

## ANÁLISIS DE DATOS (2º de Biología)

### Evaluación final (Mayo 2012)

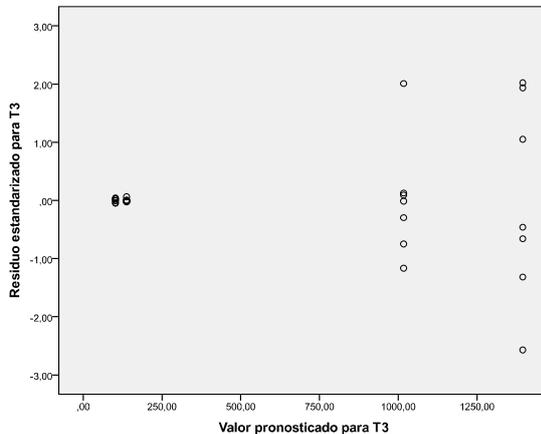
#### PRIMER EJERCICIO (5 puntos)

En un grupo de 28 ratones se ha medido la concentración en el suero de triyodotironina (T3). Catorce ratones habían recibido previamente un tratamiento con dosis elevadas de hormonas tiroideas (Trat = Si) mientras que otros catorce no habían recibido ningún tratamiento (Trat = No). De los catorce ratones que habían recibido tratamiento con hormonas, siete eran genéticamente normales (Gen = WT) y los otros siete habían sido modificados genéticamente suprimiendo el gen Dio3KO (Gen = KO). De los catorce ratones que no habían recibido tratamiento con hormonas, siete eran genéticamente normales (Gen = WT) y los otros siete habían sido modificados genéticamente suprimiendo el gen Dio3KO (Gen = KO). Los datos del experimento se analizaron con SPSS. Algunos de los resultados obtenidos se incluyen a continuación:

#### Estadísticos descriptivos

Variable dependiente:T3

Gen	Trat	Media	Desviación típica	N
KO	No	102,4286	11,44344	7
	Si	1395,5714	563,80430	7
	Total	749,0000	772,64789	14
WT	No	137,7143	9,58670	7
	Si	1017,1429	327,86350	7
	Total	577,4286	507,81579	14
Total	No	120,0714	20,93016	14
	Si	1206,3571	484,64477	14
	Total	663,2143	647,48112	28



#### Contraste de Levene sobre la igualdad de las varianzas error<sup>a</sup>

Variable dependiente:T3

F	gl1	gl2	Sig.
10,939	3	24	,000

#### Pruebas de los efectos inter-sujetos

Variable dependiente:T3

Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.
Modelo corregido	8765703,000	3	2921901,000	27,462	,000
Intersección	12315889,28	1	12315889,28	115,753	,000
Gen	206057,286	1	206057,286	1,937	,177
Trat	8260116,571	1	8260116,571	77,634	,000
Gen * Trat	299529,143	1	299529,143	2,815	,106
Error	2553555,714	24	106398,155		
Total	23635148,00	28			
Total corregida	11319258,71	27			

a. R cuadrado = ,774 (R cuadrado corregida = ,746)

Se pide:

- Escribir con detalle el modelo al que corresponden los resultados anteriores, indicando las hipótesis que se deben cumplir para que el modelo se verifique.
- ¿Cuánto vale la varianza residual en el modelo del apartado anterior? Estimar el parámetro del modelo que mide el efecto de la interacción entre el tratamiento hormonal y la modificación genética.
- ¿Podemos afirmar a nivel 0,05 que el hecho de ser o no modificado genéticamente influye significativamente en la concentración de T3? ¿Podemos afirmar a nivel 0,05 que hay interacción significativa entre los dos factores del modelo?
- Indicar a partir de la información disponible si alguna de las hipótesis mencionadas en el apartado a) no se cumple.
- Se considera un modelo que incorpore únicamente el factor *Trat* para explicar la concentración de T3. ¿Cuál es la varianza residual de este modelo?

## SEGUNDO EJERCICIO (5 puntos)

En una asignatura con 43 alumnos se realizan dos exámenes parciales (calificado cada uno de ellos de 0 a 10) y una evaluación final (calificada también de 0 a 10). Estos datos van a ser analizados con diferentes modelos estadísticos:

### Modelo de regresión lineal múltiple

Este modelo lo vamos a aplicar para intentar predecir la calificación de la evaluación final mediante las calificaciones de los dos parciales. Las siguientes tablas y gráficos son obtenidos con SPSS:

ANOVA<sup>b</sup>

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	13,889	2	6,945	3,835	,030 <sup>a</sup>
	Residual	72,435	40	1,811		
	Total	86,324	42			

