

2013

international year of
statistics



AÑO INTERNACIONAL
DE LA ESTADÍSTICA
ORGANIZACIÓN PARTICIPANTE

$\frac{(n-1)s^2}{\sigma^2}$ $1 - \frac{(X-\mu)^2}{2\sigma^2}$ $\sum_{j=1}^5 x_j = 105$ $\frac{X-\mu}{\sigma} \sim N(0,1)$

$R_c = \{Z\}$ $\frac{N_i - \hat{e}_i}{\hat{e}_i}$ $Z_{\alpha/2}$

$H_0:$ $I_{1-\alpha} = \left[X - z_{\alpha/2} \sqrt{\frac{\sigma}{n}}, X + z_{\alpha/2} \sqrt{\frac{\sigma}{n}} \right]$ $P(Z > z_{\alpha/2}) = \alpha/2$ $E(X) = \int_{-\infty}^{\infty} x f(x; \theta) dx$



UNIVERSIDAD COMPLUTENSE
ECONÓMICAS Y EMPRESARIALES

Departamento de
Estadística e Investigación Operativa II
(Métodos de Decisión)

Concurso del Año Internacional de la Estadística

Trabajos premiados



UNIVERSIDAD
COMPLUTENSE
MADRID

Índice

Prólogo	1
Estimación de Probabilidades de Obtención de Trigos de Buena Calidad: Una Aplicación al Caso Uruguayo. <i>José Ausqui</i>	3
Duración de catéteres transitorios en un servicio de diálisis: Un análisis de supervivencia. <i>Julia-Inés Fernández y Elisabet López</i>	22

Prólogo

Como directora del Departamento de Estadística e Investigación Operativa II, hay pocas cosas que me puedan producir más satisfacción que tener el honor de presentar este documento que pretende contribuir, de manera modesta pero rigurosa, a las celebraciones con motivo del Año Internacional de la Estadística 2013. Estoy segura de representar el sentir de todos los miembros que forman parte de este Departamento, si me congratulo al promover actividades como ésta que solo son posibles con el esfuerzo y la dedicación de los docentes e investigadores que lo componen, en especial del Coordinador del Seminario de Investigación David Casado de Lucas.

Con la idea de contribuir a celebrar esta efeméride, el Consejo del Departamento de Estadística e Investigación Operativa II decidió, hace ya unos meses, organizar un concurso para trabajos de investigación que cumplieran los requisitos de mostrar la utilidad e importancia de la Estadística con proyectos en los que se pudiera observar de manera palpable cómo las técnicas estadísticas son una herramienta crucial para resolver problemas teóricos o prácticos de cualquier índole y naturaleza científica. Decidimos también que este concurso tuviera un decidido carácter abierto e internacional, con objeto de que pudieran participar en él un mayor número de trabajos, tanto dirigidos como no dirigidos, provenientes de los sectores público o privado y que estuvieran redactados tanto en inglés como en español. Gracias a este carácter abierto y a una importante difusión que permitió divulgar la convocatoria por diversos medios a más de mil universidades, han sido diversos los proyectos de investigación y trabajos que nos han llegado de diferentes países y universidades tanto públicas como privadas. La calidad de casi todos ellos ha hecho muy dificultoso el fallo definitivo del Jurado, compuesto por personalidades de la talla del Dr. Alfonso Novales Cinca, Dr. Enrique García Pérez, Dr. Adolfo Hernández Estrada y Dr. Francisco Javier Martín Campo, pues tanto la presentación, la redacción, así como el valor aplicado y pedagógico de varios trabajos, hubieran requerido de más de un premio por categoría.

No obstante y a pesar de no poder otorgar nada más que los dos premios previstos, el objetivo que el Departamento se había propuesto ha sido cumplido con creces y el resultado es un equilibrado fallo del Jurado que ha decidido otorgar los galardones a los siguientes trabajos.

Primer premio en la categoría de proyectos para profesionales al trabajo titulado «Estimación de Probabilidades de Obtención de Trigos de Buena Calidad: Una Aplicación al Caso Uruguayo», presentado por José Ausqui, analista en la Unidad de Negocios Comercial, Tafilar S.A. (Uruguay). En este caso el Jurado ha valorado muy especialmente la calidad de los datos y el valor pedagógico de la propuesta de este profesional uruguayo, que analiza con rigor las probabilidades de que en las cosechas de trigo se supere el nivel de proteínas que demanda el mercado alimenticio, considerando diversas variables –el tipo de suelo, la variedad de la semilla, las lluvias, las dosis de abono, etcétera– y aplicando la regresión logística.

Primer premio a la categoría para estudiantes de Grado al trabajo titulado «Duración de catéteres transitorios en un servicio de diálisis. Un análisis de supervivencia», presentado por las estudiantes Julia-Inés Fernández y Elisabet López de la Licenciatura en Estadística de la Universidad Nacional de Rosario (Argentina). En este trabajo se ha valorado por parte del Jurado la importancia y el interés de las posibles aplicaciones de los resultados obtenidos para la mejora de los servicios médicos de diálisis, pues sus jóvenes autoras plantean un modelo de análisis en el que se tiene en cuenta la posible influencia que algunas variables –sexo, edad, técnica de conexión, lugar, motivo, etcétera– pueden tener en la duración de los catéteres colocados en un instituto de nefrología durante cierto periodo de tiempo.

Quiero finalizar felicitando a los ganadores y deseándoles que sus trabajos tengan el recorrido científico y de transferencia de resultados que se merecen. Enhorabuena, entonces, a José Ausqui, por su calidad como profesional, y a Julia-Inés Fernández y a Elisabeth López, por la capacidad intelectual que han demostrado y por un futuro lleno de buenos augurios.

Y por supuesto invitar al lector interesado a sumergirse en las páginas que siguen, pues su calidad y rigor lo merecen.

Madrid, julio de 2014

Gregoria Mateos-Aparicio Morales

Directora del Departamento

Estimación de Probabilidades de Obtención de Trigos de Buena Calidad: Una Aplicación al Caso Uruguayo

José Ausqui

Unidad de Negocios Comercial

Tafilar S.A. (Uruguay)

1. Introducción

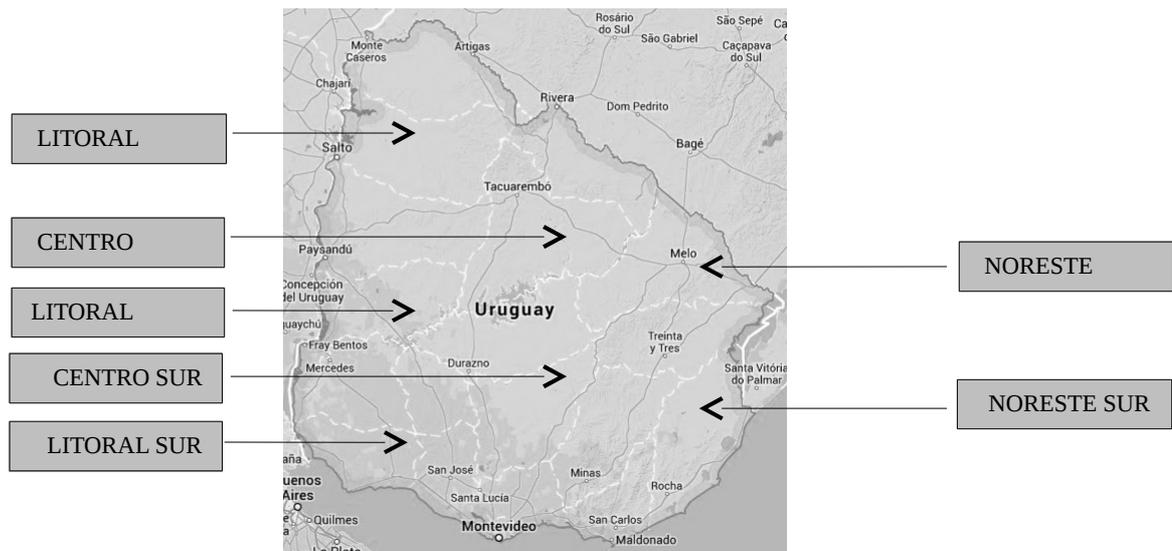
Con una extensión territorial de unos 176 mil kilómetros cuadrados y una población de poco más de 3 millones de habitantes, la economía del Uruguay está dominada por el sector agrícola y ganadero con destino exportador. Tradicionalmente el mayor peso económico se situaba en la ganadería, pero este fenómeno ha venido cambiando sostenidamente desde comienzos del nuevo siglo, siendo actualmente la agricultura el sector más relevante en la economía del Uruguay.

