

ANÁLISIS DE DATOS (2º de Biología)

Evaluación final extraordinaria (junio 2012)

PRIMER EJERCICIO (5 puntos)

En un estudio sobre el contenido proteínico en los pastos del concejo de Nava, se tomaron muestras en pastos comunales y en prados particulares tanto en el mes de mayo como en el mes de noviembre y se midió en cada muestra el porcentaje de proteínas. Un análisis con SPSS de los datos completos ha dado los resultados siguientes:

Estadísticos descriptivos

Variable dependiente: Porcentaje de contenido proteico

Mes	Caracter	Media	Desviación típica	N
Mayo	Comunal	8,2520	1,06771	10
	Particular	11,5730	,94391	10
	Total	9,9125	1,96581	20
Noviembre	Comunal	10,4750	1,21406	10
	Particular	14,6130	2,05625	10
	Total	12,5440	2,68460	20
Total	Comunal	9,3635	1,59331	20
	Particular	13,0930	2,20383	20
	Total	11,2282	2,67757	40

Contraste de Levene sobre la igualdad de las varianzas error^a

Variable dependiente: Porcentaje de contenido proteico

F	gl1	gl2	Sig.
1,632	3	36	,199

Contrasta la hipótesis nula de que la varianza error de la variable dependiente es igual a lo largo de todos los grupos.

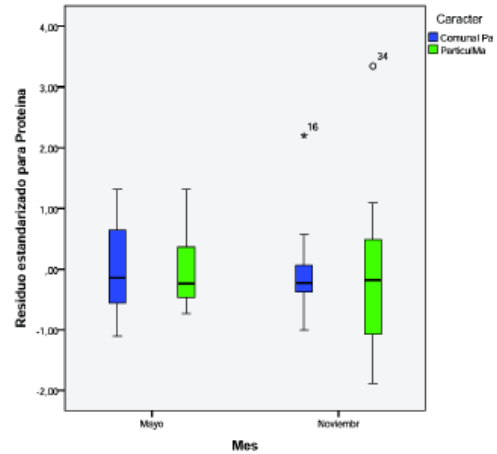
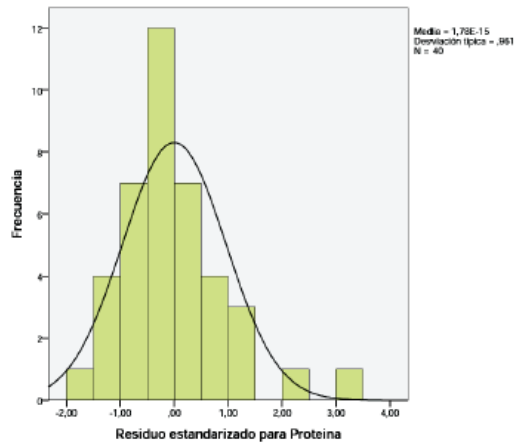
a. Diseño: Intersección + Mes + Caracter + Mes * Caracter

Pruebas de los efectos inter-sujetos

Variable dependiente: Porcentaje de contenido proteico

Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.
Modelo corregido	210,008 ^a	3	70,003	36,210	,000
Intersección	5042,944	1	5042,944	2608,508	,000
Mes	69,248	1	69,248	35,819	,000
Caracter	139,092	1	139,092	71,946	,000
Mes * Caracter	1,669	1	1,669	,863	,359
Error	69,598	36	1,933		
Total	5322,550	40			
Total corregida	279,606	39			

a. R cuadrado = ,751 (R cuadrado corregida = ,730)



a) ¿Qué modelo se ha utilizado para analizar estos datos? Expón todos sus elementos e hipótesis. No olvides decir cuál es la variable explicada, cuáles son los factores, cuáles son los parámetros del modelo y qué relaciones debe haber entre ellos. Indica los requisitos previos que deben cumplirse y realiza el diagnóstico de su cumplimiento con la información disponible.

b) Estima puntualmente los parámetros del modelo que miden el efecto de cada mes y el que mide el efecto de la interacción entre el mes de mayo y el carácter comunal ¿Cuántas réplicas se han tomado de cada combinación de niveles? ¿Es un diseño equilibrado?

c) Escribe las hipótesis nula y alternativa del contraste que estudia la interacción entre los dos factores ¿Cuál es el valor del correspondiente estadístico de contraste obtenido con esta muestra? ¿Cuál es el p-valor de este estadístico? ¿Cuál será la decisión a tomar al nivel de significación 0,05?

d) Obtén el intervalo de confianza (al 95%) para la diferencia de porcentajes medios de proteínas entre Mayo y Noviembre. ¿Conclusiones?

e) Si con los mismos datos, tuviéramos que utilizar un modelo de diseño de experimentos con un factor (mes), por haber perdido de la base de datos la información sobre el carácter comunal o particular de cada prado, ¿cuál sería la nueva tabla ANOVA? En este modelo, ¿tendría influencia el factor mes (al nivel 0,05)?

