

PROBLEMAS DE ANOVA

Problema 1

La empresa de logística y transporte **INTEMPO S.A.** desea renovar su flota de vehículos pesados. Ha consultado tres proveedores: PEGASSUS, MERCEDENS, NISSEN. El precio de adquisición es un 20% más caro para PEGASSUS que los otros dos. Desde el departamento de Marketing se aconseja, ante la nueva situación económica, implementar una estrategia agresiva de reducción de precios que se mantenga durante varios años para no disminuir cota de mercado sino, en todo caso, incrementarla. La empresa considera que el mayor gasto corresponde al carburante. Por ello, quiere adquirir camiones cuyo consumo sea lo más pequeño posible. Los tres proveedores le han proporcionado los consumos medios (en litros) de 15 rutas que realiza la empresa y que ésta ha escogido aleatoriamente. La duración de las rutas ha estado dentro de los límites marcados por la empresa. Los datos se muestran en la siguiente tabla:

	Vehículo Tipo 1	Vehículo Tipo 2	Vehículo Tipo 3
Ruta	PEGASSUS	MERCEDENS	NISSEN
1	64	59	58
2	61	58	50
3	60	60	60
4	55	56	64
5	61	54	53
6	64	64	53
7	59	62	52
8	59	60	59
9	56	50	57
10	60	50	50
11	61	58	57
12	60	50	50
13	65	58	58
14	58	58	58
15	60	52	60

A continuación se detalla el estudio que la empresa ha llevado a cabo para adoptar sus decisiones

Tabla ANOVA para combustible según tipo_vehiculo

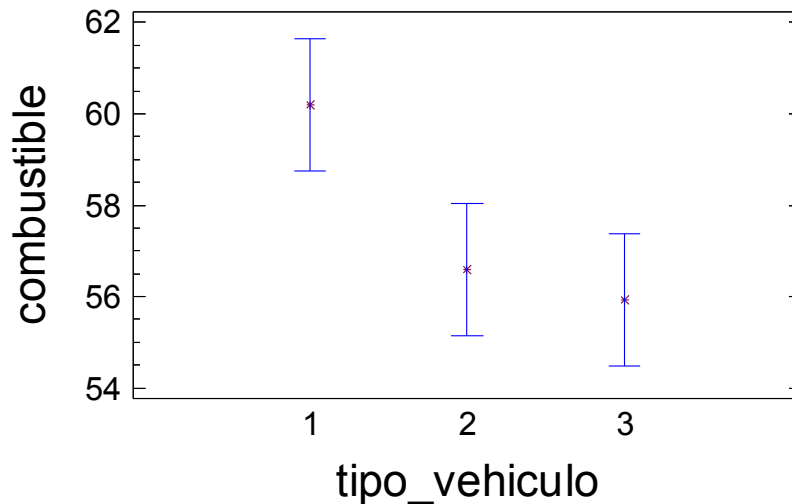
Análisis de la Varianza						
Fuente	Sumas de cuad.	Gl	Cuadrado Medio	Cociente-F	P-Valor	
Entre grupos	158,044	2			0,0102	
Intra grupos			15,4032			
Total (Corr.)	804,978	44				

Método: 95,0 porcentaje LSD

tipo_vehiculo	Frec.	Media	Grupos homogéneos		
3	15	55,9333	X		
2	15	56,6	X		
1	15	60,2	X		
Contraste			Diferencias	+/-	Límites
1 - 2			*3,6	2,89211	
1 - 3			*4,26667	2,89211	
2 - 3			0,66667	2,89211	

* indica una diferencia significativa.

