

ESTIMACIÓN DE PROBABILIDADES DE OBTENCIÓN DE TRIGOS DE BUENA CALIDAD: UNA APLICACIÓN AL CASO URUGUAYO

JOSÉ AUSQUI
TAFILAR S.A.

INTRODUCCIÓN

- ✓ El trigo es uno de los cultivos más importantes a nivel mundial, desde el punto de vista el área dedicada y el volumen producido.
- ✓ Producción mundial estimada en 2013 de casi 711 millones de toneladas y un comercio internacional de 142 millones de toneladas.
- ✓ Es uno de los granos que primero se cultivaron en la historia y al día de hoy continúa siendo uno de los principales que se utilizan como fuente de alimento humano por su capacidad nutritiva.
- ✓ También es una fuente de alimento animal, especialmente utilizada cuando la calidad del mismo no alcanza los estándares necesarios para el consumo del hombre.
- ✓ Si bien existe una correlación directa y positiva entre las toneladas cosechadas y la rentabilidad del sector, no menos importante es cumplir con requisitos de calidad de forma de cumplir con las industrias demandantes de este grano.
- ✓ En esta dirección, el porcentaje de proteína contenido en un grano de trigo es una variable fundamental a la hora de evaluar el potencial panadero del mismo.

OBJETIVO

- ✓ Análisis de las probabilidades de obtención de trigos de buena calidad, en función de un conjunto variables explicativas.
- ✓ Entenderemos en este trabajo como granos de trigo de buena calidad a aquel que se compone por un nivel de proteína de al menos 11,5% y el análisis estadístico se enfocará en analizar la probabilidad de obtención de trigo con al menos este nivel de proteína.
- ✓ Datos capturados para las zafras 2011 y 2013 por parte de la empresa Tafilar S.A.
- ✓ Se construyeron modelos logit o logísticos para cada zafra considerada individualmente y además en un contexto de datos de panel.

DATOS

Mapa del Uruguay



Zafra 2011: 3695 observaciones de cada variable, correspondientes a viajes realizados para la empresa en el período de cosecha desde los establecimientos agropecuarios hasta los lugares de destino de los granos de trigo.

Zafra 2013: 1186 viajes realizados, midiéndose las mismas variables que se utilizaron en la zafra 2011.

No se analizó la zafra 2012 en el presente trabajo ya que fue una cosecha donde el trigo tuvo un masivo éxito en la obtención de trigos con alto contenido de proteína. El 98,21% del trigo cosechado tuvo un nivel proteico de al menos 11,5%.

Las zafras 2011 y 2013 tienen 67 lotes en común, con información de las variables a analizar en este trabajo. La zafra 2012 no tiene ningún lote en común con ninguna de estas dos zafras, lo que impidió utilizarla como parte integrante de la base de datos del panel.

El panel se integró con las zafras 2011 y 2013 con 134 datos para cada variable, en función de los lotes que ambas tienen en común.

METODOLOGÍA ESTADÍSTICA

La variable a explicar es dicotómica: el trigo puede contener un nivel de proteína mayor o igual a 11,5% (buena calidad) o inferior (mala calidad).

Función de distribución logística: $Prob(Y_i = 1|X_i) = \frac{1}{1 + e^{-Z_i}}$

$$Z_i = (\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)$$

Modelo Logit: $L_i = \ln(p/q) = Z_i$

Modelo Logit: logaritmo neperiano (\ln) de la probabilidad de ocurrencia de un evento (p), dividido por la probabilidad de no ocurrencia del mismo ($q=1-p$).

Los coeficientes de la regresión refieren al cambio en el logaritmo de esta razón de probabilidades cuando una de las variables explicativas cambia una unidad, permaneciendo las demás constantes.

$$Prob(Y_{it} = 1 | X_{it}, \alpha_i) = \frac{1}{1 + e^{-Z_{it} - \alpha_i}}$$

Se asume que la probabilidad de ocurrencia de un evento es condicional a sus variables explicativas y a un efecto individual no observable que es constante en el tiempo α_i .

Si se rechaza estadísticamente la no presencia de heterogeneidad individual inobservable, se aplica el método de máxima verosimilitud condicional (MVC) propuesto por Chamberlain: utiliza las diferencias en el tiempo de los valores observados de los regresores para explicar los cambios temporales en la variable dependiente dicotómica.

$$L^c = \prod_{i=1}^n Prob(Y_{i1} = y_{i1}, \dots, Y_{iT} = y_{iT} | \sum_{t=1}^T y_{it})$$

No toma en cuenta a aquellos casos en que Y_{it} no cambió de estado en el tiempo: no contribuyen a la función de verosimilitud condicional. La muestra a utilizar será más pequeña y los resultados de las estimaciones se basarán solamente en aquellos casos en que la variable dependiente dicotómica cambió de estado en el tiempo.