SEGUNDO EJERCICIO (5 puntos)

En un estudio se analizaron las siguientes variables de nacidos en Filadelfia en 1990:

- Peso al nacer (en gramos)
- Tiempo de gestación (en semanas)
- Educación de la madre (en años de estudios)
- Si la madre fumó en el embarazo o no (1=fumadora, 0= no fumadora)

Los resultados siguientes corresponden a una muestra de 453 nacimientos:

Estadísticos descriptivos				Correlaciones			
	Media	Desviación típica	N		peso al nacer gr.	semanas gestación	años educación madre
peso al nacer gr.	3415,86	572,353	453	peso al nacer gr.	1,000	,572	,048
semanas gestación	39,40	2,017	453	semanas gestación	,572	1,000	,036
años educación madre	12,64	2,385	453	años educación madre	,048	,036	1,000

Coefficientes^a

Modelo		Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	-2565,193	434,262		-5,907	,000
	semanas gestación	156,590	10,582	,552	14,797	,000
	años educación madre	-9,271	9,278	-,039	-,999	,318
	madre fuma	-334,973	54,059	-,240	-6,196	,000

a. Variable dependiente: peso al nacer gr.

a) Indica todos los elementos del modelo utilizado y sus requisitos previos ¿En cuanto se estima que varía el peso medio al nacer si la madre es fumadora?

b) Al nivel $\alpha = 0,05$ indica las hipótesis contrastadas en la tabla de coeficientes sobre la influencia de cada una de las tres variables del modelo y las conclusiones de cada contraste. Justifica la respuesta.

Posteriormente se realizó un análisis teniendo en cuenta únicamente las madres no fumadoras.

Estadísticos descriptivos			
	Media	Desviación típica	N
peso al nacer gr.	3498,90	506,579	356
semanas gestación	39,49	1,647	356

ANOVA ^b						
Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	20380726,569	1	20380726,569	102,018	,000 ^a
	Residual	70720352,586	354	199775,007		
	Total	91101079,154	355			

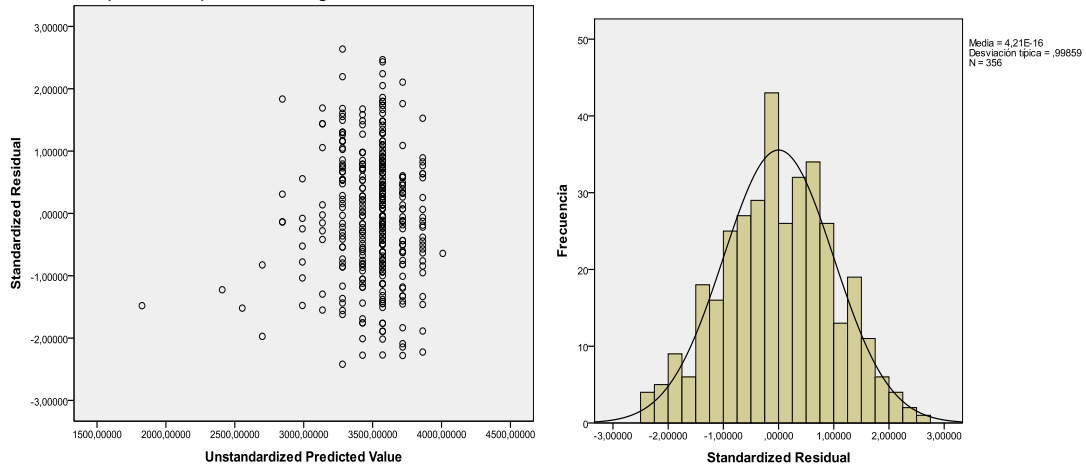
a. Variables predictoras: (Constante), semanas gestación

b. Variable dependiente: peso al nacer gr.

Coefficientes

Modelo		Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	-2246,693	569,340		-3,946	,000
	semanas gestación	145,479	14,403	,473	10,100	,000

a. Variable dependiente: peso al nacer gr.



c) Indica el modelo utilizado ahora y analiza el cumplimiento de sus requisitos previos.

d) Obtén un intervalo de confianza 0,95 para el peso medio de los nacidos con 30 semanas de gestación.

Los mismos datos de madres no fumadoras se analizan ahora con un modelo potencial ($Y = \text{constante} \cdot X^{b1}$) obteniéndose los resultados que se muestran en la siguiente tabla:

Resumen del modelo y estimaciones de los parámetros

Variable dependiente: peso al nacer gr.

Ecuación	Resumen del modelo					Estimaciones de los parámetros	
	R cuadrado	F	gl1	gl2	Sig.	Constante	b1
Lineal	,224	102,018	1	354	,000	-2246,693	145,479
Potencia	,297	149,605	1	354	,000	2,301	1,990

La variable independiente es semanas gestación.

e) Obtén con este modelo una estimación del peso al nacer con 30 semanas de gestación. ¿Qué modelo te parece mejor, el lineal o el potencial? ¿Por qué?