En el año 2000, la agricultura de secano (refiere a la agricultura que utiliza el agua que proviene de la lluvia) ocupaba menos de 350 mil hectáreas del territorio uruguayo, mientras que la cabaña de ganado vacuno alcanzaba las 14 millones de cabezas. Ya para el año 2010 la superficie sembrada de estos cultivos superaba el millón de hectáreas mientras que la cabaña de vacuno se redujo a 13,4 millones de cabezas. Los granos y derivados que representaban un 17% de las exportaciones del país medidas en dólares, representaron un 35% para el año 2010, superando por primera vez en la historia del Uruguay a las exportaciones en dólares de carne vacuna.

El segundo cultivo de mayor importancia en el Uruguay es el trigo, por detrás de la soja. En la zafra 2011 se sembraron casi 600 mil hectáreas de trigo, con un volumen cosechado superior a los 2 millones de toneladas. Más de un 10% de dicha producción correspondió a la empresa Tafilar S.A. con más de 235 mil toneladas, utilizando unas 75 mil hectáreas distribuidas a lo largo y ancho del país.

La gráfica 1 exhibe el mapa del Uruguay con las siete zonas clasificadas por Tafilar S.A. para ubicar sus campos. La empresa cosechó unas 43 mil toneladas en Centro Norte, más de 33 mil en Centro Sur, 49 mil toneladas en Litoral Centro, casi 24 mil toneladas en Litoral Norte, 27 mil toneladas en Litoral Sur, 28 mil toneladas en Noreste Norte y 31 mil toneladas en Noreste Sur.

Gráfica 1: Mapa del Uruguay



En la zafra 2012 Tafilar S.A. cosechó poco más de 100 mil toneladas de trigo. Se sembraron unas 500 mil hectáreas de trigo en el Uruguay en dicha zafra con una producción de 1,1 millones de toneladas. En esta zafra, la empresa seguía representando aproximadamente el 10% de la producción de trigo en el país.

Para la zafra de trigo del año 2013, Tafilar S.A. cosechó unas 61 mil toneladas en Uruguay. Un 14% se cosechó en Centro Norte, 5% en Centro Sur, 38% en Litoral Centro, 3% en Litoral Norte, 14% en Litoral Sur, 17% en Noreste Norte y el 8% restante en Noreste Sur. Uruguay también disminuyó las hectáreas de trigo sembradas con respecto a las dos zafras anteriores, con una superficie sembrada de unas 462 mil hectáreas y una producción estimada de 1,53 millones de toneladas. La representatividad de la empresa cayó a aproximadamente un 4% de la producción nacional.

A nivel mundial se puede decir que este cultivo es sin dudas uno de los más importantes desde el punto de vista del área dedicada y del volumen producido. Con una producción mundial estimada en 2013 de casi 711 millones de toneladas y un comercio internacional de 142 millones de toneladas de acuerdo a la FAO, en general gran parte de lo que produce cada país se consume dentro de su propia frontera.

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.

Para los productores de trigo del Uruguay, y probablemente para los productores de trigo a nivel mundial, deben ser tan relevantes las toneladas producidas de trigo como la calidad del mismo. 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, como ser la panadera, la molinera y los clientes finales en general.

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, considerándose que a mayor cantidad de proteína, mayor será la calidad del cultivo.

Contrariamente a los intereses del negocio asociado a la producción y comercialización del trigo, el conocimiento convencional establece una correlación negativa entre el rendimiento en cantidad del grano y el nivel de proteína contenida en el mismo. En condiciones de cosecha de altas cantidades de trigo con limitada absorción de nitrógeno se genera un efecto por el cual se diluye la proteína en los granos, lo que impide al cultivo almacenar suficiente cantidad de la misma.

Muchos factores inciden en la determinación del nivel de proteína del grano de trigo. Algunos pueden ser manejados por el hombre, como por ejemplo la variedad de trigo elegida, fecha de siembra o dosis de nitrógeno aplicadas, pero existen otros ajenos a su intervención, como ser el nivel de lluvias que caen en el período que transcurre entre la siembra y cosecha del grano.

Lee, Kenkel y Brorsen (2011), utilizando información del Estado de Oklahoma en los Estados Unidos de Norteamérica, encontraron una correlación positiva entre el nivel de precipitaciones previo a la cosecha y el nivel de proteína de los trigos cosechados, así como una correlación negativa entre los rendimientos cuantitativos de los granos de trigo y sus niveles de proteína.

Ereikul, Götz y Koca (2012), trabajaron analíticamente en una granja experimental de Turquía occidental, donde encontraron que el contenido proteico de los granos de trigo crece en forma paralela al crecimiento de las dosis de nitrógeno aplicadas.

García Lamothe (2006), en su análisis para el Uruguay, sostiene la importancia de la fertilización nitrogenada no solamente sobre el nivel de proteína de los granos de trigo sino también sobre el rendimiento cuantitativo de los mismos. Así mismo alerta sobre el potencial de contaminación si se incrementa la dependencia de insumos como el nitrógeno.

La importancia que conlleva tener en cuenta la calidad del trigo en los mercados es el motor que impulsa el presente trabajo. El mismo se orienta a la búsqueda y comprensión de las variables que impactan en el nivel de proteína de los granos de trigo que cosecha la empresa Tafilar S.A., de forma de avanzar en el conocimiento de esta dimensión.

2. Objetivo

El objetivo específico del trabajo consiste en analizar las probabilidades de obtención de trigos de buena calidad, en función de un conjunto de variables explicativas que se detallan en los párrafos siguientes, cuyos valores fueron capturados para las zafras 2011 y 2013 por parte de la empresa Tafilar S.A. Para ello, se construirán modelos logísticos (*logit*) para cada zafra considerada individualmente y además en un contexto de datos de panel.

Existen diversas formas de evaluar la calidad del trigo cosechado, siendo una de las más relevantes para el Uruguay la que tiene que ver con el nivel de proteína contenida en el grano. La obtención de granos de trigo con proteína de al menos 11,5% es un requisito fundamental a la hora de concretar acuerdos de compra-venta de trigos de buena calidad.

En forma restringida, entenderemos en este trabajo como 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. Conviene aclarar que existen más dimensiones a considerar a la hora de evaluar la calidad del trigo.

Con respecto a qué variables impactan en la calidad del trigo a cosechar, se establecieron aquellas que se consideraron relevantes de acuerdo al conocimiento de expertos de nuestra empresa. El conocimiento convencional establece que existe una correlación negativa entre la cantidad de trigo cosechada (denominada «rinde») y el nivel de proteína contenido en cada grano. Sin dudas que el nivel de lluvias impactará en la calidad del trigo, especialmente las lluvias de octubre. La aplicación de nitrógeno durante toda la etapa previa a la cosecha, en especial el Z30, que es la última dosis de nitrógeno que se aplica, es considerada importante para mejorar la calidad del trigo. Es de esperar que a mayor número de días transcurridos entre la aplicación de la última dosis de nitrógeno y la fecha de cosecha, menor probabilidad de obtener trigos de buena calidad. Además se utiliza en este trabajo un índice de fertilidad del suelo que se espera que esté correlacionado positivamente con la calidad del trigo.

Finalmente, se espera que las distintas variedades de semilla sembrada generen diferentes resultados en la calidad del trigo y lo mismo ocurrirá con las diferentes zonas del país donde se siembra este cultivo. Las zonas también se utilizan como niveles de la fertilidad del suelo, que es heterogénea a lo largo y ancho del Uruguay.

3. Datos

Para la zafra 2011 se obtuvieron 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. Estos viajes movieron granos de trigo por aproximadamente un 45% del total de la producción en dicha zafra.

Para la zafra 2013 se obtuvieron 1186 viajes realizados por Tafilari S.A., equivalentes a aproximadamente un 55% de la producción total en dicha zafra. Se midieron las mismas variables que se utilizaron en la zafra 2011, incorporándose además las variedades de trigo Baguette 501, Baguette 601, Baguette 701, Baguette 801 y Génesis 2359. No se utilizó en esta zafra la variedad Biointa 1006 ni la variedad Baguette 17, que sí fueron utilizadas 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. De hecho, de los 1901 viajes donde se computaron los niveles de proteína del grano, en el 98,21% de ellos el trigo cosechado tuvo un nivel proteico de al menos 11,5%.

Por otra parte, desde el punto de vista analítico existe una limitación adicional en la zafra 2012. La construcción de la base de datos para el modelo logístico en el contexto de datos de panel se edificó en base a los lotes donde se sembró trigo que tienen en común las distintas zafras. Un lote es una porción de área que constituye parte integrante de un campo o establecimiento agropecuario. Así cada campo está constituido por un número de lotes. Los viajes de la zafra 2011 están asociados a 730 lotes y los de la zafra 2013 a 182 lotes.

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. Por este motivo, 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.

