

Tabla 1. ANOVA.

Fuente	S.C	df	M.C	F0	Valor-p
Modelo	127.211	4	31.802	28.41	6.64e-10
Error	33.026	30	1.1		
Total	160.236	34			

Tabla 2. Todas las Regresiones Posibles.

# Vars	R2	adjR2	SSR	Cp	Variables
1	0.6701	0.6601	52.86031	17.0191	sabor
2	0.5463	0.5326	72.69472	35.0370	aroma
3	0.4972	0.4819	80.57274	42.1935	cuerpo
4	0.0002	-0.0301	160.20164	114.5297	fuerza
5	0.7730	0.7588	36.37441	4.0431	sabor fuerza
6	0.6856	0.6660	50.37510	16.7615	aroma sabor
7	0.6795	0.6595	51.35747	17.6539	cuerpo sabor
8	0.6198	0.5961	60.91677	26.3377	aroma cuerpo
9	0.6151	0.5911	61.67178	27.0236	aroma fuerza
10	0.5145	0.4842	77.79530	41.6705	cuerpo fuerza
11	0.7937	0.7737	33.06075	3.0329	aroma sabor fuerza
12	0.7742	0.7524	36.17620	5.8630	cuerpo sabor fuerza
13	0.6923	0.6625	49.30463	17.7891	aroma cuerpo sabor
14	0.6782	0.6471	51.56127	19.8391	aroma cuerpo fuerza
15	0.7939	0.7664	33.02453	5.0000	aroma cuerpo sabor fuerza

Tabla 3. ANOVA y Parámetros Estimados.

	SumSq	Df	MeanSq	F0	P.value
Modelo	0.16498161	3	0.05499387	6.158	1.03e-02
Error	0.09824212	11	0.00893019		

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.03775	0.05007	-0.742	0.4739
X1	2.86784	1.87789	1.527	0.1549
X2	-2.42158	0.82919	-2.920	0.0199
X3	0.04778	0.01652	2.893	0.0146

Tabla 4. Estadísticos de Salida.

	Y	yhat	se.yhat	residuals	res.estud	Cooks.D	hi1.value	Dffits
1	0.265	0.4028854	0.04024076	-0.1379	-1.7596	0.1440	0.1813	-0.8281
2	0.471	0.5292481	0.04524707	-0.0562	-0.6849	0.0367	0.2293	-0.3736
3	0.313	0.3755033	0.04695216	-0.0625	-0.7467	0.0476	0.2469	-0.4275
4	0.303	0.3808939	0.05069727	-0.0779	-0.9745	0.0964	0.2878	-0.6195
5	0.581	0.5339778	0.02834357	0.0470	0.5036	0.0067	0.0500	0.1563
6	0.606	0.5633480	0.05413438	0.0427	0.5324	0.0370	0.3282	0.3721
7	0.606	0.6952543	0.07462532	-0.0893	-1.6572	0.9817	0.6236	-2.1331
8	0.452	0.3012222	0.04066612	0.1507	0.2002	0.0577	0.2022	0.4522
9	0.548	0.4281995	0.04363968	0.1198	1.5102	0.1384	0.2133	0.7862
10	0.708	0.5895591	0.03011459	0.1183	1.5790	0.0982	0.1397	0.4765
11	0.464	0.4747666	0.03416099	-0.0108	-0.1166	0.0006	0.1307	-0.0452
12	0.724	0.6781168	0.06797959	0.0459	0.6818	0.1310	0.5175	0.7060
13	0.485	0.5870219	0.05260604	-0.1020	-1.3468	0.1896	0.3099	-0.9025
14	0.424	0.3597786	0.05558010	0.0642	0.8282	0.0934	0.3459	0.6023
15	0.500	0.4779979	0.02719051	0.0220	0.2324	0.0013	0.0828	0.0698

Tabla 5. Todas las Regresiones Posibles.

NoOfVars	SSR	Variables in model
1	0.19107	X3
1	0.20057	X2
1	0.24052	X1
2	0.11906	X2 X3
2	0.17295	X1 X2
2	0.17440	X1 X3
3	0.09823	X1 X2 X3

Tabla 6. Valores de la distribución F al nivel $\alpha = 0.05$.

