

Modelo Lineal PRACTICA 4

1. Supongamos que se plantea el modelo

$$E(Y) = \beta_0 + \beta_1 X$$

pero el verdadero modelo es

$$E(Y) = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$$

¿Cuál es el sesgo en la estimación si las observaciones se toman en $X = -3, -2, -1, 0, 1, 2, 3$?

2. En la Tabla 1 (y en el archivo anscombe.xls) se muestran 4 grupos de datos simulados por Anscombe (1973).

a) Utilizando el método de mínimos cuadrados ajuste un modelo lineal simple a cada conjunto de datos. Compare los coeficientes estimados, la suma de cuadrados de los residuos, el R^2 y el test de F obtenidos en los cuatro ajustes realizados.

¿A qué conclusión llega?

b) Calcule los residuos estandarizados para cada conjunto de datos y realice los scatter plots de residuos versus valores predichos y versus variable predictora para cada caso.

¿A qué conclusión llega?

3. Un modelo de la forma $Y = \beta_0 + \beta_1 X + \beta_2 X_2 + \beta_3 X_3 + \epsilon$ es usado para describir un proceso particular. En la Tabla 2 (y en el archivo gener1.xls) se presentan los valores observados de Y y los valores predichos \hat{Y} correspondientes a 25 observaciones.

a) Evalúe el estadístico de Durbin–Watson y utilícelo para testear la hipótesis $H_0 : \rho = 0$ versus la alternativa $H_1 : \rho \neq 0$.

b) Testee la hipótesis de independencia de los residuos utilizando el test de rachas aplicado a los residuos en el orden temporal en que se presentan en la Tabla.

4. En la Tabla 3 (y en el archivo moneda.xls) se presentan datos trimestrales correspondientes a gastos de consumo (Y) y stock monetario (X), medidos en billones de dólares en USA entre los años 1952 y 1956. Una versión simplificada de la teoría monetaria sugiere el siguiente modelo

$$Y_t = \alpha + \beta X_t + u_t$$

donde α y β son constantes desconocidas y u_t es el término de error. Los economistas están interesados en estimar β y su error standard. β es denominado el *multiplicador* y es un instrumento importante en política monetaria y fiscal.

- a) Estime por mínimos cuadrados los parámetros del modelo.
- b) Grafique los residuos estandarizados versus el tiempo, es decir versus el número de observación.
- c) Calcule el estadístico de Durbin–Watson y testee la hipótesis $H_o : \rho = 0$ versus la alternativa $H_1 : \rho > 0$.
- d) Usando el estimador de ρ ($\hat{\rho}$) transforme los datos de la siguiente manera:

$$Y_t^* = Y_t - \hat{\rho} Y_{t-1}$$

$$X_t^* = X_t - \hat{\rho} X_{t-1}$$

y halle los estimadores de mínimos cuadrados del modelo: $Y_t^* = \alpha^* + \beta^* X_t^* + u_t$.

- e) Aplique el test de Durbin–Watson al modelo transformado y grafique los residuos estandarizados versus el tiempo.
- f) Calcule los estimadores de los parámetros del modelo original en función de los estimadores de α^* y β^* .

5. Una asociación de constructores quiere estudiar la relación entre los permisos de construcción y el crecimiento poblacional. Para ello, tratan de relacionar datos anuales de permisos de construcción con el número potencial de compradores en la región. Dada la dificultad en medir ese número potencial de compradores, consideran, como representativo del mismo, el tamaño de la población de entre 22 y 44 años en la región.

En la Tabla 4 (y en el archivo housing.xls) se presentan los valores históricos de las variables en estudio en un período de 25 años.

- a) Ajuste un modelo lineal simple utilizando la variable STARTS como dependiente y POP como variable regresora. Estime por mínimos cuadrados los parámetros de este modelo. ¿Cuál es el valor de R^2 ?
- b) Grafique los residuos estandarizados versus el número de observación o sea, versus el tiempo. Aplique el test de Durbin–Watson para testear la hipótesis $H_o : \rho = 0$. ¿Le parece que el alto valor de R^2 es suficiente para indicar la calidad del ajuste?
- c) En la Tabla y en el archivo se ha incluido una tercera variable: IND, la cual es un índice que mide la disponibilidad de préstamos de dinero en la región. Incorpore esta nueva variable regresora al modelo y estime los parámetros. Aplique el test de Durbin–Watson y grafique los residuos estandarizados versus el tiempo. ¿Cuáles son sus conclusiones?
- d) Estandarice las tres variables, restándoles sus correspondientes medias y dividiéndolas por su desvío estandar, y repita el análisis sin incluir intercept en el modelo. En términos de las variables estandarizadas, ¿cuál de las variables explicativas es la más influyente sobre la variable dependiente STARTS?