4. Metodología Estadística

La variable a explicar en este caso es dicotómica: el trigo puede contener un nivel de proteína mayor o igual a 11,5% (buena calidad) o inferior (mala calidad). Dado nuestro interés en estimar probabilidades de obtención de trigos de buena calidad, utilizaremos en el análisis un modelo de regresión de respuesta cualitativa, denominado modelo logístico.

Consideremos la siguiente representación:

$$Prob(Y_i=1 | X_i) = \frac{1}{1+e^{-Z_i}} \quad (1)$$

donde $Z_i = (\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik})$.

Esta ecuación representa la probabilidad de éxito ($Y_i=1$), en el caso del presente trabajo que el trigo tenga un nivel de al menos 11,5% de proteína, siendo X_i las variables explicativas y los términos β_i los parámetros a estimar.

La ecuación (1) representa la función de distribución logística. Se puede verificar fácilmente que para cualquier valor que tome Z_i los valores que tome la probabilidad siempre van a estar entre 0 y 1. Este modelo es no lineal en los parámetros, por lo que el método tradicional de estimación de los mismos por mínimos cuadrados ordinarios no aplica.

El modelo logístico se deriva de la ecuación (1) y se define como:

$$L_i = \ln(p/q) = Z_i \quad (2)$$

Este modelo se define como el logaritmo neperiano (ln) de la probabilidad de ocurrencia de un evento (p) que se representa en la ecuación (1), dividido por la probabilidad de no ocurrencia del mismo ($q=1-p$). El mismo es lineal y se estima por el método de máxima verosimilitud, lo que nos va a permitir analizar la probabilidad de éxito para la variable de interés.

Entre las características principales de este modelo se puede mencionar que si bien el modelo logístico es lineal en los parámetros, la probabilidad que se representa en la ecuación (1) no lo es. Por otra parte, los coeficientes de la regresión refieren al cambio en el logaritmo de esta razón de probabilidades expresada en (2) cuando una de las variables explicativas cambia una unidad, permaneciendo las demás constantes.

En un contexto de datos de panel, la función de distribución logística se expresa de la siguiente forma:

$$Prob(Y_{it}=1 | X_{it}, \alpha_i) = \frac{1}{1+e^{-Z_{it}-\alpha_i}} \quad (3)$$

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 X . La maximización de la función de verosimilitud de un modelo logístico en presencia de este efecto no arrojará estimaciones consistentes de los parámetros y es por ello que se debe testear su ausencia. En caso de que se rechace la misma se deberá buscar un mecanismo para obtener estimaciones consistentes de los parámetros.

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

Bajo el método de Chamberlain se utiliza la siguiente función de verosimilitud condicional para obtener las estimaciones, en lugar de la función de verosimilitud tradicional que se utiliza, por ejemplo, en modelos logísticos de corte transversal:

$$L^C = \prod_{i=1}^{i=n} Prob(Y_{i1}=y_{i1}, \dots, Y_{it}=y_{it} | \sum_{t=1}^T y_{it}) \quad (4)$$

Una ventaja de este método es que no se necesitará hacer supuestos sobre la distribución de estos efectos individuales. Entre las desventajas del mismo se encuentran que aquellas variables que no cambian de estado de un período a otro no serán capturadas por el modelo. Por otra parte, este método no toma en cuenta a aquellos casos en que Y_{it} no cambió de estado en el tiempo. Estos no contribuyen a la función de verosimilitud condicional, y por tanto tienen que ser descartados. Esta restricción no solamente implica que la muestra a utilizar será más pequeña, sino que también 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 ji-cuadrado con K grados de libertad (equivalentes al número de parámetros a estimar) y que se formula como sigue:

$$H = (\beta_{mvc} - \beta_{mv})' (VAR_{mvc} - VAR_{mv})^{-1} (\beta_{mvc} - \beta_{mv}) \sim \chi_k^2 \quad (5)$$

Bajo la hipótesis nula que 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.

Conceptualmente se puede decir que bajo la hipótesis nula el efecto individual no observable no está correlacionado con las variables explicativas. Se espera además bajo esta hipótesis una convergencia en los resultados de las estimaciones por los dos métodos, sin embargo el método de estimación condicional no será eficiente ya que no explota toda la información, sino solamente a aquellas variables que cambiaron de estado en el tiempo.

Se utilizó la versión 12.1 del paquete estadístico Stata para realizar las regresiones logísticas en las zafra 2011 y 2013 consideradas individualmente, así como también para el análisis realizado en el contexto de datos de panel.

5. Resultados

5.1. Zafra 2011

En cada uno de los 3695 viajes se midió el porcentaje de proteína del grano. Por otra parte se registraron el rinde por hectárea y el índice de calidad del suelo para cada lote, el nivel de lluvias por zona en el período entre siembra y cosecha, las dosis de nitrógeno aplicadas en cada lote, así como las variedades de trigo utilizadas en cada uno de estos. Finalmente se asoció cada lote con la zona donde el mismo estaba ubicado.

Las variables utilizadas en este trabajo se clasifican en dicotómicas y no dicotómicas. Las dicotómicas refieren a la calidad de los trigos cosechados, tomando el valor 1 cuando el contenido de proteína del grano es mayor o igual a 11,5% y 0 en caso contrario. Cada una de las siete zonas donde opera la empresa toma un valor de 1 en caso de que el grano de trigo cosechado pertenezca a la misma y 0 en caso contrario; el mismo concepto aplica a cada una de las seis variedades de trigo aplicadas.

Los rindes están medidos en toneladas de trigo cosechadas por hectárea; los niveles de lluvias están expresados en milímetros y las dosis de nitrógeno aplicadas en cada lote están medidas en kilogramos por hectárea. Se midieron los días transcurridos desde la fecha en que se aplicó la última dosis de nitrógeno y la fecha de cosecha en cada lote. Finalmente, se utilizó una variable que captura un índice de calidad del suelo para cada lote, que toma valores que varían entre 1 (la menor calidad) y 3 (la máxima calidad).

La tabla 1 expone la información descriptiva de cada una de las variables involucradas en el análisis, incluyendo el caso de éxito que se pondrá en consideración en el modelo de regresión logístico.

El rinde promedio por lote fue de 4,02 toneladas por hectárea. Se registraron un promedio de casi 430 milímetros de lluvia en la ventana transcurrida en el período siembra-cosecha y 113 milímetros en el mes de octubre. Se aplicó en cada lote un promedio de casi 94 kilogramos por hectárea de nitrógeno, con un poco más de 33 kilogramos por hectárea correspondientes a Z30. El tiempo promedio por lote transcurrido entre la aplicación de

Z30 y la fecha de cosecha fue de 97 días, como refleja la variable «días». El índice ambiental promedió 1,85 unidades por lote.

Tabla 1: 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%

Con respecto a las variables binarias, el 43% de las 3695 observaciones correspondió a la variedad de trigo Baguette 19, seguido por el Nogal con aproximadamente un tercio de los viajes. Casi una cuarta parte de los viajes observados se originaron en los campos del Litoral Centro, siendo la zona con mayor número de viajes en la muestra. Se registraron 1377 casos donde el trigo tuvo un nivel de proteína de al menos 11,5%, equivalentes a un 37% de los viajes.

Los resultados de la regresión logística se exponen en la tabla 2. Como era de esperar se encontró una correlación negativa entre la cantidad producida de trigo y la probabilidad de obtener trigo de buena calidad. El nivel de lluvias impactó en forma negativa, sin embargo las lluvias de octubre aumentan la probabilidad de obtención de buenos trigos. Con respecto al nitrógeno, si bien ambas variables impactarían positivamente en la calidad, solamente la dosis Z30 fue significativa con un 95% de confianza. El modelo también corrobora la teoría de que a mayor cantidad de días transcurridos entre la aplicación de la última dosis de nitrógeno y la fecha de cosecha, disminuye la probabilidad de obtener trigos de buena calidad. El índice ambiental fue significativo con un 95% de confianza y con coeficiente positivo, lo que implicaría que a mayor valor del índice ambiental, aumentarán las probabilidades de que el trigo tenga al menos 11,5% de proteína.

Con respecto a las variedades de trigo, las mismas fueron representadas por variables indicadoras (*dummy*) para medir el impacto en la probabilidad de obtener trigos con buen nivel de proteína en el grano. Solamente la variedad Nogal fue significativa con un 95% de confianza. Las restantes variedades, con excepción de la Baguette 19, fueron significativas con un nivel de confianza de un 75%. Todos los coeficientes tuvieron signo positivo, siendo la variedad Nogal la de mayor coeficiente. Esto indicaría que las mayores probabilidades de obtener trigo con buena proteína se encuentran sembrando esta variedad. La variedad Baguette 17 fue quitada del análisis, ya que su contribución al modelo no iba a ser significativa debido a que se realizaron solamente 14 viajes de trigo y en apenas uno se registró trigo con al menos 11,5% de proteína.