Se utiliza un contraste del tipo Hausman para evaluar la presencia de heterogeneidad individual inobservable. Bajo la hipótesis nula el estadístico de Hausman se distribuye asintóticamente como una chi-cuadrado con K grados de libertad (equivalentes al número de parámetros a estimar):

$$H = (\beta_{mvc} - \beta_{mv})' (VAR_{mvc} - VAR_{mv})^{-1} (\beta_{mvc} - \beta_{mv}) \sim \chi_k^2$$

Bajo la hipótesis nula (rechaza la presencia de efectos individuales), el método de máxima verosimilitud es consistente y eficiente mientras que el método de máxima verosimilitud condicional es consistente pero no eficiente. Bajo la hipótesis alternativa, el estimador condicional de Chamberlain es consistente y eficiente mientras que el estimador obtenido por máxima verosimilitud es inconsistente.

ZAFRA 2011

Estadística Descriptiva

VARIABLES NO BINARIAS	Mínimo	Promedio	Máximo
Rindes	1,57	4,02	6,59
Lluvias	269	429,67	567
Lluvias Octubre	6	113,32	193
Nitrógeno/Lote	22,67	93,84	131,21
Z30	0	33,27	61,03
Días	58	96,96	161
Índice	1	1,85	3
Variedades - Binarias	Observaciones	Total	% del Total
Baguette 9	197	3695	5,33%
Baguette 11	354	3695	9,58%
Baguette 17	14	3695	0,38%
Baguette 19	1591	3695	43,06%
Biointa 1006	306	3695	8,28%
Nogal	1233	3695	33,37%
Zonas - Binarias	Observaciones	Total	% del Total
Noreste Norte	236	3695	6,39%
Litoral Centro	871	3695	23,57%
Litoral Norte	279	3695	7,55%
Centro Norte	694	3695	18,78%
Litoral Sur	376	3695	10,18%
Centro Sur	660	3695	17,86%
Noreste Sur	579	3695	15,67%
Calidad Trigo - Binaria	Observaciones	Total	% del Total
Proteína de al menos 11,5%	1377	3695	37,27%

Resultados de la Regresión Logística

Trigo Bueno	Coefficiente	Error Estándar	P-Value
Rindes	-0,286	0,0566	0,000
Lluvias	-0,006	0,0008	0,000
Lluvias Octubre	0,005	0,0017	0,003
Nitrógeno/Lote	0,002	0,0033	0,621
Z30	0,012	0,0046	0,008
Días	-0,012	0,0038	0,001
Índice Ambiental	0,227	0,1086	0,036
Baguette 11	1,232	1,0540	0,243
Baguette 19	0,402	1,0503	0,702
Nogal	2,326	1,0490	0,027
Biointa 1006	1,234	1,0640	0,246
Baguette 9	1,346	1,0657	0,207
Noreste Norte	2,448	1,2375	0,048
Litoral Centro	1,144	1,2025	0,341
Litoral Norte	1,846	1,2131	0,128
Centro Norte	2,144	1,2132	0,077
Litoral Sur	1,379	1,1903	0,247
Centro Sur	1,190	1,2008	0,322
Noreste Sur	1,033	1,2172	0,396

Bondad de Ajuste del Modelo

Clasificación del Modelo	Dato Real	
	1	0
1	726	368
0	651	1950

Porcentaje de Casos de Éxito por Zona y Variedad en la Muestra Obtenida

Variedad	Noreste Norte	Litoral Centro	Litoral Norte	Centro Norte	Litoral Sur	Centro Sur	Noreste Sur
Baguette 9	-	14,29%	100%	40,00%	60,87%	11,48%	36,84%
Baguette 11	-	40,37%	50%	66,67%	17,14%	13,41%	9,52%
Baguette 17	-	-	-	-	-	7,14%	-
Baguette 19	27,08%	23,43%	25,19%	29,60%	29,49%	4,60%	10,09%
Biointa 1006	11,54%	24,39%	57,41%	45,45%	36,11%	22,50%	16,67%
Nogal	80,86%	50,59%	72,73%	72,04%	69,84%	69,02%	47,64%

Número de Datos Observados por Zona y Variedad

Variedad	Noreste Norte	Litoral Centro	Litoral Norte	Centro Norte	Litoral Sur	Centro Sur	Noreste Sur
Baguette 9	0	28	2	45	23	61	38
Baguette 11	0	109	26	39	35	82	63
Baguette 17	0	0	0	0	0	14	0
Baguette 19	48	397	131	402	156	239	218
Biointa 1006	26	82	54	22	36	80	6
Nogal	162	255	66	186	126	184	254