F(1, 11) = 4.84	F(2, 11) = 3.98	F(3, 11) = 3.59
F(1, 12) = 4.75	F(2, 12) = 3.89	F(3, 12) = 3.49
F(1, 13) = 4.67	F(2, 13) = 3.81	F(3, 13) = 3.41



UNIVERSIDAD DE ANTIOQUIA
Facultad de Ingeniería
Parcial 2 Métodos estadísticos

Nota:

4,0

Nombre: David Felipe Rivera Arroyave

Documento: 1152219543

Nota: El examen consta de 15 numerales para ser trabajados en un tiempo máximo de 1 hora y 50 minutos. 11 incisos de selección múltiple con única respuesta y 4 en los cuales se debe presentar el procedimiento

Preguntas 1-7. Tablas 1 y 2

Se cree que la calidad del vino Pinot Noir se relaciona con sus propiedades de aroma (X_1), cuerpo (X_2), sabor (X_3) y fuerza (X_4). En un estudio realizado para confirmar esta hipótesis, se recogió información en 35 viñedos y se calificó cada una de estas propiedades. ¿Es posible concluir que los efectos parciales asociados al sabor y al cuerpo **NO** son significativos para determinar la calidad del vino? (Considerando las demás variables regresoras) $\alpha = 0,05$.

1,3 1. Complete la Tabla 1. ✓

2. La formulación de la hipótesis apropiada asociada a la pregunta en el enunciado y el estadístico de prueba son:

a) $H_0 : \beta_2 = \beta_3 = 0$; $F_0 = \frac{[SCE(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4) - SCE(\beta_0, \beta_1, \beta_2, \beta_4)]/2}{MCE(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)}$

b) $H_0 : \beta_1 = \beta_2 = 0$; $F_0 = \frac{MGB}{MCE}$

0,3 ~~c) $H_0 : \beta_2 = \beta_3 = 0$; $F_0 = \frac{[SCE(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4) - SCE(\beta_0, \beta_1, \beta_2, \beta_4)]/2}{MCE(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)}$~~

d) $H_0 : \beta_2 = 0$; $T = \frac{\hat{\beta}_2}{\sqrt{\text{var}(\hat{\beta}_2)}}$

Aroma X_1

Cuerpo X_2

Sabor X_3

Fuerza X_4

$Y = \text{Calidad}$

0,3 3. La región de rechazo apropiada es:

a) Rechazo H_0 si $F_0 > F_{0,025,2,30} = 4,18$

b) Rechazo H_0 si $|T| > T_{0,05,28} = 1,697$

~~c) Rechazo H_0 si $F_0 > F_{0,05,2,30} = 3,32$~~

d) Rechazo H_0 si $|T| > T_{0,025,30} = 2,042$

0,4 $= \frac{SCE(MR) - SCE(MF)}{MCE(MF)} = \frac{61,6 - 33,02453}{1,1} = 13,521$

0,4 4. El valor del estadístico calculado fue:

5. La conclusión obtenida fue:

a) Los efectos asociados a las propiedades sabor y cuerpo del vino **NO** son significativos para determinar la su calidad.

0,3 ~~b) Al menos uno de los efectos asociados a las propiedades sabor y cuerpo del vino son significativos para determinar la su calidad.~~

c) Los efectos asociados a las propiedades sabor y cuerpo del vino son significativos para determinar la su calidad.

d) Sólo los efectos parciales asociados a la sabor del vino son significativos para determinar la su calidad.

0,4 6. De acuerdo a la tabla 2 y los criterios MCE y C_p DOS candidatos a mejor modelo son:

0,4 MCE

$MCE = \frac{33,0676}{31} = 1,06$

$MCE = \frac{33,02453}{30} = 1,1$

Según C_p

$|5 - k| = 0$

41

$|13,0329 - 4| = 0,9671$

7. Considere $t_{0,025,30} = 2,042$, $t_{0,05,30} = 1,697$. Se desea realizar inferencias bajo las siguientes condiciones en las predictoras: calificación 5.5 en aroma, 5 en cuerpo, 4 en sabor y 3.3 en fuerza. Usando la siguiente información se tiene:

	h00.value	y0hat	se.y0hat	hmax
x	0.2152906	12.506539	0.4868220	0.403835

$$y_{\text{hat}} \pm t_{\alpha/2, n-k-1} \cdot \sqrt{\text{MCE}(1-h_{\text{max}})}$$

- a) Con una confianza del 95 % se concluye que en un experimento futuro bajo las condiciones del enunciado, la calidad promedio del vino estará entre 11,46354 y 13,54954
- b) Las condiciones del enunciado corresponden a una extrapolación
- 03 ~~c) Con una confianza del 95 % se concluye que en un experimento futuro bajo las condiciones del enunciado, la calidad promedio del vino estará entre 10,14468 y 14,86840~~
- ~~d) Ninguna es cierta.~~