Finalmente, la calidad del suelo también se aproximó a través de las distintas regiones del país donde opera la empresa, utilizando variables indicadoras. Todos los coeficientes tuvieron signo positivo. La zona Noreste Norte fue significativa con un 95% de confianza, la zona Centro Norte fue significativa con un 90% de confianza y las restantes zonas no fueron significativas a estos niveles de confianza. El coeficiente más alto se encontró en la zona Noreste Norte, seguido por el de la zona Centro Norte, lo que indicaría que aumentan las probabilidades de obtener trigos de buena calidad si se sembrara en estas zonas.

Tabla 2: Resultados de la regresión logística

Trigo Bueno	Coefficiente	Error Estándar	Valor p
Rindes	-0,286	0,5663	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

Una de las formas de medir la bondad de ajuste del modelo es a través del porcentaje de «aciertos» del mismo, que refiere al porcentaje de observaciones que se proyectaron en forma correcta. A aquellas probabilidades a las que el modelo asignó un valor igual o superior a 0,5 se les asigna un valor de 1, en caso contrario se les asigna un valor de 0. Estos resultados se comparan con los valores efectivamente observados por la variable

dicotómica que representa a la calidad del trigo y así se construye el porcentaje de aciertos.

Como se exhibe en la tabla 3, en 726 casos el modelo predijo correctamente la obtención de trigo de buena calidad y en 1950 casos el modelo predijo correctamente la obtención de trigo de mala calidad. En 1019 casos el modelo no acertó en su predicción. De esta forma, el porcentaje de aciertos fue de un 72,42%.

Tabla 3: Bondad de ajuste del modelo

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

Las siguientes 3 tablas también tienen como objetivo analizar la calidad del modelo obtenido. Se compara, para cada zona y variedad de trigo sembrada, la información descriptiva de la muestra en cuanto al porcentaje real de trigos de buena calidad con las probabilidades de éxito asignadas por el modelo logístico.

Tabla 4: 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%

Tabla 5: 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

Tabla 6: 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%

Las probabilidades de éxito del modelo logístico se obtuvieron evaluando las variables no dicotómicas en sus promedios zonales por lote. Es interesante observar que para aquellos casos con al menos 80 observaciones, el porcentaje de trigos de buena calidad por zona y variedad es cercano a las probabilidades de éxito construidas por el modelo, lo que sugiere la buena calidad del mismo.

De los 42 grupos de datos observados por zona y variedad, 16 de ellos están conformados por al menos 80 observaciones. Con excepción del caso de la variedad Nogal cosechada en la zona Centro Sur donde el modelo subestimó la probabilidad de éxito en un 22,37% con respecto a los trigos de buena calidad observados en la muestra, en cada uno de los 15 casos restantes el desvío entre las probabilidades de éxito del modelo y los datos descriptivos nunca superó el 9,23% en valores absolutos en ningún caso, con un desvío promedio de 5,64% en dichos valores.

5.2. Zafra 2013

La tabla 7 expone la información descriptiva de cada una de las variables involucradas en el análisis, incluyendo el caso de éxito que se pondrá en consideración en el modelo de regresión logístico.

El rinde promedio por lote en la zafra 2013 fue de 3,23 toneladas por hectárea, con un promedio de lluvias globales de 500 milímetros, con 89 milímetros en el mes de octubre. Se aplicaron un promedio de casi 90 kilogramos por hectárea de nitrógeno, con 33 kilogramos por hectárea correspondientes a Z30. En promedio por lote transcurrieron 88 días entre la aplicación de Z30 y la fecha de cosecha. El índice ambiental promedió 1,68 unidades por lote.

La variedad de trigo más sembrada fue Baguette 19 con poco más de la mitad de las observaciones de la muestra. Casi un tercio de los viajes se originaron en el Noreste Norte, siendo ésta la zona con mayor cantidad de movimientos en la muestra. En 945 casos el trigo tuvo un nivel de proteína de al menos 11,5%, equivalentes a casi un 80% de los viajes observados.

Los resultados de la regresión logística para esta zafra se despliegan en la tabla 8. Los rindes impactan negativamente en las probabilidades de obtener trigos de buena calidad, pero en este caso la confianza estadística es de un 78%. El mismo signo puede verse en las dos variables que miden el impacto de las lluvias con un 95% de confianza. Las variables que miden el impacto del nitrógeno en la probabilidad de obtener trigos de

buena calidad tuvieron coeficientes positivos, pero solamente la dosis Z30 fue significativa con un 84% de confianza. Al igual que en el análisis realizado para la zafra 2011, a mayor cantidad de días entre aplicación de Z30 y fecha de cosecha, menores probabilidades de obtener trigos buenos, y estas probabilidades aumentan a mayor índice ambiental.

Tabla 7: 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%

Todas las variables indicadoras asociadas a las variedades de trigo fueron significativas y con coeficientes positivos, siendo las variedades Baguette 501 y Baguette 11 las de mayores coeficientes. Las variables asociadas a las zonas de siembra tuvieron signos negativos, siendo significativas con un 95% las de Litoral Centro, Litoral Sur y Noreste Sur. Las menores probabilidades de obtener trigos de buena calidad se encuentran en las zonas Litoral Centro y Noreste Sur. No se utilizó en el análisis la zona Litoral Norte, ya que de sus 21 observaciones, 20 arrojaron trigos de buena calidad, con lo que su contribución al modelo no iba a ser significativa.

Tabla 8: Resultados de la regresión logística

Trigo Bueno	Coefficiente	Error Estándar	Valor p
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

La tabla 9 despliega la bondad de ajuste del modelo, con el mismo tipo de análisis que el exhibido en la tabla 3. En este caso, en 887 casos el modelo predijo correctamente la obtención de trigo de buena calidad y en 91 casos el modelo predijo correctamente la obtención de trigo de mala calidad. En 208 casos el modelo no acertó en su predicción, por lo que el porcentaje de aciertos fue de 82,46%.

Tabla 9: Bondad de ajuste del modelo

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

Las siguientes 3 tablas también tienen como objetivo analizar la calidad del modelo obtenido, utilizando la misma metodología que la desplegada por las tablas 4, 5 y 6 para la zafra 2011.

Tabla 10: 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%	-

Tabla 11: 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	-

Tabla 12: 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%

Existen 5 casos con más de 80 observaciones, y en todos ellos se encontraron similitudes entre los trigos de buena calidad observados en la muestra con las probabilidades de éxito obtenidas por el modelo. En ningún grupo de datos por zona y variedad con al menos 80 observaciones se encontró un desvío superior al 8,37% en valores absolutos

entre las probabilidades de éxito del modelo y los datos descriptivos, con un desvío promedio de 5,15% en dichos valores.

5.3. Datos de Panel con las Zafras 2011 y 2013

La tabla 13 muestra la información descriptiva de los 67 lotes que tienen en común las zafras 2011 y 2013. Se puede apreciar que no se replicó en forma exacta la estrategia de siembra en función de las variedades de trigo utilizadas en dichos lotes en estas dos zafras. Las únicas variables que se estructuran de la misma forma en el panel tienen que ver con las zonas asociadas a los lotes, ya que cualquiera que sea el período analizado, estos nunca cambiarán su ubicación geográfica.

Tabla 13: Estadística descriptiva

Variables No Binarias	ZAFRA 2011			ZAFRA 2013		
	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%

Dadas las muy pocas observaciones que se registran para las variedades de trigo aplicadas y las zonas donde se encuentran estos lotes, no se utilizaron estas variables indicadoras en el modelo logístico de datos de panel. La mayor cantidad de observaciones se registró para la variedad Baguette 19, con 59 datos si se suman las dos zafras. Con respecto a las zonas, se computaron apenas 34 datos en la zona Centro Norte, siendo la zona con mayor cantidad de observaciones en las dos zafras. Para las restantes variables se registraron 67 observaciones por zafra.

Se realizó el contraste de tipo Hausman para seleccionar el método de estimación a utilizar. El valor del estadístico es 7,54 y el valor p es igual a 0,2738. Esto implica que no se rechaza la hipótesis nula del contraste, estimando así el modelo por el método de máxima verosimilitud no condicional.

Conviene aclarar que no se utilizó la variable asociada a la calidad del suelo para realizar el contraste, ya que por la estructura del método de estimación condicional (trabaja con las diferencias temporales de los regresores) se distorsionaba el coeficiente asociado a dicha variable. En solamente 4 de los 67 lotes se registraron cambios en la medición de esta variable de una zafra a la otra (menos de un 6% del total de lotes), lo que implicaba que este método solo hubiera considerado 4 observaciones de la misma para la estimación. Por este motivo se decidió no utilizar dicha variable para el contraste mencionado en el párrafo anterior.