Medias y 95,0 Porcentajes Intervalos LSD



Suponiendo un $\alpha = 5\%$, se pide:

- Rellena correctamente los espacios en negros del cuadro del ANOVA.
- ¿Qué porcentaje de variabilidad detectada en el consumo de carburante se debe a elementos como la velocidad media, el estado de la carretera, el tráfico existente en ese momento, la climatología existente, el estado de los neumáticos, el estilo de conducción...?
- Señala el contraste de hipótesis que existe en el estudio realizado con ANOVA. Interpreta las conclusiones que se obtendrían del cuadro resumen del ANOVA.
- Si la empresa recibiese una subvención del 20% del precio de adquisición del modelo PEGASSUS, ¿le aconsejarías adquirir este modelo a la empresa? Justifica tu respuesta.

Problema 2

Se ha realizado un experimento con el objetivo de investigar los efectos de determinados aditivos añadidos a un producto con el fin de mejorar sus propiedades físicas. Para ello se han obtenido unas probetas en las que se ha medido la resistencia a compresión de las mismas cuyos resultados se presentan a continuación:

Tipo de aditivo	Resistencia de las probetas (Kg/cm ²)			
1	400	440	430	
2	248	125	220	150
3	160	224		
4	90	128		

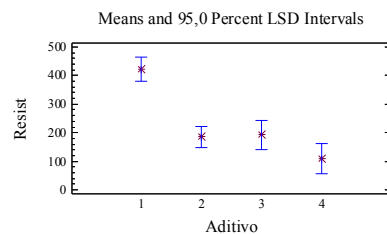
Además la empresa a partir de los datos anteriores ha elaborado las siguientes herramientas estadísticas:

ANOVA Table for Resist by Aditivo

Analysis of Variance					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Between groups	151479,0	3	50492,9	25,89	0,0004
Within groups	13653,4	7	1950,49		
Total (Corr.)	165132,0	10			

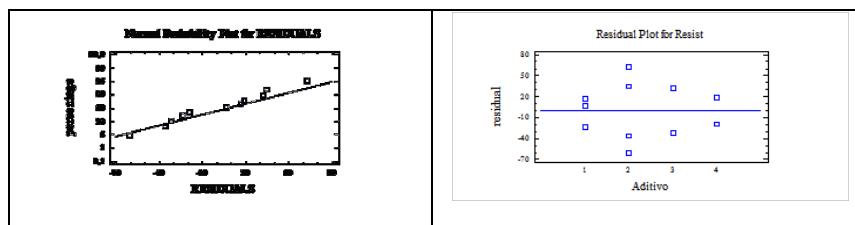
Table of Means for Resist by Aditivo with 95,0 percent LSD intervals

Aditivo	Count	Mean	Std. error (pooled s)	Lower limit	Upper limit
1	3	423,333	25,4983	380,699	465,968
2	4	185,75	22,0822	148,828	222,672
3	2	192,0	31,2289	139,784	244,216
4	2	109,0	31,2289	56,7838	161,216
Total	11	237,727			



Residuos obtenidos:

-23,3333 16,6667 6,66667 62,25 -60,75 34,25 -35,75 -32,0 32,0 -19,0 19,0



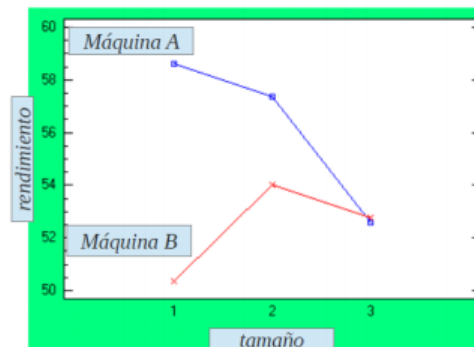
A partir de la información anterior contesta justificadamente en base a la herramienta estadística utilizada

- ¿Existen diferencias significativas entre el tipo de aditivo?
- ¿Se puede considerar los datos normales?
- ¿Podemos suponer homocedasticidad?

