Estudie las interacciones e interprete las que resulten significativas.

Modelos de regresión lineal

REGRESION SIMPLE

1. La tabla muestra los mejores tiempos mundiales en Juegos Olímpicos hasta 1976 en carrera masculina para distintas distancias.

y: tiempo (sg)	9.9	19.8	44.26	103.5	214.9	806.4	1658.4	7795
x: distancia (m)	100	200	400	800	1500	5000	10000	42196

   (a) Estimar la regresión lineal de y sobre x y calcular la varianza residual y el coeficiente de correlación.

   (b) Obtener intervalos de confianza para la pendiente y varianza residual ($\alpha = 0.01$).

   (c) Analizar si la relación lineal es adecuada, transformando las variables si es necesario.

   (d) Supóngase que en aquellas Olimpiadas hubiera existido una carrera de 500 metros. Estimar el tiempo previsto para el record olímpico en dicha carrera, dando un intervalo de confianza con $\alpha = 0.05$.

2. Según la ecuación de los gases ideales, la presión ejercida por un gas a volumen y temperatura constante es proporcional a la masa. Se puede utilizar el siguiente procedimiento para estimar el peso molecular de un gas. Se almacena el gas en un recipiente de volumen constante, y se va soltando poco a poco gas, variando la presión, pero manteniendo la temperatura constante. En la tabla adjunta se proporcionan mediciones de la presión (con respecto a la atmosférica, 1 atm = 14.7 psi) y de la masa del gas para el árgon.

Presión (psi)	Masa (g)
52	1.028
49	0.956
44	0.880
39	0.793
34	0.725
29	0.645
25	0.593
21	0.526
19	0.500
19	0.442
11	0.373
0	0.210

   (a) Para estimar el peso molecular del árgon a partir de los datos, se propone el siguiente modelo de regresión

$$P_i = \beta_0 + \beta_1 m_i + u_i \text{ con } u_i \sim N(0, \sigma^2).$$

   Estimar los parámetros del modelo y contrastar si el término independiente es significativo.

(b) Se considera el modelo alternativo

$$P_i = \alpha m_i + u_i, \text{ con } u_i \sim N(0, \sigma^2).$$

Obtener el estimador de máxima verosimilitud del parámetro α, así como su varianza.

(c) Realizar el contraste $H_0 : \alpha = 50$ frente a $H_1 : \alpha \neq 50$ con nivel de significación 0.05.

(d) Para el segundo modelo, obtener un intervalo de predicción para la presión cuando la masa es igual a 1 gramo.

(e) Obtener la varianza del estimador de $E[P_h|m_h]$, es decir del valor medio de la presión P_h para una masa dada m_h con ambos modelos. Si el modelo verdadero fuese el del primer apartado, ¿qué efecto tendría sobre la predicción adoptar el modelo alternativo?

3. Sir Francis Galton (1877) estudió la relación entre la estatura de una persona (y) y la estatura de sus padres (x) obteniendo las siguientes conclusiones:

   (a) Existía una correlación positiva entre las dos variables.

   (b) Las estaturas de los hijos cuyos padres medían más que la media era, en promedio, inferior a la de sus progenitores, mientras que los padres con estatura inferior a la media en promedio tenían hijos más altos que ellos, calificando este hecho como de "regresión" a la media.

   Contrastar ($\alpha = 0.05$) estas dos conclusiones con la ecuación $\hat{y} = 17.8 + 0.91x$ resultante de estimar un modelo de regresión lineal entre las variables (en cm.) descritas anteriormente para una muestra de tamaño 100 si la desviación típica (estimada) de $\hat{\beta}_1$ es 0.04.

4. La ley de Hubble sobre la expansión del universo establece que dadas dos galaxias la velocidad de desplazamiento de una respecto a la otra es $v = Hd$, siendo d su distancia y H la constante de Hubble. La tabla proporciona la velocidad y la distancia de varias galaxias respecto a la Via Láctea. Se pide:

Galaxia	Distancia (millones años luz)	Velocidad ($10^3 Km/s$)
Virgo	22	1.21
Pegaso	68	3.86
Perseo	108	5.15
Coma Berenices	137	7.56
Osa Mayor 1	255	14.96
Leo	315	19.31
Corona Boreal	390	21.56
Géminis	405	23.17
Osa Mayor 2	700	41.83
Hidra	1100	61.14

Tabla: Distancia y velocidad de desplazamiento de las distintas galaxias a la Via Lactea.