De todas formas, se realizó el contraste utilizando dicha variable en el modelo no condicional, obteniendo un valor del estadístico de 5,49 y un valor p de 0,4830, con lo que tampoco se hubiera rechazado la hipótesis nula.

Un aspecto interesante tiene que ver con que la dimensión de los paneles es distinta cuando se estima el modelo por el método de máxima verosimilitud con respecto al método propuesto por Chamberlain. En solamente 35 de los 67 lotes analizados se registraron cambios de estado en la variable dicotómica asociada a la calidad del trigo, lo que implica 70 observaciones por variable en dicho modelo, poco más de un 50% de las 134 observaciones del panel completo. El beneficio adicional de no haber rechazado la hipótesis nula en el contraste de tipo Hausman, tiene que ver con la posibilidad de explotar el panel en su máxima dimensión.

Los resultados de la estimación del modelo logístico por el método de máxima verosimilitud no condicional se despliegan en la tabla 14. Como se puede apreciar en la misma, con al menos un 92% de confianza fueron significativas las variables rindes, lluvias del período completo, lluvias de octubre, días transcurridos entre la aplicación de Z30 y la fecha de cosecha y el índice de la calidad del suelo. Las dos variables asociadas a los kilogramos de nitrógeno aplicados a cada lote tuvieron signo positivo pero no fueron significativas.

Las variables rindes, lluvias de octubre y días transcurridos entre aplicación de Z30 y fecha de cosecha tuvieron signo negativo, mientras que las variables lluvias del período e índice de calidad del suelo tuvieron signo positivo, implicando su impacto positivo en la probabilidad de que el trigo sea de buena calidad en la dimensión analizada.

Como se exhibe en la tabla 15, el modelo acertó en sus predicciones en 88 casos, fallando en 46 casos. De esta forma el porcentaje de aciertos fue de 65,67%, inferior a los porcentajes obtenidos en los modelos utilizados en los dos casos anteriores.

Tabla 14: Resultados de la regresión logística

Trigo Bueno	Coefficiente	Error Estándar	Valor p
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

Tabla 15: Bondad de ajuste del modelo

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

6. Conclusiones

La relevancia que está adquiriendo la calidad del trigo producido para concretar acuerdos de compra-venta constituyó el motor que impulsó el presente trabajo analítico. Las bases de datos con 3695 observaciones para la zafra 2011 y 1186 observaciones para la zafra 2013 de trigo en la empresa Tafilar S.A., constituyeron una excelente oportunidad de analizar dicha información de forma de aprender más sobre las posibilidades de obtener trigos con buen nivel de proteína.

Por otra parte, dada la extensión territorial de la empresa en el país, las conclusiones a las que se arriban para la institución analizada podrían ser tenidas en cuenta (al menos parcialmente) para el país en su conjunto.

Los resultados del modelo estadístico utilizado arrojaron luz sobre variables que forman parte del proceso de toma de decisiones humano y sobre otras variables que están fuera de dicho control, pero que describen indicios sobre qué se puede esperar con respecto al nivel de proteína de los granos de trigo a cosechar. De esta forma, se puede concluir que hubo un alto grado de cumplimiento del objetivo específico del trabajo.

De los resultados obtenidos del modelo de la zafra 2011, se desprende que sembrar la variedad de trigo Nogal incrementó las probabilidades de obtener trigos con buen nivel de proteína. Lo mismo ocurrió con los trigos sembrados en las zonas Noreste Norte y Centro Norte.

El modelo corroboró la correlación negativa entre la cantidad de trigo producida y el nivel de proteína contenida en el mismo. Resultó interesante ver los efectos contrarios que generan los niveles de lluvia en toda la etapa previa a la cosecha, con respecto a las del mes de octubre.

Finalmente, el modelo 2011 verifica la importancia de la dosis de nitrógeno Z30 y del índice de calidad del suelo en la obtención de trigos con alta proteína. La hipótesis que sostenía que las posibilidades de obtener trigos con alta proteína disminuían en la medida que la aplicación de Z30 se alejaba en el tiempo también es verificada.

Con respecto a la zafra 2013, el modelo también arrojó un signo negativo en la estimación del parámetro asociado a la variable rindes, aunque con un nivel de casi un 78% de confianza. Tanto las lluvias globales como las de octubre fueron significativas y ambas con signo negativo, a diferencia del modelo para el 2011 donde las lluvias de octubre aumentaban la probabilidad de obtener trigos de alta proteína. La aplicación de la última dosis de nitrógeno tuvo un impacto positivo y fue significativa con un nivel de confianza de un 84%.

Para este modelo 2013 se verifican las conclusiones del caso 2011 en cuanto a signos y nivel de significación de las variables que describen el índice de calidad de los suelos y el número de días que transcurren entre aplicación de Z30 y la fecha de cosecha. Finalmente, las variedades Baguette 501 y Baguette 11 son las que arrojaron coeficientes más altos y las que más ayudarían a aumentar la probabilidad de obtener buenos trigos, aunque en el caso del Baguette 501, dado que solamente se observaron 35 viajes en el Litoral Centro, no sería razonable ser definitivo en este concepto. La zona que menos contribuye en el modelo a obtener trigo con alta proteína es el Litoral Sur aunque, como en el caso de las variedades sembradas, no se puede ser concluyente en esta afirmación cuando solamente se observaron 33 datos de una única variedad, Baguette 11.

El modelo logístico utilizado en datos de panel, si bien más limitado por su menor cantidad de datos, confirma ciertos resultados obtenidos por los modelos de un solo período. Este modelo confirma lo revelado a través de los dos modelos anteriores en cuanto al impacto negativo de los rindes y de los días transcurridos entre aplicación de Z30 y fecha de cosecha en las probabilidades de éxito. También verifica este modelo el impacto positivo en la probabilidad de éxito del índice de calidad del suelo.

A pesar de sus coeficientes positivos en este tercer modelo, ninguna de las dos variables vinculadas al nitrógeno fue significativa con un elevado nivel de confianza. De todas formas, si tomamos en cuenta cada zafra a nivel individual hay una firme posibilidad de que la dosis Z30 sea influyente en la obtención de trigos de buen nivel de proteína, si nos basamos en los modelos utilizados.

Es muy difícil concluir, a partir de los tres modelos, el impacto de las lluvias del período previo a la cosecha y de octubre en las probabilidades de obtener trigos de buena calidad en el sentido analizado. Si bien todos los modelos tienen en común que estas dos variables fueron significativas, los signos difirieron.

Las estimaciones de los parámetros de cada modelo permiten obtener probabilidades de éxito, evaluando las variables explicativas en determinados niveles. En el caso del presente trabajo se evaluaron las variables explicativas no dicotómicas en sus promedios por lote para cada zona. La sensibilización de estos niveles por zona y variedad arrojará distintas probabilidades de obtención de trigos con al menos 11,5% de proteína, lo que constituye una interesante herramienta auxiliar para evaluar la estrategia de siembra considerando distintos escenarios. Estas probabilidades no se pueden sensibilizar por zona y variedad en el modelo de datos de panel, que no utilizó las variables explicativas dicotómicas en la modelización.

Finalmente, se plantea como línea de investigación futura la necesidad de analizar los determinantes de otras variables fuertemente vinculadas a la calidad del trigo, como ser el peso hectolítrico y la fuerza panadera del grano.

7. Referencias

Erekul, O., K.-P. Götz, and Y.O. Koca (2012). Effect of sulphur and nitrogen fertilization on bread-making quality of wheat (*Triticum aestivum* L.) varieties under Mediterranean climate conditions. *Journal of Applied Botany and Food Quality*, 85, 17-22.

(Disponible en <http://pub.jki.bund.de/index.php/JABFQ/article/viewFile/1994/2370>)

García Lamothe, A. (2006). El efecto de la nutrición mineral sobre el rendimiento y la calidad del grano de trigo. *INIA - Serie de Actividades de Difusión*, 444, 8-21.

Greene, W. (2002). *Econometric Analysis*. Prentice Hall (5ta edición).

Lee, B.-H., P.L. Kenkel, B.W. Brorsen (2011). *Pre-harvest Forecasting of County Wheat Yield and Wheat Quality Conditional on Weather Information*.

(Disponible en <http://purl.umn.edu/98829>)

Duración de catéteres transitorios en un servicio de diálisis: Un análisis de supervivencia

Julia-Inés Fernández y Elisabet López
Facultad de Ciencias Económicas y Estadística
Universidad Nacional de Rosario

1. Introducción

Al funcionamiento anormal de los riñones por más de tres meses o a la alteración estructural de los mismos se lo denomina enfermedad renal crónica. Esta es una enfermedad prevalente a nivel mundial que no presenta síntomas hasta estadíos avanzados en los cuales los pacientes requieren diálisis o trasplante renal. Una detección precoz puede ayudar a prevenir complicaciones y retrasar el desarrollo de la enfermedad. En el año 2007 en Argentina la tasa ajustada de pacientes en diálisis crónica fue de 608,5 por millón de habitantes.