6. Una compañía que comercializa equipos de ski en USA desea relacionar las ventas trimestrales (V) con algún indicador económico. Selecciona como indicador el ingreso personal disponible (PDI). Ambas variables se miden en billones de dólares, y los datos correspondientes se presentan en la Tabla 5 (y en el archivo ski.xls).

- a) Ajuste el modelo $V_t = \beta_0 + \beta_1 (PDI)_t + u_t$ y aplique el test de Durbin–Watson. ¿Cuál es su conclusión?
- b) Grafique los residuos estandarizados versus el tiempo, ¿qué observa?
- c) Incorpore al modelo 3 variables dummies para representar los 4 trimestres, estime los parámetros del nuevo modelo por mínimos cuadrados y realice el plot de los residuos estandarizados versus el tiempo. Aplique el test de Durbin–Watson. Compare estos resultados con los obtenidos en a).
- d) Utilice una sola variable dummy que valga 1 para el primer y el cuarto trimestre y 0 para el segundo y el tercero y repita el análisis realizado en c) con esta nueva variable dummy reemplazando a las 3 utilizadas en dicho ítem. ¿Cuál cree que es la razón que justifica la definición de esta nueva variable dummy?
- e) Analice la independencia de los residuos obtenidos con el ajuste del ítem d) y aplique el test de Kolmogorov Smirnov para estudiar su normalidad

7. Los datos que se muestran en la Tabla 6 (contenidos en el archivo viaje.xls) provienen de distintas zonas de la ciudad de Chicago. La variable X es el tiempo de viaje entre dos puntos determinados computado a partir de los horarios de los colectivos corregidos según el tiempo que se tarda en caminar desde la zona céntrica a la parada de colectivo (asumiendo que la velocidad de caminata es de 3 m.p.h.) y el tiempo que se espera el colectivo en la parada (asumido como la mitad del tiempo entre dos colectivos sucesivos). La variable Y es el promedio de los tiempos de viaje reportados al Bureau of Census por n pasajeros (n se reporta en la misma tabla).

- a) Grafique X vs. Y . ¿Qué observa?
- b) Ajuste el modelo $E(Y) = \beta_0 + \beta_1 X$ a los datos.
- c) Grafique los residuos vs. \hat{Y} . ¿Parece razonable el supuesto de homoscedasticidad?
- d) ¿Qué pesos usaría para corregir este problema?
- e) Recalcule el ajuste con las modificaciones que según su criterio sean necesarias y grafique plots adecuados basados en los residuos para verificar que su modificación ha sido apropiada.
- f) Cuando usa mínimos cuadrados pesados, ¿cuáles son los residuos que debe graficar ?

8. La Tabla 7 (archivo acelera.xls) muestra datos sobre la aceleración de diferentes vehículos (ACC), su potencia (WHP), la velocidad de desplazamiento (SP) y la pendiente del camino, G ($G=0$ implica camino horizontal).

- a) Ajuste un modelo lineal con ACC como variable dependiente usando al resto de las variables como independientes y sin hacer ninguna transformación. Realice el scatter plot de los residuos vs. \hat{y} y los plots de residuos parciales. ¿Qué diagnostica?
- b) Obtenga un ajuste bueno de los datos haciendo los cambios que considere necesarios. Realice plots adecuados para verificar que el ajuste es razonable.

9. La Tabla 8 (archivo gorgojo.xsl) muestra datos de sobrevivencia de gorgojo tropical de cereal en distintas condiciones del suelo en Tailandia (1974). Se inocularon 30 unidades experimentales con huevos de gorgojo en el mismo día. Después de 3, 6, 9, 12 y 21 días (X) se disectaron 6 plantas elegidas al azar entre las inoculadas y se contó el número de larvas vivas (Y).

- a) ¿Le parece que el supuesto de homogeneidad de varianzas es razonable? ¿Cómo verifica esto?
- b) Use el método de Box–Cox para determinar una transformación adecuada para este problema. Analice los residuos para verificar que su elección es adecuada

10. Un biólogo desea estudiar los efectos del etanol en el tiempo de sueño. Seleccionó una muestra de 20 ratas igualadas por edad y otras características y a cada rata le suministró por vía oral una concentración en particular de etanol. Se registró el movimiento ocular REM durante el sueño de cada rata durante distintos períodos. En el archivo ratas.xls se hallan las variables REM: promedio de los movimientos oculares y DOSIS que identifica las distintas dosis de etanol utilizadas en cada tratamiento.