PROBLEMA 3

En una almazara, disponen de dos màquines per etiquetar el aceite prèviament embotellat, una nova (A) i una mès antiga (B). El gerent de la planta vol determinar si el rendiment de les dos màquines (mesurat en nombre de etiquetes col·locades correctament per minut), és el mateix per tèrmino mitjà per ambdues màquines i si depèn també del tamany de la botella: petit (1), mitjà (2) i gran (3). Con aquest propòsit, se lleva a terme un experiment en el que cada prova consisteix en observar durant una hora (elegida al azar) el nombre de botelles etiquetades i calcular el nombre de etiquetes col·locades correctament per minut (variable rendiment). Se realitzen 5 proves per cada màquina i tamany de botella. Con les dades obtingudes, se han elaborat les següents eines estadístiques:

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:màquina	109.059	1	109.059	106.4	
B:tamany	45.78	2	22.89	22.33	
AB	88.9293	1*2=2	44.46	43.38	
RESIDUAL	24.60	24	1.025		
TOTAL	268.371	29			

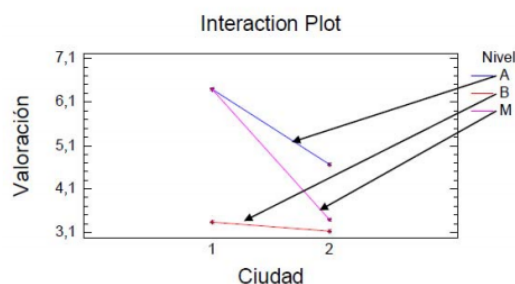
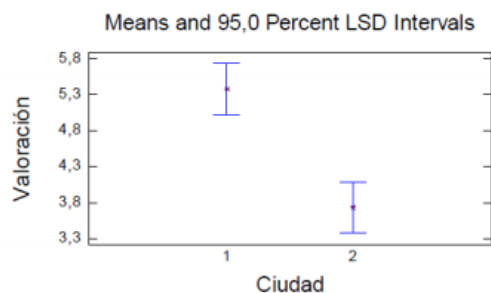
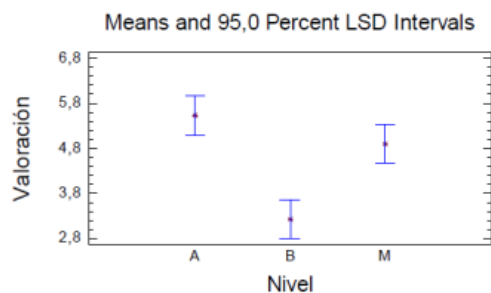


- ¿Què efectes sobre el rendiment mitjà són estadísticament significatius per a un nivell de significació del 1%?
- Indica què hipòtesis nul·la se està contrastant en el cas del factor “tamany” de la botella. Explica (sin fer cap càlcul) què mètode empraràs per determinar amb què tamany de botella se obté un rendiment significativament major i què resultat esperaràs obtenir en aquest anàlisi en cas de no rebutjar dita hipòtesis nul·la.
- Interpreta el significat de la interacció a partir del gràfic corresponent i extrae totes les conclusions que pugues a partir de ell.

PROBLEMA 4

Se ha recogido la valoración de un líder político en dos ciudades distintas 1 y 2 (Factor A), dividida cada una en tres barrios según su nivel adquisitivo (Alto, Medio y Bajo) (Factor B). Se han elaborado las siguientes herramientas estadísticas:

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
MAIN EFFECTS					
A:Ciudad	60,8444	1	60,8444	21,37	0,0000
B:Nivel	84,6889	2	42,3444	14,87	0,0000
INTERACTIONS					
AB	29,4889	2	14,7444	5,18	0,0076
RESIDUAL	239,2	84	2,84762		
TOTAL (CORRECTED)	414,222	89			