En las ciudades de Pergamino y Rojas de la provincia de Buenos Aires, República Argentina, tiene sede el Instituto de Nefrología Pergamino que desarrolla, entre otras actividades, una atención clínica nefrológica y de hemodiálisis en pacientes con insuficiencia renal crónica o aguda; la insuficiencia renal aguda es el mal funcionamiento de los riñones que se produce rápidamente, en menos de dos días. A los pacientes que sufren insuficiencia renal se les realiza hemodiálisis con el fin de eliminar del torrente sanguíneo las toxinas que el riñón no puede filtrar. Para poder llevar a cabo este proceso es necesario un acceso vascular, el cual puede ser permanente o transitorio. Un acceso vascular o catéter transitorio se utiliza cuando el ingreso del paciente a diálisis se produce de manera imprevista o existe un mal funcionamiento de un acceso vascular permanente.

Un grupo de médicos tuvo el interés de estudiar la duración de los catéteres transitorios colocados en el Instituto de Nefrología Pergamino y si la misma se ve influenciada por ciertas variables que fueron registradas en una base de datos.

2. Objetivos

A partir de las inquietudes planteadas por el equipo de profesionales del Instituto de Nefrología Pergamino, el presente trabajo tiene por objetivo analizar si el tiempo de duración de los catéteres transitorios colocados en el Instituto de Nefrología Pergamino

durante el período comprendido entre el 12/05/07 y el 26/05/13 se ve influenciado por alguna de las variables registradas.

Por otra parte, en el período en que fueron recolectados los datos el instituto cambió su técnica de conexión a hemodiálisis, lo cual se pensó que podía influir en la duración de los catéteres transitorios. Es por esto por lo que un objetivo secundario de este trabajo es evaluar si la técnica de conexión tiene influencia sobre el tiempo de duración de los catéteres transitorios.

3. Datos

La base de datos analizada contiene información acerca de 139 pacientes a los cuales se les colocaron y extrajeron catéteres transitorios en el período transcurrido entre el 12/05/07 y el 26/05/13, en el Instituto de Nefrología Pergamino. En el presente estudio sólo se analizaron los datos de los pacientes referidos al primer catéter colocado en dicho período. Se tomó esta decisión con el objeto de garantizar la independencia entre las observaciones, ya que los datos referidos a dos catéteres colocados a un mismo paciente en distintos momentos están relacionados porque ambos se ven afectados por las características de ese individuo tales como sexo, antecedentes de salud, pautas de comportamiento, etcétera. Las variables registradas en la base de datos fueron:

- Identificación del paciente.
- Sexo del paciente, clasificado en femenino (F) o masculino (M).
- Diabetes: indica si el paciente es diabético, y toma los valores sí (S) o no (N).
- Técnica de conexión: indica con qué técnica se conectó al paciente a la máquina de diálisis, si la técnica vieja, llevada a cabo por un único técnico de diálisis, o la nueva, realizada por dos técnicos. La técnica de conexión vieja se utilizó hasta el 25 de Agosto de 2008.
- Edad del paciente al momento de la colocación del catéter, medida en años.
- Sitio de colocación del catéter, clasificado en femoral derecho (FD), femoral izquierdo (FI), subclavia derecha (SD), subclavia izquierda (SI), yugular anterior derecha (YAD) y yugular anterior izquierda (YAI).
- Motivo de colocación del catéter, clasificado en las categorías: inicio de hemodiálisis crónica (IH), obstrucción del acceso vascular permanente (O), insuficiencia renal aguda (IRA), disfunción estructural (DEC), recambio de catéter por infección (RCI), disfunción del catéter (DC), pase de modalidad (PM), insuficiencia renal crónica reagudizada (IRCR), acceso definitivo (AD).
- Duración del catéter, medida en días transcurridos entre la colocación y la extracción del catéter transitorio.
- Motivo de extracción del catéter, clasificados en fallecimiento del paciente (F), recuperación de la función (RF), recambio de catéter transitorio (RCI), infección (I), disfunción (D), sospecha de infección (SI), acceso vascular definitivo (AVD), derivado a otro centro (DC), trasplante (TX) y desprendimiento espontáneo del catéter (SS).

4. Métodos Estadísticos

En primera instancia se llevó a cabo un análisis descriptivo univariado de las variables registradas en la base de datos. Se construyeron tablas de frecuencias para las variables categóricas, mientras que para las variables continuas se construyeron histogramas y se calcularon medidas de posición central y de dispersión.

La variable duración mide el tiempo hasta la ocurrencia de un evento definido como la extracción del catéter transitorio por mal funcionamiento del mismo. Al tiempo hasta la ocurrencia de un evento se lo denomina dato de supervivencia (*survival*) y para analizar datos de este tipo se utilizan técnicas conocidas como análisis de supervivencia. Por lo tanto, en una segunda instancia se aplicaron algunas técnicas no paramétricas para analizar la duración de los catéteres transitorios con fines descriptivos, como la estimación de funciones de supervivencia por el método de Kaplan-Meier y el uso del contraste del rango logarítmico (*log-rank*) para comparar las experiencias de supervivencia en dos o más grupos definidos a partir de los niveles de las variables categóricas.

Por último, se estimó un modelo paramétrico de riesgos (*hazard*) proporcionales para estudiar cómo la duración de los catéteres transitorios de los pacientes del Instituto de Nefrología Pergamino depende de los valores de las variables registradas al momento de la colocación de los catéteres. En dicho modelo se asumió que los tiempos de duración siguen una distribución Weibull, en base a lo observado en el denominado gráfico del logaritmo de la función de riesgo acumulada.

Una característica de los datos de supervivencia es que para muchos pacientes no es posible observar el evento de interés durante el periodo en estudio, en este caso, la extracción del catéter por mal funcionamiento del mismo. Los tiempos de duración de los catéteres en dichos pacientes son tiempos censurados.

Para definir cuáles son los pacientes que presentan tiempos censurados se utilizaron las categorías de la variable motivo de extracción. Cuando el motivo de extracción del catéter transitorio fue acceso vascular definitivo, trasplante, recuperación de la función renal, fallecimiento o derivación del paciente a otro centro, se consideraron los datos como censurados, ya que la extracción del mismo no se produjo en estos casos por un mal funcionamiento del catéter.

Todos los cálculos utilizados en el análisis de los tiempos de duración de los catéteres transitorios fueron realizados con el software estadístico libre R.3.0.1.

4.1. Análisis Descriptivo

Se analizaron de manera individual cada una de las variables registradas en la base de datos (tabla 1), y se observó que el 61% de los pacientes son hombres, el 76% de los pacientes no son diabéticos, el 80% fue conectado utilizando la nueva técnica de conexión a diálisis, en el 80% de los casos el catéter transitorio se colocó en la yugular anterior derecha y que los motivos más frecuentes de colocación del catéter son por insuficiencia renal aguda (39%), ingreso a hemodiálisis crónica (30%) y obstrucción del acceso vascular permanente (19%).

Las variables edad y duración del catéter son continuas. Para dichas variables se realizaron los correspondientes histogramas de frecuencias en los cuales se observan distribuciones asimétricas (gráficos 1 y 2), razón por la cual se analizan la mediana como medida de posición central y el rango intercuartílico como medida de dispersión. El 50% de los pacientes tiene 64 años o menos y la distancia que existe en el 50% central de la distribución de la edad es de 25 años (tabla 2). El 50% de los catéteres transitorios duró 16 días o menos y la distancia que existe en el 50% central de la distribución de la duración es de 35 días (tabla 2).

