

**Taller #1 Diseño de Experimentos 3007340 - Semestre 02 2021: Fecha de entrega Noviembre 16 de 2021**

**Nota: Además de lo que pida cada problema, especifique claramente lo siguiente:**

1. Característica o variable respuesta estudiada y unidad de medida
2. Estructura de tratamientos (factores de diseño controlables, tipos de efectos y niveles)
3. Unidad experimental y estructura de diseño (completamente aleatorizada o diseño en bloques y quién es el factor de bloqueo)
4. La presentación del taller deberá realizarse usando la plantilla para informes de talleres, siguiendo estrictamente las directrices que allí se dan.

1. Se desea investigar la influencia de la temperatura en el rendimiento de un proceso químico, en particular interesa investigar un rango de temperatura entre 60 a 120° C. Se tienen recursos para realizar 20 corridas experimentales:

- a. Los niveles de temperatura con los que se experimentan son 60, 65, 70 y 120° C, haciendo cinco replicaciones en cada nivel. ¿Considera que es adecuado el diseño experimental usado? Argumente, y de ser necesario proponga alternativas.
- b. El orden en que se decidieron hacer las corridas experimentales para facilitar el trabajo experimental fue: primero las cinco del nivel bajo de temperatura, luego las cinco del siguiente y así hasta finalizar ¿Es correcto lo que hicieron? Argumente.
- c. Para hacer el análisis estadístico, compararon mediante una prueba **t de Student** de dos en dos las medias en los niveles de temperatura y obtuvieron conclusiones ¿Es adecuado el análisis? Argumente y en su caso proponga alternativas.

2. En un centro de investigación se realiza un estudio para comparar varios tratamientos que al aplicarse previamente a los frijoles crudos reduzca su tiempo de cocción. Estos tratamientos son:

- T1: El control; no aplicar ningún tratamiento
- T2: Remojo en agua con bicarbonato de sodio
- T3: Remojo en agua con sal común
- T4: Remojo en agua con una combinación de bicarbonato y sal en proporciones iguales.

La variable respuesta fue el tiempo de cocción en minutos. Los datos recolectados se dan a continuación:

T1	T2	T3	T4
213	76	57	84
214	85	67	82
204	74	55	85
208	78	64	92
212	82	61	87
200	75	63	79
207	82	63	90

- a. ¿De qué manera el experimentador debe aleatorizar los experimentos y el material experimental?
- b. Dé ejemplos de factores que deben estar fijos durante las pruebas experimentales para que estos no afecten los resultados ni a las conclusiones.
- c. Obtenga los boxplots comparativos y el gráfico de medias. Interprete.
- d. Formule el modelo ANOVA, formule y pruebe la hipótesis de que los tiempos promedios de cocción de los tratamientos son iguales. Calcule la potencia alcanzada usando un nivel de significancia de 0.05.
- e. Haga agrupamiento de medias de tratamientos mediante Tukey, Duncan, LSD ¿se llegan a los mismos resultados? ¿cuál grupo contiene los mejores tratamientos?
- f. ¿Hay algún tratamiento mejor? ¿Cuál es el tiempo de cocción esperado para el mejor tratamiento? Formule y pruebe las comparaciones de medias o contrastes que considere pertinentes para responder a esta pregunta.
- g. Estime los efectos de los tratamientos, interprete el signo del efecto.
- h. Pruebe si se cumplen los supuestos del modelo: Use test de Levene y test de Bartlett para evaluar la varianza constante así como el debido análisis de residuales, gráfico de probabilidad y Shapiro Wilk para probar la normalidad

3. Los datos de la siguiente tabla representan el número de horas de alivio que proporcionan cinco marcas diferentes de tabletas contra el dolor de cabeza que se administran a 25 sujetos que sufren fiebres de 38°C o más.

Tableta				
A	B	C	D	E
5.2	9.1	3.2	2.4	7.1
4.7	7.1	5.8	3.4	6.6
8.1	8.2	2.2	4.1	9.3
6.2	6.0	3.1	1.0	4.2
3.0	9.1	7.2	4.0	7.6

Responda a las preguntas a) a h) planteadas en el problema anterior, pero formuladas en términos de los tratamientos y respuesta correspondientes a los datos aquí presentados.

4. Un ingeniero civil está interesado en estudiar la absorción media de humedad en cinco mezclas de concreto, denotadas de 1 a 5, que contienen diferentes porcentajes de impermeabilizante. Para realizar este estudio se ejecuta el siguiente experimento: i. se toman aleatoriamente 6 placas de cada tipo de mezcla; ii. las 30 placas seleccionadas se colocan aleatoriamente en un recinto y se exponen

durante 48 horas a la humedad, procurando condiciones de humedad homogéneas para las placas; iii. al finalizar las 48 horas, se mide la cantidad de humedad absorbida por cada placa. Los resultados obtenidos aparecen en la siguiente tabla.

Mezcla				
1	2	3	4	5
551	595	639	417	563
457	580	615	449	631
450	508	511	517	522
731	583	573	438	613
499	633	648	415	656
632	517	677	555	679

- Identifique la variable respuesta, la variable que define el factor y los tratamientos. Muestre cómo pudo proceder el experimentador en la distribución aleatoria de las placas dentro del recinto (haga un bosquejo detallado asumiendo un diseño físico particular del recinto y muestre la asignación aleatoria).
- Obtenga los boxplots comparativos y el gráfico de medias. Interprete.
- Utilice el modelo de análisis de varianza apropiado para probar diferencias entre las medias de tratamiento. Estime los efectos de cada uno de los tratamientos, interprete el signo del efecto y utilice esta información para seleccionar el tratamiento óptimo, si existe.
- Determine si el tamaño de muestra usado fue adecuado para detectar diferencias entre medias de al menos 5 unidades con una probabilidad del 90%, usando una significancia de 0.05.
- Agrupe las medias de tratamiento que tienen en promedio la misma respuesta utilizando los métodos vistos en clase
- Pruebe algunos contrastes de interés.
- Valide supuestos. Use test de Levene y test de Bartlett para evaluar la varianza constante así como el debido análisis de residuales, gráfico de probabilidad y Shapiro Wilk para probar la normalidad

5. Se realizó un experimento para determinar cuál tipo de batería no recargable para linternas era la más económica. En particular el experimentador estaba interesado en *comparar los tiempos de vida por unidad de costo* de la marca particular que él a menudo compraba con la marca del almacén donde usualmente compra. También deseaba conocer si valía la pena pagar dinero extra por baterías alcalinas respecto a las de larga duración. Dado que la vida de la bombilla de una linterna es considerablemente mayor a la de cualquier batería, se decide usar la misma bombilla para probar las baterías del experimento. Se mantuvieron aproximadamente constantes variables de ruido como la temperatura ambiental y las baterías fueron compradas con distintas fechas de fabricación y en distintos lugares con el fin de dar una representación amplia de las fechas de manufactura, así la variabilidad causada por tal factor sería medida como parte de la variabilidad natural. Las condiciones de corrida se mantuvieron constantes y cada batería fue usada en forma continua hasta agotarse, el tiempo fue registrado en minutos. El diseño se realizó en forma completamente aleatorizado y se tomaron como tratamientos los siguientes:

**Nivel tratamiento**

- alcalina, marca A, costo=\$0.985 dólares/unidad
- alcalina, marca del almacén, costo=\$0.935 dólares/unidad
- de larga duración, marca A, costo=\$0.520 dólares/unidad
- de larga duración, marca del almacén, costo=\$0.495 dólares/unidad

Los espacios de tiempo o turnos de prueba fueron asignados aleatoriamente a los tipos de batería para así determinar el orden en el cual las baterías iban a ser observadas, y se hicieron cuatro replicaciones de cada nivel. Se montó un circuito que conectaba una batería a una bombilla de linterna y se colocó un pequeño reloj en el circuito el cual paraba cuando se agotaba la batería. El tiempo transcurrido en el reloj fue tomado como la medición de la vida de la batería (minutos). Los datos según el orden de observación se muestran a continuación:

Tipo de batería	Tiempo de vida
1	602
2	863
1	529
4	235
1	534
1	585
2	743
3	232
4	282
2	773
2	840
3	255
4	238
3	200
4	228
3	215