Nota: Obsérvese que según el modelo de Hubble la regresión debe pasar por el origen. Tómese 1 año luz = 300 000 Km/seg x 31 536 000 seg = 9.46 10^{12} Km.

(a) Estimar por regresión la constante de Hubble.

(b) Como $T = d/v = d/Hd = 1/H$, la inversa de la constante de Hubble representa la edad estimada del Universo. Construir un intervalo de confianza del 95% para dicha edad .

5. Para establecer la relación entre el alargamiento en mm (Y) producido en un cierto material plástico sometido a tracción y la tensión aplicada en toneladas por cm^2 (X) se realizaron 10 experimentos cuyos resultados se muestran en la tabla

x_i	0.20	0.50	0.60	0.70	0.90	1.00	1.20	1.50	1.60	1.70
y_i	23	20	33	45	67	52	86	74	98	102

Tabla: Alargamiento y_i (mm) producidos por la tensión x_i (Tm/cm^2).

(a) Ajustar el modelo de regresión lineal $E(Y|x) = \beta_0 + \beta_1 x$ y contrastar ($\alpha = 0.01$) la hipótesis de que, en promedio, por cada Tm/cm^2 de fuerza aplicada es de esperar un alargamiento de 50 milímetros, sabiendo que la desviación típica residual vale 10.55.

(b) Si el límite de elasticidad se alcanza cuando $x = 2.2$ Tm/cm^2, construir un intervalo de confianza al 95% para el alargamiento medio esperado en ese punto.

(c) Teniendo en cuenta que el alargamiento esperado cuando la fuerza aplicada es nula debe ser nulo también, estimar el nuevo modelo $E[Y|x] = \beta x$ con los datos anteriores ¿Cuál es el sesgo del estimador del parámetro de la pendiente si se estima según el modelo del apartado 1?

6. Estimar por mínimos cuadrados los parámetros a y b de la ecuación $y = a + bx^2$ con la muestra de tres puntos siguientes (y, x) : (3, -1); (4, 0); (6,1).

7. La ecuación de regresión entre las ventas de un producto y y su precio x es $\hat{y} = 320 - 1.2x$, $\hat{s}_R = 2$ y $\hat{s}_y = 4$. Si el número de datos ha sido $n = 50$, contrastar $H_0 : \beta_1 = -1$ frente a la alternativa $H_1 : \beta_1 < -1$.

8. Se estudia la relación entre el tiempo de reparación (minutos) de ordenadores personales y el número de unidades reparadas en ese tiempo por un equipo de mantenimiento con los resultados mostrados en la siguiente tabla

unidades reparadas	1	3	4	6	7	9	10
tiempo de reparación	23	49	74	96	109	149	154

Se pide:

(a) Construir la recta de regresión para prever el tiempo de reparación y utilizarla para construir un intervalo de confianza ($\alpha = 0.01$) para el tiempo medio de reparación de 8 unidades.

(b) Construir un intervalo de confianza ($\alpha = 0.01$) del tiempo de reparación para un lote de 14 unidades.

(c) Si los tiempos de reparación fuesen medias de 10 datos. ¿Cual sería la recta de regresión?

REGRESION MULTIPLE

9. En la tabla se muestran los costes financieros mensuales en miles de euros (y) de 16 delegaciones de una gestora de inversiones, además se proporciona el número de nuevos préstamos del mes (x_1) y el número de préstamos pendientes (x_2).

n	x_1	x_2	y
1	80	8	2256
2	93	9	2340
3	100	10	2426
4	82	12	2293
5	90	11	2330
6	99	8	2368
7	81	8	2250
8	96	10	2409
9	94	12	2364
10	93	11	2379
11	97	13	2440
12	95	11	2364
13	100	8	2404
14	85	12	2317
15	86	9	2309
16	87	12	2328

(a) Estima la ecuación de regresión

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \text{ con } u_i \sim N(0, \sigma^2)$$

incluyendo la varianza del modelo.

(b) Realizar los contrastes individuales e interpretar los coeficientes.

(c) Realiza el contraste general de regresión o contraste de la F. Proporciona el p-valor.

(d) Proporciona la tabla con valores previstos y residuos.

(e) Comprueba las hipótesis del modelo.