PROBLEMA 5

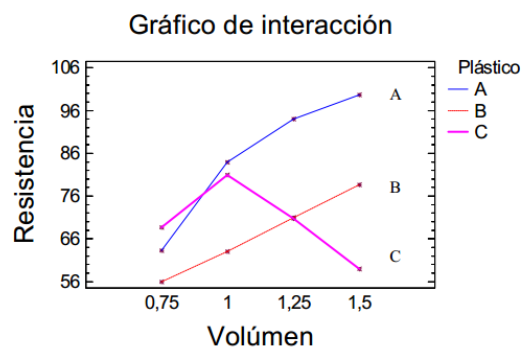
A partir de la información anterior, ¿a qué conclusiones llegarías? En una fábrica de botellas de plástico se quiere decidir qué producto resulta más resistente. Para ello se supone que tanto el tipo de plástico utilizado como materia prima, como el volumen de las botellas, pueden afectar a dicha resistencia. Se estudiaron tres tipos de plástico, (A;B;C) y 4 volúmenes diferentes (0,75; 1; 1,25 y 1,5), midiendo la resistencia de 3

botellas elegidas al azar para cada posible combinación de tipo de plástico y volumen (se analizaron 36 botellas en total).

- a) Completar la tabla del ANOVA de dos factores que resultó de dicho experimento

Fuente	SC	GI	CM	F _c
Plástico				
Volúmen	1613,64			
Plástico x Volúmen	2284,61			
Residual	639,33			
Total	6824,75			

- b) En vista de la tabla anterior, ¿qué se puede decir de la significación de los efectos de los factores estudiados? ¿Qué significado tiene en este caso concreto de estudio la interacción? Tomar $\alpha = 0.05$
- c) A partir del gráfico de interacción, ¿qué combinación de tipo de plástico y volumen da una botella más resistente? Si por motivos económicos el único tipo de plástico que se puede utilizar es el C, ¿qué volumen de botella da una mayor resistencia?



SOLUCIONES

PROBLEMA 1.-

a)

Tabla ANOVA para combustible según tipo_vehículo					
Análisis de la Varianza					
Fuente	Sumas de cuad.	Gl	Cuadrado Medio	Cociente-F	P-Valor
Entre grupos	158,044	2	79,0222	5,13	0,0102
Intra grupos	646,933	42	15,4032		
Total (Corr.)	804,978	44			

b) Esa variabilidad es la que se recoge en la parte de intragrupos.

$$646,933/804,978 \Rightarrow 80,37\%$$

c) Si m_i = consumo medio con el vehículo i

$$H_0: m_1 = m_2 = m_3$$

$$H_a: \text{algún } m_1 \neq m_2 \neq m_3$$

Como el p-valor es inferior al 5% no podemos aceptar la hipótesis nula. Por tanto, si que existe diferencias estadísticamente significativas en el consumo de combustible si las rutas se realizan con un tipo de vehículo u otro.

d) Analizando los intervalos LSD observamos que entre los vehículos donde existen diferencias significativas en cuanto al consumo medio de combustible es entre el Pegassus y el Mercedens y entre el Pegassus y el Nissen. Como el objetivo es reducir costes aunque todos me costasen igual no se debería adquirir el Pegassus porque es el que presenta mayor consumo medio de combustible y justamente mi objetivo es reducir costes en carburante.

PROBLEMA 2

Resolución:

a) En este experimento se ha utilizado un diseño factorial desequilibrado, en el que se estudia un único factor (Tipo de aditivo) a cuatro niveles con distinto número de pruebas en cada nivel.

b) Para ello realizamos un ANOVA con los datos del problema:

$$\begin{array}{lll} \bar{x}_1 = 423.33 & T_1 = 1270 & \bar{x} = 237.727 \\ \bar{x}_2 = 185.75 & T_2 = 743 & TG = 2615 \\ \bar{x}_3 = 192.0 & T_3 = 384 & SG = 2615^2/11 = 621656.82 \end{array}$$

$$\bar{x}_4 = 109.0 \quad T_4 = 218 \quad SC_{Aditivo} = \frac{1270^2}{3} + \frac{743^2}{4} + \frac{384^2}{2} + \frac{218^2}{2} - SG = 151479$$

$$SC_{Residual} = SC_{Total} - SC_{Aditivo} = 13653$$

La tabla resumen del ANOVA es la siguiente:

ANOVA Table for Resist by Aditivo

Analysis of Variance					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Between groups	151479,0	3	50492,9	25,89	0,0004
Within groups	13653,4	7	1950,49		
Total (Corr.)	165132,0	10			

Por lo que se puede afirmar que sí existen diferencias estadísticamente significativas entre los tipos de aditivos en cuanto a la resistencia media de las probetas.

