

## TRABAJO PRÁCTICO N° 8 REGRESIÓN Y CORRELACIÓN

*En todos los problemas, aun cuando no esté especificado, suponer que las variables en estudio siguen distribución normal.*

*En todos los ejercicios definir claramente las variables aleatorias utilizadas e, indicar las suposiciones necesarias*

### Ejercicio 8.1 PC:

Con el objeto de estudiar el efecto que el tratamiento con estrona produce en ratas hembras ovariectomizadas, una muestra aleatoria de 14 ratas se dividió aleatoriamente en 4 grupos y se administró a cada grupo una dosis diferente. Después de un tiempo se observó el peso del útero, expresado en mg. Los resultados fueron:

Dosis en $\mu\text{g}$	12	14	16	20
Peso del útero	35	56	61	96
	34	45	56	88
	28	53	60	92
	22		69	

Efectuar un diagrama de dispersión para observar si es razonable suponer que el peso del útero depende linealmente de la dosis de estrona administrada, en el rango observado de dosis.

**Instrucciones para Excel:** Pintar las dos columnas llamadas Dosis y Peso-> Insertar gráfico -> XY (Dispersión). En el paso 3 se pueden seleccionar opciones para título, ejes, líneas de división, etc. Cuando el gráfico está terminado, con el botón derecho del mouse sobre uno de los rótulos de los datos se puede Agregar línea de tendencia.

### Ejercicio 8.2 PC:

En un estudio sobre la destrucción in vitro de adrenalina por el tejido hepático de rata, se efectuaron 3 determinaciones replicadas de la concentración Y de adrenalina en 5 momentos diferentes X, y se obtuvo:

X (min)	6	18	30	42	54
Y (ng/ml)	30,0	8,9	4,1	1,8	0,8
	28,6	8,0	4,6	2,6	0,6
	28,5	10,8	4,7	2,2	1,0

Efectuar un diagrama de dispersión y observar si es razonable suponer que la concentración Y depende linealmente del tiempo X.

### Ejercicio 8.3:

En un estudio sobre la toxicidad del veneno del escorpión *Tytilus serrulatus*, se midió el tiempo de supervivencia de un grupo de ratas a las que se administraron diferentes dosis de veneno. Se obtuvieron los resultados siguientes:

X (dosis en $\mu\text{g}$ )	30	40	60	80
Y (Tiempo de sobrevida en minutos)	61	45	42	22
	63	54	31	12
	78	65	27	10

- Ajustar una recta de regresión por el método de cuadrados mínimos según el modelo  $Y_{ij} = \alpha + \beta x_i + E_{ij}$ .
- Si se le administra a una rata 50  $\mu\text{g}$  de veneno, estimar puntualmente el tiempo de sobrevida.
- Construir un intervalo de confianza del 95% para la pendiente  $\beta$  de la recta.
- Del resultado obtenido en c), ¿se puede afirmar que la regresión es significativa?
- Efectuar un test de hipótesis con nivel de significación 0,05 para decidir si la regresión es significativa. Comparar el resultado con d). Calcular el nivel justo de significación.
- Obtener un intervalo de confianza del 95% para el parámetro  $\alpha$  de la recta de regresión.
- Estimar, mediante un intervalo de confianza del 95%, el tiempo medio de sobrevida que corresponde a una dosis de 50  $\mu\text{g}$ .
- Hallar un intervalo de predicción del 95% para el tiempo de sobrevida de una rata que recibe una dosis de 50  $\mu\text{g}$ . Compararlo con el intervalo hallado en f).
- Ponga a prueba la significación de la regresión mediante el estadístico  $CM_{\text{reg}}/CM_{\text{res}}$ . Compare con el estadístico obtenido en e). Interprete.

### Ejercicio 8.4:

Un espectrofotómetro se calibra con 10 soluciones estándar para medir la concentración de una droga en el suero, obteniéndose:

Solución	1	2	3	4	5	6	7	8	9	10
Concentración X(en mg/ml)	20	20	30	30	30	40	40	60	60	60
Lectura Y	19	21	24	27	27	29	31	35	36	37

- Ajustar una recta por el método de cuadrados mínimos
- Si se coloca una solución de concentración 35 mg/ml. ¿Qué lectura se espera en el espectrofotómetro?.

- c) Si una solución da una lectura 30 en el espectrofotómetro, ¿cuál es su concentración?
- d) Hallar un intervalo del 95% de confianza para el parámetro  $\beta$  de la recta de regresión.
- e) Efectuar el test para decidir si la regresión es significativa con nivel de significación 0,05.
- f) Hallar un intervalo del 95% de confianza para el parámetro  $\alpha$  de la recta de regresión.

**Ejercicio 8.5:**

Los datos siguientes corresponden a la presión sanguínea diastólica Y tomada durante e sueño en 5 tiempos t medidos desde el comienzo del sueño:

<b>T (minutos)</b>	5	10	15	20	25
<b>Y (mmHg)</b>	72	66	70	62	64

- a) Ajustar una recta por el método de cuadrados mínimos según el modelo  $Y = \alpha + \beta t + E$
- b) Sabiendo que  $SCRes = 28,8$ , hallar un intervalo del 95% de confianza para la pendiente  $\beta$  de la recta de regresión y utilizando ese intervalo decidir si la regresión es significativa con nivel de significación 0,05.
- c) Hallar un intervalo del 90% de confianza para el parámetro  $\alpha$  de la recta de regresión.
- d) Hallar un intervalo de confianza del 95% para  $E(Y)$  correspondiente a  $t = 10$  min.