- a) Si se aplica un ANOVA 1 factor para determinar si hay diferencia entre las medias de los promedios de movimientos REM para las distintas dosis utilizadas, ¿es razonable el supuesto de normalidad de los errores?
- b) Si no es razonable el supuesto de normalidad, ¿cómo testaría la igualdad de varianzas?
- c) ¿Qué alternativa tiene ante la falta de normalidad para responder al test planteado en a)? Utilice el método que le parece más adecuado para comparar las medias. Si los datos son significativos, realice comparaciones múltiples para hallar intervalos de confianza para los pares de efectos debidos a los observadores al nivel 0.05.

Tabla 1. Datos de Anscombe

<i>obs</i>	<i>x1</i>	<i>y1</i>	<i>x2</i>	<i>y2</i>	<i>x3</i>	<i>y3</i>	<i>x4</i>	<i>y4</i>
1	10.000	8.0400	10.000	9.1400	10.000	7.4600	8.0000	6.5800
2	14.0000	9.9600	14.0000	8.1000	14.0000	8.8400	8.0000	5.7600
3	5.0000	5.6800	5.0000	4.7400	5.0000	5.7300	8.0000	7.7100
4	8.0000	6.9500	8.0000	8.1400	8.0000	6.7700	8.0000	8.8400
5	9.0000	8.8100	9.0000	8.7700	9.0000	7.1100	8.0000	8.4700
6	12.000	10.840	12.000	9.1300	12.000	8.1500	8.0000	7.0400
7	4.0000	4.2600	4.0000	3.1000	4.0000	5.3900	8.0000	5.2500
8	7.0000	4.8200	7.0000	7.2600	7.0000	6.4200	19.0000	12.500
9	11.000	8.3300	11.000	9.2600	11.000	7.8100	8.0000	5.5600
10	13.000	7.5800	13.000	8.7400	13.000	12.740	8.0000	7.9100
11	6.0000	7.2400	6.0000	6.1300	6.0000	6.0800	8.0000	6.8900

Tabla 2. Datos generados

obs	Y	\hat{Y}	obs	Y	\hat{Y}
1	10.98	10.86	14	9.57	9.18
2	11.13	10.48	15	10.94	10.61
3	12.51	11.79	16	9.58	9.49
4	8.4	8.73	17	10.09	9.49
5	9.27	9.13	18	8.11	8.3
6	8.73	8.2	19	6.83	6.51
7	6.36	6.18	20	8.88	8.56
8	8.5	8.4	21	7.68	7.74
9	7.82	8.05	22	8.47	9.38
10	9.14	9.22	23	8.86	9.78
11	8.24	9.64	24	10.36	11.01
12	12.19	11.54	25	11.08	11.76
13	11.88	11.6			

Tabla 3. Gastos de consumo

obs	año	trimestre	gasto	stock	obs	año	trimestre	gasto	stock
1	1952	1	214.6	159.3	11	1954	3	238.7	173.9
2	1952	2	217.7	161.2	12	1954	4	243.2	176.1
3	1952	3	219.6	162.8	13	1955	1	249.4	178
4	1952	4	227.2	164.6	14	1955	2	254.3	179.1
5	1953	1	230.9	165.9	15	1955	3	260.9	180.2
6	1953	2	233.3	167.9	16	1955	4	263.3	181.2
7	1953	3	234.1	168.3	17	1956	1	265.6	181.6
8	1953	4	232.3	169.7	18	1956	2	268.2	182.5
9	1954	1	233.7	170.5	19	1956	3	270.4	183.3
10	1954	2	236.5	171.6	20	1956	4	275.6	184.3

Tabla 4. Permisos de construcción

obs	<i>POP</i>	<i>STARTS</i>	<i>IND</i>	obs	<i>POP</i>	<i>STARTS</i>	<i>IND</i>
1	2.2	0.0909	0.03635	14	2.433	0.11604	0.05282
2	2.222	0.08942	0.03345	15	2.482	0.11688	0.05473
3	2.244	0.09755	0.0387	16	2.532	0.12044	0.05531
4	2.267	0.0955	0.03745	17	2.58	0.12125	0.05898
5	2.28	0.09678	0.04063	18	2.605	0.1208	0.06267
6	2.289	0.10327	0.04237	19	2.631	0.12368	0.05462
7	2.289	0.10513	0.04715	20	2.658	0.12679	0.05672
8	2.29	0.1084	0.04883	21	2.684	0.12996	0.06674
9	2.299	0.10822	0.04836	22	2.711	0.13445	0.06451
10	2.3	0.10741	0.0516	23	2.738	0.13325	0.06313
11	2.3	0.10751	0.04879	24	2.766	0.13863	0.06573
12	2.34	0.11429	0.05523	25	2.793	0.13964	0.07229
13	2.386	0.11048	0.0477				