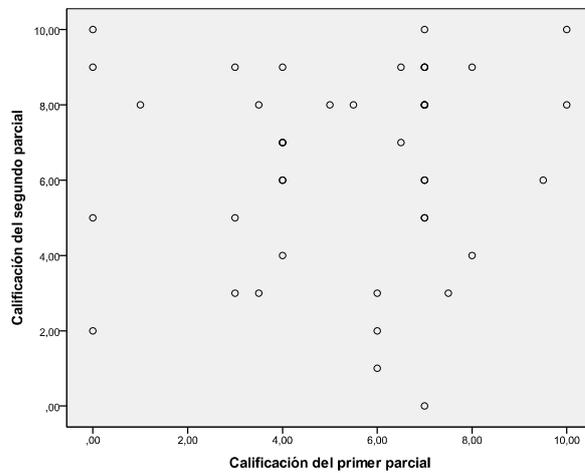
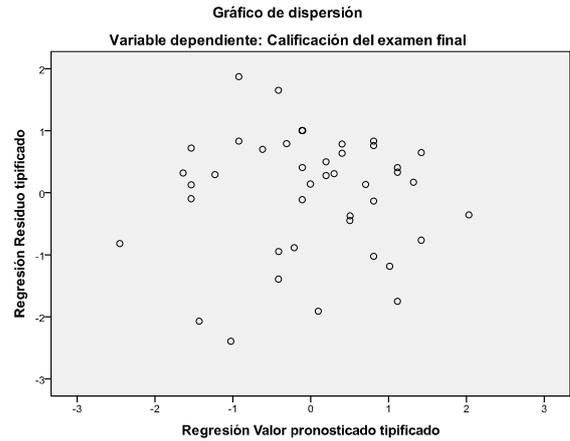
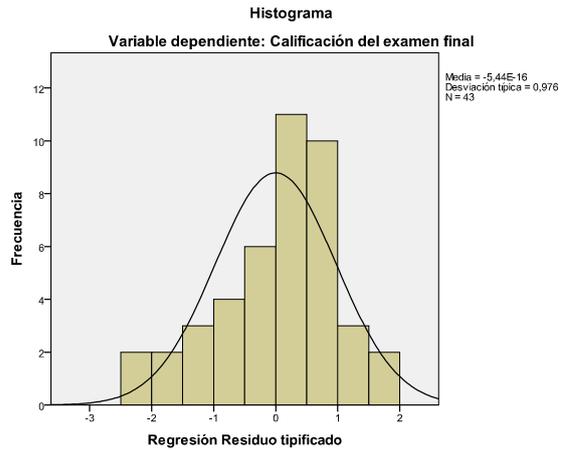
a. Variables predictoras: (Constante), Calificación del segundo parcial, Calificación del primer parcial

b. Variable dependiente: Calificación del examen final

Coefficientes<sup>a</sup>

Modelo		Coefficients no estandarizados		Coefficients tipificados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	5,050	,671		7,526	,000
	Calificación del primer parcial	,117	,079	,214	1,473	,148
	Calificación del segundo parcial	,176	,079	,324	2,231	,031

a. Variable dependiente: Calificación del examen final



- Plantear todos los elementos e hipótesis previas del modelo, y diagnosticar dichas hipótesis utilizando los gráficos disponibles.
- ¿El modelo es conjuntamente explicativo? ¿Las calificaciones de cada parcial son individualmente explicativas? Plantear los contrastes de hipótesis adecuados, responder razonadamente (con un nivel de significación del 5%), y obtener las conclusiones que parezcan oportunas.

### Modelo de regresión lineal simple

Este modelo lo vamos a aplicar para intentar predecir la calificación de la evaluación final utilizando solamente la calificación del segundo parcial. Las siguientes tablas son obtenidas con SPSS:

**Coefficientes<sup>a</sup>**

Modelo	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
	B	Error típ.	Beta		
1 (Constante)	5,623	,555		10,141	,000
Calificación del segundo parcial	,184	,080	,340	2,312	,026

a. Variable dependiente: Calificación del examen final

**ANOVA<sup>b</sup>**

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	9,958	1	9,958	5,346	,026 <sup>a</sup>
	Residual	76,366	41	1,863		
	Total	86,324	42			

a. Variables predictoras: (Constante), Calificación del segundo parcial

b. Variable dependiente: Calificación del examen final

**Estadísticos descriptivos**

	Media	Desviación típica	N
Calificación del examen final	6,8116	1,43364	43
Calificación del segundo parcial	6,4419	2,63947	43

- (c) ¿La calificación del segundo parcial tiene una influencia significativa sobre la calificación de la evaluación final? Plantear el contraste de hipótesis adecuado, y dar una respuesta razonada (con un nivel de significación del 5%).
- (d) Predecir (mediante un intervalo al 95%) la calificación de la evaluación final de un único alumno que haya obtenido un 6 en el segundo parcial.

**Modelo de regresión logística**

Este modelo lo vamos a aplicar para intentar calcular (aproximadamente) la probabilidad de que un alumno obtenga más de un 7 en la evaluación final a partir de las calificaciones de los dos parciales. Para hacer esto, codificaremos con Y=1 a los alumnos que obtienen más de 7 en la evaluación final, y con Y=0 a los que obtienen menos. La siguiente tabla se obtiene con SPSS:

**Variables en la ecuación**

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 <sup>a</sup> Parcial1	,212	,135	2,474	1	,116	1,236
Parcial2	,247	,133	3,434	1	,064	1,280
Constante	-2,573	1,203	4,570	1	,033	,076

a. Variable(s) introducida(s) en el paso 1: Parcial1, Parcial2.

- (e) Utilizar el modelo ajustado para calcular la probabilidad de que un alumno, que ha obtenido un 7 en el primer parcial y un 6 en el segundo, obtenga más de un 7 en la evaluación final.