8. Las principales fuentes de multicolinealidad entre las predictoras de un modelo de RLM son:

- a) El método de recolección de los datos, restricciones en el modelo, correcta especificación del modelo, un modelo bien definido.
- b) El método de recolección de datos, restricciones en el modelo, incorrecta especificación del modelo, un modelo con más observaciones que variables regresora.
- 03 ~~c) El método de recolección de datos, restricciones en el modelo, incorrecta especificación del modelo, un modelo sobredefinido.~~
- d) El método de recolección de datos, restricciones en el modelo, correcta especificación del modelo, un modelo con más variables regresoras que observaciones

9. De acuerdo al análisis de los índices de condición. ¿Cuál de las siguientes afirmaciones es FALSA?

- a) Si $10 < \sqrt{\kappa_j} < 31,62$ entonces hay multicolinealidad moderada
- 03 ~~b) Si al menos un $\sqrt{\kappa_j} < 10$ entonces no hay problemas de multicolinealidad~~
- c) Si para todo j , $\sqrt{\kappa_j} < 10$ entonces no hay problemas de multicolinealidad
- d) Si al menos un $\sqrt{\kappa_j} > 31,62$ entonces hay multicolinealidad severa

10. ¿Cuál de las siguientes definiciones de residuales, es la CORRECTA? (H es la matriz "hat")

a) Residuales Estudentizados: $r_i = \frac{e_i}{\sqrt{(1-H)\text{MCE}}}$

b) Residuales Estandarizados: $d_i = \frac{e_i}{\sqrt{\text{SCE}_i}}$

03 ~~c) Residuales Crudos: $e_i = \hat{y}_i - \bar{y}_i$~~

~~d) Residuales Estudentizados: $r_i = \frac{e_i}{\sqrt{\text{MCE}(1-h_{ii})}}$~~

Tenga en cuenta el siguiente enunciado y las Tablas 3 a 6, para responder las preguntas 11 a 15.

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 anotaciones de campo (X_1), la proporción de tiros de tres puntos hechos por el equipo contrario (X_2) y la cantidad de recuperaciones hechas por el equipo contrario (X_3). Considere el modelo $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$.

11. Con base en la tabla ANOVA y usando un nivel de significancia del 5 %, se puede CONCLUIR que:

- a) Ninguna de las variables predictoras es significativa
- b) Todos los parámetros son diferentes de cero
- 03 ~~c) Ninguno de los parámetros es diferente de cero~~
- ~~d) Al menos una de las variables predictoras es significativa~~

7. Considere $t_{0,025,30} = 2,042$, $t_{0,05,30} = 1,697$. Se desea realizar inferencias bajo las siguientes condiciones en las predictoras: calificación 5.5 en aroma, 5 en cuerpo, 4 en sabor y 3.3 en fuerza. Usando la siguiente información se tiene:

	h00.value	y0hat	se.y0hat	hmax
x	0.2152906	12.506539	0.4868220	0.403835

$$y_{\text{hat}} \pm t_{\alpha/2, n-k} \cdot \sqrt{\text{MCE}(1-h_{00})}$$

- a) Con una confianza del 95 % se concluye que en un experimento futuro bajo las condiciones del enunciado, la calidad promedio del vino estará entre 11,46354 y 13,54954
b) Las condiciones del enunciado corresponden a una extrapolación

~~c) Con una confianza del 95 % se concluye que en un experimento futuro bajo las condiciones del enunciado, la calidad promedio del vino estará entre 10,14468 y 14,86840~~

~~d) Ninguna es cierta~~

8. Las principales fuentes de multicolinealidad entre las predictoras de un modelo de RLM son:

a) El método de recolección de los datos, restricciones en el modelo, correcta especificación del modelo, un modelo bien definido.

b) El método de recolección de datos, restricciones en el modelo, incorrecta especificación del modelo, un modelo con más observaciones que variables regresora.

~~c) El método de recolección de datos, restricciones en el modelo, incorrecta especificación del modelo, un modelo sobredefinido.~~

d) El método de recolección de datos, restricciones en el modelo, correcta especificación del modelo, un modelo con más variables regresoras que observaciones

9. De acuerdo al análisis de los índices de condición. ¿Cuál de las siguientes afirmaciones es FALSA?

a) Si $10 < \sqrt{\kappa_j} < 31,62$ entonces hay multicolinealidad moderada

~~b) Si al menos un $\sqrt{\kappa_j} < 10$ entonces no hay problemas de multicolinealidad~~

c) Si para todo j , $\sqrt{\kappa_j} < 10$ entonces no hay problemas de multicolinealidad

d) Si al menos un $\sqrt{\kappa_j} > 31,62$ entonces hay multicolinealidad severa

10. ¿Cuál de las siguientes definiciones de residuales, es la CORRECTA? (H es la matriz "hat")

a) Residuales Estudentizados: $r_i = \frac{e_i}{\sqrt{(1-H)\text{MCE}}}$

b) Residuales Estandarizados: $d_i = \frac{e_i}{\sqrt{\text{SCE}_i}}$

c) Residuales Crudos: $e_i = \hat{y}_i - \bar{y}_i$

~~d) Residuales Estudentizados: $r_i = \frac{e_i}{\sqrt{\text{MCE}(1-h_{ii})}}$~~

Tenga en cuenta el siguiente enunciado y las Tablas 3 a 6, para responder las preguntas 11 a 15.