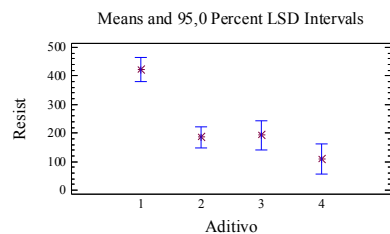
c) Para comprobar entre qué tipos de aditivo existen diferencias, se calcula mediante los intervalos LSD:

$$\bar{x}_i \pm 0,707 t_{glres}^{(d/2)} \sqrt{CMR / N_i}$$

Table of Means for Resist by Aditivo
with 95,0 percent LSD intervals

Aditivo	Count	Mean	Std. error (pooled s)	Lower limit	Upper limit
1	3	423,333	25,4983	380,699	465,968
2	4	185,75	22,0822	148,828	222,672
3	2	192,0	31,2289	139,784	244,216
4	2	109,0	31,2289	56,7838	161,216
Total	11	237,727			

A la vista de los resultados obtenidos, se puede considerar que el tipo de aditivo 1 es claramente diferente a los otros tres (pues no hay solapamiento) y como es el que, en promedio, más resistencia causa en las probetas es el que se debería utilizar atendiendo a este criterio. En cuanto a los otros tres aditivos, con la información disponible no se puede considerar que existan diferencias entre ellos, ya que existe un claro solapamiento en sus intervalos LSD.

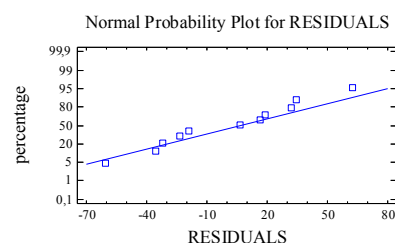


d) Las hipótesis en que se basa el ANOVA realizado en este problema son la independencia de las observaciones, la homocedasticidad y la normalidad de las poblaciones estudiadas. Las hipótesis aplicadas a este ejemplo consiste en considerar que las probetas utilizadas en el experimento han sido extraídas independientemente de la población considerada, es decir, se han obtenido mediante una muestra aleatoria simple de cada población y además las probetas entre cada tratamiento son también independientes. Consideramos también que la varianza de las cuatro poblaciones estudiadas es la misma (homocedasticidad). Y, por último, la resistencia de las probetas sigue una distribución normal para cada población estudiada.

e) Mediante el análisis de los residuos se puede detectar si hay algún dato anómalo, también la hipótesis de normalidad de las poblaciones y adicionalmente la hipótesis de homocedasticidad. Para ello se calculan los residuos de los datos obtenidos:

-23,3333 16,6667 6,66667 62,25 -60,75 34,25 -35,75 -32,0 32,0 -19,0 19,0

Al representarlo en un ppn se puede asumir que la resistencia de las probetas sigue una distribución normal para las cuatro poblaciones estudiadas.



En la representación gráfica de los residuos obtenidos para los distintos tipos de aditivo se observa que los residuos correspondientes al aditivo tipo 2 están más dispersos que el resto. Esto haría sospechar de una posible varianza de la resistencia superior a las de los otros aditivos. Este posible efecto del tipo de aditivo sobre la varianza de la resistencia (falta de homocedasticidad de las poblaciones) se podría estudiar estadísticamente realizando un ANOVA utilizando como variable respuesta los cuadrados de los residuos.