Probabilidad de Éxito del Modelo Logístico por Zona y Variedad

Variedad	Noreste Norte	Litoral Centro	Litoral Norte	Centro Norte	Litoral Sur	Centro Sur	Noreste Sur
Baguette 9	52,27%	33,94%	51,68%	51,02%	45,78%	24,70%	23,07%
Baguette 11	49,42%	31,43%	48,83%	48,17%	42,97%	22,64%	21,11%
Baguette 17	-	-	-	-	-	-	-
Baguette 19	29,89%	16,67%	29,40%	28,86%	24,74%	11,32%	10,46%
Biointa 1006	49,49%	31,49%	48,90%	48,24%	43,04%	22,69%	21,16%
Nogal	74,48%	57,80%	74,03%	73,53%	69,24%	46,65%	44,43%

ZAFRA 2013

Estadística Descriptiva

Variables No Binarias	Mínimo	Promedio	Máximo
Rindes	2,063	3,23	5,01
Lluvias	25	500,06	719
Lluvias Octubre	0	88,92	187
Nitrógeno/Lote	22,88	89,55	195,87
Z30	0	33,10	130,23
Días	62	87,93	116
Índice	1	1,68	3
Variedades - Binarias	Observaciones	Total	% del Total
Baguette 9	56	1186	4,72%
Baguette 11	207	1186	17,45%
Baguette 19	610	1186	51,43%
Baguette 501	35	1186	2,95%
Baguette 601	25	1186	2,11%
Baguette 701	11	1186	0,93%
Baguette 801	13	1186	1,10%
Génesis 2359	53	1186	4,47%
Nogal	176	1186	14,84%
Zonas - Binarias	Observaciones	Total	% del Total
Noreste Norte	379	1186	31,96%
Litoral Centro	247	1186	20,83%
Litoral Norte	21	1186	1,77%
Centro Norte	265	1186	22,34%
Litoral Sur	33	1186	2,78%
Centro Sur	81	1186	6,83%
Noreste Sur	160	1186	13,49%
Calidad Trigo - Binaria	Observaciones	Total	% del Total
Proteína de al menos 11,5%	945	1186	79,68%

Resultados de la Regresión Logística

Trigo Bueno	Coficiente	Error Estándar	P-Value
Rindes	-0,227	0,1860	0,223
Lluvias	-0,005	0,0010	0,000
Lluvias Octubre	-0,006	0,0028	0,026
Nitrógeno/Lote	0,003	0,0043	0,551
Z30	0,012	0,0082	0,160
Días	-0,041	0,0082	0,000
Índice Ambiental	0,695	0,2993	0,020
Baguette 9	9,776	1,7006	0,000
Baguette 11	11,249	1,6733	0,000
Baguette 19	8,728	1,6369	0,000
Baguette 501	11,583	1,8105	0,000
Baguette 601	10,160	1,7339	0,000
Baguette 701	10,698	1,899	0,000
Baguette 801	11,166	1,9020	0,000
Génesis 2359	10,195	1,7223	0,000
Nogal	9,564	1,5981	0,000
Noreste Norte	-1,522	1,0974	0,166
Litoral Centro	-3,343	1,0510	0,001
Centro Norte	-1,02	1,0584	0,336
Litoral Sur	-5,99	1,1683	0,000
Centro Sur	-1,220	1,1052	0,270
Noreste Sur	-3,205	1,0740	0,003

Bondad de Ajuste del Modelo

Clasificación del Modelo	Dato Real	
	1	0
1	887	150
0	58	91

Porcentaje de Casos de Éxito por Zona y Variedad en la Muestra Obtenida

Variedad	Noreste Norte	Litoral Centro	Litoral Norte	Centro Norte	Litoral Sur	Centro Sur	Noreste Sur
Baguette 9	-	100%	-	92,50%	-	100%	66,67%
Baguette 11	95,52%	92,31%	-	100%	42,42%	-	100%
Baguette 19	54,40%	56,82%	95,24%	87,08%	-	80,77%	75,28%
Baguette 501	-	94,29%	-	-	-	-	-
Baguette 601	-	80,00%	-	-	-	-	93,33%
Baguette 701	-	90,91%	-	-	-	-	-
Baguette 801	-	92,31%	-	-	-	-	-
Génesis 2359	-	100%	-	-	-	-	92,31%
Nogal	86,92%	100%	-	73,68%	-	95,45%	-

Número de Datos Observados por Zona y Variedad

Variedad	Noreste Norte	Litoral Centro	Litoral Norte	Centro Norte	Litoral Sur	Centro Sur	Noreste Sur
Baguette 9	-	6	-	40	-	7	3
Baguette 11	67	65	-	28	33	-	14
Baguette 19	182	88	21	178	-	52	89
Baguette 501	-	35	-	-	-	-	-
Baguette 601	-	10	-	-	-	-	15
Baguette 701	-	11	-	-	-	-	-
Baguette 801	-	13	-	-	-	-	-
Génesis 2359	-	14	-	-	-	-	39
Nogal	130	5	-	19	-	22	-