- Postule el modelo ANOVA, haga la tabla ANOVA y pruebe la significancia del modelo. Determine si el tamaño de muestra usado era suficiente para detectar diferencias entre las medias de 250 minutos/\$ con un nivel de significancia de 0.05.
- Valide los supuestos del modelo. Tenga en cuenta que como se conoce el orden de las corridas, debe probar primero incorrelación, use para ello test Ljung-Box (consulte sobre esta prueba para que pueda formularla y aplicarla).
- Determine si ¿las baterías alcalinas son más económicas que las de larga duración? ¿Existe diferencias entre la vida medias por unidad de costo de las baterías de la marca A vs. las de marca del almacén? En este último ¿cuáles son más económicas?
- Verifique que para todas las comparaciones por pares, Tukey da I.C más cortos que el método de Scheffé (consulte sobre este método y vea manual de la librería agricolae) y haga el agrupamiento de medias por cada uno de estos dos métodos.
- Considere ahora como respuesta la vida de las baterías sin importar su costo. Suponiendo un modelo ANOVA de efectos fijos de un solo factor en un DCA, pruebe la normalidad y la igualdad de las varianzas. ¿Será que la transformación raíz cuadrada estabiliza varianza y nos lleva a las mismas conclusiones sobre cuál es la mejor batería cuando el análisis se hace en la escala original de los tiempos de vida?

6. Se utilizan cuatro laboratorios para realizar análisis químicos. Muestras del mismo material se mandan a los laboratorios para su análisis como parte del estudio para determinar si en promedio, dan los mismos resultados. Los resultados analíticos para los cuatro laboratorios son los siguientes:

Laboratorio			
A	B	C	D
58.7	62.7	55.9	60.7
61.4	64.5	56.1	60.3
60.9	63.1	57.3	60.9
59.1	59.2	55.2	61.4
58.2	60.3	58.1	62.3
$\bar{Y}_{1\bullet} = 59.66$	$\bar{Y}_{2\bullet} = 61.96$	$\bar{Y}_{3\bullet} = 56.52$	$\bar{Y}_{4\bullet} = 61.12$
$S_1^2 = 1.983$	$S_2^2 = 4.668$	$S_3^2 = 1.352$	$S_4^2 = 0.592$

- Identifique la estructura de tratamientos y de diseño de este estudio. Escriba el modelo ANOVA identificando claramente qué representa cada uno de sus términos, dé los supuestos y restricciones. ¿Qué se desea probar o comparar?
- Analice los boxplots comparativos de la distribución de las observaciones por niveles del factor. Extraiga conclusiones respecto a los laboratorios.
- Calcule la tabla ANOVA y realice el test de hipótesis asociado. ¿Qué se concluye acerca de los laboratorios?
- Utilice la prueba de Bartlett en el nivel de significancia del 0.05 para probar que las varianzas dentro de los laboratorios no son significativamente diferentes. Enuncie claramente las hipótesis de la prueba, el valor del estadístico de prueba, cómo se distribuye dicho estadístico, cómo se calcula el correspondiente valor P y a cuánto asciende tal valor. Concluya.
- Valide el supuesto de normalidad mediante el test Shapiro wilk. Enuncie claramente las hipótesis de la prueba, dé el valor observado del estadístico de la prueba y el correspondiente valor P. Concluya. Evalúe también con gráfico de probabilidad.
- Estime las medias de los análisis por laboratorio, y los efectos de cada laboratorio sobre el promedio global de los análisis, obtenga intervalos de confianza del 95% para estos efectos. Interprete los efectos estimados. ¿Si algún efecto no es significativo qué se puede decir acerca de la media de los resultados del respectivo laboratorio?
- Usando un nivel de significancia de 0.05, compare y agrupe los laboratorios usando el método de Tukey, el método LSD y el método de Duncan. ¿Se llega a las mismas conclusiones con estos tres métodos? Nota:  $q_{0.05}(4,16) = 4.046093$ ,  $t_{0.025,16} = 2.119905$  y valores críticos de rangos significativos de Duncan son:  $r_{0.05}(2,16) = 2.998$ ,  $r_{0.05}(3,16) = 3.144$ ,  $r_{0.05}(4,16) = 3.235$
- Formule y pruebe la significancia de los siguientes contrastes usando tanto la prueba F y la prueba t. Indique en cada método de prueba cómo se obtiene el estadístico de prueba, el valor de los estadísticos de prueba y el criterio de decisión que se debe usar:
  - Promedio de los resultados del laboratorio B vs. el promedio conjunto de los laboratorios A, C, y D;
  - Promedio de los resultados del laboratorio C vs. el promedio conjunto de los laboratorios A, y D;
  - Promedio de los resultados del laboratorio A vs. el promedio del laboratorio D.

7. La Cía Kenton Food deseaba probar 4 diseños de empaque para un nuevo cereal. 20 almacenes con aproximadamente igual volumen de ventas fueron seleccionados como las unidades experimentales. A cada almacén se le asignó aleatoriamente uno de los diseños de empaque, con cada diseño asignado a 5 almacenes. Se eligieron los almacenes de modo que resultaran comparables en ubicación, volumen de ventas, precios, cantidad y ubicación de espacios para estantes y esfuerzos promocionales especiales. Se observó las ventas durante el período de estudio y los resultados aparecen a continuación:

Diseño empaque	Almacén				
	1	2	3	4	5
1	11	17	16	14	15
2	12	10	15	19	11
3	23	20	18	17	18
4	27	33	22	26	28

- Analice gráficamente los datos.
- Realice el análisis de varianza o ANOVA y calcule la potencia alcanzada con un nivel de significancia de 0.05.
- Estime los efectos del factor de tratamientos y concluya.
- Haga las comparaciones de medias por Tukey y Duncan y concluya.
- Valide supuestos.
- ¿Son los diseños 3 y 4 mejores que los diseños 1 y 2? Formule el contraste apropiado y el test de hipótesis que corresponda a lo que se pregunta. Pruebe usando una significancia de 0.05.
- ¿Cuál diseño se debe recomendar y por qué?

8. Contenido de sodio (efectos fijos): Un investigador estudió el contenido de sodio en la cerveza fermentada en 6 marcas de cerveza manufacturadas en cierta compañía. Para realizar su estudio seleccionó aleatoriamente 8 latas o botellas de 12 onzas de cada marca y midió el contenido de sodio en miligramos de cada lata o botella.

Marca	Contenido de sodio (mg/botella)							
1	24.4	22.6	23.8	22.0	24.5	22.3	25.0	24.5
2	10.2	12.1	10.3	10.2	9.9	11.2	12.0	9.5
3	19.2	19.4	19.8	19.0	19.6	18.3	20.0	19.4
4	17.4	18.1	16.7	18.3	17.6	17.5	18.0	16.4
5	13.4	15.0	14.1	13.1	14.9	15.0	13.4	14.8
6	21.3	20.2	20.7	20.8	20.1	18.8	21.1	20.3

- Clasifique el diseño empleado en la investigación anterior, defina la variable respuesta, el factor y los niveles del factor.
- Plantee, realice e interprete la prueba de hipótesis central del experimento, con base en la tabla ANOVA
- Valide los supuestos del modelo. Use las gráficas de residuales, el test de Levene, el test de Bartlett y el test Shapiro Wilk
- Emplee el criterio de Tukey y de Duncan para formar grupos con medias estadísticamente iguales. Tenga en cuenta que  $q_{0.05}(6,42) = 4.221779$  y los valores críticos de rangos significativos de Duncan son  $r_{0.05}(2,42) = 2.8539$ ,  $r_{0.05}(3,42) = 3.001$ ,  $r_{0.05}(4,42) = 3.0973$ ,  $r_{0.05}(5,42) = 3.1668$ ,  $r_{0.05}(6,42) = 3.2199$ . ¿se llega a los mismos grupos de medias estadísticamente iguales? Considerando el análisis de los bloxplots ¿cuál método cree que da la agrupación de medias correcta?
- El fabricante de la cerveza marca 2 asegura que el contenido promedio de sodio de la cerveza que produce es menor que 9.5 miligramos por lata. A un nivel de significancia 0.05 ¿es cierta esta afirmación?
- El fabricante también afirma que el contenido promedio de las cervezas marca 3 y 4 es en promedio igual al de la marca 6. A un nivel de significancia 0.05 ¿es cierta esta afirmación?
- Estime los efectos y medias de contenido de sodio para las marcas 1, 2 y 6, dé los respectivos I.C del 95% e interprete.
- Determine la potencia alcanzada con un nivel significancia de 0.05 y determine si el tamaño de muestra usado era suficiente para rechazar el test ANOVA con una potencia de 0.9 cuando la diferencia entre cualesquier par de medias es de 10mg/botella.