**Ejercicio 8.6:**

En una investigación con aves acuáticas se intenta establecer una relación mediante la cual pueda predecirse el tiempo de reproducción en días (Y), conociendo el fotoperíodo (X = número de horas de luz por día) bajo el cual se inició la reproducción. Se obtuvieron los siguientes datos observando el comportamiento de 11 patos buceadores:

<b>X</b>	12,8	13,9	14,1	14,7	15,0	15,1	16,0	16,5	16,6	17,2	17,9
<b>Y</b>	110	54	98	50	67	58	52	50	43	15	28

- a) Suponiendo que es razonable ajustar una recta de regresión, estimar los parámetros por el método de cuadrados mínimos.
- b) Estimar, mediante un intervalo de confianza del 95%, la duración media del período de reproducción cuando el fotoperíodo bajo el cual se inicia la reproducción es de 14 hs.
- c) ¿La ecuación hallada en b), puede emplearse correctamente para estimar la duración media del período de reproducción si ésta se inició cuando había 9 horas de luz por día?

**Ejercicio 8.7:**

Los siguientes datos se obtuvieron en una investigación para determinar la capacidad de cortes de hígado de cobayos de edades diferentes para conjugar fenoltaleína con ácido glucorónico (Y):

<b>X (edad en días)</b>	1	3	5	6	10	11	14	15	21
<b>Y (nmoles conjugados)</b>	5,6	12	18	31	38	22	37	46	54
	8,8				44				

- Ajustar una recta por el método de cuadrados mínimos según el modelo  $Y = \alpha + \beta x + E$
- Hallar un intervalo de confianza del 95% para  $E(Y)$  correspondiente a  $X = \bar{x}$
- Hallar intervalos de confianza del 95% para  $E(Y)$  correspondientes a  $x = 15$ ;  $x = 18$ ;  $x = 2$ ;  $x = 5$

**Ejercicio 8.8:**

En una experiencia para calibrar un instrumento para medir la resistencia eléctrica de cierto material se obtuvieron las siguientes mediciones. X es la resistencia eléctrica (en ohm) determinada por un método suficientemente exacto como para ser considerado sin error, e Y es la medición leída en el instrumento.

<b>X</b>	<b>60</b>	<b>70</b>	<b>80</b>	<b>190</b>
	38	64	72	110
<b>Y</b>	44	70	76	118
	50		82	

- Hallar un intervalo de confianza del 95% para el valor esperado de Y cuando  $X = 90$  ohm
- Si se toma una nueva porción de material, se mide su resistencia eléctrica y ésta resulta 85 ohm, hallar un intervalo de predicción del 95% para la medición aún no observada.
- Suponiendo que se hace una nueva lectura, independiente de las anteriores, y resulta igual a 73, hallar el valor de la resistencia eléctrica que le corresponde a ese material. Este es el objeto fundamental del experimento, pues se desea calibrar el instrumento de manera que se mida Y y con esto se pueda determinar la resistencia eléctrica X.

**Ejercicio 8.9:**

En una investigación sobre mujeres menopáusicas con hipercolesterolemia (colesterol total  $> 240$  mg/dL), se les indicó una dieta rica en soja (tomar al menos un litro de leche de soja por día). Se tomó una muestra aleatoria de estas mujeres de acuerdo con el tiempo (X) de tratamiento, en meses, que llevaban, y se les midió el colesterol total (Y) en mg/dL. Se obtuvieron los siguientes datos:

Tiempo	Colesterol Total (mg/dL)				
<b>0</b>	270	260	275	302	243
<b>3</b>	231	255	230	280	260
<b>6</b>	229	249	210	268	242
<b>9</b>	239	238	260	219	203
<b>12</b>	212	190	218	216	238

- Graficar la nube de puntos y ver si es razonable ajustar una recta de regresión.
- Según el resultado de a) ajustar el modelo  $Y = \alpha + \beta X$  por el método de cuadrados mínimos y graficar la recta obtenida.
- Interpretar la pendiente y hallar un intervalo de confianza del 95% para ella.
- Estimar, mediante un intervalo de confianza del 95%, el colesterol total medio que podría esperarse con 8 meses de tratamiento.
- Predecir el colesterol total que podría esperarse en una mujer con 8 meses de tratamiento mediante un intervalo con una probabilidad del 95%. Comparar con el resultado de d).
- Graficar las bandas de confianza y de predicción del 95%.
- ¿La ecuación hallada en b), puede emplearse correctamente para estimar el colesterol total medio que podría esperarse con 15 meses de tratamiento?
- Hallar el coeficiente de determinación.
- Análizar el gráfico de los residuos y el gráfico de probabilidad normal para decidir si se cumplen las suposiciones del modelo lineal.

### Instrucciones para Excel:

En la ventana Regresión tildar Gráfico de Residuales y Gráfico e probabilidad normal.

En una columna colocar valores de t de 0 a 12. En las columnas posteriores calcular, mediante las fórmulas correspondientes, los valores estimados y los extremos de los intervalos de confianza y de predicción. Con los datos de las 6 columnas crear un gráfico de dispersión con subtipo con líneas suavizadas y sin marcadores de datos.

Si se quiere agregar los datos originales, pararse sobre el gráfico, hacer clic con botón derecho, datos de origen, Serie, Agregar. Colocar en X e Y los datos originales y marcar sólo puntos sin líneas.