

Estadística II - Taller 06 Semestre: 2024-01

Profesores: Johnatan Cardona Jimenez, Freddy Hernández Barajas, Raul Alberto Perez

Monitor: Ronald Palencia

Parte teorica

Para el modelo de RLM, $Y = X\beta + \epsilon$, con k variables predictoras. Considere los diferentes tipos de pruebas presentados a continuación. ¿Cuál estadístico de prueba está bien definido?

(a) Para probar $H_0 : \beta_j = 0$, un estadístico de prueba es, $T_{0,j} = \frac{\hat{\beta}_j}{\sqrt{\text{MSE}}}$.

(b) Para probar la significancia de la regresión el estadístico es, $F_0 = \frac{\text{SSR}/k}{\text{MSE}}$.

(c) Para probar $H_0 : L\beta = 0$ donde L es una matriz $m \times p$ de constantes con r filas linealmente independientes, se usa el estadístico de prueba $F_0 = \frac{\text{SSH}/k}{\text{MSE}}$.

(d) Para probar $H_0 : \beta_j = 0$, un estadístico de prueba es, $F_{0,j} = \frac{\text{SSE}(\hat{\beta}_0, \dots, \hat{\beta}_j = 0, \dots, \hat{\beta}_k) - \text{SSE}(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)}{\text{MSE}}$.

¿Cuál de las siguientes definiciones de residuales, es correcta?. Acá \mathbf{H} es la matriz "hat", \mathbf{I} es la matriz identidad y h_{ii} es el i -ésimo elemento de la diagonal de la matriz "hat".

(a) Residuales studentizados: $e_i = \frac{d_i}{\sqrt{(1-h_{ii})\text{MSE}}}$

(b) Residuales estandarizados: $d_i = \frac{e_i}{\sqrt{\text{MSE}h_{ii}}}$

(c) Residuales crudos: $e_i = y_i - \bar{y}_i$

(d) Residuales studentizados: $r_i = \frac{e_i}{\sqrt{(1-h_{ii})\text{MSE}}}$

De los siguientes criterios para decidir si una observación es atípica, punto de balanceo o influyente, señale cuál es correcto:

(a) $|r| > 3$

(b) $\lambda_i < 1$

(c) $|DFFITs_i| > 2\sqrt{\frac{p}{n}}$

(d) $h_{ii} > \frac{2k}{n}$

Parte practica

La Asociación Nacional de Baloncesto (NBA) lleva un registro de diversos aspectos estadísticos de cada equipo. Cuatro de estos datos son: la proporción de juegos ganados (Y), la proporción de acierto en intentos de campo (X_1), la proporción de acierto en intentos de tres puntos hechos por el equipo contrario (X_2), y la cantidad de recuperaciones hechas por el equipo contrario (X_3). Considera el modelo:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

	Sum_of_Squares	DF	Mean_Square	F_Value	P_value
Model	0.4574	3	0.1525	5.2792	0.0101
Error	0.4621	16	0.0289		

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.0537	1.1998	-0.8762	0.3928
X1	1.4927	2.5394	0.5878	0.5649
X2	-0.4586	1.0815	-0.4240	0.6772
X3	0.0793	0.0244	3.2541	0.0050

	Y	Yhat	residuals	res.stud	Cooks.D	hii.value	Dfitts
1	0.7508	0.5265	0.2243	1.4418	0.1005	0.1620	0.6581
2	0.6910	0.5973	0.1063	0.6776	0.0192	0.1436	0.2730
3	0.4808	0.4899	-0.0091	-0.0619	0.0002	0.1610	-0.0184
4	0.7624	0.5963	0.1651	1.0476	0.0551	0.1829	0.4166
5	0.5103	0.6970	-0.1868	-1.3402	0.1083	0.1501	-0.6833
6	0.6083	0.4097	0.1986	1.2749	0.1501	0.1580	0.6882
7	0.9237	0.6970	0.2267	1.5375	0.0895	0.1024	0.8630
8	0.8296	0.7397	0.0899	0.5062	0.0089	0.1228	0.1595
9	0.5708	0.5399	0.0309	0.1983	0.0029	0.0641	0.0535
10	0.5462	0.2869	0.2602	1.8941	0.4753	0.1331	1.2757
11	0.8540	0.7197	0.1343	0.8962	0.0405	0.1257	0.2723
12	0.9152	0.7777	0.1375	0.7474	0.0308	0.1555	0.3455
13	0.9134	0.6970	0.2634	1.7574	0.2747	0.1207	1.1424
14	0.9624	0.8587	0.1037	0.5859	0.0146	0.1601	0.2309
15	0.5008	0.4693	0.0315	0.2113	0.0015	0.1125	0.0496
16	0.6068	0.4937	0.1131	0.6532	0.0240	0.1430	0.2325
17	0.9261	0.8587	0.0674	0.4427	0.0129	0.0905	0.1688
18	0.5014	0.4065	0.0949	0.6151	0.0123	0.0915	0.2181
19	0.4352	0.5874	-0.1522	-1.0207	0.0512	0.1601	-0.6461
20	0.6635	0.7702	-0.1067	-0.6923	0.0257	0.1768	-0.3153

TABLA 3. Estadísticos para estimación/predicción

h00.value	y0hat	se.y0hat
0.2354	0.7729	0.0824

- (a) En algunos textos estadísticos se usa como criterio para detectar observaciones atípicas si: $|r| > 2$. Con base en la TABLA 2, usando este criterio anterior, se puede CONCLUIR que:
- a) Las demás afirmaciones son incorrectas.
 - b) No hay observaciones atípicas.
 - c) Solo la observación 10 es atípica.
 - d) Las observaciones 10 y 13 son atípicas.
- (b) Con base en la TABLA 2, se puede CONCLUIR que:
- a) Las observaciones 14 y 15 son puntos de balanceo.
 - b) Las demás afirmaciones son incorrectas.

- c) Solo la observación 14 es punto de balanceo.
- d) No hay puntos de balanceo.
- (c) Se desea realizar inferencias bajo las siguientes condiciones en las predictoras: $X_{01} = 0.4511$ la proporción de acierto en intentos de campo, $X_{02} = 0.3657$ la proporción de acierto de intentos de tres puntos hechos por el equipo contrario y $X_{03} = 16.656$ la cantidad de recuperaciones hechas por el equipo contrario. **En este caso, se TIENE que:** (use la TABLA 3)
- a) Las condiciones del enunciado corresponden a una extrapolación.
- b) Con una confianza del 89% se concluye que bajo las condiciones del enunciado el promedio de la proporción de juegos ganados estará en:

$$[0.4532, 1.0926]$$

- c) Una estimación puntual de un valor futuro de la proporción de juegos ganados bajo las condiciones del enunciado es: 0.6729.
- d) Con una confianza del 89% se concluye que en un experimento futuro bajo las condiciones del enunciado la proporción de juegos ganados estará en:

$$[0.4532, 1.0926]$$

$$(1-0.89)/2$$

$$[1] \ 0.055$$

$$0.7729 + qt(0.055, 16, lower.tail = F) * 0.0824$$

$$[1] \ 0.9123318$$

$$0.7729 - qt(0.055, 16, lower.tail = F) * \sqrt{0.0289 + (0.0824^2)}$$

$$[1] \ 0.4532266$$

$$qt(0.04, 38, lower.tail = F)$$

$$[1] \ 1.79878$$

Respuestas

- 1) b
- 2) d
- 3) c
- 4) b
- 5) b
- 6) a