9. En un experimento para estudiar el efecto de la cantidad de polvo de hornear en una masa de galletas sobre el aumento de la altura de las galletas, se probaron cuatro niveles de polvo de hornear y se hicieron cuatro galletas replicadas con cada nivel en un orden aleatorio. Los resultados se muestran a continuación

Cantidad de polvo de hornear (en cucharaditas)			
0.25	0.5	0.75	1.0
11.4	27.8	47.6	61.6
11.0	29.2	47.0	62.4
11.3	26.8	47.3	63.0
9.5	26.0	45.5	63.9

- Haga un análisis descriptivo de los datos
- Escriba un modelo estadístico para este experimento, explique los términos. ¿Cuáles son las suposiciones necesarias para un análisis de varianza válido?
- Calcule el análisis de varianza y pruebe con una significancia de 0.05 la hipótesis de interés en este estudio. Escriba claramente las hipótesis, el estadístico de prueba y su valor, y el valor P.
- Realice análisis de residuos. Valide supuestos
- Formule un contraste de medias de tratamientos para probar la hipótesis de que el aumento en la altura de las galletas es una función lineal del incremento de la cantidad de polvo de hornear en la masa de las galletas, y pruebe esta hipótesis.
- Si la masa de las galletas se prepara por lotes y las cuatro mediciones con cada nivel de cantidad de polvo de hornear vinieran de un mismo lote ¿cómo cambiaría su respuesta en a)? ¿podrían estos datos analizarse de la misma manera si éste fuera el caso?
- Determine la potencia alcanzada con un nivel significancia de 0.05 y determine si el tamaño de muestra usado era suficiente para rechazar el test ANOVA con una potencia de 0.9 cuando la diferencia entre cualesquiera pares de medias es de 15 unidades?

10. Un investigador estudió el contenido de sodio en la cerveza rubia, seleccionando una muestra aleatoria de seis marcas de cervezas de E.U y Canadá vendidas en áreas metropolitanas. El investigador entonces eligió aleatoriamente de expendios al detal en el área, 8 envases o botellas de 12 onzas de cada marca seleccionada, y midió a cada una el contenido de sodio en mg.

Marca	Contenido de sodio (mg/botella)							
1	24.4	22.6	23.8	22.0	24.5	22.3	25.0	24.5
2	10.2	12.1	10.3	10.2	9.9	11.2	12.0	9.5
3	19.2	19.4	19.8	19.0	19.6	18.3	20.0	19.4
4	17.4	18.1	16.7	18.3	17.6	17.5	18.0	16.4
5	13.4	15.0	14.1	13.1	14.9	15.0	13.4	14.8
6	21.3	20.2	20.7	20.8	20.1	18.8	21.1	20.3

- Explique (proponga) el plan experimental para este problema
- Escriba el modelo de análisis de varianza apropiado para el experimento, identifique sus términos y los supuestos necesarios.
- Con base en el gráfico de cajas haga un análisis descriptivo y formule conclusiones preliminares sobre el efecto de las marcas en la variabilidad del contenido de sodio de las cervezas.
- Teniendo en cuenta el diseño aplicado construya la tabla ANOVA apropiada, y realice la prueba de hipótesis pertinente para este experimento.
- Estime él o los parámetros que en este problema pueden resultar de interés y construya en cada caso un intervalo de confianza del 95%. Interprete estas estimaciones.
- Valide supuestos

11. La cantidad de oxígeno disuelto en el agua es una medida de la contaminación de un río. A mayor cantidad de oxígeno menor contaminación. Para estudiar la variabilidad en la contaminación de un río producida por una planta industrial en un trayecto de 4 kilómetros se seleccionaron aleatoriamente 4 puntos, que resultaron ubicados así: el 1ro y el 2do, antes del sitio en el cual la planta industrial arroja al río el agua residual, a 1.70 y a 0.88 kilómetros respectivamente; el 3° y el 4° después del sitio de descarga, a 0.29 y a 1.45 kilómetros respectivamente. En cada uno de los 4 puntos se tomaron aleatoriamente 5 muestras de agua.

Punto	Contenido medio de oxígeno disuelto				
L_1	5.9	6.1	6.3	6.1	6.0
L_2	6.3	6.6	6.4	6.4	6.5
L_3	4.8	4.3	5.0	4.7	5.1
L_4	6.0	6.2	6.1	5.8	5.9

Resuelva los mismos literales a) a f) formulados en el problema anterior, pero en términos del factor y respuesta del presente conjunto de datos. Adicionalmente, explique en detalle cómo se podría haber llevado a cabo un DCA con efectos fijos en este problema y discuta sobre si es más apropiado o no aplicar un diseño con efectos fijos en lugar del diseño con efectos aleatorios seguido en este experimento.

12. En las pruebas de anticuerpos de VIH en muestras sanguíneas de pacientes, se utiliza un espectrómetro el cual determina la densidad óptica de cada muestra. La densidad óptica se mide como la absorbencia de luz en una longitud de onda particular. La muestra sanguínea es positiva si excede cierto valor máximo que se determina mediante las muestras de control para esa prueba. Los investigadores se interesan en la comparación de la variabilidad del laboratorio para los valores de control positivos. Los siguientes datos representan los valores de control positivos para 10 diferentes pruebas en cuatro laboratorios seleccionados al azar.

	Laboratorio			
corrida	1	2	3	4
1	0.888	1.065	1.325	1.232
2	0.983	1.226	1.069	1.127
3	1.047	1.332	1.219	1.051
4	1.087	0.958	0.958	0.897
5	1.125	0.816	0.819	1.222
6	0.997	1.015	1.140	1.125
7	1.025	1.071	1.222	0.990
8	0.969	0.905	0.995	0.875
9	0.898	1.140	0.928	0.930
10	1.018	1.051	1.322	0.775

Resuelva los mismos literales a) a f) formulados en el problema 11, pero en términos del factor y respuesta del presente conjunto de datos.

13. Un fabricante sospecha que los lotes de materia prima despachados por su proveedor difieren significativamente en el contenido de calcio. Actualmente en la bodega hay en inventario un gran número de lotes almacenados de este material. El fabricante decide escoger al azar cinco de estos lotes para un estudio. Un químico realiza cinco determinaciones del contenido de calcio sobre cada uno de los lotes y obtiene los siguientes datos:

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	23.39	23.38

- Proponga el plan experimental para este problema
- Escriba el modelo de análisis de varianza apropiado para el experimento, identifique sus términos y los supuestos necesarios.
- Con base en el gráfico de cajas haga un análisis descriptivo y formule conclusiones preliminares sobre la posible significancia de la variación en el contenido de calcio, de lote a lote.
- Realice el test anova pertinente a este experimento y concluya sobre significancia de la variación en el contenido de calcio, de lote a lote
- Halle un intervalo de confianza del 95% para proporción de varianza debida a los lotes e interprete los resultados
- Valide supuestos

14. Realice el siguiente problema: Los siguientes datos muestran la producción de cuatro operadores seleccionados aleatoriamente, en una máquina particular:

Operador			
1	2	3	4
175.4	168.5	170.1	175.2
171.7	162.7	173.4	175.7
173.0	165.0	175.7	180.1
170.5	164.1	170.7	183.7
$\bar{Y}_{1\bullet} = 172.650$	$\bar{Y}_{2\bullet} = 165.075$	$\bar{Y}_{3\bullet} = 172.475$	$\bar{Y}_{4\bullet} = 178.675$
$S_1^2 = 4.40333$	$S_2^2 = 6.109167$	$S_3^2 = 6.682500$	$S_4^2 = 16.069167$

- Haga un análisis gráfico de los datos mediante box plots comparativos de la producción según operadores seleccionados.
- ¿Cuál es el modelo ANOVA apropiado para estos datos? Escríbalo identificando cada término y los supuestos correspondientes.
- Pruebe la hipótesis asociada al modelo correspondiente al 5% de significancia. Especifique claramente la hipótesis nula y alternativa, el estadístico de prueba y su valor, el criterio de decisión y concluya en términos del problema.
- Estime la componente de varianza de los operadores y la componente de varianza del error experimental.
- Verifique los supuestos respecto a los errores del modelo.
- Determine si se tienen muestras suficientemente grandes para asegurar con una probabilidad de al menos 0.95 una componente de varianza significativa debida a los operadores, si la  $\sigma_\alpha^2$  real es  $1.5\sigma^2$ ? Si no, ¿cuántas corridas se necesitan por operador? Utilice una significancia de 0.05.
- Valide supuestos sobre los errores del modelo.