Tabla 1

Variable	Categoría	Porcentaje
Sexo	F	38,85%
	M	61,15%
Diabetes	N	76,26%
	S	23,74%
Conexión	N	79,86%
	V	20,14%
Sitio de colocación	FD	2,88%
	FI	2,16%
	SD	3,60%
	SI	2,15%
	YAD	79,86%
	YAI	9,35%
Motivo de colocación	DEC	2,16%
	AD	0,72%
	DC	0,72%
	IH	30,22%
	IRA	39,57%
	IRCR	5,76%
	OFAV	18,71%
	PM	0,72%
	RCI	1,44%

Gráfico 1

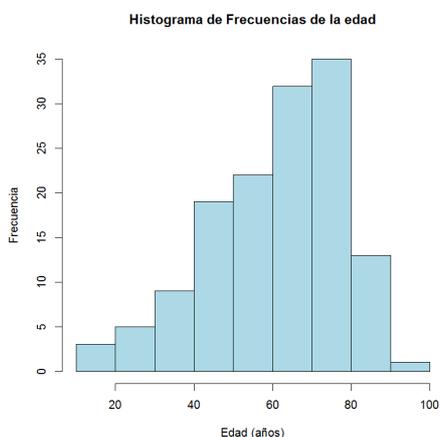


Gráfico 2

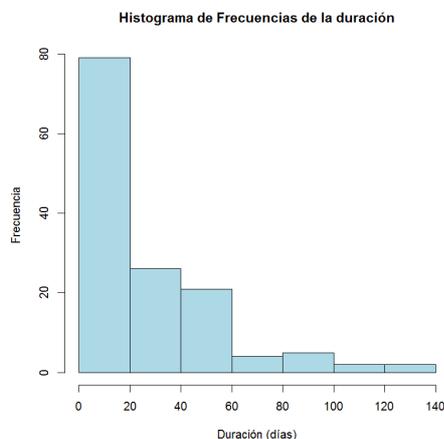


Tabla 2

	Edad	Duración
Mediana	64	16
Mínimo	13	1
Máximo	94	133
Rango Intercuartílico	25	35

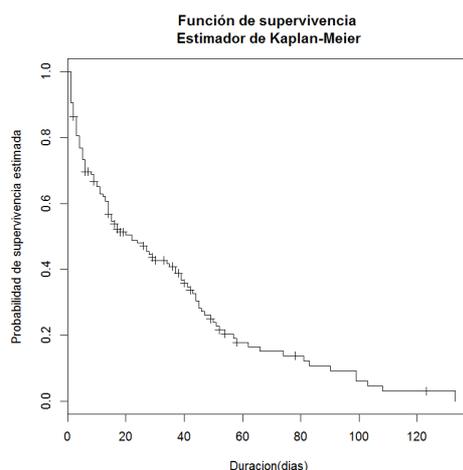
4.2. Análisis de Supervivencia

Para estudiar el comportamiento de la duración de los catéteres transitorios se realizó una estimación de la curva de supervivencia de los mismos utilizando el método de Kaplan-Meier (gráfico 3). La función o curva de supervivencia para un tiempo de duración determinado representa la probabilidad de que un catéter transitorio tenga una duración mayor o igual a ese tiempo, por ejemplo un catéter tiene aproximadamente un 18% de probabilidad de durar 60 días o más.

Existen distintos métodos para estimar una curva de supervivencia. Uno de los métodos más utilizados es el de Kaplan-Meier; esta técnica no paramétrica permite incorporar la información que proporcionan aquellos tiempos de duración de los catéteres que fueron censurados.

El 50% de los catéteres transitorios duran 22 días o más (gráfico 3). Con una confianza del 95% es de esperar que el tiempo mediano de duración de los catéteres se encuentre entre los 12 y 32 días.

Gráfico 3



Antes de plantear un modelo para datos de supervivencia puede evaluarse si cada una de las variables registradas influyen sobre los tiempos de duración de los catéteres transitorios de manera individual. Pueden compararse los tiempos de supervivencia entre

los grupos definidos por cada una de dichas variables utilizando una prueba de hipótesis, por ejemplo comparar la duración de catéteres colocados a hombres y a mujeres. Con este objetivo se llevó a cabo el contraste del rango logarítmico para evaluar si las experiencias de supervivencia son las mismas en los distintos grupos definidos por las variables registradas.

Para este contraste y para el modelo paramétrico de riesgos proporcionales, se recategorizaron las variables sitio y motivo de colocación dado que algunas categorías presentaban frecuencias bajas. Las nuevas categorías para sitio de colocación son yugular anterior derecha y otros sitios, mientras que para motivo de colocación las nuevas categorías son ingreso a hemodiálisis crónica, insuficiencia renal aguda, obstrucción del acceso vascular permanente y otros motivos. Además, para realizar el contraste del rango logarítmico se creó una variable categórica para la edad: se agruparon los datos en cuatro intervalos tomando como puntos de corte los cuartiles de dicha variable.

Con una significación del 5% se puede concluir que las experiencias de supervivencia de los catéteres difieren sólo para los distintos motivos de colocación (tabla 3).

Tabla 3

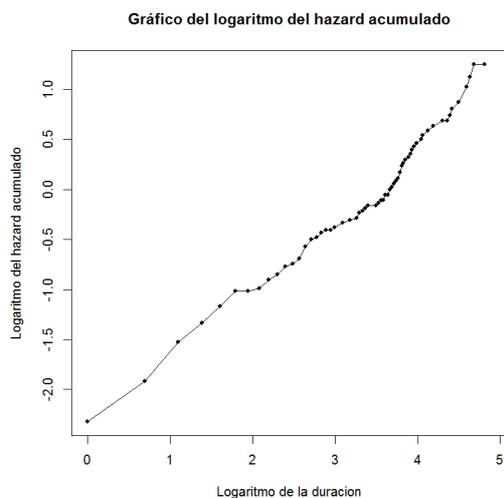
Variable	Estadística	Grados de libertad	Probabilidad asociada
Sexo	0,3	1	0,563
Diabetes	0,3	1	0,563
Conexión	2,1	1	0,152
Edad	1,3	3	0,729
Sitio	0,2	1	0,677
Motivo de colocación	57,5	3	<0,0001

Para evaluar si las variables registradas influyen de manera conjunta sobre los tiempos de duración de los catéteres transitorios se ajustó un modelo de riesgos proporcionales. Estos modelos pueden ser paramétricos o no paramétricos; en un modelo paramétrico de riesgos proporcionales se supone que los tiempos de duración de los catéteres siguen una determinada distribución de probabilidad. La función de riesgo en un momento determinado indica el riesgo de que sea necesario extraer un catéter en ese momento. El gráfico del logaritmo de la función de riesgo acumulada permite determinar si los tiempos de duración siguen algún tipo de distribución de probabilidad.

El gráfico del logaritmo de la función de riesgo acumulada, estimada para el conjunto de datos analizados, representa aproximadamente una recta (gráfico 4), lo cual conduce a pensar que la distribución de los tiempos de duración de los catéteres es una Weibull. En consecuencia se ajustó un modelo paramétrico de riesgos proporcionales.

Para determinar cuáles de las variables explicativas consideradas influyen sobre la duración de los catéteres transitorios se llevó a cabo un proceso de selección del modelo que consiste en comparar las verosimilitudes de distintos modelos paramétricos de riesgos proporcionales estimados considerando una distribución Weibull de los tiempos de duración.

Gráfico 4



Si se desea comparar dos modelos y uno de ellos contiene solamente un subconjunto de los parámetros que contiene el otro, se dice que los modelos están anidados. Para evaluar si uno de dos modelos anidados presenta un mejor ajuste de los datos se calcula menos dos veces el logaritmo de la verosimilitud estimada, $-2\log(L)$, para cada modelo y se comparan dichos valores. Es posible demostrar que la diferencia entre el valor de la cantidad $-2\log(L)$ para el modelo con menos parámetros y para el modelo con más parámetros sigue una distribución ji-cuadrado bajo la hipótesis de que los parámetros que no son compartidos por ambos modelos son iguales a cero, con tantos grados de libertad como la diferencia entre la cantidad de parámetros en los dos modelos.

Esta estadística con distribución ji-cuadrado se utilizó para realizar el proceso de selección de modelos, que consistió en seguir los siguientes pasos (tabla 4):

- Se ajustaron modelos con cada una de las variables explicativas de manera individual, y cada uno de ellos se comparó con el modelo nulo (ninguna variable explicativa) utilizando un nivel de significación del 10% para determinar cuáles son significativas por sí solas; en este caso, la técnica de conexión y el motivo de colocación del catéter.
- Se ajustó un modelo con las variables que resultaron significativas en el paso anterior, que fueron conexión y motivo. Se agregaron a este modelo, de a una, las variables que no resultaron significativas en el primer paso y se determinó que ninguna mejoraba significativamente el ajuste del modelo.
- Se comparó el modelo que tenía técnica de conexión y motivo de colocación con los modelos que tenían solo conexión y sólo motivo, y se concluyó que el modelo con ambas variables es el que presenta el mejor ajuste.
- Se ajustó un modelo incluyendo la interacción de las variables técnicas de conexión y motivo de colocación del catéter y se concluyó que la misma no es significativa.

Tabla 4

Variables en el modelo	2 log(L)
Sexo	982,8
Diabetes	983,0
Conexión	980,6
Edad	983,2
Sitio	983,2
Motivo	932,0
Conexión y Motivo	928,2
Conexión, Motivo y Sexo	926,0
Conexión, Motivo y Diabetes	927,0
Conexión , Motivo y Edad	927,8
Conexión, Motivo y Sitio	928,0
Conexión, Motivo e interacción	920,4

En conclusión, el proceso de selección condujo a un modelo que contiene como variables explicativas a la técnica de conexión y el motivo de colocación del catéter.