10. Los fabricantes que utilizan rodamientos en sus productos tienen interés en la fiabilidad de estos componentes. La medida básica de fiabilidad se denomina rating life, y consiste en el número de revoluciones que soporta el 90% de los rodamientos antes de la fractura, a esto se denota por L10. Los modelos teóricos indica que este valor está relacionado con la carga (P) a la que se somete el rodamiento, el diámetro (D) del rodamiento y el número de bolas (Z) del mismo, mediante la ecuación:

$$L10 = \left(\frac{kZ^a D^b}{P}\right)^3.$$

Se desea comprobar experimentalmente esta ecuación, para lo cual se realizó un experimento con rodamientos de distintos fabricantes y tipos. Los datos se encuentran en el archivo (ballbearing.txt), en la tabla 1 se muestra los 10 primeros datos. La información que contiene es la siguiente:

Com: Codigo de empresa 1, 2, and 3

N: Número de ensayo (en cada empresa)

Year: Año del ensayo NA = No disponible

NB : Número de Rodamiento

P: Carga

Z: Número de bolas

D: Diámetro

L10: Percentil 10

L50: Percentil 50

Slope: Parámetro de la distribución Weibull

Btype: Tipo de rodamiento 1, 2, y 3 in la empresa 2; 0 en los demás casos.

Com	N	Year	NB	P	Z	D	L10	L50	Slope	Btype
1	1	1936	24	4240	8	.68750	19.200	84.50	1.27	0
1	2	1937	20	4240	8	.68750	26.200	74.20	1.81	0
1	3	1937	14	4240	8	.68750	11.100	68.10	1.04	0
1	4	1937	19	4240	8	.68750	11.800	66.80	1.09	0
1	5	1937	18	4240	8	.68750	13.500	79.40	1.06	0
1	6	1938	21	2530	9	.50000	5.800	25.70	1.27	0
1	7	1938	28	4240	8	.68750	18.300	44.70	2.10	0
1	8	1938	27	4240	8	.68750	5.620	73.20	0.73	0
1	9	1940	20	4240	8	.68750	15.800	82.70	1.14	0
1	10	1940	22	4240	8	.68750	8.700	41.60	1.20	0
...	...	...	...	...	...	...	...	...	...	...

(a) Estima el modelo

$$\log(L10_i) = \beta_0 + \beta_1 \log(Z_i) + \beta_2 \log(D_i) + \beta_3 \log(P_i) + u_i \text{ con } u_i \sim N(0, \sigma^2),$$

y realiza los contrastes individuales y el contraste general.

(b) Según el modelo, $\beta_3 = -3$. Realiza el contraste

$$\begin{aligned} H_0 &: \beta_3 = -3 \\ H_1 &: \beta_3 \neq -3 \end{aligned}$$

Proporciona el p-valor del contraste.

(c) Da un intervalo de confianza para los parámetros a y b del modelo teórico.

(d) Se definen las variables ficticias T_2 y T_3 para identificar los rodamientos tipo 2 y 3 del segundo fabricante (información en la variable Btype). Estima e interpreta el siguiente modelo de regresión:

$$\begin{aligned} \log(L10_i) = \; & \beta_0 + \beta_1 \log(Z_i) + \beta_2 \log(D_i) + \beta_3 \log(P_i) + \\ & \alpha_2 T_{2i} + \gamma_2 T_{2i} \times \log(Z_i) + \delta_2 T_{2i} \times \log(D_i) + \\ & \alpha_3 T_{3i} + \gamma_3 T_{3i} \times \log(Z_i) + \delta_3 T_{3i} \times \log(D_i) + u_i \end{aligned}$$

(e) Compara el modelo del apartado 1 con el modelo del apartado 4.

11. La matriz de varianzas de tres variables estandarizadas es la siguiente

$$\begin{bmatrix} 1 & 0.8 & 0.6 \\ 0.8 & 1 & 0.2 \\ 0.6 & 0.2 & 1 \end{bmatrix}$$

Calcular la ecuación de regresión de la primera variable respecto a las otras dos.

12. Dos variables x_1 y x_2 tienen la siguiente matriz de varianzas

$$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

y las regresiones simples con y son $\hat{y} = 0.75x_1$; $\hat{y} = 0.6x_2$. Calcular la regresión múltiple entre y y las dos variables x_1, x_2 sabiendo que la variable y tiene media cero y varianza unidad.

13. Para establecer la relación entre el voltaje de unas baterías y la temperatura de funcionamiento se han hecho unos experimentos cuyos resultados se muestran en la siguiente tabla

Batería	1	2	3	4	5	6	7	8
Temperatura	10	10	20	20	30	30	40	40
Voltaje	7.2	7.7	7.3	7.4	7.7	9.4	9.3	10.8

Se pide:

(a) Contrastar la hipótesis ($\alpha = 0.05$) de que no existe relación lineal entre el voltaje y la temperatura.