15. Una aleación de metal se manufactura en un proceso de moldeo a alta temperatura; cada molde se descompone en barras individuales más pequeñas que se usan en aplicaciones que requieren menores cantidades de la aleación. La fuerza de tensión de la aleación es crítica para su uso futuro, por lo que el proceso de moldeo está diseñado para producir barras con una fuerza de tensión superior en promedio a las especificaciones mínimas. Es aceptable alguna variación en esta fuerza sólo cuando un pequeño número de las barras no cumple con dichas especificaciones. Sin embargo, una variación excesiva tendrá como resultado que una cantidad inaceptable de las barras no las cumplan. Dos componentes aportan a la variabilidad total de la fuerza de tensión de las barras fabricadas: la variabilidad entre los moldes de fabricación y las alteraciones en el proceso de moldeo que afectan a las barras del mismo molde. Para mantener el control sobre la variación se requiere conocer las contribuciones a la variabilidad de cada parte del proceso. Se planeó un experimento para aislar la variación en la fuerza de tensión debida a los efectos de los distintos moldes de la atribuible a las alteraciones dentro del mismo molde. Se tomaron los moldes a alta temperatura de la aleación de tres corridas de producción elegidas aleatoriamente en la misma fábrica; cada molde se descompuso en barras individuales y se obtuvieron las mediciones destructivas de la fuerza de tensión en una muestra aleatoria de 10 barras de cada uno de los tres moldes. Los datos aparecen a continuación en psi:

MOLDE		
1	2	3
88.0	85.9	94.2
88.0	88.6	91.5
94.8	90.0	92.0
90.0	87.1	96.5
93.0	85.6	95.6
89.0	86.0	93.8
86.0	91.0	92.5
92.9	89.6	93.2
89.0	93.0	96.2
93.0	87.5	92.5

Los tres moldes usados en el estudio representan una muestra de la población potencial de moldes que pueden producirse en esa fábrica; los investigadores estaban interesados en **la variación** de la fuerza de tensión entre los moldes producidos en ella; la preocupación no está en los tres moldes específicos del experimento. Por ello, consideraron los moldes sólo como una muestra aleatoria de los producidos en la instalación. Las inferencias se extenderán a la población de moldes que pudieran producirse en esa fábrica.

- ¿Puede el anterior experimento modelarse como un modelo ANOVA de un factor efectos fijos? ¿por qué? En caso de que el modelo ANOVA no sea de efectos fijos. ¿Cómo lo re enunciaría para que lo fuera?
- ¿Sería apropiado llevar a cabo un test de Tukey o formular contrastes de medias o cualquier tipo de comparación de medias, así como estimar efectos del factor involucrado en este experimento? ¿por qué?
- Calcule el ANOVA, analice según el modelo apropiado.
- Usando los resultados del experimento calcule la potencia alcanzada, con una significancia del 0.05. Interprete este resultado. Calcule además el tamaño de muestra necesario para alcanzar una potencia de 0.90 a un nivel de significancia de 0.05 si se quiere rechazar el test ANOVA cuando  $\sigma_a^2 = 1.5\sigma^2$ .
- ¿En qué se diferencian las ANOVA's para el modelo de efectos fijos de la del modelo de efectos aleatorios? Estime los parámetros de interés en este problema e interprételos.
- Realice el análisis de residuales y valide supuestos

16. Un patólogo de plantas tomó 4 muestras de 3 lbs. cada una, de lotes de 50 toneladas de semilla de algodón acumulada en varias cosechas durante la temporada de limpia. Las muestras se analizaron en el laboratorio para buscar aflatoxin, que es una toxina producida por organismos asociados con las semillas. Los datos se muestran a continuación:

	Concentración de Aflatoxin presente(pp billón)			
	muestra			
Lotes	1	2	3	4
LT1	39	57	63	66
LT2	56	13	25	31
LT3	64	83	88	71
LT4	29	55	21	51
LT5	38	66	53	81
LT6	11	49	34	10
LT7	23	0	5	20
LT8	10	11	23	37

- Suponga que los lotes y sus muestras son obtenidos aleatoriamente. Escriba el modelo para el estudio, explique los términos, muestre los cuadrados medios esperados y haga la prueba de hipótesis pertinente a este diseño
- Analice los box plots comparativos. ¿Qué puede concluirse a partir de éste?
- Estime las componentes de la varianza para los lotes y para las muestras. Calcule la varianza total estimada para una observación individual y la proporción de la variación total de aflatoxin que se puede atribuir a la variación entre lotes y entre muestras respectivamente.
- Usando los resultados del experimento calcule la potencia alcanzada, con una significancia del 0.05. Interprete este resultado. Calcule además el tamaño de muestra necesario para alcanzar una potencia de 0.90 a un nivel de significancia de 0.05 si se quiere rechazar el test ANOVA cuando  $\sigma_a^2 = 1.5\sigma^2$
- ¿Cuál es el valor estimado de la respuesta promedio global? ¿Cómo se distribuye dicho estimador? ¿Cuál es la estimación del error estándar de este estimador?

17. Un investigador estudió los efectos de tres dietas experimentales que varían los contenidos de grasa sobre el nivel total de lípidos en el plasma.

El nivel total de lípidos es una variable predictora de enfermedades coronarias ampliamente usada. Quince sujetos varones quienes estaban dentro del 20% del peso corporal ideal fueron agrupados en cinco bloques de acuerdo a la edad. Dentro de cada bloque las tres dietas experimentales fueron aleatoriamente asignadas a los tres sujetos. Los datos sobre la reducción del nivel de lípidos (en gramos por litro) obtenida después de que los sujetos estuvieron a dieta por un periodo de tiempo fijo se muestran a continuación:

	Contenido de Grasa en la Dieta		
Edades	Extremadamente Baja	Medianamente Baja	Moderadamente Baja
15 - 24	0.73	0.67	0.15
25 - 34	0.86	0.75	0.21
35 - 44	0.94	0.81	0.26
45 - 54	1.40	1.32	0.75
55 - 64	1.62	1.41	0.78

- Describa claramente la estructura de tratamientos, la estructura de diseño aplicado y la variable respuesta medida. ¿Se debe considerar o no interacción entre la dieta y las edades? .
- Haga un análisis descriptivo de los datos (según dieta, y según edades).
- Escriba un modelo estadístico para este experimento, explique los términos. ¿Cuáles son las suposiciones necesarias para un análisis de varianza válido?
- Calcule el análisis de varianza y pruebe con una significancia de 0.05 la hipótesis de interés en este estudio. Escriba claramente las hipótesis, el estadístico de prueba y su valor, y el valor P.
- Valide los supuestos de varianza constante con gráficos de residuales estandarizados (vs. valores ajustados, dietas y edades) y normalidad con test de Shapiro Wilk y gráfico de probabilidad con residuales estandarizados. Escriba las hipótesis y pruebas correspondientes. Concluya.
- Agrupe las medias del factor de interés mediante el método de Tukey. Interprete en términos del problema.
- Evalúe la eficiencia del diseño empleado mediante la pseudo prueba F sobre el factor de bloqueo al 5% de significancia y mediante la fórmula de la EER. ¿Valió la pena bloquear? Explique qué se ganó o qué se perdió, según la eficiencia del diseño.
- Comparada con las otras dos dietas, ¿es conveniente una dieta extremadamente baja en grasas? Plantee y estime el contraste apropiado y haga la prueba pertinente para responder a la pregunta. Dé conclusiones y recomendaciones de acuerdo al problema y objetivos del estudio.

