

Estadística II - Taller 07 Semestre: 2024-01

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

Monitor: Ronald Palencia

Parte teorica

Pregunta 1: De acuerdo con las principales causas de la multicolinealidad, ¿cuál de las siguientes opciones describe correctamente un escenario que podría dar lugar a este problema?

- (a) La utilización de variables predictoras no relacionadas entre sí.
- (b) Agregar términos polinomiales a un modelo cuando el rango de la predictora es amplio.
- (c) El uso de un modelo sobredefinido donde se toman muchas predictoras para una muestra pequeña.
- (d) La recolección de datos de múltiples fuentes independientes.

Pregunta 2: ¿Qué efecto de la multicolinealidad se asocia con la inflación de las varianzas de los estimadores?

- (a) Valores de los coeficientes con signo contrario a lo esperado.
- (b) Regresión significativa pero ninguna variable individualmente significativa.
- (c) β_j muy grandes en términos absolutos.
- (d) Inflación de los valores c_{jj}^* en las varianzas de los estimadores.

Pregunta 3: ¿Cuál de los siguientes es un síntoma común de la multicolinealidad grave en un modelo de regresión?

- (a) Todos los coeficientes tienen el signo esperado.
- (b) El modelo de regresión es significativo, pero ninguna variable predictora es individualmente significativa.
- (c) β_j con valores pequeños en términos absolutos.

- (d) La traza de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$ es pequeña.

Pregunta 4: ¿Cuál de las siguientes afirmaciones es correcta acerca de la matriz de correlaciones entre las predictoras para diagnosticar la multicolinealidad?

- (a) Una correlación alta entre dos predictoras siempre indica la presencia de multicolinealidad.
- (b) Una correlación baja entre dos predictoras garantiza la ausencia de multicolinealidad.
- (c) La matriz de correlaciones es útil para detectar multicolinealidad solo si están involucradas más de dos variables.
- (d) Valores bajos de correlación no necesariamente implican ausencia de multicolinealidad.

Pregunta 5: ¿Qué valor del Factor de Inflación de Varianza (VIF) indica un problema grave de multicolinealidad?

- (a) $\text{VIF} \leq 5$.
- (b) $5 < \text{VIF} \leq 10$.
- (c) $\text{VIF} > 10$.
- (d) $\text{VIF} = 0$.

Pregunta 6: Según el análisis de los valores propios de la matriz $\mathbf{X}'\mathbf{X}$, ¿cuándo se considera que existe un problema moderado de multicolinealidad?

- (a) Cuando $\sqrt{\kappa} \leq 10$.
- (b) Cuando $10 < \sqrt{\kappa} \leq 31.62$.
- (c) Cuando $\sqrt{\kappa} > 31.62$.
- (d) Cuando $\kappa = 0$.

Pregunta 7: ¿Cuál de las siguientes afirmaciones es correcta sobre las proporciones de descomposición de varianza en el diagnóstico de multicolinealidad?

- (a) Las proporciones de descomposición de varianza son irrelevantes para diagnosticar multicolinealidad.
- (b) Proporciones bajas ($\pi_{ij} < 0.5$) para varios coeficientes de regresión asociados con un mismo valor propio pequeño son evidencia de multicolinealidad.
- (c) Proporciones altas ($\pi_{ij} > 0.5$) para dos o más coeficientes de regresión asociados con un mismo valor propio pequeño son evidencia de multicolinealidad.
- (d) Las proporciones de descomposición de varianza no están relacionadas con los valores propios de la matriz $\mathbf{X}'\mathbf{X}$.

Parte Práctica

El director de una universidad administró cuatro exámenes de conocimiento a cada uno de los 25 estudiantes que aplicaban a un programa de becas. Para el propósito del estudio, todos los estudiantes fueron aceptados en el programa independientemente de sus puntajes en los exámenes. Después de un semestre, cada estudiante fue evaluado en su desempeño académico. El puntaje obtenido en el desempeño académico Y y los puntajes en los cuatro exámenes X_1, X_2, X_3, X_4 .

Tabla 1

	Y	X_1	X_2	X_3	X_4
Y	1.0000	0.6200	0.5100	0.8500	0.8800
X_1	0.6200	1.0000	0.2000	0.3000	0.4000
X_2	0.5100	0.2000	1.0000	0.5400	0.4600
X_3	0.8500	0.3000	0.5400	1.0000	0.7900
X_4	0.8800	0.4000	0.4600	0.7900	1.0000

Tabla 2

	Estimation	Coef.Std	Limit_2.5%	Limit_97.5%	VIF
(Intercept)	-130.4821	0.0000000	-160.0000	-100.0000	0.00000
X_1	0.3100	0.320000	0.210000	0.410000	1.15000
X_2	0.0500	0.045000	-0.080000	0.170000	1.40000
X_3	1.2500	0.600000	0.980000	1.520000	3.20000
X_4	0.5000	0.280000	0.250000	0.750000	2.85000