(b) Las lecturas 1,3,5 y 7 fueron realizadas con unas baterías de Cadmio y las 2,4, 6 y 8 con baterías de Zinc. Introducir en el análisis anterior una variable cualitativa que tenga en cuenta los dos tipos de baterías y contrastar si es significativa al 95%.

(c) Dar un intervalo de confianza para el voltaje de una batería de Cadmio que va a trabajar a 35º centígrados. (Utilizar el modelo estimado en el apartado 2).

(d) Comprobar que se cumplen las hipótesis del modelo construido en los apartados anteriores.

14. La variable y se relaciona con las variables x_1 y x_2 según el modelo $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$; no obstante se estima el siguiente modelo de regresión que no incluye la variable x_2

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i}.$$

Justificar en qué condiciones el estimador $\hat{\beta}_1$ es centrado.

15. Se efectúa una regresión con dos variables explicativas $E[y] = \beta_0 + \beta_1 x_1 + \beta_2 x_2$. La matriz de varianzas de x_1 y x_2 es

$$\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

¿Cuál de los dos estimadores $\hat{\beta}_1$ y $\hat{\beta}_2$ tendrá menor varianza?

16. Con los datos de la tabla, se pide:

x	-2	-2	-1	-1	0	0	1	1	2	2	3	3
y	1.1	1.3	2.0	2.1	2.7	2.8	3.4	3.6	4.0	3.9	3.8	3.6

(a) Estimar un modelo de regresión simple con y como variable dependiente y x como regresor. Indicar si el modelo es apropiado, justificando la respuesta.

(b) Estimar el modelo

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i$$

y realizar el contraste $H_0 : \beta_2 = 0$.

(c) Estimar el modelo

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + u_i$$

Realizar el contraste general de regresión con $\alpha = 0.01$. Seleccionar entre los tres el modelo más adecuado, justificando la respuesta.

17. Una de las etapas de fabricación de circuitos impresos requiere perforar las placas y recubrir los orificios con una lámina de cobre mediante electrólisis. Una característica esencial del proceso es el grosor de la capa de cobre. Se han realizado 12 experimentos para evaluar el efecto de 7 variables, X_1: Concentración de Cobre, X_2: Concentración de Cloruro, X_3: Concentración de Ácido, X_4: Temperatura, X_5: Intensidad, X_6: Posición y X_7: Superficie de la placa. Cada variable se ha estudiado a dos niveles. Las condiciones experimentales y los resultados de cada experimento se muestran en la tabla.

X_1	X_2	X_3	X_4	X_5	X_6	X_7	Y
1	1	-1	1	1	1	-1	2.13
1	-1	1	1	1	-1	-1	2.15
-1	1	1	1	-1	-1	-1	1.67
1	1	1	-1	-1	-1	1	1.53
1	1	-1	-1	-1	1	-1	1.49
1	-1	-1	-1	1	-1	1	1.78
-1	-1	-1	1	-1	1	1	1.80
-1	-1	1	-1	1	1	-1	1.93
-1	1	-1	1	1	-1	1	2.19
1	-1	1	1	-1	1	1	1.61
-1	1	1	-1	1	1	1	1.70
-1	-1	-1	-1	-1	-1	-1	1.43

Responder a las siguientes preguntas aplicando el modelo de regresión múltiple: matriz identidad de 8×8.

(a) Estimar el modelo de regresión múltiple

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7i} + u_i.$$

Obtener la descomposición de la variabilidad del modelo y realizar el contraste

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$$

frente a la hipótesis alternativa H_1: algún β_j es distinto de cero.

(b) Realizar cada uno de los contrastes individuales e indicar qué variables tienen efecto significativo.

(c) Eliminar del modelo del apartado 1 todas las variables no significativas. Estimar el modelo y contrastar sus coeficientes. Interpretar los resultados del experimento.

18. El molibdeno se añade a los aceros para evitar su oxidación, pero en instalaciones nucleares presenta el inconveniente de ser el causante de gran parte de los productos radioactivos. Se ha realizado un experimento para determinar el grado de oxidación del acero en función del porcentaje de molibdeno. Además se ha tenido en cuenta el efecto del tipo de refrigerante utilizado (R_1, R_2). Los resultados se muestran en la tabla.