18. En una empresa lechera se tienen cinco silos para almacenar leche (cisternas de 60000 litros). Un aspecto crítico para que se conserve la leche es la temperatura de almacenamiento. Se sospecha que en algunos silos hay problemas, por ello durante cinco días se decide registrar la temperatura a cierta hora crítica. Obviamente la temperatura de un día a otro es una fuente de variabilidad que podría impactar la variabilidad total. Los datos se dan a continuación,

	Día				
Silo	1: Lunes	2: Martes	3: Miércoles	4: Jueves	5: Viernes
A	4.0	4.0	5.0	2.5	3.0
B	5.0	6.0	2.0	4.0	4.0
C	4.5	4.0	3.5	2.0	3.0
D	2.5	4.0	6.5	4.5	4.0
E	4.0	4.0	3.5	2.0	4.0

- Describa claramente la estructura de tratamiento, la estructura de diseño aplicado y la variable respuesta medida. ¿Se debe considerar o no interacción entre los silos y los días? Realice además una descripción de cómo se llevaría a cabo este experimento.
- Haga un análisis descriptivo de los datos (según silo, y según días).
- Escriba un modelo estadístico para este experimento, explique los términos. ¿Cuáles son las suposiciones necesarias para un análisis de varianza válido?
- Calcule el análisis de varianza y pruebe con una significancia de 0.05 la hipótesis de interés en este estudio. Escriba claramente las hipótesis, el estadístico de prueba y su valor, y el valor P.
- Valide los supuestos de varianza constante con gráficos de residuales estandarizados (vs. valores ajustados, silos y días) y normalidad con test de Shapiro Wilk y gráfico de probabilidad con residuales estandarizados. Escriba las hipótesis y pruebas correspondientes. Concluya.
- Agrupe las medias del factor de interés mediante el método de Tukey. Interprete en términos del problema.
- Evalúe la eficiencia del diseño empleado mediante la pseudo prueba F sobre el factor de bloqueo al 5% de significancia y mediante la fórmula de la EER. ¿Valió la pena bloquear? Explique qué se ganó o qué se perdió, según la eficiencia del diseño.
- Realice un contraste para comparar la temperatura de almacenamiento de los silos B y D vs. los silos A, C y E. Dé conclusiones y recomendaciones de acuerdo al problema y objetivos del estudio teniendo en cuenta además que la temperatura de refrigeración no debe exceder los 4°C.



19. Una compañía farmacéutica realizó un experimento para estudiar los tiempos promedios (en días) necesarios para que una persona se recupere de los efectos y complicaciones que siguen a un resfriado común. En este experimento se compararon los efectos de distintas dosis diarias de vitamina C (Tratamiento A, B, C). Para hacer el experimento se contactó a un número determinado de personas que en cuanto les daba el resfriado empezaban a recibir algún tipo de dosis. Se consideró la edad de éstas como una posible fuente de variabilidad, así que las personas fueron clasificadas en cuatro grupos. A continuación se muestran los datos.

Grupo de Edad	Tratamiento (dosis diaria de vitamina C)		
	A	B	C
1	3	7	4
2	4	9	6
3	2	3	3
4	6	10	7

- Describa claramente la estructura de tratamiento, la estructura de diseño aplicado y la variable respuesta medida. ¿Se debe considerar o no interacción entre la dosis de vitamina C y la edad? Realice además una descripción de cómo se llevaría a cabo este experimento.
- Haga un análisis descriptivo de los datos (según dosis de vitamina C, y según grupo de edad).
- Escriba un modelo estadístico para este experimento, explique los términos. ¿Cuáles son las suposiciones necesarias para un análisis de varianza válido?
- Calcule el análisis de varianza y pruebe con una significancia de 0.05 la hipótesis de interés en este estudio. Escriba claramente las hipótesis, el estadístico de prueba y su valor, y el valor P.
- Valide los supuestos de varianza constante con gráficos de residuales estandarizados (vs. valores ajustados, dosis de vitamina y grupo de edad) y normalidad con test de Shapiro Wilk y gráfico de probabilidad con residuales estandarizados. Escriba las hipótesis y pruebas correspondientes. Concluya.
- Agrupe las medias del factor de interés mediante el método de Tukey. Interprete en términos del problema.
- Evalúe la eficiencia del diseño empleado mediante la pseudo prueba F sobre el factor de bloqueo al 5% de significancia y mediante la fórmula de la EER. ¿Valió la pena bloquear? Explique qué se ganó o qué se perdió, según la eficiencia del diseño.
- Determine si con el tratamiento B la recuperación es más lenta en relación a los otros dos tratamientos, usando un contraste apropiado. Dé conclusiones y recomendaciones de acuerdo al problema y objetivos del estudio.

20. Una empresa contable, previamente a la introducción en la firma de un entrenamiento generalizado en muestreo estadístico para auditoria, probó tres métodos de entrenamiento: (1) Estudio en casa con material de entrenamiento programado, (2) sesiones de entrenamiento en oficinas locales conducidos por la administración local, y (3) Sesión de entrenamiento en Chicago conducido por la administración nacional. Los 30 auditores a entrenar, fueron agrupados en 10 bloques cada uno de tres auditores, de acuerdo al tiempo transcurrido desde el grado universitario (el bloque 1 consiste de auditores graduados más recientemente, y el bloque 10 consiste de aquellos auditores graduados más distancientemente). Al final del entrenamiento, cada auditor fue evaluado en el análisis de un caso complejo que involucraba aplicaciones estadísticas; una medida del nivel de competencia basada en este análisis fue obtenida por cada auditor. Los resultados se muestran en la siguiente tabla:

bloque	Método		
	1	2	3
1	73	81	92
2	76	78	89
3	75	76	87
4	74	77	90
5	76	71	88
6	73	75	86
7	68	72	88
8	64	74	82
9	65	73	81
10	62	69	78

- ¿Por qué piensa que el factor de bloque "tiempo transcurrido desde el grado universitario" fue empleado?
- Escriba el modelo apropiado, explique sus términos y supuestos
- Analice los gráficos sobre los boxplots según método y según bloques de auditores. Establezca conclusiones preliminares respecto a los métodos de entrenamiento y los auditores
- Construya la tabla Anova para este experimento. Pruebe la significancia de los efectos de los métodos de entrenamiento indicando claramente la hipótesis nula y alternativa, el estadístico de prueba y el criterio de rechazo.
- Evalúe la eficiencia del bloqueo aplicado usando la pseudo prueba F (prueba de que existen efectos de bloque) y calculando la eficiencia relativa del diseño (fórmula de la EER). Concluya.
- Escriba la expresión general para el cálculo de las respuestas predicha y de los residuales. Realice análisis gráfico de residuales estandarizados y establezca si los supuestos del modelo sobre los errores se cumplen.
- Haga todas las comparaciones de pares de medias de los métodos de entrenamiento usando la prueba de Tukey. Establezca los grupos de medias y recomiende con base en estos resultados el método de entrenamiento más apropiado.