Tabla 5. Ventas de equipos de ski

obs	<i>Ventas</i>	<i>PDI</i>	obs	<i>Ventas</i>	<i>PDI</i>
1	37	109	21	44.9	153
2	33.5	115	22	41.6	156
3	30.8	113	23	44	160
4	37.9	116	24	48.1	163
5	37.4	118	25	49.7	166
6	31.6	120	26	43.9	171
7	34	122	27	41.6	174
8	38.1	124	28	51	175
9	40	126	29	52	180
10	35	128	30	46.2	184
11	34.9	130	31	47.1	187
12	40.2	132	32	52.7	189
13	41.9	133	33	52.2	191
14	34.7	135	34	47	193
15	38.8	138	35	47.8	194
16	43.7	140	36	52.8	196
17	44.2	143	37	54.1	199
18	40.4	147	38	49.5	201
19	38.4	148	39	49.5	202
20	45.4	151	40	54.3	204

Tabla 6. Tiempos de viaje

obs	n	X	Y	obs	n	X	Y
1	1	26	35.0	17	5	25	34.0
2	1	40	57.0	18	4	29	32.5
3	7	32	34.3	19	3	24	28.3
4	3	36	38.3	20	6	34	40.8
5	2	27	37.5	21	7	28	29.3
6	4	39	36.3	22	5	21	26.0
7	4	29	31.3	23	4	40	47.0
8	3	22	35.0	24	2	35	40.0
9	1	34	30.0	25	3	24	30.0
10	1	25	30.0	26	2	35	45.0
11	10	37	40.5	27	2	31	37.5
12	2	36	47.5	28	9	21	25.0
13	1	20	30.0	29	17	36	51.1
14	2	26	40.0	30	2	36	42.5
15	2	31	30.0	31	1	35	45.0
16	3	22	26.7	32	3	24	25.0

Tabla 7. Aceleración

obs	<i>ACC</i>	<i>WHP</i>	<i>SP</i>	<i>G</i>	obs	<i>ACC</i>	<i>WHP</i>	<i>SP</i>	<i>G</i>
1	8.0	20.5	7.5	0	26	1.4	40.0	45.0	2
2	5.0	20.5	22.5	0	27	1.0	40.0	55.0	2
3	5.0	20.5	30.0	0	28	6.7	40.0	7.5	6
4	4.0	20.5	40.0	0	29	3.7	40.0	22.5	6
5	3.0	20.5	50.0	0	30	0.7	40.0	35.0	6
6	2.0	20.5	60.0	0	31	0.5	40.0	45.0	6
7	7.8	20.5	7.5	2	32	0.2	40.0	55.0	6
8	4.6	20.5	22.5	2	33	2.0	84.5	7.5	0
9	4.2	20.5	35.0	2	34	1.0	84.5	22.5	0
10	3.4	20.5	45.0	2	35	1.0	84.5	30.0	0
11	2.4	20.5	55.0	2	36	0.6	84.5	40.0	0
12	6.7	20.5	7.5	6	37	0.2	84.5	50.0	0
13	3.7	20.5	22.5	6	38	1.6	84.5	7.5	2
14	3.4	20.5	35.0	6	39	0.6	84.5	22.5	2
15	2.5	20.5	45.0	6	40	0.6	84.5	35.0	2
16	1.5	20.5	55.0	6	41	0.2	84.5	45.0	2
17	8.0	40.0	7.5	0	42	0.7	84.5	7.5	6
18	5.0	40.0	22.5	0	43	2.0	257.0	7.5	0
19	2.0	40.0	30.0	0	44	1.0	257.0	22.5	0
20	1.8	40.0	40.0	0	45	0.8	257.0	30.0	0
21	1.5	40.0	50.0	0	46	0.4	257.0	40.0	0
22	0.7	40.0	60.0	0	47	1.6	257.0	7.5	2
23	7.8	40.0	7.5	2	48	0.6	257.0	22.5	2
24	4.6	40.0	22.5	2	49	0.3	257.0	35.0	2
25	1.6	40.0	35.0	2	50	0.7	257.0	7.5	6

Tabla 8. Sobrevida de gorgojo tropical

obs	X	Y	obs	X	Y
1	3	17	16	9	3
2	3	22	17	9	5
3	3	26	18	9	4
4	3	20	19	12	14
5	3	11	20	12	8
6	3	14	21	12	4
7	6	37	22	12	6
8	6	26	23	12	3
9	6	24	24	12	3
10	6	11	25	21	10
11	6	11	26	21	13
12	6	16	27	21	5
13	9	8	28	21	7
14	9	5	29	21	3
15	9	12	30	21	4

Tabla 9. Estudio del Sueño

REM *DOSIS*

88.6	1
73.2	1
91.4	1
68.0	1
80.2	1
63.0	2
53.9	2
69.2	2
50.1	2
71.5	2
44.9	3
59.5	3
40.2	3
56.3	3
38.7	3
31.0	4
39.6	4
45.3	4
25.2	4
22.7	4