Refrig.	Molibdeno (%) 0.5%	1%	1.5%	2%	Medias
R_1	26.2	23.4	20.3	23.3	23.3
R_2	34.8	31.7	29.4	26.9	30.7
R_1	33.2	31.3	28.6	29.3	30.6
R_2	43.0	40.0	31.7	33.3	37.0
Media	34.3	31.6	27.5	28.2	30.4

(a) Escribir un modelo de regresión que incluya el porcentaje de molibdeno y el tipo de refrigerante como regresores; estimar el modelo e indicar qué parámetros son significativos ($\alpha = 0.05$)).

(b) Los experimentos relativos a las dos primeras filas se realizaron en un tipo de instalación y los correspondientes a las dos últimas en otra distinta. Escribir un nuevo modelo que incluya este aspecto. Comprobar que este nuevo regresor está incorrelado con los dos anteriores. Estimar el nuevo modelo.

(c) Demostrar que en un modelo con los regresores incorrelados, la eliminación de uno de ellos no influye en el valor de los estimadores $\hat{\beta}_i, (i \neq 0)$ restantes. ¿ Influye en la varianza residual y en los contrastes ? Explicar este efecto en función de que el parámetro β del regresor eliminado sea o no nulo.

19. Sea x_1 la altura del tronco de un árbol y x_2 el diámetro del mismo en su parte inferior. El volumen y del tronco de árbol puede ser calculado aproximadamente con el modelo

$$y_i = \alpha x_{1i} x_{2i}^2 + u_i,$$

según el cual, el volumen del tronco es proporcional al volumen de un cono con las medidas x_{1i}, x_{2i}, siendo α el parámetro (desconocido) de proporcionalidad, más una componente de error aleatorio u_i. La tabla siguiente contiene los datos (en metros y metros cúbicos) correspondientes a una muestra aleatoria de 15 troncos de una variedad de pino.

Obs.	x_{1i}	x_{2i}	y_i
1	10,1	0,117	0,062
2	11,3	0,130	0,085
3	20,4	0,142	0,204
4	14,9	0,193	0,227
5	23,8	0,218	0,470
6	19,5	0,236	0,484
7	21,6	0,257	0,623
8	22,9	0,269	0,722
9	19,8	0,297	0,821
10	26,8	0,328	1,280
11	21,0	0,351	1,034
12	27,4	0,376	1,679
13	29,0	0,389	2,073
14	27,4	0,427	2,022
15	31,7	0,594	4,630

(a) Estimar α por máxima verosimilitud suponiendo que las variables u_i tienen distribución normal de media cero, con la misma varianza e independientes.

(b) Un tronco tiene una altura de 20 metros y un diametro de 0.25 metros, dar un intervalo de predicción de su volumen (95% de confianza).

(c) En el análisis de los residuos se observa que la varianza de los errores crece con el volumen del tronco. Para obtener homocedasticidad se propone el siguiente modelo transformado utilizando logaritmos neperianos,

$$\log y_i = \beta_0 + \beta_1 \log x_{1i} + \beta_2 \log x_{2i} + u_i$$

Contrastar (nivel de significación 0.05) si estos dos valores son aceptables.

(d) Con este modelo, dar un intervalo de predicción (95% de confianza) para el volumen del tronco del apartado 2.

20. Ciertas propiedades del acero se mejoran sumergiéndolo a alta temperatura ($T_0 = 1525$ ^oF) en un baño templado de aceite ($t_0 = 95$ ^oF). Para determinar la influencia de las temperaturas del acero y del baño de aceite en las propiedades finales del material se han elegido tres valores de la temperatura del acero y tres del baño de aceite,

$$\text{Temperatura acero } (T) \begin{cases} 1450\ ^oF \\ 1525\ ^oF \\ 1600\ ^oF \end{cases} \qquad \text{Temperatura aceite } (t) \begin{cases} 70\ ^oF \\ 95\ ^oF \\ 120\ ^oF \end{cases}$$

y se han realizado los siguientes experimentos:

x_{1i}	0	0	0	0	-1	1	-1	1	0	0	-1	1
x_{2i}	0	0	0	0	-1	-1	1	1	-1	1	0	0
y_i	49.2	49.4	47.0	49.5	28.2	88.6	54.9	31.3	59.2	43.6	41.9	58.0

dónde se ha utilizado la siguiente transformación (para simplificar cálculos)

$$x_{1i} = \frac{T_i - 1525}{75} \qquad \text{y} \qquad x_{2i} = \frac{t_i - 95}{25}.$$

Estimar el modelo de regresión

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + u_i$$

e indicar qué parámetros son significativos para nivel de significación 0.05. Estimar y contrastar el modelo anterior empleando las variables originales T_i y t_i.