- h. Estime los efectos del factor de tratamientos mediante contrastes de medias, construya los intervalos de confianza del 95% para cada uno, y con base en estos pruebe la significancia de cada efecto. Dé conclusiones y recomendaciones de acuerdo al objetivo de este estudio

**21.** Un fabricante de cierta aleación de aluminio produce refinadores de textura en forma de lingotes. Los ingenieros de proceso sospechan (y desean comprobar) que la velocidad de agitación influye en el tamaño de grano del producto. La compañía manufactura el producto en cuatro hornos. Se sabe que cada horno tiene sus propias características de operación, de modo que los hornos se considerarán una variable problemática o de ruido en cualquier corrida experimental en la fundición que implique más de un horno. Cada horno puede operarse a cuatro velocidades de agitación distintas. Los datos de tamaño de grano resultantes son como sigue:

Velocidad de agitación (rpm)	Horno			
	1	2	3	4
5	16	7	6	10
10	20	17	13	10
15	21	11	12	17
20	24	11	16	17

- Describa claramente la estructura de tratamiento, la estructura de diseño aplicado, las unidades experimentales y la variable respuesta medida. ¿Se debe considerar o no interacción entre la velocidad de agitación y los hornos? Realice además una descripción de cómo se llevaría a cabo este experimento.
- Haga un análisis descriptivo de los datos (según velocidad de agitación y según hornos).
- Escriba un modelo estadístico para este experimento, explique los términos. ¿Cuáles son las suposiciones necesarias para un análisis de varianza válido?
- Calcule el análisis de varianza y pruebe con una significancia de 0.05 la hipótesis de interés en este estudio. Escriba claramente las hipótesis, el estadístico de prueba y su valor, y el valor P.
- Valide los supuestos de varianza constante con gráficos de residuales estandarizados (vs. valores ajustados, velocidad de agitación y horno) y normalidad con test de Shapiro Wilk y gráfico de probabilidad con residuales estandarizados. Escriba las hipótesis y pruebas correspondientes. Concluya.
- Defina los efectos de tratamientos como un contraste de medias de tratamientos y estímelos junto con un I.C.
- Asumiendo válidos supuestos sobre el error, agrupe las medias del factor de interés mediante el método de Tukey. Interprete en términos del problema.
- Evalúe la eficiencia del diseño empleado mediante la pseudo prueba F sobre el factor de bloqueo al 5% de significancia y mediante la fórmula de la EER. ¿Valió la pena bloquear? Explique qué se ganó o qué se perdió, según la eficiencia del diseño.
- ¿Qué recomendaría al ingeniero de proceso, si un tamaño de grano pequeño es deseable? Formule un contraste de medias de tratamiento para probar su recomendación.

**22.** Una empresa dedicada a la prueba de productos de consumo, realizó un experimento para comparar el consumo anual de energía de cinco marcas de deshumidificadores. Como el consumo de energía depende del nivel de humedad presente, se decidió monitorear cada marca en cuatro niveles de humedad (desde humedad moderada hasta intensa, codificadas por 1, 2, 3, y 4). Dentro de cada nivel de humedad se asignaron al azar las marcas entre cinco lugares seleccionados. Las observaciones resultantes (KWh anuales) fueron los siguientes:

Marcas	Nivel de humedad			
	1	2	3	4
1	685	792	838	875
2	722	806	893	953
3	733	802	880	941
4	811	888	952	1005
5	828	920	978	1023

- Describa claramente la estructura de tratamiento, la estructura de diseño aplicado, las unidades experimentales y la variable respuesta medida. ¿Se debe considerar o no interacción entre las marcas de los deshumidificadores y el nivel de humedad?
- Haga un análisis descriptivo de los datos (según marcas y según nivel de humedad).
- Escriba un modelo estadístico para este experimento, explique los términos. ¿Cuáles son las suposiciones necesarias para un análisis de varianza válido?
- Calcule el análisis de varianza y pruebe con una significancia de 0.05 la hipótesis de interés en este estudio. Escriba claramente las hipótesis, el estadístico de prueba y su valor, y el valor P.
- Valide los supuestos de varianza constante con gráficos de residuales estandarizados (vs. valores ajustados, marcas y nivel de humedad) y normalidad con test de Shapiro Wilk y gráfico de probabilidad con residuales estandarizados. Escriba las hipótesis y pruebas correspondientes. Concluya.
- Defina los efectos de tratamientos como un contraste de medias de tratamientos y estímelos junto con un I.C.
- Asumiendo válidos supuestos sobre el error, agrupe las medias del factor de interés mediante el método de Tukey. Interprete en términos del problema. ¿cuál marca recomendaría?
- Evalúe la eficiencia del diseño empleado mediante la pseudo prueba F sobre el factor de bloqueo al 5% de significancia y mediante la fórmula de la EER. ¿Valió la pena bloquear? Explique qué se ganó o qué se perdió, según la eficiencia del diseño.

- i. Formule un contraste apropiado y desarrolle test para probar que los deshumidificadores de marcas 4 y 5 exceden el consumo de los deshumidificadores de las marcas 2, y 3 en más de 60KWh anuales.

### Grupos y asignaciones:

APELLIDOS Y NOMBRE	CORREO	GRUPO	Asignación problemas
Galeano Arenas, Juan José	igaleanoa@unal.edu.co	1	1, 2, 10, 19
Galeano Muñoz, Simón Pedro	sigaleanom@unal.edu.co		
Gaviria Sanchez, Sebastian	segavirias@unal.edu.co		
Zuluaga Ayala, Santiago	sazuluagaay@unal.edu.co		
Alvarez Morales, Guillermo	gualvarezmo@unal.edu.co	2	1, 3, 11, 17
Betancur Palacio, Alejandra	albetancurpa@unal.edu.co		
Cañizalez Gil, Leyder Dain	ldcanizalezg@unal.edu.co		
Vásquez Gómez, Kleider Stiven	klsvasquezgo@unal.edu.co		
Cordoba Espinosa, Catherine Andrea	cacordobaec@unal.edu.co	3	1, 4, 12, 20
Duque Calle, David	daduqueca@unal.edu.co		
Gonzalez Ruiz, Kevin Mauricio	kmgonzalezr@unal.edu.co		
Naranjo Garcia, Yised Katerine	yknananjog@unal.edu.co		
Amaya Muñoz, Karen Andrea	kaamayam@unal.edu.co	4	1, 5, 13, 22
García Muñoz, Jhonatan Smith	jhsgarciamu@unal.edu.co		
Sánchez Henao, Ana María	amsanchezh@unal.edu.co		
Vergara Clavijo, Carolina	cavergaracl@unal.edu.co		
Agudelo Carmona, Valentina	vaagudeloc@unal.edu.co	5	1, 6, 14, 21
Palacios Duque, Sara	spalaciosd@unal.edu.co		
Rojas Bolaños, Carlos Arturo	caarojasbo@unal.edu.co		
Vasco Ruiz, Daniela	dvasco@unal.edu.co		
Bula Charry, Valentina	vbulac@unal.edu.co	6	1, 7, 15, 20
González Gallego, Santiago	sagonzalezga@unal.edu.co		
Pabón Palacio, Antonio	apabonp@unal.edu.co		
Zuluaga Serna, Juliana	jzuluagase@unal.edu.co		
Bula Isaza, Juan Daniel	jdbulai@unal.edu.co	7	1, 8, 16, 19
Carvajal Torres, Santiago	scarvajal@unal.edu.co		
Franco Valencia, Santiago	sfrancov@unal.edu.co		
Jaramillo Calle, David	djaramillo@unal.edu.co		
Florez Sanchez, Luisa Fernanda	lflorezs@unal.edu.co	8	1, 9, 14, 18
Pulgarin Carmona, Juan Camilo	jucpulgarinca@unal.edu.co		
Rios Garcia, Jhon Alexander	jhrhos@unal.edu.co		
Serrano Santos, Andrea	aserrano@unal.edu.co		
Gutiérrez Pulgarin, Alejandro	agutierrezpu@unal.edu.co	9	1, 5, 10, 22
Tineo Puerta, Alvaro Norbey	antineop@unal.edu.co		

**Advertencia:** Si se detecta copia entre grupos que tienen puntos en común, a todos los involucrados se les anulará esta evaluación (nota 0.0).