Tabla 3

Eigen_Value	Condition_Index	Intercept	X_1	X_2	X_3	X_4
5.000000	1.00000	0.000280	0.001300	0.007000	0.001000	0.001200
0.030000	9.0000	0.002200	0.780000	0.140000	0.003000	0.001500
0.012000	10.0000	0.084000	0.170000	0.790000	0.015000	0.036000
0.007000	40.0000	0.620000	0.000200	0.008000	0.009500	0.290000
0.001700	68.0000	0.310000	0.050000	0.075000	0.970000	0.670000

Tabla 4

k	R_{sq}	adj_R_{sq}	SSE	C_p	$Variables_in_model$
1	0.800	0.790	1800.500	85.500	X_3
2	0.750	0.735	2200.300	112.000	X_4
3	0.235	0.225	6700.000	380.000	X_1
4	0.245	0.235	6900.200	395.000	X_2
5	0.930	0.950	610.500	20.000	X_1X_3
6	0.880	0.865	1120.000	55.000	X_2X_3
7	0.810	0.790	1730.500	88.500	X_2X_4
8	0.770	0.755	1950.100	100.000	X_1X_4
9	0.955	0.920	350.000	30.00	X_3X_4
10	0.865	0.845	1100.000	45.000	X_1X_2
11	0.965	0.960	335.000	3.000	$X_2X_3X_4$
12	0.900	0.950	595.000	20.000	$X_1X_3X_4$
13	0.875	0.860	1200.000	65.000	$X_1X_2X_3$
14	0.870	0.860	1450.000	70.000	$X_1X_2X_4$
15	0.960	0.955	340.000	5.000	$X_1X_2X_3X_4$

Preguntas a responder con los resultados anteriores.

Pregunta 1: Según la matriz de correlación entre las variables bajo estudio, ¿cuál de las siguientes afirmaciones es correcta?

- (a) La variable X_1 tiene la correlación más baja con la variable Y .
- (b) La variable X_3 tiene la correlación más alta con la variable Y .
- (c) La variable X_4 tiene una correlación más fuerte con Y que X_3 .
- (d) La correlación entre X_2 y X_3 es la más alta entre todas las variables.

Pregunta 2: De la Tabla de correlaciones entre las predictoras y factores de inflación de varianza (VIF), ¿cuál de las siguientes afirmaciones es correcta?

- (a) El análisis de los VIF sugiere que NO hay multicolinealidad fuerte.
- (b) Si al menos una correlación entre pares de predictoras tiene un valor absoluto mayor que 0.7, entonces se tiene multicolinealidad fuerte.
- (c) Como al menos un VIF es mayor a 3, entonces se tiene multicolinealidad moderada.
- (d) En una RLM entre X_3 actuando como respuesta, en función de las demás variables como predictoras, se tiene un R^2 muy bajo.

Pregunta 3: De la tabla de análisis de los valores propios de la matriz $\mathbf{X}'\mathbf{X}$ para los datos originales, ¿cuál de las siguientes afirmaciones es correcta?

- (a) Al analizar el número de condición, como $\sqrt{\kappa} = 68$, entonces se tiene en el modelo un grado de multicolinealidad fuerte.
- (b) Del análisis de las proporciones de descomposición de varianza se observa que las variables Intercept, X1 y X4 están asociadas a una multicolinealidad fuerte.
- (c) En total se tienen 2 relaciones de multicolinealidad (1 grado fuerte y 1 grado moderado).
- (d) Todas las anteriores son incorrectas

Pregunta 4: De la tabla de análisis de los valores propios de la matriz $\mathbf{X}'\mathbf{X}$ para los datos originales, ¿cuál de las siguientes afirmaciones es correcta?

- (a) Al analizar el número de condición, como $\sqrt{\kappa} = 20$, entonces se tiene en el modelo un grado de multicolinealidad moderado.
- (b) La traza de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$ indica que no hay problemas significativos de multicolinealidad.
- (c) Del análisis de las proporciones de descomposición de varianza se observa que las variables, X3 y X4 están asociadas a una multicolinealidad fuerte.
- (d) La proporción de descomposición de varianza para la variable X1 sugiere una baja colinealidad con las demás variables.

Pregunta 5: Use el método de todas las regresiones posibles, para seleccionar el “mejor” submodelo.

- (a) Al analizar el R^2 tenemos que el mejor modelo es $Y = X1 + X3 + X4$
- (b) Al analizar el $R_a^2 dj$ tenemos que el mejor modelo es $Y = X1 + X2 + X4$
- (c) Todas son falsas
- (d) Analizando el R^2 , $R_a^2 dj$ y C_p se llega a que el mejor modelo es Y vs $X2, X3, X4$

Respuestas

Parte teorica

- 1) c
- 2) d
- 3) b
- 4) d

- 5) c
- 6) b
- 7) c

Parte practica

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