Probabilidad de Éxito del Modelo Logístico por Zona y Variedad

Variedad	Noreste Norte	Litoral Centro	Litoral Norte	Centro Norte	Litoral Sur	Centro Sur	Noreste Sur
Baguette 9	82,79%	83,16%	-	94,79%	14,42%	94,23%	91,06%
Baguette 11	95,45%	95,56%	-	98,75%	42,35%	98,62%	97,80%
Baguette 19	62,77%	63,38%	-	86,44%	5,58%	85,14%	78,11%
Baguette 501	96,70%	96,78%	-	99,11%	50,65%	99,01%	98,41%
Baguette 601	87,59%	87,87%	-	96,39%	19,83%	96,00%	93,73%
Baguette 701	92,36%	92,54%	-	97,86%	29,74%	97,62%	96,24%
Baguette 801	95,08%	95,20%	-	98,65%	40,34%	98,50%	97,61%
Génesis 2359	87,97%	88,25%	-	96,51%	20,40%	96,13%	93,93%
Nogal	79,55%	79,97%	-	93,63%	11,99%	92,96%	89,17%

DA TOS DE PANEL

Estadística Descriptiva

		ZAFRA 2011			ZAFRA 2013		
Variables No Binarias	Mínimo	Promedio	Máximo	Mínimo	Promedio	Máximo	
Rindes	1,94	3,83	5,78	2,06	3,17	4,26	
Lluvias	331	450,17	538	25	489,84	707	
Lluvias Octubre	75	115,16	193	0	78,09	187	
Nitrógeno/Lote	54,26	88,39	116,82	39,15	80,72	175,98	
Z30	0	29,6	56,3	16,05	28,85	65,13	
Días	75	102,65	156	63	86,66	116	
Índice	1,01	1,73	3	1,01	1,68	2,32	
Variedades - Binarias	Observaciones	Total	% del Total	Observaciones	Total	% del Total	
Baguette 9	1	67	1,49%	7	67	10,45%	
Baguette 11	10	67	14,93%	13	67	19,40%	
Baguette 19	32	67	47,76%	27	67	40,30%	
Baguette 601	-	67	-	1	67	1,49%	
Baguette 701	-	67	-	1	67	1,49%	
Biointa 1006	10	67	14,93%	-	67	-	
Génesis 2359	-	67	-	5	67	7,46%	
Nogal	14	67	20,90%	13	67	19,40%	
Zonas - Binarias	Observaciones	Total	% del Total	Observaciones	Total	% del Total	
Noreste Norte	16	67	23,88%	16	67	23,88%	
Litoral Centro	12	67	17,91%	12	67	17,91%	
Litoral Norte	-	67	-	-	67	-	
Centro Norte	17	67	25,37%	17	67	25,37%	
Litoral Sur	6	67	8,96%	6	67	8,96%	
Centro Sur	11	67	16,42%	11	67	16,42%	
Noreste Sur	5	67	7,46%	5	67	7,46%	
Calidad Trigo - Binaria	Observaciones	Total	% del Total	Observaciones	Total	% del Total	
Proteína de al menos 11,5%	26	67	38,80%	57	67	85,07%	

Resultados de la Regresión Logística

Trigo Bueno	Coefficiente	Error Estándar	P-Value
Rindes	-0,503	0,2759	0,068
Lluvias	0,003	0,0018	0,068
Lluvias Octubre	-0,009	0,0052	0,080
Nitrógeno/Lote	0,012	0,0126	0,342
Z30	0,015	0,0177	0,399
Días	-0,022	0,0117	0,062
Índice	1,313	0,5259	0,013

Bondad de Ajuste del Modelo

	Dato Real	
Clasificación del Modelo	1	0
1	66	29
0	17	22

CONCLUSIONES

- ✓ Se corroboró la correlación negativa entre la cantidad de toneladas cosechadas y la calidad del trigo obtenido.
- ✓ Se verificó la importancia del NZ30 y del Índice de Calidad del Suelo.
- ✓ Las probabilidades de obtener trigos con alta proteína disminuyen en la medida que la aplicación de Z30 se aleja en el tiempo.
- ✓ Tanto las lluvias del período siembra-cosecha como las de octubre resultaron significativas en las dos zafras pero no se puede concluir definitivamente sobre la dirección del impacto.
- ✓ Se construyó una herramienta auxiliar para analizar la sensibilidad de estas probabilidades ante cambios en las variables explicativas.

MUCHAS GRACIAS