### Programas R lectura y gráfica boxplots datos problemas 2 a 22

```
library(gmodels)
library(multcomp)
library(daewr)
library(car)
library(outliers)
library(lsmmeans)
library(agricolae)
library(lme4)

rm(list=ls(all=TRUE))
datos2=data.frame(tratamiento=factor(factor(rep(c("T1","T2","T3","T4"),times=7))),Tiempo=scan())
213 76 57 84
214 85 67 82
204 74 55 85
208 78 64 92
212 82 61 87
200 75 63 79
207 82 63 90

attach(datos2)
mediast=sapply(split(Tiempo,tratamiento),mean)
boxplot(Tiempo~tratamiento,boxwex=0.5)
lines(mediast,col=2,type="b",pch=19)

#escriba aquí el código R que sea necesario

detach(datos2)
```

```
rm(list=ls(all=TRUE))
datos3=data.frame(Tableta=factor(rep(c("A", "B", "C", "D", "E"), times=5)), horas=scan())
5.2 9.1 3.2 2.4 7.1
4.7 7.1 5.8 3.4 6.6
8.1 8.2 2.2 4.1 9.3
6.2 6.0 3.1 1.0 4.2
3.0 9.1 7.2 4.0 7.6
```

```
datos3
attach(datos3)
mediasTab=sapply(split(horas, Tableta), mean)
boxplot(horas~Tableta, boxwex=0.5)
lines(mediasTab, col=2, type="b", pch=19)
```

#escriba aquí el código R que sea necesario

---

```
detach(datos3)
rm(list=ls(all=TRUE))
datos4=data.frame(Mezcla=factor(rep(1:5, times=6)), humedad=scan())
551 595 639 417 563
457 580 615 449 631
450 508 511 517 522
731 583 573 438 613
499 633 648 415 656
632 517 677 555 679
```

```
datos4
attach(datos4)
mediasMez=sapply(split(humedad, Mezcla), mean)
boxplot(humedad~Mezcla, boxwex=0.5)
lines(mediasMez, col=2, type="b", pch=19)
```

#escriba aquí el código R que sea necesario

---

```
detach(datos4)
rm(list=ls(all=TRUE))
datos5=data.frame(scan(what=list(Tipo="", Vida=0)))
1 602
2 863
1 529
4 235
1 534
1 585
2 743
3 232
4 282
2 773
2 840
3 255
4 238
3 200
4 228
3 215
```

```
datos5
attach(datos5)
costo=ifelse(Tipo=="1", 0.985, ifelse(Tipo=="2", 0.935, ifelse(Tipo=="3", 0.520, 0.495)))
Vida.costo=Vida/costo
raizVida=sqrt(Vida)
```

```
mediasbat1=sapply(split(Vida.costo, Tipo), mean)
boxplot(Vida.costo~Tipo)
lines(mediasbat1, col=2, type="b", pch=19)
```

```
mediasbat2=sapply(split(Vida, Tipo), mean)
boxplot(Vida~Tipo)
lines(mediasbat2, col=2, type="b", pch=19)
```

```
mediasbat3=sapply(split(raizVida, Tipo), mean)
boxplot(raizVida~Tipo)
lines(mediasbat3, col=2, type="b", pch=19)
```

#escriba aquí el código R que sea necesario

---

```
detach(datos5)
rm(list=ls(all=TRUE))
datos6=data.frame(Laboratorio=factor(rep(c("A", "B", "C", "D"), times=5)), medida=scan())
58.7 62.7 55.9 60.7
61.4 64.5 56.1 60.3
60.9 63.1 57.3 60.9
59.1 59.2 55.2 61.4
58.2 60.3 58.1 62.3
```

```
datos6
attach(datos6)
medias=sapply(split(medida, Laboratorio), mean); medias
vars=sapply(split(medida, Laboratorio), var); vars
```

```
boxplot(medida-Laboratorio,boxwex=0.3,xlab="Laboratorio")
lines(medias,col=2,type="b",pch=19)
```

#escriba aquí el código R que sea necesario

```
detach(datos6)
rm(list=ls(all=TRUE))
datos7=data.frame(diseño=factor(rep(1:4,each=5)),ventas=scan())
11 17 16 14 15
12 10 15 19 11
23 20 18 17 18
27 33 22 26 28
```

```
datos7
attach(datos7)
medias=sapply(split(ventas,diseño),mean);medias
```

```
boxplot(ventas-diseño,boxwex=0.3,xlab="diseño del empaque")
lines(medias,col=2,type="b",pch=19)
```

#escriba aquí el código R que sea necesario

```
detach(datos7)
rm(list=ls(all=TRUE))
datos8=data.frame(Marca=factor(rep(1:6,each=8)),ContSodio=scan())
24.4 22.6 23.8 22.0 24.5 22.3 25.0 24.5
10.2 12.1 10.3 10.2 9.9 11.2 12.0 9.5
19.2 19.4 19.8 19.0 19.6 18.3 20.0 19.4
17.4 18.1 16.7 18.3 17.6 17.5 18.0 16.4
13.4 15.0 14.1 13.1 14.9 15.0 13.4 14.8
21.3 20.2 20.7 20.8 20.1 18.8 21.1 20.3
```

```
datos8
attach(datos8)
mediasM=sapply(split(ContSodio,Marca),mean);mediasM
boxplot(ContSodio-Marca)
lines(mediasM,col=2,type="b",pch=19)
```

#escriba aquí el código R que sea necesario

```
detach(datos8)
rm(list=ls(all=TRUE))
datos9=data.frame(CantPolvoHornear=factor(rep(c(0.25,0.5,0.75,1.0),times=4)),AumentoAltura=scan())
11.4 27.8 47.6 61.6
11.0 29.2 47.0 62.4
11.3 26.8 47.3 63.0
9.5 26.0 45.5 63.9
```

```
datos9
attach(datos9)

mediasCantPolvoHornear=sapply(split(AumentoAltura,CantPolvoHornear),mean)
mediasCantPolvoHornear
```

```
boxplot(AumentoAltura-CantPolvoHornear,boxwex=0.5,xlab="CantPolvoHornear",col="bisque")
lines(mediasCantPolvoHornear,col=2,type="b",pch=19)
```

#escriba aquí el código R que sea necesario

```
detach(datos9)
rm(list=ls(all=TRUE))
datos10=data.frame(Marca=factor(rep(1:6,each=8)),ContSodio=scan())
24.4 22.6 23.8 22.0 24.5 22.3 25.0 24.5
10.2 12.1 10.3 10.2 9.9 11.2 12.0 9.5
19.2 19.4 19.8 19.0 19.6 18.3 20.0 19.4
17.4 18.1 16.7 18.3 17.6 17.5 18.0 16.4
13.4 15.0 14.1 13.1 14.9 15.0 13.4 14.8
21.3 20.2 20.7 20.8 20.1 18.8 21.1 20.3
```

```
datos10
attach(datos10)
mediasM=sapply(split(ContSodio,Marca),mean);mediasM
boxplot(ContSodio-Marca)
lines(mediasM,col=2,type="b",pch=19)
```

#escriba aquí el código R que sea necesario

```
detach(datos10)
rm(list=ls(all=TRUE))
datos11=data.frame(Punto=factor(rep(c("L1","L2","L3","L4"),each=5)),ContO2=scan())
5.9 6.1 6.3 6.1 6.0
6.3 6.6 6.4 6.4 6.5
4.8 4.3 5.0 4.7 5.1
6.0 6.2 6.1 5.8 5.9
```

```
datos11

attach(datos11)
mediasP=sapply(split(ContO2,Punto),mean);mediasP
```

```
boxplot(ContO2~Punto)
lines(mediasP,col=2,type="b",pch=19)
```

#escriba aquí el código R que sea necesario

---

```
detach(datos11)
rm(list=ls(all=TRUE))
datos12=data.frame(Lab=factor(rep(1:4,times=10)),VCposit=scan())
0.888 1.065 1.325 1.232
0.983 1.226 1.069 1.127
1.047 1.332 1.219 1.051
1.087 0.958 0.958 0.897
1.125 0.816 0.819 1.222
0.997 1.015 1.140 1.125
1.025 1.071 1.222 0.990
0.969 0.905 0.995 0.875
0.898 1.140 0.928 0.930
1.018 1.051 1.322 0.775
```

```
datos12
attach(datos12)
mediasL=sapply(split(VCposit,Lab),mean);mediasL
boxplot(VCposit~Lab)
lines(mediasL,col=2,type="b",pch=19)
```

#escriba aquí el código R que sea necesario

---

```
detach(datos12)
rm(list=ls(all=TRUE))
datos13=data.frame(Lote=factor(rep(1:5,times=5)),ContCalcio=scan())
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 23.39 23.38
```

```
datos13
attach(datos13)
mediasLot=sapply(split(ContCalcio,Lote),mean);mediasLot
boxplot(ContCalcio~Lote)
lines(1:5,mediasLot,col=2,type="b",pch=19)
```