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 anotaciones de campo (X_1), la proporción de tiros de tres puntos hechos por el equipo contrario (X_2) y la cantidad de recuperaciones hechas por el equipo contrario (X_3). Considere el modelo $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$.

11. Con base en la tabla ANOVA y usando un nivel de significancia del 5 %, se puede CONCLUIR que:

a) Ninguna de las variables predictoras es significativa

b) Todos los parámetros son diferentes de cero

c) Ninguno de los parámetros es diferente de cero

~~d) Al menos una de las variables predictoras es significativa~~

12. En la tabla de parámetros estimados, considere el valor p más grande en las pruebas t. Con esta información se puede AFIRMAR que: (USE $\alpha = 0,05$)

a) Se garantiza que la correspondiente variable es significativa

b) Todas las afirmaciones son ciertas

0.3 ☒ c) El correspondiente valor t_0 es el de menor magnitud

d) El efecto parcial de la proporción de anotaciones de campo es significativo

$$\frac{2 \cdot \sqrt{k+1}}{n} = 0,55$$

$$2 \cdot \sqrt{\frac{k+1}{n}} = 1,032$$

13. Con base en la Tabla 4, se puede CONCLUIR que:

a) No se detectan puntos de balanceo

0.3 ☒ b) Sólo la observación 7 es un punto de balanceo

c) Las observaciones 7 y 12 son puntos de balanceo

d) Sólo la observación 12 es un punto de balanceo

14. Con base en la Tabla 4, se puede CONCLUIR que:

a) No se detectan puntos influyentes

0.10 ☒ b) Las observaciones 1, 7 y 13 son influyentes

☒ c) Sólo la observación 7 es influyente

d) Sólo la observación 13 es influyente

15. Use el método forward para determinar cuales variables deben estar incluidas en el modelo. Use la tabla 6 para tomar las decisiones.

0.2 ☒ a) Paso desde X_3

$H_0: \beta_2 = 0$ vs $H_a: \beta_2 \neq 0$

$$F = \frac{0,19107 - 0,11906}{0,11906 / (15 - (2+1))} = 7,257$$

con esto se rechaza

Como $F^* > 4,75$ no tengo
argumentos para rechazar la
hipótesis nula y β_2 no
es significativo para el modelo

12. En la tabla de parámetros estimados, considere el valor p más grande en las pruebas t. Con esta información se puede AFIRMAR que: (USE $\alpha = 0,05$)

- a) Se garantiza que la correspondiente variable es significativa
 b) Todas las afirmaciones son ciertas
 0,3 ~~c) El correspondiente valor t₀ es el de menor magnitud~~
 d) El efecto parcial de la proporción de anotaciones de campo es significativo

$$\frac{2 \cdot (k+1)}{n} = 0,53$$

$$2 \cdot \sqrt{\frac{k+1}{n}} = 1,032$$

13. Con base en la Tabla 4, se puede CONCLUIR que:

- a) No se detectan puntos de balanceo
 0,3 ~~b) Sólo la observación 7 es un punto de balanceo~~

- c) Las observaciones 7 y 12 son puntos de balanceo
 d) Sólo la observación 12 es un punto de balanceo

14. Con base en la Tabla 4, se puede CONCLUIR que:

- a) No se detectan puntos influyentes
 0,0 ~~b) Las observaciones 1,7 y 13 son influyentes~~

- ~~c) Sólo la observación 7 es influyente~~
 d) Sólo la observación 13 es influyente

15. Use el método forward para determinar cuales variables deben estar incluidas en el modelo. Use la tabla 6 para tomar las decisiones.

0,2 1 Paso desde X_3

$H_0: B_2 = 0$ vs $H_a: B_2 \neq 0$

$$F^* = \frac{0,19107 - 0,11906}{0,11906 / (15 - (2+1))} = 7,257$$

con esto se rechaza

Como $F^* > 4,75$ no tengo
 argumentos para rechazar la
 hipótesis nula y B_2 no
 es significativo para el modelo