A partir de un modelo de supervivencia pueden estimarse razones de riesgo que permiten comparar el riesgo de que sea necesaria la extracción del catéter por mal funcionamiento del mismo para distintas categorías de las variables explicativas. Las razones de riesgo toman valores positivos; si toman el valor uno significa que el riesgo es el mismo en los dos grupos que se están comparando, mientras que un valor mayor a uno implica un riesgo mayor para alguno de los grupos. Además de las estimaciones puntuales pueden construirse intervalos de confianza para estas razones de riesgo.

5. Resultados

Luego de ajustar el modelo paramétrico de riesgos proporcionales con una distribución Weibull de los tiempos de duración de los catéteres transitorios, se estimaron las razones de riesgo con sus respectivos intervalos de confianza y se obtuvieron las siguientes conclusiones:

- Cuando el catéter transitorio fue colocado por una insuficiencia renal aguda el riesgo de que sea necesario extraerlo es aproximadamente seis veces mayor que el riesgo cuando el motivo de colocación del catéter es el de ingreso a hemodiálisis crónica, si se considera la misma técnica de conexión.
- Cuando el catéter transitorio fue colocado por una causa distinta a la de insuficiencia renal aguda o a la obstrucción del acceso vascular permanente, el riesgo de que sea necesario extraerlo es aproximadamente tres veces mayor que el riesgo cuando el motivo de colocación del catéter es el de ingreso a hemodiálisis crónica, considerando la misma técnica de conexión.

- El riesgo de que haya que extraer el catéter es el mismo para los pacientes conectados con ambas técnicas de conexión cuando el motivo de colocación es el mismo.

Además, el modelo ajustado permitió obtener las siguientes estimaciones (Anexo):

- Cuando los pacientes son conectados con la técnica nueva y el motivo de colocación del catéter transitorio es el ingreso a hemodiálisis crónica, el 50% de los catéteres dura 41 días o más.
- Cuando los pacientes son conectados con la técnica nueva y el motivo de colocación del catéter transitorio es una insuficiencia renal aguda, el 50% de los catéteres dura 8 días o más.
- Cuando los pacientes son conectados con la técnica nueva y el motivo de colocación del catéter transitorio es la obstrucción del acceso vascular permanente, el 50% de los catéteres dura 27 días o más.
- Cuando los pacientes son conectados con la técnica nueva y el motivo de colocación del catéter transitorio es distinto a los mencionados anteriormente, el 50% de los catéteres dura 16 días o más.

6. Conclusiones

A través de la aplicación de un análisis de supervivencia se pudieron cumplir los objetivos planteados, ya que puede concluirse que, de las variables registradas, el motivo de colocación del catéter tiene una marcada influencia sobre la duración de los catéteres transitorios para el período analizado en el Instituto de Nefrología de Pergamino. Cuando el motivo de colocación es el ingreso a hemodiálisis los resultados parecen indicar que la duración del catéter transitorio es mayor que la duración de catéteres colocados por otros motivos. Además, la técnica empleada para conectar al paciente no afecta el riesgo de que sea necesaria la extracción de un catéter transitorio.

Gran parte de la literatura encontrada acerca de catéteres transitorios para hemodiálisis estudia la infección de los mismos en lugar de estudiar su duración, independientemente del motivo por el cual haya sido necesario extraerlo.

Se observó que la duración media estimada en el presente trabajo es similar a la estimada en trabajos anteriores en los cuales se estudió la duración de catéteres transitorios.

Un aspecto a destacar de los resultados presentados es que cuando el motivo de colocación del catéter transitorio es el ingreso a hemodiálisis crónica, es decir, cuando hay un control previo del estado del paciente, la duración media del acceso vascular es mayor que la estimada para los otros motivos de colocación considerados.

No se hallaron trabajos previos acerca de la duración de catéteres transitorios realizados en Argentina; por este motivo los resultados obtenidos pueden ser un aporte de importancia en el proceso de toma de decisiones en un servicio de nefrología.

Cabe destacar que en el presente estudio no se registró ninguna variable acerca del cuidado que cada paciente da al catéter; este tipo de variables podría enriquecer el análisis acerca de la duración de los catéteres transitorios.

Por otra parte, las conclusiones obtenidas pueden verse afectadas por el hecho de que algunos de los pacientes registrados en la base de datos ya habían utilizado previamente catéteres transitorios. Si bien se consideró el primer catéter registrado durante el período en estudio, sería conveniente analizar los datos de pacientes que utilizan por primera vez un catéter transitorio.

7. Referencias

Collett, D. (2003). *Modelling Survival Data in Medical Research*. Chapman & Hall / CRC.

Pérez Delgado Y., Y. Sotolongo Molina, M. Muradás Augier, L. Vigoa Sánchez y E. Lugo López (2006). Supervivencia y complicaciones de los catéteres para hemodiálisis: nuestra experiencia. *Revista Cubana de Cirugía*, 45 (3-4).

(Disponible en <http://scielo.sld.cu/pdf/cir/v45n3-4/cir08306.pdf>)

Naumovic RT, DB Jovanovic, and LJ Djukanovic (2004). Temporary vascular catheter for hemodialysis: a 3 year prospective study. *The International Journal of Artificial Organs*, 27 (10), 848-854.

Ministerio de Salud de la Nación (2010). Guía de Práctica Clínica sobre Prevención y Detección Precoz de la Enfermedad Renal Crónica en Adultos en el Primer Nivel de Atención.

(Disponible en http://www.msal.gov.ar/images/stories/bes/graficos/0000000069cnt-2012-08-02_guia-prevencion-deteccion-precoz-enfermedad-renal-cronica-adultos.pdf)

Sarkar, D. (2008) *Lattice: Multivariate Data Visualization with R*. Springer.

Therneau, T.M., and P.M. Grambsch (2000). *Modeling Survival Data: Extending the Cox Model*. Springer.

Hubeaux, S., and K. Rufibach (2014). *SurvRegCensCov: Weibull Regression for a Right-Censored Endpoint with an Arbitrarily Censored Covariate*. R package, version 1.0.

(Disponible en <http://cran.r-project.org/web/packages/SurvRegCensCov/index.html>)

Anexo

El modelo estimado (tabla I) tiene como variables explicativas la técnica de conexión y el motivo de colocación del catéter:

$$\hat{h}_i(t) = \exp(\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4) \hat{\lambda} \hat{\gamma} t^{\hat{\gamma}-1}, \quad i=1, \dots, 139 \quad (1)$$

donde:

- t es la duración del catéter transitorio en días
- $X_1 = \begin{cases} 1 & \text{si la técnica de conexión es la vieja} \\ 0 & \text{si la técnica de conexión es la nueva} \end{cases}$
- $X_2 = \begin{cases} 1 & \text{si el motivo de colocación es insuficiencia renal aguda} \\ 0 & \text{en otro caso} \end{cases}$
- $X_3 = \begin{cases} 1 & \text{si el motivo de colocación es obstrucción del acceso vascular permanente} \\ 0 & \text{en otro caso} \end{cases}$
- $X_4 = \begin{cases} 1 & \text{si el motivo de colocación fue clasificado como otro} \\ 0 & \text{en otro caso} \end{cases}$
- $\hat{\lambda}$ es la estimación del parámetro de escala de la distribución Weibull de los tiempos de duración
- $\hat{\gamma}$ es la estimación del parámetro de forma de la distribución Weibull de los tiempos de duración
- $\exp(\hat{\beta}_j)$, $j=1,\dots,4$, permite estimar la razón de los riesgos de extracción del catéter transitorio entre dos categorías de alguna de las variables en el modelo

Tabla I

Parámetro	Estimación	Desvío Estándar	Estadística de Wald	Probabilidad Asociada
β_1	0,43	0,23	7,8	0,03
β_2	1,72	0,25	6,94	<0,001
β_3	0,43	0,29	1,52	0,06
β_4	0,94	0,39	2,41	0,008
λ	0,02	0,006	-	-
Υ	1,003	0,07	-	-

En función del modelo seleccionado se calcularon las medianas de los tiempos de duración de los catéteres teniendo en cuenta distintas categorías de las variables incluidas en el mismo. Solo se consideró la nueva técnica de conexión debido a que la vieja ya no es utilizada.

La fórmula utilizada para calcular estas medianas es:

$$\tilde{\chi}_t = \left\{ \frac{\log(2)}{\hat{\lambda} \exp(\hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4)} \right\}^{1/\hat{\gamma}} \quad (2)$$