#escriba aquí el código R que sea necesario

---

```
detach(datos13)
rm(list=ls(all=TRUE))
datos14=data.frame(operador=factor(rep(1:4,times=4)),producción=scan())
175.4 168.5 170.1 175.2
171.7 162.7 173.4 175.7
173.0 165.0 175.7 180.1
170.5 164.1 170.7 183.7
datos14
```

```
attach(datos14)
medias=sapply(split(producción,operador),mean);medias
vars=sapply(split(producción,operador),var);vars

boxplot(producción~operador,boxwex=0.3,xlab="operador")
lines(medias,col=2,type="b",pch=19)
```

#escriba aquí el código R que sea necesario

---

```
detach(datos14)
rm(list=ls(all=TRUE))
datos15=data.frame(MOLDE=factor(rep(1:3,times=10)),TENSION=scan())
88.0 85.9 94.2
88.0 88.6 91.5
94.8 90.0 92.0
90.0 87.1 96.5
93.0 85.6 95.6
89.0 86.0 93.8
86.0 91.0 92.5
92.9 89.6 93.2
89.0 93.0 96.2
93.0 87.5 92.5
```

```
datos15
attach(datos15)
medias=sapply(split(TENSION,MOLDE),mean);medias

boxplot(TENSION~MOLDE,boxwex=0.3,xlab="Molde")
lines(medias,col=2,type="b",pch=19)
```

#escriba aquí el código R que sea necesario

---

```
detach(datos15)
```

```

rm(list=ls(all=TRUE))
datos16=data.frame(Lote=factor(rep(c("LT1", "LT2", "LT3", "LT4", "LT5", "LT6", "LT7", "LT8"), each=4)), Aflatoxin=scan())
39 57 63 66
56 13 25 31
64 83 88 71
29 55 21 51
38 66 53 81
11 49 34 10
23 0 5 20
10 11 23 37

datos16
attach(datos16)
medias=sapply(split(Aflatoxin,Lote), mean);medias

boxplot(Aflatoxin~Lote, boxwex=0.5, ylab="Aflatoxin (ppb)", xlab="Lote")
lines(medias, type="b", lty=2, pch=19)

#escriba aquí el código R que sea necesario

detach(datos16)

```

---

```

rm(list=ls(all=TRUE))
datos17=data.frame(ContGrasa=factor(rep(c("ExtBaja", "MedBaja", "ModBaja"), times=5)), Edades=factor(rep(c("15-24", "25-34", "35-44", "45-54", "55-64"), each=3)), ReducLipidos=scan())
0.73 0.67 0.15
0.86 0.75 0.21
0.94 0.81 0.26
1.40 1.32 0.75
1.62 1.41 0.78

datos17

attach(datos17)

mediasContGrasa=sapply(split(ReducLipidos, ContGrasa), mean);mediasContGrasa
mediasEdades=sapply(split(ReducLipidos, Edades), mean);mediasEdades

boxplot(ReducLipidos~ContGrasa, boxwex=0.5, xlab="Contenido Grasa")
lines(mediasContGrasa, type="b", lty=2, pch=19)

boxplot(ReducLipidos~Edades, boxwex=0.5, xlab="Edades")
lines(mediasEdades, type="b", lty=2, pch=19)

#escriba aquí el código R que sea necesario

detach(datos17)

```

---

```

rm(list=ls(all=TRUE))
datos18=data.frame(Silos=factor(rep(c("A", "B", "C", "D", "E"), each=5)), Día=factor(rep(c("1", "2", "3", "4", "5"), times=5)), temperatura=scan())
4.0 4.0 5.0 2.5 3.0
5.0 6.0 2.0 4.0 4.0
4.5 4.0 3.5 2.0 3.0
2.5 4.0 6.5 4.5 4.0
4.0 4.0 3.5 2.0 4.0
datos18

attach(datos18)

mediasSilos=sapply(split(temperatura, Silos), mean);mediasSilos
mediasDía=sapply(split(temperatura, Día), mean);mediasDía

boxplot(temperatura~Silos, boxwex=0.5, xlab="Silos")
lines(mediasSilos, type="b", lty=2, pch=19, col=2)

boxplot(temperatura~Día, boxwex=0.5, xlab="Día")
lines(mediasDía, type="b", lty=2, pch=19, col=2)

#escriba aquí el código R que sea necesario

detach(datos18)

```

---

```

rm(list=ls(all=TRUE))
datos19=data.frame(DosisVitaC=factor(rep(c("A", "B", "C"), times=4)), GrupoEdad=factor(rep(1:4, each=3)), DiasRecup=scan())
3 7 4
4 9 6
2 3 3
6 10 7

datos19

attach(datos19)

mediasDosisVitaC=sapply(split(DiasRecup, DosisVitaC), mean);mediasDosisVitaC
mediasGrupoEdad=sapply(split(DiasRecup, GrupoEdad), mean);mediasGrupoEdad

boxplot(DiasRecup~DosisVitaC, boxwex=0.5, xlab="Dosis de Vitamina C")
lines(mediasDosisVitaC, type="b", lty=2, pch=19, col=2)

```

```
boxplot(DiasRecup~GrupoEdad,boxwex=0.5,xlab="Grupo de edad")
lines(mediasGrupoEdad,type="b",lty=2,pch=19,col=2)
```

#escriba aquí el código R que sea necesario

```
detach(datos19)
rm(list=ls(all=TRUE))
datos20=data.frame(Método=factor(rep(1:3,times=10)),bloque=factor(rep(1:10,each=3)),NivelCompet=scan())
73 81 92
76 78 89
75 76 87
74 77 90
76 71 88
73 75 86
68 72 88
64 74 82
65 73 81
62 69 78
```

```
datos20
```

```
attach(datos20)
```

```
mediasMétodo=sapply(split(NivelCompet,Método),mean);mediasMétodo
mediasbloque=sapply(split(NivelCompet,bloque),mean);mediasbloque
```

```
boxplot(NivelCompet~Método,boxwex=0.5,xlab="Método")
lines(mediasMétodo,type="b",lty=2,pch=19,col=2)
```

```
boxplot(NivelCompet~bloque,boxwex=0.5,xlab="Bloque")
lines(mediasbloque,type="b",lty=2,pch=19,col=2)
```

#escriba aquí el código R que sea necesario

```
detach(datos20)
rm(list=ls(all=TRUE))
datos21=data.frame(Velocidad=factor(rep(c(5,10,15,20),each=4)),Horno=factor(rep(1:4,times=4)),tamañograno=scan())
16 7 6 10
20 17 13 10
21 11 12 17
24 11 16 17
```

```
datos21
```

```
attach(datos21)
```

```
mediasVelocidad=sapply(split(tamañograno,Velocidad),mean)
mediasVelocidad
mediasHorno=sapply(split(tamañograno,Horno),mean)
mediasHorno
```

```
boxplot(tamañograno~Velocidad,boxwex=0.5,xlab="Velocidad",col="bisque")
lines(mediasVelocidad,type="b",lty=2,pch=19,col=2)
```

```
boxplot(tamañograno~Horno,boxwex=0.5,xlab="Horno",col="bisque")
lines(mediasHorno,type="b",lty=2,pch=19,col=2)
```

#escriba aquí el código R que sea necesario

```
detach(datos21)
rm(list=ls(all=TRUE))
datos22=data.frame(Marcas=factor(rep(1:5,each=4)),Nivelhumedad=factor(rep(1:4,times=5)),KWh=scan())
685 792 838 875
722 806 893 953
733 802 880 941
811 888 952 1005
828 920 978 1023
```

```
datos22
```

```
attach(datos22)
```

```
mediasMarcas=sapply(split(KWh,Marcas),mean)
mediasMarcas
mediasNivelhumedad=sapply(split(KWh,Nivelhumedad),mean)
mediasNivelhumedad
```

```
boxplot(KWh~Marcas,boxwex=0.5,xlab="Marcas",col="bisque")
lines(mediasMarcas,col=2,lty=1,type="b",pch=19)
```

```
boxplot(KWh~Nivelhumedad,boxwex=0.5,xlab="Nivelhumedad",col="bisque")
lines(mediasNivelhumedad,col=2,lty=1,type="b",pch=19)
```

#escriba aquí el código R que sea necesario

```
detach(datos22)
```